UNIVERSITY OF TARTU
Faculty of Social Sciences
School of Economics and Business Administration

Zulfiyya Piriyeva, Nazan Hajiyeva

Data Literacy in Organizations

Master's thesis

Supervisors: Lennart Kitt, Associated professor Juliia Trabskaia

Tartu 2023

We have written this master's thesis independently. Any ideas
or data taken from other authors or other sources have been fully referenced.

List of Content

List of Figures

List of Tables

Abstract

The world today generates, analyzes, and utilizes data to understand the main patterns behind it. However, lack of data literacy is a global problem, which creates a significant disconnect between data literacy's capability and its actual utilization. Data literacy is essential in business and must become the "second language" of corporate organizations if they are to gain a competitive advantage. The purpose of this study is to develop a data literacy competency framework, coordination, and measurement at an organizational level. This will be achieved through an overview of the literature view on data literacy competences, skills, insight on current needs in an organization, proposing a framework which describes the main data skills to support the development of an organizational culture that embraces the use of data, and identifying how to coordinate and measure it within an organization.

Keywords: Data, Literacy, competency, organizational culture.

Introduction

Data is everywhere. It is hard to imagine today's world without generating, analyzing and using it to understand the main pattern behind it. Every person, company, device produces data. Here companies and organizations should be underlined, as they need to utilize data for growing and establishing competitiveness in the market. The economist and director of the Massachusetts Institute of Technology Center for Digital Business, Erik Brynjolfsson, was quoted in the article of New York Times as saying: "We're rapidly entering a world where everything can be monitored and measured, but the big problem is going to be the ability of humans to use, analyze, and make sense of the data."

United Nation conducted a report "A World that Counts: Mobilizing the Data Revolution for Sustainable Development", and identified data revolution as:

New technologies are leading to an exponential increase in the volume and types of data available, creating unprecedented possibilities for informing and transforming society and protecting the environment. Governments, companies, researchers and citizen groups are in a ferment of experimentation, innovation and adaptation to the new world of data, a world in which data is bigger, faster and more detailed than ever before. (United Nations, 2014a, p. 2)

This report also emphasizes that the global problem of inadequate data literacy demands immediate attention and prompt measures to be taken in order to address its profound impact on various domains.

According to studies, a significant portion of our population is unaware of what data literacy is and what is expected of it. It causes a significant disconnect between data literacy's capability and its actual utilization. Developing strategies to explain and communicate data handling and conversion abilities to audiences with various backgrounds and skill levels is essential. It is even more important for businesses to implement appropriate training approaches because one of the biggest gaps in most organizations is how to educate the employees in data literacy. Today's world citizens are considered individuals who have a high capability to interact with and utilize technology, think critically, make connections, and apply these skills within social and professional context [Chinien and Boutin, 2011]. It is a fact that almost every area of the economy requires employees who have data literacy.

Despite this is a hot topic and investigated upon many different fields, such as Data Literacy in education [Ridsdale et al., 2015], in universities as an undergraduate curriculum (Carlson et al.), in public services, etc., there is still room for research about Data Literacy in organizations and the workforce. If the workforce does not have sufficient data literacy skills, then this puts economic growth at risk in the long term. The professional diversity and poor data literacy understanding within organizations lead to communication challenges, especially between business and IT units. According to Gartner, as the new key skill of the digital era, many C-level executives and professionals struggle to speak data fluently. Business organizations are particularly affected by the data literacy gap because it limits their capacity to understand the data and use it for decision-making. Data literacy is essential in business and must become the "second language" of corporate organizations if they are to gain a competitive advantage [Logan, 2018].

The master thesis is the result of collaboration between Tartu University and SEB Bank. In the initial stages, we aimed to understand why SEB Bank was interested in this topic and what they

expected from the partnership. To achieve this, it was studied existing literature and interviewed relevant stakeholders to gain insights into SEB Bank's current state and identify opportunities for improvement. SEB Bank's participation in this research highlights their dedication to staying ahead of industry developments and enhancing organizational performance through data literacy.

The purpose of this study is to develop a data literacy competency framework, coordination and measurement at an organizational level. The goal will be achieved through the accomplishment of the following tasks:

1. Provide an overview of the literature view on data literacy competences, skills
2. Provide insight on current needs in an organization
3. Propose a framework which describes the main data skills to support the development of an organizational culture that embraces the use of data
4. Identify how to coordinate it within organization
5. Identify how to measure it within an organization

The present thesis aims to provide an in-depth analysis of data literacy and its key competencies, as well as to propose a comprehensive framework for its development and evaluation. In order to achieve this goal, the thesis is structured as follows.

Section 2 is aimed at a theoretical review of the literature on data literacy, including definitions, conceptualizations, and related concepts. Drawing on a wide range of sources from various disciplines, the chapter provides a comprehensive overview of the current state of knowledge on the topic. Section 3 outlines the selection of qualitative and quantitative methods, as well as the procedures used to collect and analyze the data. Specifically, the qualitative method involves conducting interviews with a sample of participants, while the quantitative method involves the extraction and analysis of data from LinkedIn job advertisements. The findings of the data analysis are used to identify the key competencies that comprise the proposed data literacy framework. Section 3 also presents the proposed data literacy framework and its competencies in detail. The section explains the significance of each competency, as well as its practical relevance in the contemporary data-driven world. Furthermore, the section provides a discussion of how the competencies are coordinated and evaluated, including the methods used to assess the development of data literacy skills. Finally, Section 4 offers a summary of the overall findings, limitations, and potential extensions of the research. The section synthesizes the results of the study, highlighting the most significant contributions and implications for future research.

CERCS research specifications: S190

1. Literature Review

1.1 Definition of Data Literacy

In its widest definition, literacy refers to certain ways of approaching and engaging with reading and writing. Nowadays, this term is also used to outline competency in a particular field. For instance, information literacy, statistics literacy, data literacy, etc. are some of many areas. When it comes to the topic of data literacy, there are a plethora of definitions that exist. However, the most recognized definition encompasses the competencies that individuals must possess in order to effectively interact with data. The most known and cited definitions were presented in table 1.

Table 1. Definitions of Data Literacy

The authors	**Description of Data Literacy**
Ridsdale et al. (2015)	"...Data literacy is the ability to collect, manage, evaluate, and apply data, in a critical manner."
Qin & D'Ignazio (2010)	"…Knowledge and skills in collecting, processing, managing, evaluating, and using data for scientific inquiry."
Wolff et al. (2016)	"...Data literacy is the ability to ask and answer real-world questions from large and small data sets through an inquiry process, with consideration of ethical use of data.."
Love (2004)	"…The ability to examine multiple measures and multiple levels of data, to consider the research, and to draw sound inferences."
Carlson et al. (2011)	"…Understanding what data means, including how to read graphs and charts appropriately, draw correct conclusions from data, and recognize when data are being used in misleading or inappropriate ways."

Source: Developed by authors

The authors mainly proposed to define data literacy as the ability to read, work with, analyze, and argue with data. Each of these four abilities can be further broken down into specific skills and competencies:

Reading data: involves understanding what data is, its sources and limitations, and what it represents. It requires the ability to interpret charts, graphs, and other data visualizations to extract meaningful insights.

Working with data: involves acquiring, cleaning, and managing data to ensure it is accurate, complete, and reliable. It requires proficiency in using software tools and techniques to organize and structure data in a way that is usable for analysis.

Analyzing data: involves applying statistical and other analytical methods to data to uncover patterns, trends, and relationships. It requires the ability to filter, sort, aggregate, and compare data to identify key insights and draw meaningful conclusions.

Arguing with data: involves using data to support a larger narrative intended to communicate some message to a particular audience. It requires the ability to identify key insights and present them in a compelling and understandable way, using data visualizations and other communication techniques.

According to a workshop report that developed ODI [Oceans of Data Institute, 2016], building global interest in data literacy requires a dialogue that involves various stakeholders, including individuals, governments, businesses, and educational institutions. Report states that education is key to building data literacy. It is important to promote the importance of data literacy in all educational institutions, from primary schools to universities. Governments play a critical role in promoting data literacy, government leaders can help to emphasize the importance of data literacy, which can lead to funding for initiatives that promote it. A report emphasizes the value of data can help to promote interest in data literacy. It is important to highlight the ways in which data can be used to solve problems, make decisions, and improve outcomes. Building global interest in data literacy requires collaboration across various sectors and disciplines, fostering collaboration can help to bring together stakeholders with different perspectives and expertise to work towards a common goal.

1.2 Data Literacy skills and competences

Data literacy is a critical skill in today's digital world as it enables individuals to understand, work with, and make reasonable decisions based on data. This literature review explores the various perspectives and efforts aimed at defining and promoting data literacy.

What does it mean to be data literate? The question is vital and also challenging. It is very essential to identify the elements of data literacy. If the elements of data literacy are understandable then it will be easy to build the model of it. Since there is no unified definition of data literacy these elements can be changed depending on context nevertheless mainly it consists of skills and competences described below. There are different researches to clarify the skills and competencies of data literacy, some of them mainly focus on technical skills and others more basic and soft skills. Further the implementation of data literacy differs depending on sectors which can be schools, university, public or private organizations. One of the well-known data literacy frameworks proposed by Ridsdale (2015). The developed model by Ridsdale et al. (2015) constitutes a data literacy competencies matrix which is composed of 5 core aspects: Data, Collection, Management, Evaluation and Application (Table 2). Within these 5 core aspects presented 22 competencies which cover various skills, knowledge and tasks. Data manipulation, data ethics, data visualization and various fields of data are considered as data competency and cleaning data, identifying useful data, working with data in an ethical manner are considered as skills/tasks.

Table 2. Data Literacy Framework

Conceptual framework	Data Collection	
Introduction to Data	Data Discovery and Collection	Evaluating and ensuring Quality if Data and Sources

Data Management					
Data Organization	Data Manipulation	Data Conversion (from format to format)	Metadata creation and use	Data curation, Security, Re-use	Data preservation

Data Evaluation						
Data Tools	Basic Data Analysis	Data Interpretation (Understanding Data)	Identifying Problems Using Data	Data visualization	Presenting Data (verbally)	Data Driven Decision Making (DDDM) (Making decisions based on data)

Data Application					
Critical Thinking	Data Culture	Data Ethics	Data Citation	Data Sharing	Evaluating Decisions Based on Data

Legend:	Conceptual competencies	Core Competencies	Advanced Competencies

Source: Ridsdale (2015)

Grillberg and Romeike (2018) developed a data literacy competency model which consists of 2 main parts-content areas which addressed competencies and process areas which emphasize the practical skills. To avoid overlaps and recombination the author introduced the main 4 content areas (data and information, data storage and access, data analysis, data ethics and protection) and described them. Those content areas represent different aspects of computer science. The process areas are based on the data life cycle (gathering, modeling and cleansing; implementing and optimizing; analyzing, visualizing and interpreting; sharing, archiving and erasing). At this point mainly given the key skills that needed how people come across with data and how to handle it. Each process area has a link to all content areas and vice versa. As a result, the content and process areas are combined. Content areas mainly included theoretical background and scientific notions whereas processes indicate practical skills that how individuals can communicate with data, how to manage and analyze it.

The emphasis of Wolff et al. (2016) is on developing data literacy for a data-driven society. According to the authors, data literacy should be viewed as a lifelong learning process that is

interwoven into all facets of life and education. The author suggests that data literacy should be recognized as a 21st-century ability, alongside traditional literacy, and numeracy, in order to prepare the next generation for a data-rich future. The article highlights the importance of data literacy for people, organizations, and society as a whole in order for people to make informed decisions, for businesses to make data-driven decisions, and for society as a whole to hold corporations accountable for their data practices In order to elucidate the significance of data literacy, the author draws on several research and papers that are cited to support her claims. The article also looks at obstacles to teaching data literacy, such as a lack of funding and the difficulty of creating a thorough curriculum. Wolff et al. (2016) mentions data literacy can be taught using the PPDAC (Plan, Prepare, Data, Analyze, and Communicate) paradigm. The PPDAC framework offers an organized method for dealing with data and is intended to aid people in understanding and analyzing data in a methodical and efficient way. Plan, Prepare, Data, Analyze, and Communicate are the five phases of the PPDAC framework. Individuals decide their data analysis's goal and desired outcomes in the first phase, Plan. They collect and sanitize the required data in the second stage, Prepare. The third phase, Data, is where they explain and condense the information they have gathered. To address the queries, they raised in the first phase, they undertake additional in-depth studies, such as statistical tests, in the fourth step, Analyze. They convey their findings in a clear and succinct manner in the last phase, Communicate, emphasizing the most important discoveries and their implications.

The author Sternkopf and Mueller (2018) have developed a maturity model for data literacy in non-governmental organizations, which considers 11 competencies at both the organizational and individual levels. The first two competencies refer to the organizational level and the rest refer to the individual level. These competencies include: (1) data culture, (2) data ethics and security, (3) the ability to ask questions and define data, (4) finding data, (5) acquiring data, (6) verifying data, (7) cleaning data, (8) analyzing data, (9) visualizing data, (10) communicating data insights, and (11) assessing and interpreting data. The data literacy model has four stages, including (1) uncertainty, (2) enlightenment, (3) certainty, and (4) data fluency, due to the difficulty of increasing the number of levels. The model targets organizations that have very little or no experience with data to organizations that have established a data-informed culture.

Another definition and approach are given by Mandinach and Gummer (2013). They mainly emphasize data literacy as supporting data-driven decision making in the education field and identify certain competencies. According to the authors, data literacy is composed of a particular skill set and knowledge base that enables educators to transform data into information and ultimately into actionable knowledge. These skills include knowing how to identify, collect, organize, analyze, summarize, and prioritize data and also how to develop hypotheses, identify problems, interpret data, set up plans, implement and monitor courses of action [Mandinach and Gummer (2013)]. Educators need to develop these skills. The authors also mentioned that these competences can be extended depending on context. For example, administrative data literacy needs leadership and management competency besides others.

For understanding the core of data information literacy, we examined the standards of ACRL (Association of College and Research Libraries). According to these standards data literate people can [ACRL (2000)]:

1. Determine the extent of information needed.

2. Access needed information efficiently and effectively.

3. Evaluate information and its sources critically.

4. Incorporate selected information into one's knowledge base

5. Use information effectively to accomplish a specific purpose.

6. Understand the economic, legal, and social issues surrounding the use of information, and access and use information ethically and legally.

Considering these standards Carlson et al. (2011) proposed core competencies to expand data information literacy for librarians and graduate students which consists of elements of Table 3. To gather information from students and research professors at four schools, the authors applied a mixed-methods strategy that included surveys and interviews. The outcomes indicated that both groups had limited understanding of and experience with data management techniques, and that further in-depth training courses were required. To fill in the gaps in the current educational programs, comprehensive data management training programs can be created using the identified core competencies as a framework. The study emphasizes the significance of dealing with data information literacy in higher education and highlights the need for more coordinated efforts by librarians, educators, and IT experts.

Table 3. Framework proposed by Carlson et al. (2011)

Data processing and Analysis	Data Curation and Re-use
Data Management and Organization	Data Conversion and Interoperability
Data Preservation	Data Visualization and Representation
Databases and Data formats	Discovery and Acquisition
Ethics, including citation of data	Metadata
Data Quality and Documentation	Culture of Practices

Source: Carlson et al. (2011)

The key pillars of data literacy were explained in the report of "Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data." which developed by Harvard Humanitarian Initiative, MIT Media Lab and Overseas Development Institute (2015) as : *data education*- involves the knowledge and understanding of the basic concepts, terminology, and tools used in data analysis; *data visualization*-refers to the ability to represent and communicate data effectively using visual tools and techniques, understanding how to create charts, graphs, and other visual representations that accurately convey data insights to others; *data modeling*- involves the ability to create and use models to analyze data and make predictions and understanding how to use statistical methods and machine learning algorithms to identify patterns and trends in data; and *data participation*-refers to the ability to effectively collaborate with others in a data-driven environment and understanding how to communicate data insights to others, as well as how to work effectively with data analysts, scientists, and other stakeholders in the data analysis process.

Prado et al. (2018) proposes a framework for developing a data literacy curriculum that includes four core competencies: data management, data analysis, data visualization, and data ethics. These competencies provide learners with the knowledge and skills required to access, analyze, and comprehend data. To design a successful data literacy curriculum, Prado et al. (2018) suggests assessing students' learning needs, selecting the appropriate teaching methods and materials, and integrating the curriculum into the present information literacy program. This curriculum consists of seven modules, each covering different aspects of data literacy. The first module exposes students to the essential competencies and covers the fundamentals of data literacy. The second module instructs students on data administration, emphasizing efficient data collection and storage. Data analysis is covered in the third module, which instructs students on how to analyze and understand data using the right tools and methods. The fourth module instructs students on data visualization and teaches them how to present data in a captivating and understandable manner. In the fifth module, which addresses data ethics, the necessity of responsible data usage is emphasized, as are issues of data ethics. The seventh module gives students the opportunity to employ their data literacy abilities in a final project. The sixth module focuses on applying data literacy skills to real-world circumstances. Work by Prado et al. (2018) emphasizes the value of data literacy skills and provides a framework for including them in information literacy training. By describing the essential skills of data literacy and stressing the role of librarians, this research is an important resource for educators working to prepare students for success in a data-driven future.

The article written by Sapp Nelson (2017) proposes the development of a pilot competency matrix for data management skills as a step towards the creation of systematic data information literacy programs. The matrix is intended to provide a framework for identifying and developing the skills necessary for managing and using data effectively and can be tailored to the specific needs of organizations or programs. Data planning, data collection, data organization, data documentation, data processing and analysis, data preservation, and data sharing were the seven categories that the author selected after conducting a review of the literature. Subsequently, a team of data management experts examined these skills through a survey. Over a certain period of time (undergraduate, graduate, and data steward), competencies that can be applied at the personal level, within a team, and across a research enterprise were prioritized over those that are specific to a particular learning domain. The developed competency matrix served as a beneficial tool for assessing data management competencies and highlighting potential development areas. It can be used by educators, researchers, and librarians to build training courses and seminars that assist people in refining their data management competencies.

1.3 The role of Data Literacy in organizations

As data increasingly serves as a strategic asset, it is crucial in organizations. Data literacy fosters a culture of evidence-based decision-making by empowering employees at all levels to make knowledgeable choices based on data-driven insights. Organizations can utilize the full potential of their data, stimulate innovation, and gain a competitive edge in the market by providing staff with data literacy training. “Impact of Data Democratization and Data Literacy

on Employee Productivity" [Bandari, 2020] highlights significance of these two factors enhancing productivity of employees in organization. Data Democratization refers to the process of making data accessible and available for a wider range of individuals within an organization beyond the traditional data experts. By breaking down the barriers to data access and empowering employees at all levels, organizations can leverage the collective intelligence of their workforce and make informed decisions based on data-driven insights. The study emphasizes that data democratization alone is insufficient for optimizing productivity. It stresses the importance of data literacy, which refers to the ability to understand, analyze, and interpret data effectively. Employees need to possess the necessary skills and knowledge to navigate through the vast amount of available data, identify relevant information, and draw meaningful conclusions. The research suggests that when employees have access to data and possess data literacy skills, several positive outcomes can be observed. First, it enables employees to make data-informed decisions, which can lead to better problem-solving and more efficient workflows. Second, it promotes collaboration and knowledge sharing among teams, as individuals can understand and communicate insights derived from data. Third, it fosters a culture of innovation, where data-driven ideas can be generated and tested.

Another research shows the concepts of data governance, data literacy, and the management of data quality as discussed in the paper conducted by Koltay (2016). The study highlights the significance of these three factors in the effective management and utilization of data within organizations. Data governance refers to the overall management of data assets within an organization. It encompasses the establishment of policies, processes, and procedures to ensure the integrity, availability, and usability of data. Effective data governance ensures that data is managed in a consistent and controlled manner, promoting data quality, privacy, and compliance. The management of data quality is another essential aspect highlighted in the research. Data literacy, on the other hand, pertains to the ability of individuals to understand, analyze, and interpret data effectively. It is crucial for employees at all levels to leverage data as a valuable asset and contribute to the organization's objectives. Data literacy enables employees to understand and utilize data within the parameters set by data governance, enhancing decision-making and productivity.

In conclusion, the research by Koltay highlights the importance of data governance, data literacy, and data quality management for maximizing the value of data assets, driving informed decision-making, and ensuring organizational success. Simultaneously, Bandari emphasizes that data democratization and data literacy are crucial for optimizing employee productivity, enabling access to data and empowering employees with the skills to analyze and apply it effectively, leading to improved decision-making, collaboration, and innovation. Together, these studies emphasize the significance of these factors in harnessing the power of data for competitive advantage and organizational achievement.

1.4 Data Literacy matching skills to the workforce

As known, every human, company, device produces data. The opportunities presented by this data are manifold and can contribute significantly to the growth of a business, including but not limited to reducing operational costs and optimizing advertising, service, product, sale strategies to capture a greater market share. The literature research demonstrates that the

demand of data literacy in the workplace beyond the data experts' occupations of data analyst/scientists. Data literate employees assist moving forward faster in today's enormous business world. Data scientists have a crucial role, however other roles also require DL skills and these competencies are needed within many units of organization. The elements of data literacy become the most valuable competences in the labor market. Nevertheless, data literacy should not be only the required skills of companies' recruitment processes or job advertisement, it also should be part of culture across organizations. The Winterberry Group Report (2018) provides an example of the increasing demand for data literacy within organizations, underscoring the challenges that businesses face in becoming data-centric due to a scarcity of qualified talent; the need to provide training and development opportunities to employees possessing the appropriate skillset.

Pothier and Condon (2019) mentioned that graduates of business faculties must be adequately prepared to meet the demands of the current job market. They mainly focused on the syllabus of business schools and investigated how data literacy education integrated. Authors believe that to fill the lack of data literacy needs to be started from universities and shortage of data literate candidates in the markets is caused by gaps in the curriculum.

Table 4 provided by Skills Online Vacancy Analysis Tool for Europe (OVATE) shows that various occupations demand data literacy skills, indicating the increasing importance of data in various industries and sectors [Leon-Urrutia et al., 2020]. The European Commission (EC) emphasizes taking steps to address the data literacy gap in the region. One of the initiatives promoted by EC is the DEDALUS [Walker et al., 2019] project that aims to bridge the gap between universities and industry by providing a training program for European professionals and higher education students to address data literacy. The project's ultimate aim of providing a training program for European professionals and higher education students that specifically addresses data literacy is crucial in equipping individuals with the skills and knowledge needed to navigate the data landscape. By identifying gaps and areas for improvement in current data literacy education and training programs, the project can design a program that meets the needs of European professionals and higher education students and prepares them for the demands of the job market. The need for data literacy courses is becoming increasingly important for European enterprises. In order to recruit potential employees and business partners who can effectively understand and utilize data, it is essential to identify successful approaches for incorporating data literacy into university programs. It is also important to gain a comprehensive understanding of the unmet data training needs of small to medium businesses in Europe. By investing in data literacy education, European enterprises can stay competitive and adapt to the rapidly changing digital landscape. According to a DEDALUS research report the challenges that small and medium-sized enterprises (SMEs) face regarding data literacy are multifaceted. While they recognize the importance of having at least basic skills in this area, they also face barriers in terms of high training costs and the lack of availability of training programs that meet their needs. This means that many employees must learn on the job, which can be time-consuming and inefficient. At the same time, some companies are concerned that a focus on data literacy may become too burdensome for them, similar to the challenges they have faced with General Data Protection Regulation (GDPR) compliance.

Table 4. Occupations require Data literacy.

Data literacy related skill	**Occupation**	**%Online advertisement for careers containing the selected skills**
Perform data analysis	Financial analysts	50.5
	Life science technicians	43.7
	Database and network professionals	39.9
	Mathematicians, actuaries, and statisticians	35.5
Manage data, information and digital content		
	Pawnbrokers and money-lenders	82.2
	Credit and loan officers	63.9
	Advertisement and marketing professionals	53.7
	Management and organizational analysts	50.2
	Announcers on radio, television and other media	48.8

Source: Leon-Urrutia et al. (2020)

Walker et al. (2019) offers a thorough framework for fostering data literacy in both higher education and the workforce. The authors list the essential skills for data literacy, such as knowing how to handle data, analyze data, and communicate data. Also, they go over the significance of context, including knowing how the usage of data will impact society and ethics.

From higher education to the workforce, the authors provide a paradigm for data literacy. The DEDALUS competence framework divides competencies into two aspects and two levels as follows: Basic Digital skills which are meant to be learned in school or university. Advanced Digital skills, Basic Data Literacy and Advanced Data Literacy which can be learned through job experience. They use Figure 1 to demonstrate how DL is taught in higher education institutions across Europe by faculties through academic fields, which also demonstrates how progressively more computer science disciplines are being provided to students in a variety of university faculties. This shows that data literacy and computational prowess are relevant across a variety of areas.

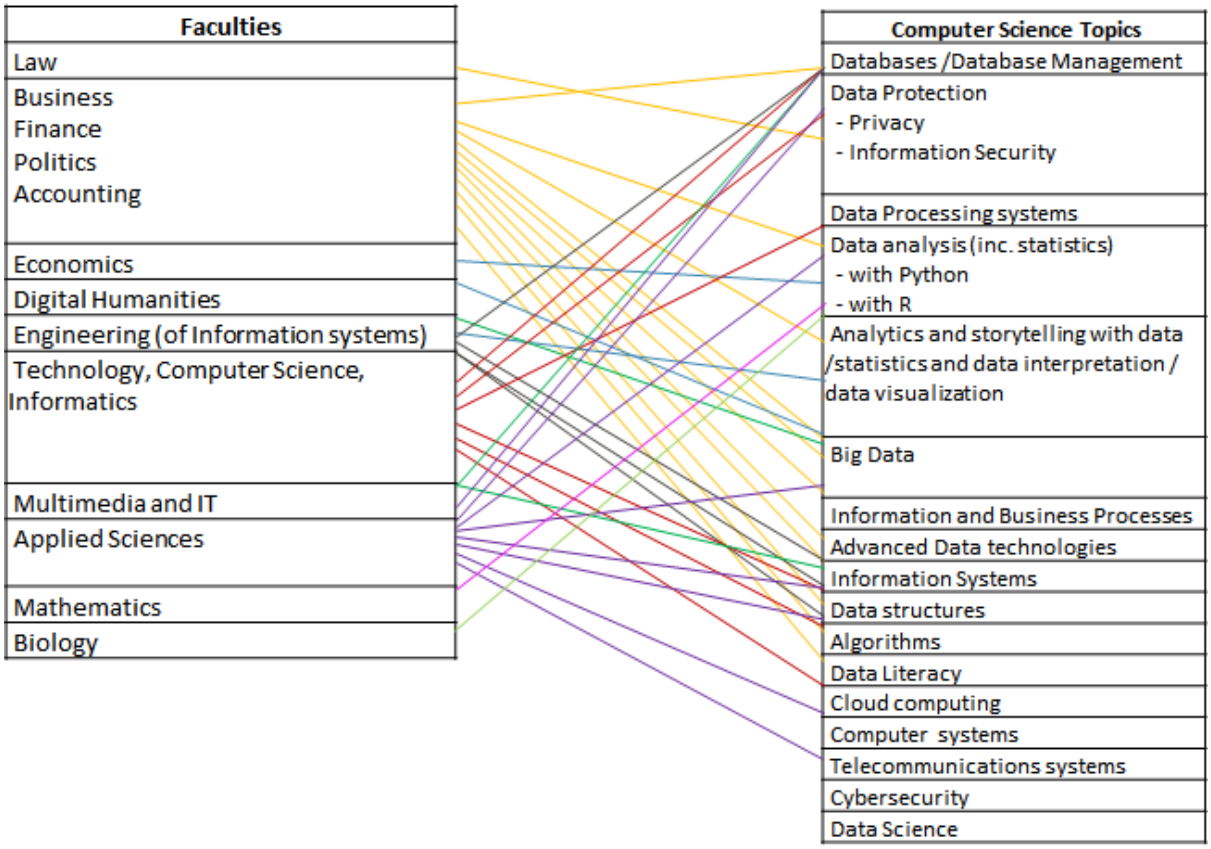

Figure 1. Students in the university faculties on the left and frequently offered modules in the computer science topics linked on the right [Walker et al., 2019]

The researchers then go on to present a thorough study of the five major skill clusters required for data literacy. The capacity to gather, analyze, and understand data is a component of data skills. The term "computational abilities" refers to the capacity to modify and display data using programming and other tools. In order to make wise judgments, critical thinking abilities require the capacity to analyze and comprehend facts. Understanding the context in which the data is being utilized, such as in healthcare or finance, is referred to as domain knowledge. Lastly, good communication abilities are essential for sharing data insights with a non-technical audience.

Although the various frameworks on data literacy were taken different approaches, they share a significant overlap in terms of the competencies they include, as demonstrated in Table 5. These competencies encompass the essential skills required to access data, manipulate it, assess its quality, perform analysis, interpret the results, and, in most cases, utilize data ethically, either at an individual or organizational level. Collectively, the frameworks illustrate that data literacy is a multifaceted concept that involves a range of competencies, which vary in proficiency levels. The complex nature of data literacy underscores the need for a comprehensive approach to training and education to ensure individuals and organizations can effectively navigate data-driven environments.

Table 5. Comparison of papers

	Ridsdale et al. (2015)	Databilities by Data To The People (2018)	Wolff et al. (2016)	Sternkopf and Mueller (2018)	Grillenberger and Romeike (2018)	Carlson et al. (2011)	Mandinach and Gummer (2013)
Data perception	A	N/A	A	N/A	N/A	A	N/A
Data ethics	A	N/A	A	A	A	A	N/A
Data tools	A	N/A	A	N/A	A	N/A	N/A
Data gathering	A	A	A	A	A	A	N/A
Data management and organization	A	A	N/A	N/A	A	A	N/A
Explore data(cleaning, manipulating data)	A	A	A	A	A	A	N/A
Assuring quality of data and sources	A	A	A	A	A	A	N/A
Basic data analysis	A	A	A	A	A	A	N/A
Data visualization	A	A	N/A	A	A	A	N/A
Data storytelling	A	A	N/A	A	N/A	N/A	N/A
Identifying problems using data	A	A	A	A	N/A	N/A	A

	Ridsdale et al. (2015)	Databilities by Data To The People (2018)	Wolff et al. (2016)	Sternkopf and Mueller (2018)	Grillenberger and Romeike (2018)	Carlson et al. (2011)	Mandinach and Gummer (2013)
Data driven decision making (DDDM)	A	A	A	N/A	N/A	N/A	A
Metadata	A	A	N/A	N/A	N/A	A	N/A
Data curation and re-use	A	A	N/A	N/A	A	A	N/A
Data preservation	A	N/A	N/A	N/A	A	A	N/A

Source: Developed by authors

In conclusion, literature review highlights the various competencies that are required to become proficient in working with data. While some researchers, such as Ridsdale et al. (2015), emphasized all these competencies, others had different priorities. For instance, Data To The People (2018) covered most of the competencies but missed out on data ethics, tools, and gathering, while Wolff et al. (2016) overlooked data management, organization, visualization, storytelling, metadata, curation, and preservation. Moreover, Carlson et al. (2011), only mentioned basic data analysis, data visualization, metadata, data curation, and preservation while Mandinach and Gummer (2013) only identified competencies related to problem identification and data-driven decision making. The literature review highlights the wide range of competencies that are necessary for working with data effectively. Different researchers have emphasized different competencies, with some researchers covering all the competencies, while others had a narrower focus. The review demonstrates the importance of having a comprehensive understanding of all the competencies required to work with data effectively. Furthermore, organizations need to ensure that their employees have the necessary skills to work with data and handle any challenges that may arise.

1.5 How to measure Data Literacy

Measuring data literacy is essential for evaluating the competencies of individuals and the state of a company. This assessment can provide valuable insights into areas that require development, the most effective places to focus training, and the progress that has been made. "The Data Literacy Index" [Qlik, 2018] is a research initiative that aims to measure the level of data literacy in organizations and understand its impact on business performance. The initiative was launched by Qlik (2018), a data analytics company, in collaboration with the Wharton School at the University of Pennsylvania. The definition of corporate data literacy

[Qlik, 2018] is given as "the ability of an organization to read, analyze, utilize for decisions, argue with, and communicate data throughout the organization" captures the key elements of what it means to be data literate in a corporate context. It is not limited to scientific or technical organizations, nor is it simply achieved through hiring data specialists.In fact, corporate data literacy involves the entire organization, from senior leaders to front-line employees. It requires a culture that values data-driven decision-making, as well as the necessary tools, training, and resources to support it. To be data literate, organizations must be able to read and interpret data, analyze it to extract insights and make informed decisions, use it to drive innovation and growth, and effectively communicate the results to stakeholders. Additionally, they must be able to engage in constructive debates and challenge assumptions based on data analysis. Corporate data literacy is critical to achieving a competitive advantage in today's data-driven business landscape. The Data Literacy Index consists of three main components:

1. A survey of business decision-makers: This survey measures the level of data literacy in organizations by asking decision-makers about their knowledge of data and their ability to use it to drive business outcomes.
2. An assessment of data-driven culture: This component evaluates the extent to which organizations prioritize and promote data literacy, as well as the availability of data-related resources and support.
3. A financial performance analysis: This component measures the correlation between data literacy and business performance by analyzing financial metrics such as revenue growth and return on assets.

In general, the goal of The Data Literacy Index [Qlik, 2018] is to help organizations understand the importance of data literacy and provide a framework for measuring and improving their data literacy capabilities. By doing so, organizations can leverage the power of data to drive innovation, make better decisions, and improve their bottom line.

Developing and Testing a Comprehensive Data Literacy Model for Business Users by Smolnikova et al. (2021) creates a complete data literacy model for business users. The authors argue that data literacy is essential for business users to make informed decisions and effectively communicate insights and ideas with stakeholders. To do this, they provide a comprehensive model of data literacy that focuses on five competencies: data awareness, data skills, data management, data attitude, and data ethics. Data awareness is the knowledge of the many types and sources of data that are available to business users, as well as its limitations and biases. Having data skills means being able to manipulate and analyze data using a range of tools and techniques. Data management comprises securely arranging and storing data while maintaining data security and quality. The capacity to critically analyze data, think creatively about it, and pose insightful questions are all traits of the data mentality. Last but not least, data ethics entails adhering to ethical data practices as well as being aware of the ethical issues raised by the gathering and use of data. The authors created a prototype of a measuring instrument consisting of four questionnaires intended to evaluate the skills suggested in the model in order to verify the validity and reliability of the proposed model. Participants in these questionnaires rank their level of proficiency on a scale of 1 to 5, which is a self-assessment technique. Although the authors are aware of the possible drawbacks of self-assessment, they

believe that it is a useful method for evaluating data literacy in a commercial setting where more unbiased assessments might not be viable.

One of the most known online self-assessment tools is myDatabilities which developed based on the Ridsdale et al. (2015) framework. Respondents are prompted to select the option that best represents them for each inquiry. The structure is the same for each skill, and the answers are displayed from lowest to highest proficient in each capability which the explanation is given in the table 6. Upon completion of the survey a personalized report is generated that demonstrates to responders exactly where they are succeeding and where they need more assistance. In addition to the online survey that mentioned, the corporation also performs organizational evaluations, which include extra questions specifically suited to the organization, department, and position of the person answering the survey. Moreover, to identify groups who might profit from data literacy skill training, such an evaluation would show where in an organization there are high and low concentrations of data literacy abilities. In the organization being evaluated, the company makes recommendations for the kind of training that would be beneficial to particular groups of people. Data to the People has been hired to carry out data literacy surveys at several businesses throughout the world, including certain government agencies.

Table 6. Assessment spectrum from Level 1 to Level 6 in myDatabilities

Level	**Criteria**
Level 1	An individual can complete simple tasks with instruction
Level 2	An individual can complete simple tasks on their own, with guidance where needed
Level 3	An individual can complete well defined tasks on their own
Level 4	An individual can complete well defined problems and tasks on their own
Level 5	An individual can guide and assist others to complete simple tasks and problems
Level 6	An individual can guide and assist others to complete complex problems

Source: ***myDatabilities***

The Open Data Institute (ODI) [Inverarity et al., 2021], a nonprofit organization with headquarters in the UK, provides consulting and training in the creation and application of open data. It published a survey of government workers all over the world, focusing on the skills participants said they wanted to learn (rather than attempting to gauge individual data literacy skills), the type of training they would rather receive (e.g., face-to-face, online), and the factors influencing the use of open data in the respondent's job. In addition, the ODI created what it

refers to as a skills framework for data literacy that is supposed to help people gauge their data literacy development. The Data Skills Framework demonstrates how the right combination of technical data skills and other competencies, such as service design, data innovation, and change leadership, is necessary to guarantee that data initiatives have a positive social and economic impact on all parties involved.

The evaluation tool for the data literacy maturity model, created by Sternkopf and Mueller (2018), is accessible on the website of the German nonprofit Datenschule [School of Data]. Nine competencies are covered by the tool's questions. Each question has four possible responses that each describe one of four levels of proficiency; the respondent is asked to select one. The author has identified five major evaluation criteria for the model, including: correctness, flexibility, usability, implement ability, and economic efficiency. The evaluation was conducted by using self-assessment questionnaires, test results, and feedback from interviews.

In summary of an extensive examination of the literature, our research has revealed that there exists a wide range of assessment tools that can be used to evaluate data literacy at both the individual and team levels. Depending on the specific objectives of the evaluation and the target audience, either or both approaches can be effectively implemented as a means of measuring data literacy within an organization. It is significant of note that obtaining a suitable evaluation method should be based on a thorough understanding of the organization's particular requirements, objectives, and resources. It is important to undertake an in-depth examination of the organization's data processes and identify any potential gaps or areas for development in order to decide on the most effective strategy of action. In order to maintain the organization's performance and competitiveness, the execution of an effective assessment strategy will ultimately require detailed preparation, constant evaluation, and a dedication to continuous development.

2. Methodology

The present master thesis is a product of a collaborative effort between Tartu University and SEB Bank. SEB (Skandinaviska Enskilda Banken AB) is a Nordic banking and financial services company headquartered in Stockholm, Sweden. It has operations in several Northern European countries, including Sweden, Norway, Denmark, Finland, Estonia, Latvia, and Lithuania. SEB has more than 15 thousand employees working across its various operations. During the initial stages of this thesis, we sought to understand the underlying reasons for SEB Bank's interest in this topic and their specific expectations for this partnership. To accomplish this, we conducted a thorough exploration of the existing literature and interviewed relevant stakeholders to gain insight into the current state of affairs at SEB Bank, including identifying potential gaps and opportunities for improvement in organizational level. As a large and reputable organization, SEB Bank recognizes the immense potential of data literacy and its pivotal role in driving informed decision-making and business success. Their interest in participating in this research underscores their commitment to staying at the forefront of industry developments and continuously improving their organizational performance.

As a methodology for this research two different approaches are selected.

1. *Qualitative:* Semi-structured interviews are conducted with SEB Bank representatives. Semi-structured interviews are a flexible and effective way to collect data directly from individuals, and interviewing SEB Bank representatives can provide valuable insights into the bank's operations and policies.
2. *Quantitative:* Descriptive statistics methods applied to extracted LinkedIn data. As LinkedIn is the largest social network for both job seekers and employers, we gained valuable insights into job requirements in the current job market.

Three representatives of SEB Bank respectively HR, Business and Data Science were interviewed for this study. The interview questions were designed to elicit detailed information about the participants' experiences and perspectives related to data literacy in their respective roles. The interviews were conducted via Microsoft Team, recorded, and transcribed verbatim. Data analysis was conducted using a thematic approach to identify common themes and patterns across the participants' responses. The interview questions were formulated based on the survey questions that were utilized in the DEDALUS [Walker et al., 2019] and synthesized them with our own research problems to create a comprehensive set of questions that would enable us to gather meaningful data. Drawing from the DEDALUS study, we incorporated previously tested and validated survey questions into our interview structure, ensuring a certain level of reliability and validity. Interview questions are given in table 7.

Table 7. Interview questions

Questions	**Developed by**
How would you describe a data literate person?	Thesis Authors
Which Data Literacy (DL) competencies and skills are expected from employees, units-IT, business units, etc.? What about your own unit/ position?	Thesis Authors
What data literacy competencies are being taught in your organization? Do you have any training program or professional development course to boost the data literacy competencies of the staff?	Thesis Authors
How important is the possibility to assess and give evidence to the competence development(certificate)?	DEDALUS [Walker et al. (2019)]
What do you think about how often data literacy evaluation should be carried through a company? What is the frequency of DL evaluation in your company?	Thesis Authors
Are you satisfied with the data literacy competencies of employees?	DEDALUS [Walker et al. (2019)]
In your opinion, who should carry these training and evaluations? HR or a particular unit?	Thesis Authors
Do you think that your company would benefit from employees with stronger data competences? How?	Thesis Authors
What kind of data related competences do you look for when hiring a new employee?	DEDALUS [Walker et al. (2019)]
What other soft competences (personal, social and organizational) a new employee should have?	DEDALUS [Walker et al. (2019)]
Where do you think they should ideally have acquired these competencies? (University, previous experience, cpd training)	DEDALUS [Walker et al. (2019)]
In addition to the data literacy training and work done now, what do you think can be done better? or what are the gaps that have room for improvement? Is there anything else you would like to discuss about training and professional development for data literacy in your company or in general?	Thesis authors

Data scraping was performed by using the LinkedIn API to extract data related to predefined job occupations. The data was retrieved in the form of raw text which was then cleaned and preprocessed using Python programming language. This involves removing duplicates, irrelevant data, and formatting the data in a structured format for analysis. The script that was developed successfully computed the percentage frequency of each predetermined competence pertaining to data literacy within the processed dataset.

The proposed model developed as the result of reviewing and evaluating interview outcomes and analyzing acquired data. The results of this study will provide valuable insights into the current state of data literacy within the organization and inform the development of effective strategies for improving data literacy at all levels.

2.1 Data Life Cycle

To develop a data literacy model, it is needed to understand the data life cycle and key concepts related to it, for instance data management, data analysis, and data science. Each step of the life cycle requires competencies. The illustrated data life cycle model presented in Figure 2 is based on the model proposed by Grillenberger and Romeike (2018), with some minor modifications made by us to enhance its comprehensiveness. While the cleansing and modeling steps were distinct in their model, in ours, we have grouped them together under the processing stage. Additionally, we have introduced interpretation as an extra step, which was not part of their original model.

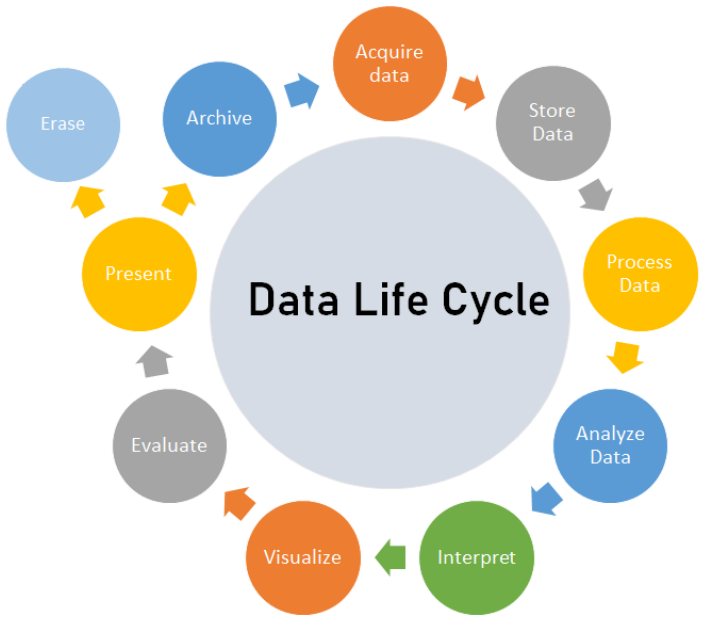

Figure 2. Data Life-Cycle

Source: Developed by authors

The term "acquisition" refers to a certain procedure that may be used to create, gather, or get access to existing data sets in order to gather information. A database, file system, or directory are examples of repositories for persistently storing data sets. In most cases acquired data is unstructured and raw in the source of the system. To analyze data, it is essential to process it

to the structured form which is straightforward to understand and utilize. Finding patterns in data that has been gathered, determining what the shapes could indicate is the process of data analysis. Trying to explain the patterns that were found and driving business insight is the practice of interpreting the data. Visualization is a method of graphical information representation that assists the reader gain rapid insights by showing patterns and trends in data. Evaluation is the practice that covers making strategic decisions based on interpreted data. In accordance with reusability of data it can be archived or destroyed. The data life-cycle steps, as illustrated in Figure 4, are interdependent and demand a unique assortment of expertise and proficiency in each step.

2.2 Analysis

The study involved the extraction of data from LinkedIn for a period of one week, specifically between the dates of 01.04.2023- 08.04.2023 in the EU country context. The data was obtained for four specific positions that regularly interact with data, despite the fact that members of these positions do not possess a predominantly technical background. For each position data set contains roughly 600 job advertisements. The selected positions for the study were Business Analyst, Risk Analyst, Financial Analyst, Marketing Specialist, and Data Analyst- chosen for the purpose of comparison. Despite job searches made using these keywords, the retrieved data consisted of all related job openings, such as Portfolio Analyst and Equity Analyst in the case of a Financial Analyst search result. All of these positions involve making strategic decisions based on data analysis. These positions require analytical and critical thinking skills, as well as a deep understanding of the business context, which makes them suitable for exploring the importance of data literacy in various organizational contexts. Additionally, to ensure a focused and detailed examination of data literacy requirements and competencies within organizations, we chose to limit our analysis to the most common roles, such as financial analyst, risk analyst, marketing specialist, and business analyst.

The primary competencies analyzed in this study were related to Business Intelligence tools such as Tableau, Power BI, SQL, Excel, and Data in general, which encompassed all data-related competencies, including but not limited to Data Management, Data Protection, Big Data, and Data Operation. These competencies are illustrated in Figure 3 and were analyzed in

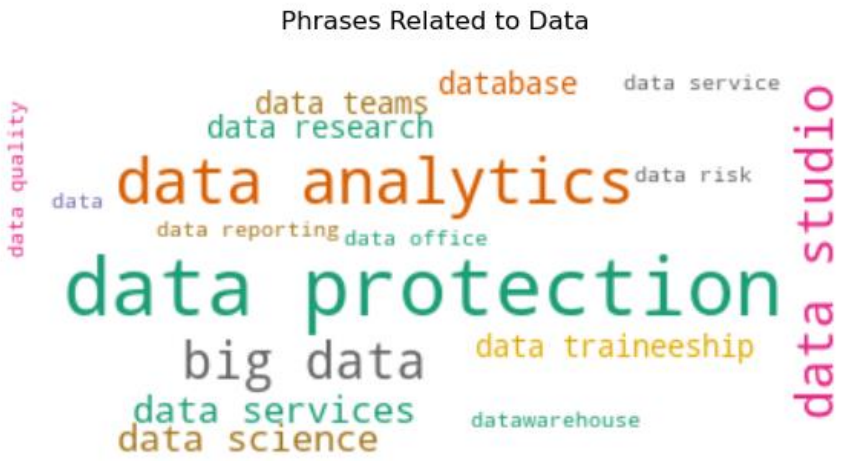

Figure 3. The most used phrases related to Data in job advertisements

Source: Developed by authors

order to gain a comprehensive understanding of the data-related skills required for these particular positions.

Based on the data acquired, it has been determined that the job position with the highest requirement for data literacy is that of Financial Specialist. Specifically, 59.4% of job openings

seeking Financial Specialists require data literacy competencies. Within this category, Excel is the most sought-after competency, with 41.6% of job openings seeking this

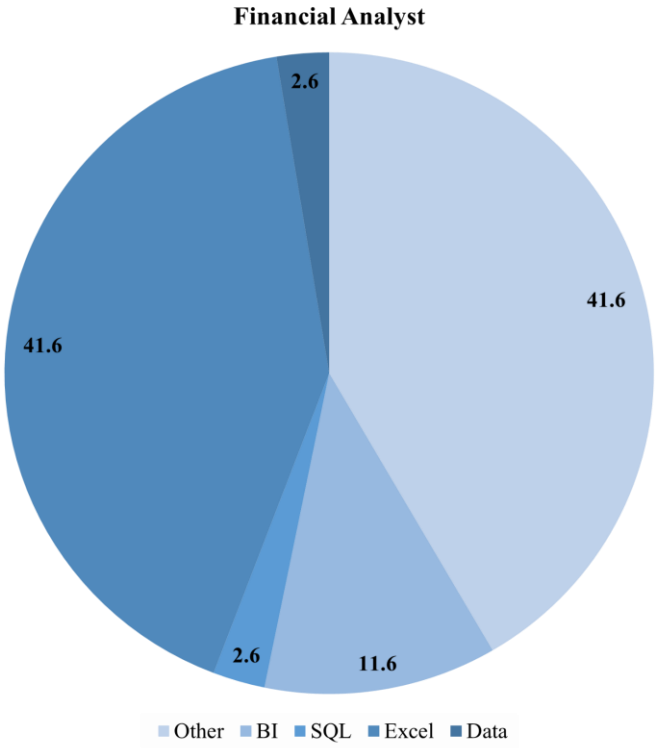

Figure 4. Distributions of competencies for Financial analysts

Source: Developed by authors

skill. Meanwhile, Business Intelligence (BI) tools were required by 11.6% of job openings seeking Financial Specialists, as illustrated in Figure 4.

Meantime, Risk Analysts occupy the second position among the job positions requiring the most data literacy. Based on the analyzed data, half of the job advertisements demand for DL competent candidates. Among these competencies, Excel still remains at the top of the list with 34.5% and followed by Structured Query Language (SQL) with 7.1% (Figure 5).

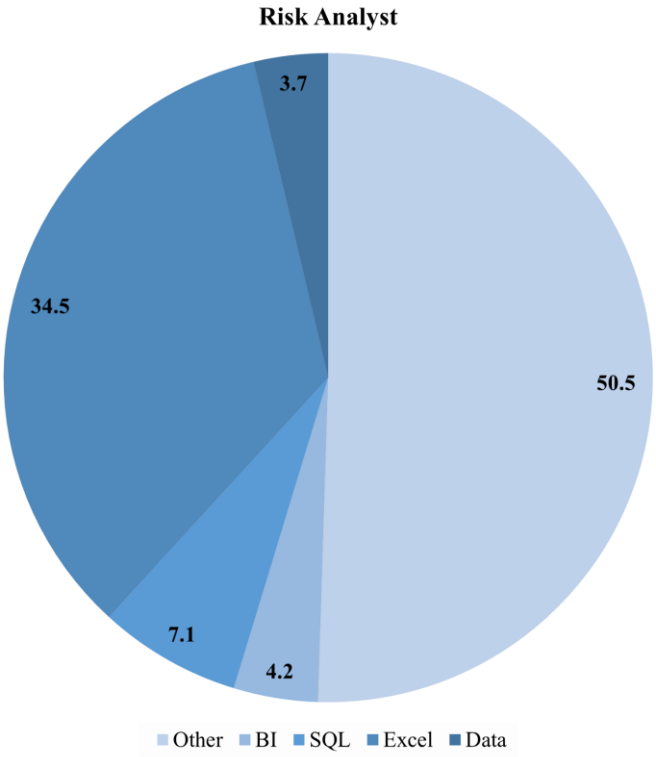

Figure 5. Distributions of competencies for Risk analysts

Source: Developed by authors

Business Analysts are the next position on the list, with a requirement for 32.4% data literacy competency. In this case, the most in-demand skill is related to Business Intelligence (BI) tools, accounting for one-third of all data literacy skills required (11.2%). SQL and Excel follow the list with a demand of 9.9% and 6.1% respectively (Figure 6).

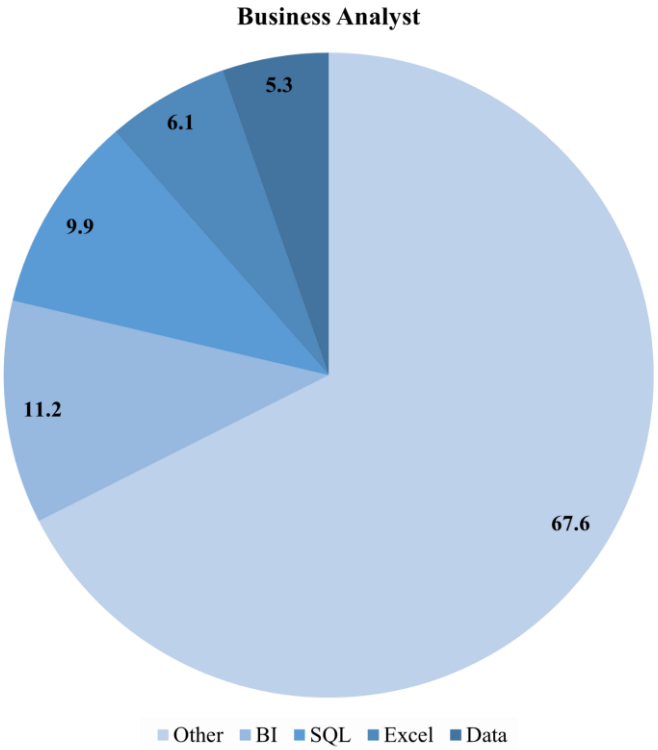

Figure 6. Distributions of competencies for Business analysts

Source: Developed by authors

Finally, the last position on the list is occupied by Marketing Specialists, with only a 15.6% requirement for data literacy. This requirement is distributed among three competencies: Excel, Data, and SQL, with a demand of 13.5%, 1.9%, and 0.2% respectively, as illustrated in Figure 7.

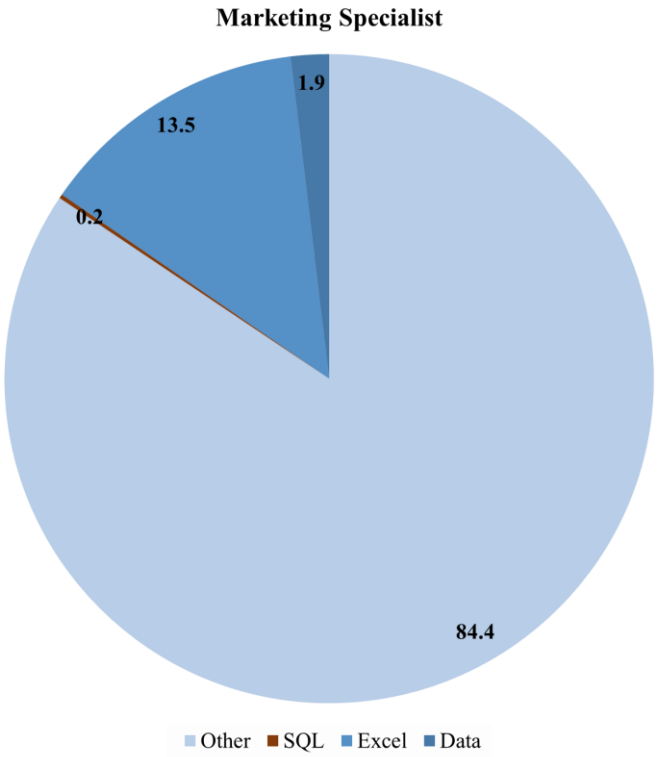

Figure 7. Distributions of competencies for Marketing analysts

Source: Developed by authors

Based on the data acquired, it has been determined that the job position with the highest requirement for data literacy is that of Data Analyst. Specifically, 69.4% of job openings seeking Data Analyst require data literacy competencies, naturally. Within this category, Data contained keywords are the most sought-after competency, with 31.8% of job openings seeking these skills. Meanwhile, Business Intelligence (BI) tools were required by 19.7% of job openings seeking, followed by SQL (11.3%) as illustrated in Figure 8.

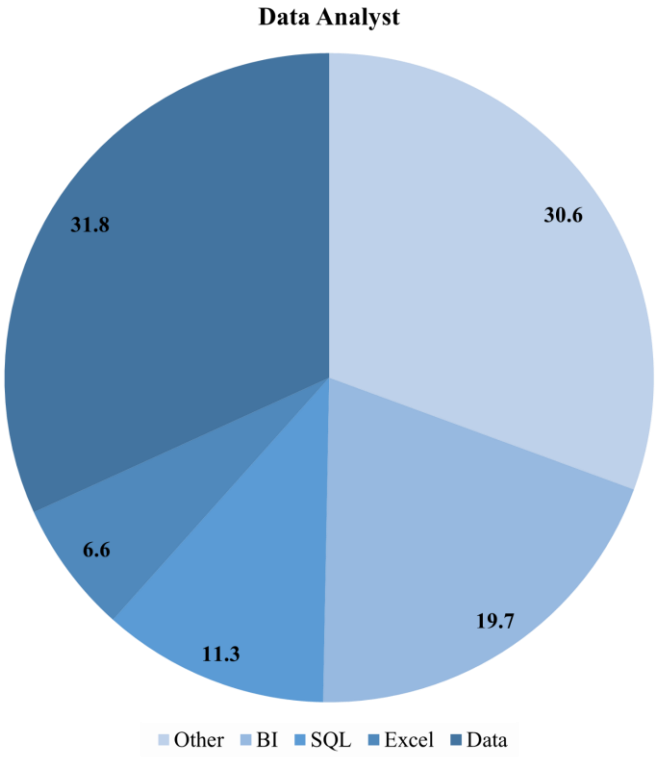

Figure 8. Distributions of competencies for Data analysts

Source: Developed by authors

When comparing the data literacy requirements of the Data Analyst and Business Analyst job positions, several differences become apparent. Firstly, the percentage of data literacy competencies required for Data Analyst positions is significantly higher than that required for Business Analyst positions. Specifically, the Data Analyst positions require a higher percentage of BI (19.7% vs. 11.2%), SQL (11.3% vs. 9.9%), and Data (31.8% vs. 5.3%) competencies.

One possible explanation for these differences could be the nature of the roles themselves. Data Analysts typically deal with large volumes of data and perform advanced data analysis and modeling. As a result, a high level of technical skills is required, including proficiency in BI tools, SQL, and data-related competencies such as data management and data protection. On the other hand, Business Analysts tend to focus on the business side of data analysis, including identifying trends, creating reports, and making strategic decisions based on data insights. While technical skills are still required for Business Analyst positions, the level of technical proficiency may not be as high as that required for Data Analyst positions, resulting in a lower percentage of data literacy competencies required.

When it comes to risk analysts, the data literacy skills requirements are quite different. There is a much higher emphasis on Excel skills, with a staggering 34.5% of job advertisements demanding candidates with proficiency in this area. The reason behind this can be attributed to the fact that risk analysis often involves managing large amounts of data in spreadsheets, where

Excel is a dominant tool. In addition, there is a relatively low emphasis on BI and data skills, as risk analysts often rely on more traditional statistical methods and models for their analysis.

Moving onto financial analysts, there is again a higher emphasis on Excel skills, with 41.6% of job advertisements requiring proficiency in this area. This is not surprising, as financial analysts often work with large amounts of financial data, which is often managed in Excel spreadsheets. Interestingly, there is a relatively low emphasis on SQL skills, which is surprising given that financial analysts often need to access and manipulate data from various sources.

Finally, the marketing specialist position has a much lower overall data literacy skills requirement, with an emphasis on Excel skills, followed by data and other skills. Moreover, marketing specialists often work with data generated from marketing campaigns and digital media platforms, which are often provided in Excel-compatible formats.

In conclusion, the differences in data literacy skills requirements among these positions can be attributed to several factors, including the nature of the work involved, the types of data being managed, and the tools and technologies used for data management and analysis. Additionally, the educational and professional backgrounds of individuals in these positions can also play a role in determining the specific data literacy skills required.

The interview conducted with SEB Bank representatives from the Business, Human Resources, Data science department aimed to understand the views of different units within the organization on the importance of data literacy and training. The HR unit mentioned that there is no mandatory organizational level data literacy training, and self-assessment exists. However, they believe it is essential for business people to know about data to communicate effectively with tech people. Business people are not required to have data literacy skills during the hiring process, but they expect tech background people to possess such skills. The unit agreed that they need more data literacy within the organization. Meanwhile, the business unit agreed that it is crucial for new candidates to understand what type of data is required for different topics and work with basic programs such as Excel and Access. Employees need to understand the underlying meaning behind the numbers to interpret customer behavior, trends, and drive visualization. They think that there should be mandatory data literacy training to some extent, with different levels for different positions. Basic knowledge of DL in the business side can also help communicate better between tech-side stakeholders. Likewise, the data scientist interviewee also stated that they expect better questions and requirements from the business unit side to understand the problem better. They agree that training or a course could be beneficial in asking better questions and requirements for verifying those problems and establishing better communication.

In summary, the framework was developed based on the analysis of LinkedIn data and incorporating the key points mentioned by the interviewees. The analysis of the LinkedIn data revealed that proficiency in spreadsheet programs, including Excel and Google Sheets, as well as an understanding of database components and business intelligence, are critical competencies that must be met. Based on the interviews conducted, it can be concluded that data literate employees are expected to possess requirement gathering and analytical skills. The results of this study highlight the importance of data literacy in today's workplace and the need

for organizations to provide adequate training and resources to ensure that their employees possess the necessary skills to succeed in data-driven environments.

2.3 Data Literacy Framework

After conducting extensive literature review and comparison of selected papers it was concluded that the most comprehensive and known frameworks were developed by Ridsdale et al. (2015), Grillenberger and Romeike (2018), and Carlson et al. (2011). They provide an exhaustive overview of the literature and represent the skills and proficiencies. These frameworks were chosen due to their exhaustive overview of literature and their ability to represent the necessary skills and proficiencies required for data literacy. Ridsdale et al. (2015) developed a framework that encompasses the full data lifecycle, from data collection to dissemination, providing a comprehensive and holistic view of data literacy. Grillenberger and Romeike (2018) focused on the practical aspects of data literacy, including data processing and analysis, making their framework a valuable resource for individuals seeking to improve their data management and analysis skills. Carlson et al. (2011) conducted a study to determine the data literacy needs of both students and research faculty, resulting in a framework that represents the necessary skills and competencies required for effective data use across various disciplines. By combining the insights and frameworks presented by Ridsdale et al. (2015), Grillenberger and Romeike (2018), and Carlson et al. (2011), this chapter aims to provide a comprehensive and integrated view of data literacy. By drawing on the strengths of each framework, we can develop a nuanced and thorough understanding of the necessary skills and competencies required for effective data use. Ultimately, this approach will serve to improve data literacy across various disciplines and support more informed decision-making in both academic and professional contexts. Therefore, this chapter will be guided by the combination of Ridsdale et al. (2015), Grillenberger and Romeike (2018), and Carlson et al. (2011).

It is a well-established fact that a lack of data literacy constitutes a significant challenge for various units outside of the data-related departments. Data is becoming more and more essential in many industries, and individuals across departments need to have a certain level of data literacy to effectively manage and interpret data. It is the main reason why a lack of data literacy is a significant challenge for various units outside of the data-related departments. However, non-data departments often lack the resources and training required to acquire these skills. Considering that this framework is intended to be applicable to individuals from different backgrounds, the competencies outlined therein are not solely representative of those possessing a computer science-based degree. As this framework targets people with different backgrounds, the competency levels do not represent individuals who hold computer science based degrees. Some advanced skills can be core for Data scientist/analysts/engineers. Since the framework is not targeting mainly them, it has been considered as an advanced level of knowledge.

According to report published by Qlik (2018) these are all important points to consider when promoting data literacy:

1. *Agile and adaptive promotion:* It's crucial to recognize that technology and data are constantly evolving, so promoting data literacy should focus on building adaptive capacities and resilience, rather than just teaching specific platforms or technical

languages. This means focusing on developing critical thinking and problem-solving skills that can be applied to different contexts and technologies.

2. *Literacy as a continuum:* Data literacy should be viewed as part of a larger continuum of literacy that includes foundational skills like reading, writing, and numeracy. By building on these foundational skills, individuals can more easily develop data literacy competencies.
3. *Empowering individuals:* Data literacy promotion should focus on empowering individuals to navigate their own ecosystems in ways that are meaningful and effective for them. This involves recognizing and respecting different cultural, social, and economic contexts, and providing support and resources that are tailored to the needs of different communities and individuals.
4. *Multiple pathways:* Promoting data literacy should involve providing multiple pathways for individuals with different needs and capacities to interact within a complex system. *This means recognizing that not everyone needs the same level of data literacy, and providing different options for training, education, and support that can be customized to meet the needs of different groups and individuals.*

When we developed the data literacy framework we took into consideration these 4 main points.

A more data literate organization requires employees to have a clear understanding of how and by whom data is used, what it is used for, and to use data in an ethical and equitable manner to fulfill their goals. If we break down this statement a bit more:

1. *Understanding of how and by whom data is used*: Employees need to understand how data is collected, processed, and analyzed, and who is responsible for these activities within the organization. This includes an understanding of data governance policies, data security protocols, and data sharing practices.
2. *Understanding of what data is used for*: Employees should also have a clear understanding of what the data is being used for and how it is being used to inform business decisions, strategy, and operations. This includes understanding how data is used to measure performance, identify trends, and inform decision-making.
3. *Using data in an ethical manner*: Employees should use data in a responsible and ethical manner, ensuring that data is used in compliance with legal and regulatory requirements, and that privacy and security protocols are followed. Employees should also use data in an equitable manner, considering the potential impact of data use on different stakeholders and taking steps to mitigate any potential negative consequences.

By promoting these requirements and features of a more data literate organization, employees can work together more effectively to leverage data to drive business success while also ensuring that data is used in a responsible and ethical manner.

According to the interview results of the DEDALUS project the most organizations feel more comfortable with the data literacy of their technical and/or senior staff, however less with their junior and generalist staff. Technical/senior employees typically have specialized training and experience in working with data and are therefore better equipped to understand and utilize it effectively. Nevertheless, some junior or generalist staff may not need the same level of

technical expertise as their more specialized colleagues, they still require a minimum level of it to be effective in their roles [Walker et al., 2019].

After conducting an in-depth analysis of various literature, we have created a detailed plan to assist individuals in obtaining data literacy. This plan is comprised of seven main categories and subcategories, where each category is associated with a particular level of expertise, represented by a distinct color. The framework outlines the necessary knowledge and skills required to reach each level, as depicted in Figure 9.

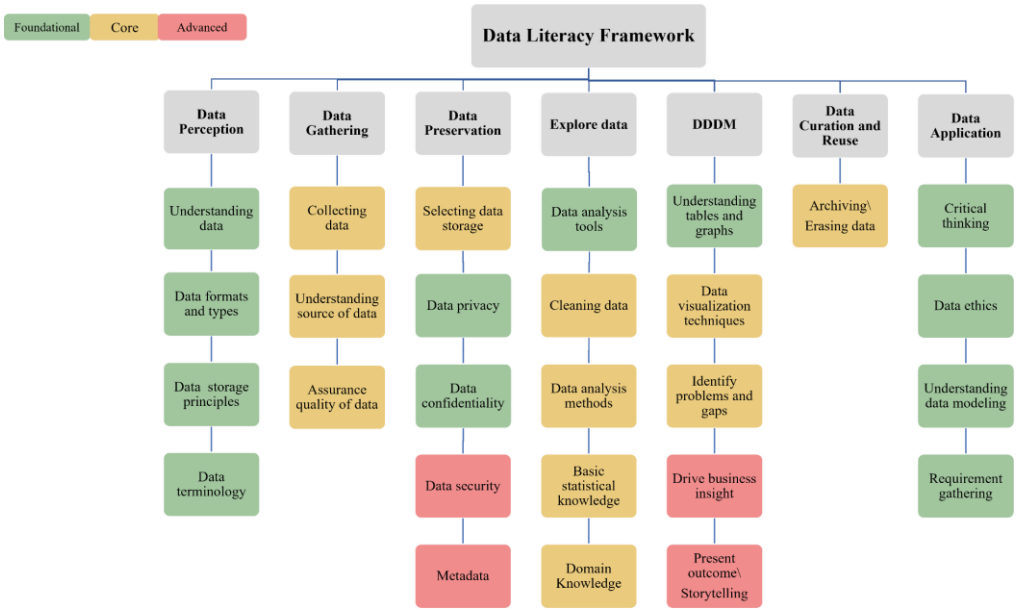

Figure 9. The Data Literacy Framework

Source: Developed by authors

The first and primary level of the framework focuses on the perception and understanding of data and its role in society. This stage requires a solid foundation of knowledge in general concepts such as data formats, types, storage methods, and basic terminology. These building blocks are essential for the development of data literacy.

Before embarking on any data collection or acquisition efforts, it is crucial to have a clear understanding of the purpose behind it. Data can provide answers to numerous questions, but only if it contains the relevant variables and records. Thus, defining the purpose of data collection and identifying the valuable data, its sources, and evaluating their credibility beforehand will steer literate people in their data journey and equip them with the necessary skills for future endeavors. Upon acquiring a basic conception of data, the next step is to understand the principles of data management which includes data gathering and data preservation in our case. This includes sub-groups on data gathering and preservation. In the sub-group on data gathering, candidates learn about various methods for collecting data from different sources. It is imperative to have confidence in the quality and methodology of data collection to ensure accuracy and unbiased analysis in subsequent stages. Data collection can differ from rather fundamental and basic methodologies, such as just carrying surveys to collect

people's favorite food, to complex ones which may demand a need for proficiency in software and technology or a deep understanding of programming and mathematics. Thus, data collection is considered as one of the core competencies. At the initial point of data collection, it is necessary to understand the source of data and how it originated directly affects reliability of data. Besides reliability of sources the main properties of data quality which are data accuracy and completeness must be fulfilled.

The following sub-group encompasses the broad domain of data preservation and demands a specialized understanding of technical concepts related to computer science. The initial stage of the framework is the Data Perception block, which comprises four essential terms to understand data efficiently. Introduction to data identifies basic building blocks of the field. Facts, figures, and other information that may be gathered, saved, and analyzed to draw conclusions or make choices are referred to as data. Data can be of many various forms and formats, including numerical and textual data, organized and unstructured data, and more. Data can be saved in a variety of forms and kinds, including text, picture, audio, and video files. To guarantee effective storage and analysis of diverse types of data, it is crucial to grasp the proper format and type. Besides, to guarantee that data is not destroyed and can be retrieved when necessary, a variety of data storage guidelines, including redundancy, backups, and disaster recovery plans, should be followed. Data sets, variables, and observations are just a few of the numerous terminologies that are used in the subject of data. To deal with data efficiently, it's crucial to comprehend these words.

The scope of Data gathering and preservation is wide and even in some literature there is another section which is called Data Management Literacy [Ridsdale et al., 2015]. It is linked to the first and second steps of the data life cycle. Starting with *Data Collection*: Data can be gathered from a variety of sources, including surveys, experiments, and data mining. Making sure data is gathered in a trustworthy and methodical manner is crucial. Data collection techniques can be selected depending on the research topics, data quantity and target population. One of the biggest challenges is to determine how to measure the data quality. Answering this question requires understanding of data quality metrics [Pipino, 2020]. The main data quality metrics are accuracy, completeness, consistency, timeliness, validity, and uniqueness, however there are plenty of metrics that can be applied to data depending on specific use [Gabr et al., 2021]. In "Data Quality Dimensions, Metrics, and Improvement Techniques" [Gabr et al., 2021] 63 data quality metrics and handling techniques were presented. Knowing the source of the data is essential to determining its accuracy and dependability. Data of organization is generated as a result of business processes. If critical data are flawed or missing undesired results are very likely to occur. Therefore, data quality monitoring is essential for the effectiveness of organizations. Assuring data quality, data cleansing, validation, and verification procedures can be used to verify the accuracy, consistency, and completeness of the data.

The third module is Data Preservation. The second phase of the data life cycle is storage. Preservation of data means that storing data in intact form. The difficulty of effectively storing and managing organizational data is one that businesses face in consideration of the expansion of digital data and the growing demand for data-driven decision-making. Important changes have occurred in the data storage industry, including the increase of cloud storage, big data, and the demand for real-time data analytics. While selecting the optimal data storage solution

could be challenging, it is crucial to guarantee that the data of your corporation is secure and easy to access. Selecting the right data storage depends on several factors such as the size of organization, the level of security, performance speed, budget, etc. A key factor to take into account when selecting a data storage solution for your business is security. To prevent unauthorized access, it is crucial to employ a data storage solution with strong authentication and encryption protection. The protection of data from unauthorized access, use, disclosure, modification, or destruction is referred to as data security. In order to prevent data breaches and preserve the integrity and availability of data, this requires the implementation of security measures such as encryption, firewalls, and access controls. Diego Lopez, a senior researcher at the networking department of Telefónica I+D in Spain, has written extensively on the topic of data security [Lopez, 2016]. In his work, he emphasizes the importance of data security as a critical aspect of information technology, as well as the need for strong security measures to protect against data breaches and cyberattacks. He stresses that data security is a complex and multifaceted issue that requires a range of technical, organizational, and policy measures to be effective. He highlights the importance of implementing strong encryption protocols, firewalls, and access controls to protect against unauthorized access, as well as the need for regular vulnerability assessments and security audits to identify and address potential weaknesses in security systems. Additionally, Lopez emphasizes the role of organizational policies and procedures in promoting data security, such as establishing clear guidelines for data handling and access, ensuring employee training and awareness, and implementing incident response and recovery plans in the event of a security breach.

Protecting sensitive information in the digital world requires both data security and data privacy, two ideas that are closely tied to one another. Data privacy refers to the proper handling and use of personal data, as well as the right of individuals to control their own personal information. Effective data privacy measures, such as transparency about data collection and use, data minimization, and obtaining user consent for data use, help to ensure that personal data is collected, processed, and used in a responsible and ethical manner. Since the General Data Protection Regulation (GDPR) became law on May 25, 2018, maintaining privacy and understanding how personal data is handled have become fundamental rights, at least inside the European Union. The paper “Data Privacy” developed by Flores-Román et al. (2020) provides an overview of data privacy, which they define as the ability of individuals to control their personal information, including how it is collected, used, shared, and stored. They emphasize that data privacy is becoming increasingly important in the digital age, as personal information is being collected and used by individuals, organizations, and governments in new and complex ways. Authors highlight several key principles of data privacy, including transparency, data minimization, and user control. It has been stressed the importance of informing individuals about data collection and use practices, limiting the collection of unnecessary data, and giving individuals the ability to control how their personal data is used and shared. Additionally, authors discuss the regulatory landscape of data privacy, mentioning key laws and regulations such as the GDPR in the European Union and the California Consumer Privacy Act (CCPA) in the United States. They note that compliance with these regulations is essential for organizations that collect and use personal data. Another paper “The role of data privacy in marketing” written by Martin and Murphy (2017) focusing on the importance of protecting consumer privacy while still leveraging consumer data to create effective marketing strategies. They note that data privacy is becoming increasingly important in the digital age, as consumer data is being collected and used by marketers to drive

personalized marketing campaigns. However, they also emphasize the need to respect consumer privacy and ensure that data is collected and used in a responsible and ethical manner. Martin and Murphy highlight several key principles of data privacy in marketing, including transparency, choice, and control. They stress the importance of informing consumers about data collection and use practices, giving them the ability to control how their data is used, and obtaining their consent for data collection and use. Additionally, the authors discuss the benefits of adopting a data privacy-focused approach to marketing, including increased consumer trust and loyalty, as well as improved regulatory compliance. They note that by prioritizing data privacy, marketers can build stronger relationships with consumers and create more effective, targeted marketing campaigns.

The next component of data preservation is data confidentiality. Sometimes data privacy and data confidentiality are confused. Data confidentiality refers to the protection of sensitive data from unauthorized access, use, disclosure, and modification. It is concerned with ensuring that only authorized individuals or systems can access sensitive information, and that the information remains confidential and secure. In other words, data privacy is focused on protecting personal data, while data confidentiality is focused on protecting sensitive data. While they share some common goals and principles, such as the need for access controls and encryption, they differ in their scope and focus. In “Cybersecurity service level agreements: Understanding government data confidentiality requirements” [Nugraha and Martin, 2020] examined the requirements for cybersecurity service level agreements (SLAs) in the context of government data confidentiality. Specifically, they focus on the need for SLAs to comply with government regulations and requirements related to data confidentiality and security. The authors discuss the importance of SLAs in ensuring that service providers meet the necessary standards for data confidentiality and security, as well as the need for clear and measurable metrics for assessing compliance. They also mention the potential risks and consequences associated with non-compliance, including data breaches and other security incidents. The article has also emphasized the importance of understanding the specific data confidentiality requirements and regulations that apply to government agencies, and the need for SLAs to be tailored to meet these requirements. They point out that SLAs must be clear, concise, and specific, with well-defined metrics and measurable outcomes to ensure compliance.

In data preservation the role of metadata is essential. Data that offers details about other data is referred to as metadata. In order to make other data easier to understand and manage, it describes or summarizes the other data. The size of a file, the author's name, the file format, and other specifics that assist identify and categorize the data are all examples of the types of information that can be included in metadata. In the article "Metadata as Data Intelligence," [Greenberg, 2016] mentioned how metadata can be used to provide data intelligence for organizations. The authors argue that metadata plays a critical role in enabling organizations to understand and use their data effectively. The article provides an overview of the key concepts and practices related to metadata and data intelligence. They describe several case studies that illustrate how metadata can be used to support data-driven decision-making, improve data quality, and facilitate data sharing and reuse. It was pointed out that the challenges and opportunities associated with metadata and data intelligence and the importance of metadata standards and interoperability, as well as the need for metadata management policies and practices.

The fourth module is data exploration which mainly covers processing and analyzing phases of data life cycle. After acquiring data from reliable sources the next step is to derive meaningful information from it. Data analysis is the process of answering questions by examination and interpretation of data. The main scope of it is to solve business problems and create new opportunities. Analysis can require several steps depending on complexity of data. It is essential to make data free of errors and remove undesired data or merge various datasets. In today's business world there are plenty of data analysis tools. The most known of them are Microsoft Power BI, SAP BO, Qlik, SPSS and Tableau. To find an appropriate one it is needed to consider corporate business demand, who will use it, the tool's data modeling capabilities and price. If it targets the nontechnical users, then selected tools must be more user friendly for analyzing. For better analyzing the user must be knowledgeable regarding data context, basic statistics and methods. According to Hamed (2020) the main data analyzing techniques are categorized in 6 groups: Descriptive, Exploratory, Inferential, Predictive, Explanatory and Mechanistic.

The next module is making decisions according to analyzed data. Data-driven decision making (DDDM) has become increasingly important in organizations. This group mainly covers usage of basic visualization tools and avoiding misleading representation and being knowledgeable about various types of visualization, for example, graphs, maps, charts. The ability to read, understand and construct data visualizations is counted as a core skill in today's job market and very closely related to data literacy. In the paper "A principled way of assessing visualization literacy" [Boy, 2014] the definition of data visualization literacy is given as: *the ability to confidently use a given data visualization to translate questions specified in the data domain into visual queries in the visual domain, as well as interpreting visual patterns in the visual domain as properties in the data domain*. As seen from the definition that visualization not only covers the graphical representation but also covers interpret and evaluate graphical messages, the goal is insight of it. In the article "The effects of visualization on judgment and decision-making: a systematic literature review" written by Eberhard (2023). It was displayed that visualizations can positively affect decision-making by enhancing comprehension, reduction of cognitive load, and increasing engagement; it can lead to more informed, accurate, and efficient decision-making, especially in complex and data-rich contexts. The paper demonstrates that data visualization can increase decision speed and quality, with mixed effects on other factors, such as decision confidence. Identifying problems and gaps in data is also crucial for making informed decisions. In the study "Data Science and its Relationship to Big Data and Data-Driven Decision Making" [Provost and Fawcett, 2013] examined the importance of addressing biases in data and understanding the limitations of the models and algorithms used to make decisions. By driving business insight, data-driven decision making can lead to better outcomes for organizations. In another study "Data Analytics for Supply Chain Management" [Li et al., 2006] it was found that data-driven decision making can lead to improved forecasting accuracy and inventory management, resulting in lower costs and better customer satisfaction. Finally, presenting outcomes through storytelling can assist stakeholders understand the importance of data-driven insights. In the empirical research [Daradkeh, 2021] was discussed the relationship between data storytelling competency and business performance. It analyzes data collected from business analytics practitioners in various organizations to examine the mediating role of decision-making quality in the relationship between data storytelling competency and business performance. The results recommend that data storytelling competency is positively associated with decision-making quality and

business performance. The effective articulation, presentation and communication of insights to the audience at the right time in the right format is essential to guide the decision-making process and influence actions that create value to business performance and also effective storytelling can make data more accessible and understandable to stakeholders [Daradkeh, 2021].

The following module is data curation and re-use which covers the last phases of its life cycle. Organizations store incredibly large volumes of data. Due to advancements in automated data acquisition, the amount of data that is acquired, analyzed, evaluated, and stored has greatly expanded. The process of creating, arranging, and managing data sets so that users seeking information can access and use them is known as data curation. Data curation is an important part of an enterprise data strategy from a business perspective since it ensures that the organization can use its data effectively and adhere to data-related regulatory and security obligations. As last phase data can be archived for future use or be deleted. Data archiving is the process of securing and storing data and storing it in a secure and accessible manner. It is crucial for ensuring the long-term preservation and accessibility of data in addition to enabling re-use. In this case study [Fitzgerald, 2011] explores the use of data archiving tools to preserve archival records in business systems. The study found that implementing a data archiving tool helped to manage the growth of data and ensure the preservation of important records, while also freeing up space in the business system. Additionally, the use of metadata allowed for easier retrieval of archived data when needed. The case study highlights the benefits of using data archiving tools for preserving important records in business systems and suggests that such tools can be a valuable asset in managing data growth and ensuring long-term preservation of important records. The study also underscores the importance of considering the selection and implementation of data archiving tools as part of an organization's overall data management strategy.

The last module of the framework is data application which also covers some soft skills related to data, mainly critical thinking. Critical thinking is an essential competence in the workplace since it allows employees to break down complex problems into smaller, more manageable parts and then analyze those parts to identify potential solutions or areas for improvement. Gathering and analyzing data, identifying patterns and trends, and using logic and reasoning to reach conclusions are all aspects of critical thinking. It involves problem solving, decision making, strategic planning, data analysis and also foster innovation and creativity.

As is known, data science provides efficiency and productivity for organization, however these opportunities are also associated with ethical challenges. A new area of ethics known as "data ethics" studies and assesses moral problems connected to data, mainly how to gather, protect and use personal identifiers. Data ethics are crucial because there needs to be a precise set of rules governing what corporations can and cannot do with the personal information they acquire from customers. The article Ethics-Aware Data Governance (Vision Paper) [Tanca, and Atzeni, 2018 discusses the importance of ethics in data governance. The authors argue that ethical considerations should be an integral part of data governance frameworks to ensure that data is used in a responsible and sustainable manner. It presents a vision for ethics-aware data governance, which includes principles such as transparency, fairness, accountability, and privacy. The authors also discuss the challenges of implementing an ethics-aware data governance framework, such as the lack of clear ethical guidelines and the difficulty of

balancing ethical considerations with business objectives. As a conclusion by emphasizing the need for collaboration between stakeholders, including data owners, data users, policymakers, and the public, to establish ethical guidelines for data governance.

Building business intelligence (BI) and data warehousing systems with considerable business value starts with a solid data model. A data model is a description of the data structures and business requirements that correspond to the demands of the business. When developing applications or transforming data from source to data warehouse the first thing is to gather business requirements from business stakeholders and document them. Data modeling tools also provide visual representation of data models which helps to understand data components and their relationship for analyzing. There are 3 levels of data models which are conceptual, logical and physical. The high-level business view is represented by the conceptual data model. In order to comprehend and identify the business processes within the organization, this type of communication with the business is quite productive. An organized business view of the data needed to support business processes, document business events, and monitor associated performance indicators is provided by the conceptual data model. This model focuses on identifying the data that is used in the business, but not on how that data is processed or how it is physically developed. The most common data model utilized when creating BI applications is logical data modeling. It expands on the requirements placed out by the business group and provides an extra level of specificity that supports both the data requirement and company system-related requirements. Understanding the features and performance limitations of the database system being used is necessary for implementing the physical data model. A relational database is frequently used, so it's important to comprehend how the tables, columns, data types, and relationships between the tables and columns are implemented in the particular relational database product. The data modeling workflow progresses from business requirements to physical implementation of the database. As seen from the description another main point is gathering requirements. The practice of obtaining and documenting information from stakeholders to describe the requirements and limitations of a project or system is known as requirement gathering. This is also mentioned by interviewees that the main misconception between business units and tech stakeholders is caused by lack of requirements. It entails determining a project's main objectives, aims, and expected outcomes as well as the particular features and functionality necessary to achieve them. Depending on context, requirement gathering can be technical or business requirements. Effective requirement gathering is essential to a project's success since it ensures that the system or product meets the needs and expectations of all stakeholders. It also reduces the possibility of scope creep or other problems that could develop if requirements are not precisely established and accepted from the beginning. The article “Systematic Review of Requirement Elicitation Techniques" [Khan et al., 2019] provides a comprehensive overview of various requirement elicitation techniques used in software development projects. The research highlights the importance of effective requirement elicitation in software development projects and the challenges associated with it. The authors discuss various techniques such as interviews, questionnaires, focus groups, prototyping, and observation, and compare their effectiveness in terms of their ability to elicit accurate and complete requirements, user satisfaction, and the overall success of the project. They also identify several factors that can influence the effectiveness of requirement elicitation techniques, such as project size, complexity, stakeholder involvement, and communication channels.

2.4 Coordination

In order to distribute data throughout the corporate culture, individuals must coordinate, collaborate, and engage with one another within the organization. To reduce false beginnings, it is essential to be clear about the issues being solved for as well as the needs and drivers to be addressed with a data literacy program. To implement the data literacy framework, the following steps should be considered:

1. *Assess the current data literacy levels of individuals and the organization*. This can help identify areas where more training and support is needed.
2. *Develop a training program that covers the main points of the data literacy framework.* The training program should be tailored to the needs of the individuals and the organization.
3. *Provide opportunities for hands-on learning and practice.* This can help individuals apply what they have learned in the training program and build their confidence in working with data.
4. *Encourage data sharing and collaboration among individuals and teams.* This can help foster a culture of data-driven decision-making and enable individuals to learn from each other.
5. *Continuously monitor and evaluate the effectiveness of the data literacy framework.* This can help identify areas for improvement and ensure that the training program is meeting the needs of the individuals and the organization.

Moreover, according to the report [Hart et al., 2022] published by Deloitte important points that can be applied by organizations when establishing program for data literacy include:

1. *Engage Senior Leaders with Clear Roles and Expectations*. Clearly defining individual and collective responsibilities for leadership helps support the success of the data literacy program led by the chief data officer, human capital officials, and other relevant personnel.
2. *Clarify Target Competencies and Personas for Actionable Gap Analysis.* To build effective data literacy programs, it's important to assess the current competencies, capacity, and skills of the workforce using data, and identify areas of improvement based on organizational priorities.
3. *Cultivate a Common, Shared Language.* Using relatable language to discuss data within an organization can help increase the use and application of data skills and expertise, resulting in a lasting impact.
4. *Improve Data Accessibility.* To foster a data-informed culture across the organization, data literacy programs must address any unintentional restrictions on data accessibility.
5. *Align Data Governance and Data Literacy.* As organizations evolve their data governance policies and practices, it's important to adjust data literacy programs accordingly, and vice versa. Data governance work products, such as a data catalog, glossary, and data dictionary, serve as foundational components for successful data literacy initiatives, creating critical dependencies between the two efforts.
6. *Encourage the Use of Data in Decision-Making.* As organizations develop data literacy programs and enhance data governance processes, senior leaders are in a position to

emphasize the explicit value and impact of data in decision-making at all levels of the organization.

Based on the interviews conducted with representatives from SEB, it can be inferred that establishing a data literacy culture within an organization requires effective cross-functional collaboration. Furthermore, it is important to appropriately distribute responsibilities among the relevant units to ensure the success of the initiative, all relevant parties should be involved as drivers. According to the interviews, it is recommended that senior members of units who frequently work with data are actively involved in the delivery process of the data literacy training or practice program. In addition, the human resource team should prioritize the proper distribution of relevant materials to ensure the success of the initiative.

2.5 Evaluation

Measuring data literacy is crucial for determining the existing state of the company, the areas that require development the most, the best places to focus on training and support, and the progress that has been accomplished. By implementing a data literacy framework that includes these steps, individuals and organizations can improve their data literacy skills and make better decisions based on data insights. The process of evaluating data literacy entails determining the degree to which people, teams, and the organization as a whole possess the necessary knowledge and abilities. Evaluation should be performed both before and after the implementation of data literacy programs to measure their effectiveness in improving data literacy levels. In order to produce accurate and trustworthy results, the evaluation process should be carried out in a controlled and methodical way using a variety of techniques. Additionally, the evaluation of data literacy needs to be in line with the organization's particular goals and objectives. This necessitates creating relevant measures to assess how well the program's use of data literacy contributes to the desired results. A thorough assessment of data literacy can reveal significant details about the capabilities and limitations of the organization's data-related knowledge and abilities. This can assist in identifying areas for improvement in data governance and management as well as assist in making decisions about where to concentrate future training and development efforts. In the end, data literacy evaluation could assist companies in developing a culture of data-driven decision-making, which will boost their operational effectiveness, consumer satisfaction, and competitiveness.

Self-assessment checklists can be an effective tool for employee evaluation, and there are several reasons why they are advantageous. Firstly, they allow employees to assess their own performance in a comfortable and non-threatening environment, without the pressure of being evaluated by someone else. As a conclusion of conducted interviews, it can be stated that self-assessment tools enable more honest and accurate self-assessment, as employees are more likely to identify their own strengths and weaknesses when they are not worried about how others will perceive them. Moreover, self-assessment checklists can help to identify areas where an employee may need additional support or training, which can be useful for managers to tailor training programs and professional development opportunities to the specific needs of each employee. By providing employees with a clear understanding of their strengths and areas for improvement, self-assessment checklists can motivate them to take ownership of their own learning and development, which can lead to better overall performance. In addition, self-assessment checklists are a quick and easy way to evaluate employee performance and require

less time and effort than formal evaluations. This can be particularly beneficial for organizations with limited resources or time constraints, as they can provide a cost-effective way to gather feedback on employee performance. Overall, while self-assessment checklists can be a useful tool for evaluating data literacy competency, they should be used in conjunction with other evaluation methods to provide a more comprehensive and accurate assessment of an employee's skills. By combining multiple evaluation methods, organizations can ensure that they are effectively evaluating employee data literacy skills, identifying areas for improvement, and providing the necessary training and support to help employees reach their full potential.

In the field of data literacy, there exist specific traits that are widely accepted as indicative of a data-literate employee. These traits can be categorized into two main categories: knowledge and competencies. The proposed checklist outlines these key traits and is based on a framework for assessing the data literacy of individuals within an organization. To properly evaluate data literacy, it is important to use various techniques to support the proposed checklist.

The following check-list developed by authors taking into consideration the proposed data literacy framework.

Self-assessment Checklist for Data Literacy

Data Literate employee possesses following knowledge and skills:

Knowledge:

☐ Understanding data formats and types

☐ Understanding how data is collected from data sources

☐ Understanding data storage principles

☐ Being familiar with basic data terminologies

☐ Understanding data privacy, security, and confidentiality principles

☐ Understanding the advantages and disadvantages of various data visualization techniques

☐ Understanding the audience and tailor the storytelling approach to their needs and preferences

☐ Identifying the entities, attributes, and relationships that need to be represented in the data model

Skills:

☐ Ensure adherence of data quality standards

☐ Ability to create metadata that accurately describes datasets and their attributes

☐ Implement suitable data analysis methods

☐ Visualize analyzed data with an appropriate technique

☐ Examine data sets to find pertinent information, patterns, and trends

☐ Drive business insight from analyzed data

☐ Choose a suitable archiving\deleting method for the specified data

☐ Consider reflection of ethical norms and its impact

☐ Use a range of techniques to elicit requirements

☐ Document and validate requirements in a clear and concise manner

Discussion

The results from this study have provided insight into the main research question of this study, which sought to address how to improve data literacy competencies of employee groups who have not a computer science background throughout organizations.

The findings of this thesis highlight a significant gap in the current policies and training available for employees to develop data literacy competencies. The lack of coordination and mandatory training can lead to suboptimal decision-making processes and misunderstandings of data, ultimately hindering organizational performance. Furthermore, stakeholders expect better questions from business units to understand problems better, emphasizing the need for improved data literacy competencies.

The competencies of our proposed framework are closely aligned with those of Ridsdale et al. (2015), indicating that both frameworks share many similarities in terms of competencies listed in Ridsdale et al. (2015) as well as division to different levels. However, our framework differs from Ridsdale et al. (2015) in terms of its focus area. While Ridsdale et al. (2015) focused on the general population, our framework specifically targets non-technical employees across all levels of an organization. In other words, our framework is designed to be more relevant and applicable to employees who do not possess technical backgrounds. Grillberg and Romeike (2018) take a different approach to data literacy than our proposed framework does. They focus on aspects of data literacy as process and content, whereas our framework does not differentiate between different aspects. However, despite this difference in approach, most of the competencies identified by Grillberg and Romeike (2018) overlap with those identified by Ridsdale et al. (2015). This implies that proposed framework by thesis owners covers all of the skills and competencies that are proposed by Grillberg and Romeike (2018), as well as those identified by Ridsdale et al. (2015). While Carlson et al. (2011) focused on developing data literacy abilities for librarians and graduate students in preparation for the job market, proposed framework by authors takes a specific approach, targeting non-technical employees across all levels of an organization. Despite the difference in target audience, our framework's list of competencies aligns with those identified by Carlson et al. (2011), indicating the relevance and importance of these competencies.

It is worth noting that our proposed framework additionally includes several crucial skills that were identified through a combination of interviews and analysis. These skills: "Data Privacy," "Requirement Gathering," "Data Confidentiality," and "Data Security," were added to the framework to ensure its relevance and comprehensiveness in today's data-driven workplace.

The primary improvement of this thesis is the proposal of a comprehensive framework for data literacy that covers all seven key competencies required for employees with no prior background in data. This framework provides a systematic approach to the development of data literacy competencies and can facilitate the establishment of a data-driven culture within organizations. The recommendations for implementing the framework, including the development of a comprehensive training program and evaluation system, can support organizations in equipping their employees with the necessary skills to make informed decisions based on data.

Overall, this thesis contributes to the field of data literacy by highlighting the critical importance of this skill set in the modern workplace and proposing a comprehensive framework to develop these competencies. By doing so, it was aimed to support organizations in building a data-driven culture and ensuring that their employees possess the necessary skills to succeed in an increasingly data-driven world.

Conclusion

In conclusion, conducted research by thesis authors has highlighted the importance of data literacy knowledge within organizations and the potential impact that the lack of it can have on communication and decision-making processes. The findings have shown that there is currently no common policy or mandatory training in place to ensure that employees possess the necessary data literacy competencies. The only training that is available is self-learning, which can be inadequate, especially for employees with no prior background in data. Furthermore, the lack of training and coordination can result in misunderstandings and misinterpretations of data, leading to ineffective decision-making processes. Stakeholders expect better questions from business units to understand problems better.

The main contribution to this field is the proposal of a framework for data literacy that encompasses all key competencies required for employees with no prior background in data. It has been provided recommendations on how to organize and evaluate the implementation of this framework by authors. By doing so, it was aimed to facilitate the development of a data-driven culture within organizations and ensure that employees are equipped with the necessary skills to make informed decisions based on data. The framework is based on seven key competencies: Data perception, data gathering, data preservation, exploring data, data-driven decision making, data curation and reuse, and data application.

To implement the framework, it is recommended that organizations develop a comprehensive data literacy training program that encompasses all seven key competencies. The program should be designed to meet the specific needs of each employee, taking into account their level of expertise and the specific roles they perform within the organization. We also recommend that organizations establish a system for evaluating the effectiveness of the training program to ensure that employees are developing the necessary skills.

However, there is limitations in conducted research. It was only interviewed one company, which limited our ability to provide a comprehensive mapping of specific roles and required competencies. The conducted research has also focused on the financial sector, and therefore, our findings may not be representative of other industries. Future research can build upon our findings by collecting data from a more diverse range of organizations and sectors. This would enable us to provide a more in-depth understanding of the specific competencies required for different roles and industries.

The conducted research has highlighted the importance of data literacy and proposed a framework for ensuring that employees possess the necessary competencies to effectively interpret and communicate data within organizations. The findings underscore the need for organizations to prioritize data literacy training and coordination to improve communication and decision-making processes. We hope that our research can serve as a starting point for future studies and initiatives aimed at improving data literacy within organizations.

Reference

1. C. Chinien and F. Boutin (2011). Defining essential digital skills in the Canadian workplace: Final Report.
2. Chantel Ridsdale, James Rothwell, Michael Smit, Hossam Ali-Hassan, Michael Bliemel, Dean Irvine, Daniel Kelley, Stan Matwin, and Bradley Wuetherick (2015). Strategies and Best Practices for Data Literacy Education: Knowledge Synthesis Report.
3. V. Logan (2018, September 21), Information as a Second Language: Enabling Data Literacy for Digital Society, Gartner
4. Building Global Interest in Data Literacy: A Dialogue (2016). Oceans of Data Institute
5. A. Grillenberger and R. Romeike (2018). Developing a theoretically founded data literacy competency model. In Proceedings of the 13th Workshop in Primary and Secondary Computing Education, Potsdam, Germany, October 4 to 6, 2018. New York: ACM
6. Wolff, A., D. Gooch, J.J. Cavero Montaner, U. Rashid, and G. Kortuem (2016). Creating an understanding of data literacy for a data-driven society. The Journal of Community Informatics 12 (3): 9–26.
7. Helena Sternkopf and Roland M. Mueller (2018). Doing Good with Data: Development of a Maturity Model for Data Literacy in Non-governmental Organizations.
8. Mandinach, E.B. and Gummer, E.S. (2013), A systemic view of implementing data literacy in educator preparation, Educational Researcher, Vol. 42 No 1, pp. 30-37.
9. Association of College and Research Libraries (ACRL) (2000). Information literacy competency standards for higher education.
10. Carlson J, Fosmire M, Miller C, et al. (2011). Determining data information literacy needs: A study of students and research faculty. portal: Libraries and the Academy 11(2): 629–657.
11. Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data. Data-Pop Alliance White Paper Series. Data-Pop Alliance (Harvard Humanitarian Initiative, MIT Media Lab and Overseas Development Institute) and Internews. September 2015.
12. Javier Calzada Prado (2018). Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents. Journal of Education for Library and Information Science, 59(1), 48-62.
13. M. R. Sapp Nelson (2017). A Pilot Competency Matrix for Data Management Skills: A Step toward the Development of Systematic Data Information Literacy Programs, Journal of eScience Librarianship 6(1)
14. V. Bandari (2020). Impact of Data Democratization and Data Literacy on Employee Productivity. University of South Australia.
15. T. Koltay (2016). Data governance, data literacy and the management of data quality. International Federation of Library Associations and Institutions, 42(4), 303–312.
16. Winterberry Group. (2018, February). The Data-Centric Organization 2018, (A Winterberry Group Research Report).
17. W. G. Pothier, P. B. Condon (2019). Towards data literacy competencies: Business students, workforce needs, and the role of the librarian. Journal of Business & Finance Librarianship, 25(3–4), 123–146.
18. M. Leon-Urrutia, D. Taibi, V. Pospelova, S. Splendore, L. Urbsiene, and U. Marjanovic (2020). Data Literacy: An essential skill for the industry, 18th Int. Sci. Conf. Ind. Syst. Ind. Innov. Digit. Age, pp. 1–6.

19. Johanna Walker, Manuel Leon-Urrutia, Sarah Hewitt, Davide Taibi, Vera Pospelova, Sergio Splendore, Andrea Nelson Mauro, Tim Scholze, Laimute Urbsiene, & Ugljesa Marjanovic (2019). Data Literacy: From Higher Education to the Workplace, DEDALUS, Erasmus+ Programme. Home - DEDALUS European Project (cnr.it)
20. Qlik, (2018). The Data Literacy Index, Available at: https://thedataliteracyproject.org/
21. Smolnikova, M., Potančok, M., & Chalupová, H. (2021). Building and testing a comprehensive data literacy model for business users. Journal of Business Research, 133, 195-208.
22. Pipino, Leo L., Yang W. Lee, and Richard Y. Wang (2002). Data quality assessment, Communications of the ACM 45.4: 211-218.
23. Gabr, Menna Ibrahim, Helmy Yehia Mostafa, and Elzanfaly Doaa Saad (2021). Data Quality Dimensions, Metrics, and Improvement Techniques. Future Computing and Informatics Journal: Vol. 6: Iss. 1, Article 3.
24. D. Lopez (2016). Data security: A critical aspect of information technology. IEEE Communications Magazine, 54(6), 48-54.
25. Flores-Román, L. H. (2020). Data Privacy. Journal of Law, Communications and New Technologies, (22), 1-12.
26. Martin, K. D., Murphy P. E. (2017). The role of data privacy in marketing. Journal of the Academy of Marketing Science, 45(2), 135-155.
27. Nugraha, Y., Martin A. (2020). Cybersecurity service level agreements: Understanding government data confidentiality requirements. Computers & Security, 89, 101660.
28. Greenberg, J., Wu M., and Liu W. (2018). Metadata as Data Intelligence. Journal of Data and Information Science, 3(2), 3-18.
29. Hamed Taherdoost (2020). Different types of data analysis. Data analysis methods and techniques in research projects. International Journal of Academic Research in Management (IJARM) Vol. 9, No. 1, Page: 1-9, ISSN: 2296-1747.
30. Boy J, Rensink RA, Bertini E, Fekete J-D (2014) A principled way of assessing visualization literacy. IEEE Trans Vis Comput Graph 20:1963–1972.
31. K. Eberhard (2023). "The effects of visualization on judgment and decision-making: a systematic literature review". Management Review Quarterly:73, 167–214.
32. F. Provost and T. Fawcett (2013). Data science and its relationship to big data and data-driven decision making. Big Data 1 (1): 51-59.
33. Li, S., Ragu-Nathan, B., Ragu-Nathan, T.S., and Rao, S.S. (2006). Data Analytics for Supply Chain Management.Journal of Business Logistics:107-124.
34. Daradkeh, M. K. (2021). An empirical examination of the relationship between data storytelling competency and business performance: The mediating role of decision-making quality. Journal of Business Research, 133, 106-117.
35. N. Fitzgerald (2011). Using data archiving tools to preserve archival records in business systems – a case study. Journal of Business & Finance Librarianship, 16(1), 1-12.
36. Tanca, L., & Atzeni, P. (2018). Ethics-aware data governance (Vision Paper). ACM Journal on Data and Information Quality, 9(4), 1-10.
37. Khan, S., Dulloo, A. B., & Verma, M. (2019). Systematic review of requirement elicitation techniques. International Journal of Software Engineering and Knowledge Engineering, 29(05), 621-656
38. C. Inverarity, D. Tarrant, E. Forrest & P. Greenwood (2021). Towards Benchmarking Data Literacy. Data Science Journal, 20(1), 26
39. N. Hart, A. Karkera & V. Logan (2022). Data Literacy for the Public Sector: Lessons from Early Pioneers in the U.S. [Report]. Deloitte.
40. mydatabilities.com

Resümee

Andmekirjaoskus organisatsioonides

Tänapäeva maailm loob, analüüsib ja kasutab andmeid, et mõista nende seoseid. Andmepädevuse puudumine on aga ülemaailmne probleem, mis loob olulise takistusi andmepädevuse võimaluste ja selle tegeliku kasutamise vahel. Andmekirjaoskus on ettevõtluses vajalik ja sellest peab saama organisatsioonide "teine keel", kui nad soovivad saada konkurentsieelist. Käesoleva töö eesmärk on välja töötada andmekirjaoskuse pädevusraamistik, koordineerimine ja mõõtmine organisatsiooni tasandil. See saavutatakse analüüsides olemasolevat kirjandust, mis puudutab andmekirjaoskust ja organisatsiooni vajadusi, pakkudes välja raamistiku, mis kirjeldab peamisi andmeoskusi, et toetada andmete kasutamist toetava organisatsioonikultuuri arengut, ning selgitada välja, kuidas seda organisatsiooni sees koordineerida ja mõõta.

Märksõnad: Andmed, Kirjaoskus, pädevus, organisatsioonikultuur.

Non-exclusive licence to reproduce thesis and make thesis public

Zulfiyya Piriyeva, Nazan Hajiyeva herewith grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright, my thesis *Data Literacy in Organizations* supervised by Lennart Kitt and Juliia Trabskaia.

2. I grant the University of Tartu a permit to make the work specified in point 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in points 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Zulfiyya Piriyeva, Nazan Hajiyeva

18/05/2023