BUSINESS INTELLIGENCE

BUSINESS ANALYTICS SKILLS FOR THE FUTURE-PROOF SUPPLY CHAINS

Authors:
Dario Šebalj
Dejan Mirčetić
Michał Adamczak

Dario Šebalj, Dejan Mirčetić, Michał Adamczak

BUSINESS INTELLIGENCE

Poznan 2025

Foreword

In an age defined by rapid digital transformation and data-driven decision-making, the ability to navigate, interpret, and utilize vast volumes of data is no longer optional - it is essential. This book stands out as a timely and valuable resource for students, educators, and professionals who seek to develop or deepen their competencies in business analytics, especially within the context of supply chain and logistics management.

Developed as part of the Erasmus+ co-funded BAS4SC project (Business Analytics Skills for Future-proof Supply Chains), this book addresses the identified gap between academic curricula and industry demands across Europe. It reflects a rigorous research process involving analysis of international study programmes and direct engagement with over a hundred stakeholders from academia and industry. The result is a comprehensive selection of the most critical analytical skills needed for future supply chain professionals.

The book begins with the foundations of understanding and interpreting data, then advances through essential topics including business data analytics, data mining, machine learning, business process management, and information systems. Special attention is given to emerging areas such as e-logistics, GIS, and data ethics - topics that are increasingly shaping the competitive environment of modern supply chains.

We hope this book serves as both a learning tool and a professional guide. Whether you are a student preparing to enter the field or a practitioner looking to adapt and thrive in a dynamic environment, the insights and skills presented herein will empower you to make smarter, faster, and more responsible decisions.

Dario Šebalj

Dejan Mirčetić

Michał Adamczak

Content

INTRODUCTION

This textbook, entitled Business Intelligence, is the second in a series developed as part of the Business Analytics Skills for Future-proof Supply Chains (BAS4SC) project. Several preliminary research activities were conducted to determine this textbook's content. First, a comprehensive study was conducted to analyse the business analytics courses in the curriculums of the different study programmes in European Union, United States, and United Kingdom which deal with logistics & supply chain management. The methodology of the research includes reviews of the publicly available curriculums and descriptions of different study programmes (bachelor or master studies). The focus of the research was on study programmes in the area of business economics and applied sciences. This analysis revealed a gap between the logistics knowledge and business intelligence skills required in the field and those currently offered to students. Through comprehensive interviews and questionnaires with university teaching staff, students, and industry professionals, more than 100 essential business analytics skills were identified. Applying the ABC ranking classification method, 33 of these skills were selected for inclusion in this book, with a strong focus on areas such as data management, business process analysis, machine learning, data visualisation and business information systems. These selected skills and identified needs informed the creation of ten content chapters, each addressing the most crucial skills required in the field.

This book begins by laying the groundwork in Understanding and Interpreting Data, a foundational chapter that explores essential data concepts like data types, quality, and sources. It explains how effective decision-making is rooted in robust data understanding, enabling professionals to extract meaningful insights and avoid biases that can cloud business judgment. Topics such as the data-information-knowledge-wisdom (DIKW) pyramid, as well as methods for identifying trends and correlations in business data, offer a solid base for readers to navigate complex data landscapes.

The second chapter, Business Data Analytics, contextualizes big data within business environments, explaining how organizations can leverage large-scale data analytics to enhance decision-making. By covering the "Five V's" of big data—volume, velocity, variety, veracity,

and value—this chapter helps readers understand how big data analytics supports operational and strategic goals. It includes case studies to illustrate practical applications, demonstrating how insights from big data can lead to competitive advantages in dynamic markets.

In Data Mining and Knowledge Discovery, readers are introduced to the core techniques for extracting valuable patterns from large datasets. This chapter covers data mining concepts, such as clustering, classification, association rule mining, and anomaly detection, which help businesses uncover actionable insights. It highlights the critical steps in the knowledge discovery process, from data cleaning to pattern interpretation, ensuring that insights derived from data mining are both relevant and applicable in real-world decision-making.

The Machine Learning chapter explores the role of machine learning in predictive analytics and business intelligence. It explains the fundamentals of machine learning, including supervised and unsupervised learning, and how these techniques allow computers to identify patterns, make predictions, and adapt based on new data. Covering machine learning applications like recommendation systems, demand forecasting, and anomaly detection, the chapter demonstrates how machine learning algorithms empower organizations to automate decision-making, enhance customer experience, and optimize operations.

Following this, the Business Process Management (BPM) and Process Mining chapter examines the structures and strategies behind efficient business processes. This section outlines how organizations can analyse and refine their processes, leveraging BPM and process mining to align operations with strategic objectives. The chapter includes tools and frameworks for evaluating business processes, explaining how process mining can uncover bottlenecks and inefficiencies that impede performance.

Moving into specialized systems, Information Systems in Logistics offers an overview of critical IT systems, such as ERP (Enterprise Resource Planning), WMS (Warehouse Management Systems), and TMS (Transportation Management Systems). This chapter discusses how these systems integrate various logistics functions, streamline data flow, and improve efficiency in managing resources across supply chains. It emphasizes the importance of these systems in providing real-time data and enabling responsive, data-driven logistics management.

The chapter on E-Logistics explores the transformative impact of digital technologies on logistics processes, emphasizing the seamless integration of material and digital flows.

It begins by defining e-logistics and situating it within the broader concept of e-business, highlighting how digital processes support functions like material acquisition, warehousing, and transportation, especially in the context of e-commerce. The chapter outlines the development of e-logistics from early material planning systems to modern ERP, WMS, and TMS solutions, which enable real-time data sharing across supply chains. With the rapid growth of the digital economy, e-logistics has become essential for organizations aiming to improve competitive positioning and streamline their logistics operations.

Geographic Information Systems (GIS) in Logistics provides insights into spatial analysis and its role in optimizing logistics networks. The chapter covers how GIS applications enable companies to visualize, analyse , and manage geographic data, supporting tasks like route optimization, site selection, and risk assessment. By leveraging GIS, businesses can make more informed decisions that take into account geographic and environmental factors, ultimately improving efficiency and customer service.

The chapter on Data Visualization methods underscores the importance of transforming complex data into visual formats that facilitate understanding and decision-making. This chapter walks readers through effective visualization techniques, from simple charts and graphs to advanced interactive dashboards. It discusses tools and best practices that make data accessible to stakeholders, enabling businesses to communicate insights clearly and drive informed actions.

Adding an ethical perspective, Data Ethics and Information Security addresses the ethical considerations of data use in business contexts. This chapter discusses privacy laws, data protection practices, and the ethical implications of handling personal data, particularly in supply chain analytics. By covering topics such as data security protocols and ethical frameworks, the chapter equips readers to uphold standards of integrity and trust in data-driven environments.

1. UNDERSTANDING AND INTERPRETING DATA

Author: Dario Šebalj

This chapter explains the fundamental concepts of understanding and interpreting data. Data is the foundation of effective decision-making and plays a crucial role in driving organizational success. By gaining proficiency in data analysis and interpretation, individuals (data analysts, managers, business professionals, data enthusiasts, etc.) can acquire the skills necessary to extract valuable insights from the vast ocean of available information. Data provides analysis with a factual foundation. It enables organizations to make decisions that are more objective and grounded in facts by enabling them to go beyond assumptions and intuition. Organizations can use data to find correlations, trends, and patterns that they might miss otherwise.

Finding organizational inefficiencies and areas for improvement is another benefit of data analysis. Organizations can find bottlenecks, streamline workflows, and enhance the general efficiency of business processes by evaluating operational data. Additionally, data analysis makes it easier to assess the success of projects or strategies that have been put into action, which speeds up decision-making and improves planning.

Prior to the data analysis procedure, it is crucial to describe the various types and sources of data, to discuss data modelling, and to emphasize the significance of data quality. This chapter will lay a solid foundation of Business Intelligence, providing the users with the tools to unlock the power of data and make informed decisions.

1.1. Data, information, knowledge, wisdom

When defining the term data, the familiar data-information-knowledge-wisdom pyramid (DIKW) or DIKW hierarchy is frequently used as a starting point. Rowley (2007) states that the hierarchy serves the purposes of identifying and describing the processes involved in the transformation of an entity at a lower level in the hierarchy (such as data) into an entity at a higher level in the hierarchy (such as information), as well as contextualizing data,

information, knowledge, and occasionally wisdom in relation to one another. Each level of the pyramid builds on the level below it, and for data-driven decision making to be effective, all four levels are required (Cotton, 2023). Figure 1.1 shows the DIKW pyramid.
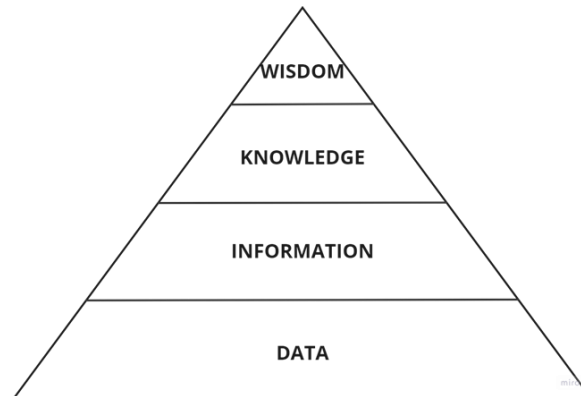
Figure 1.1 DIKW pyramid

Source: Rowley (2007).

Data represents raw material that has no meaning. It is content that is directly observable or verifiable, an unorganized fact that is out of context and is difficult to understand (Brackett, 2015; Dalkir, 2023). Data can be in the form of numbers, text, images, etc. Without interpretation, data will remain meaningless. Example of data: a dataset which contains temperature readings collected from weather stations.

A collection of data in context that is relevant to one or more persons at a given time or for a certain amount of time is called **information.** It is processed, organized, structured, and contextualized data. Answers to common inquiries like "who," "what," "where," and "when" can be found through information. (Brackett, 2015; Cotton, 2023). Example of information: By analysing the temperature data, it can be seen that the average temperature in the past month is higher than in the same period last year.

Cotton (2023) asserts that **knowledge** is the outcome of information analysis and interpretation, which reveals patterns, trends, and connections. It offers insight into "how" and "why" particular occurrences take place. It entails a greater understanding of the fundamental ideas. Chaffey & Wood (2005, cited in Rowley, 2007) define knowledge as "the combination of data and information, to which is added expert opinion, skills, and experience, to result in a valuable asset which can be used to aid decision making". Example

of knowledge: With the knowledge gained from analysing historical temperature data, a meteorologist can predict the weather conditions for the upcoming week.

The capacity to understand the underlying facts and make well-informed decisions and effective actions is known as **wisdom** (Cotton, 2023). Example of wisdom: Using weather forecasts and understanding local climatic conditions, a farmer can make a decision to plant a certain type of crop.

Understanding the interplay between data, information, knowledge, and wisdom forms the fundamental basis for harnessing the potential of Business Intelligence. Moving forward in this book, this understanding will serve as the basis upon which the power of data can be harnessed to transform it into actionable insights, driving organizations toward success in the ever-evolving landscape of the information age.

1.2. Data sources and data types

In the modern era, data is often referred to as the "new oil" – a valuable resource that fuels innovation and decision-making. At the heart of every data-driven endeavor lies a spectrum of data types, each with its unique characteristics and significance. Every day a massive amount of data is produced. Current projections say that there are 97 zettabytes of data in the world and each day more than 2.5 quintillion bytes of data is created. 90% of the data in the world was generated over the last two years (Marr, n.d.).

According to Kenett & Shmueli (2016), „data can arise from different collection instruments: surveys, laboratory tests, field experiments, computer experiments, simulations, web searches, mobile recordings, etc. Data can be primary, collected for the purpose of the study, or secondary, collected for a different reason. Data can be univariate or multivariate, discrete, continuous or mixed. Data can contain semantic unstructured information in the form of text, images, audio and video. Data can have various structures, including cross-sectional data, time series, panel data, networked data, geographic data, etc. Data can include information from a single source or from multiple sources. Data can be of any size and any dimension".

Data is generated more quickly and continuously. New data is generated by all of these and more, and for some reason, it needs to be kept somewhere. Social networking, smartphones, and imaging technology utilized in medical diagnosis are a few examples. Devices and sensors

automatically create diagnostic data, which needs to be analyse d and stored right away. Maintaining this enormous volume of data is difficult enough, but analysing it in order to identify trends and extract valuable information is far more difficult, particularly when the data does not adhere to traditional notions of data structure (EMC Education Services, 2015).

According to Blazquez & Domenech (2018), more and more, technologies related to internet, smartphones and smart sensors are being incorporated into the majority of business and personal daily operations. For example, a lot of businesses use social media to promote their brands, offer products online, use smartphones to track the routes taken by salespeople, or use specialized sensors to record the operation of machinery. On the other hand, people use computers, smartphones, and tablets to browse the internet for products, communicate with friends, share opinions, and find their way around. Additionally, sensors positioned throughout cities, on highways, and in public areas like supermarkets record the daily movements and activities of the citizens. Because of this, a massive amount of newly digitized and fresh data about people's and businesses' activities are being produced by all of these technologies. When properly analyse d, this data may help identify trends and track the economic, industrial, and social behaviours.

Sherman (2015) emphasizes that it can be a problem when an enterprise has more data than it can manage. Through their daily interactions with clients, partners, and suppliers, they gather enormous volumes of data both internally and externally. They conduct market research and keep tabs on information about their rivals. Websites with tracking codes enable them to track the precise number of visitors and their origins. They keep and manage the data needed for industry initiatives and governmental requirements. These days, there's the Internet of Things (IoT), which gathers data from sensors incorporated into real-world items like dog collars, pacemakers, and thermostats. It is a deluge of data. According to EMC Education Services (2015), among the Big Data sources with the fastest rate of growth are social media and genetic sequencing, which are instances of non-traditional data sources being utilized for analysis.

Howson (2014) believes that the success of a business intelligence (BI) initiative is contingent upon the availability of high-quality and relevant data, encompassing a wide range of data sources necessary to inform decision-making processes.

Data comes in various forms. The main data types include (Figure 1.2):

1. **Quantitative (numerical) data** – data which are expressed in numbers. They can be further divided into discrete and continuous data. Discrete data is data that can have only a certain value (e.g. number of employees). Continuous data can have an infinite number of possible values (e.g. price of the product).

2. **Qualitative data** – this type of data can be divided into categorical and ordinal data. Categorical data represent text data which can be grouped into distinct categories (e.g. birthplace, region, product category), while ordinal data can be ranked or ordered (e.g. customer satisfaction – very unsatisfied, unsatisfied, neutral, satisfied, very satisfied; education level – elementary, high school, bachelor's degree, master's degree, PhD).
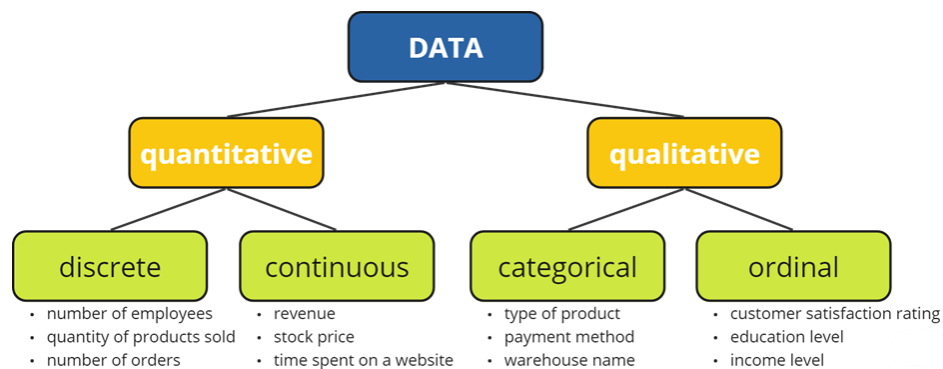
Figure 1.2 Main data types

Source: Author.

Another classification of data is structured, semi-structured and unstructured.

Structured data can be stored, processed and manipulated in a traditional relational database management system. This data comes from a variety of sources, including clickstreams, web-based forms, point-of-sale transactions, sensors, and machines. It can be produced by humans or machines. These data have a predetermined format, kind, and organization (EMC Education Services, 2015; Person & Porway, 2015).

Semi-structured data is organized by tags that help give the data a hierarchy and order even if it doesn't fit into a structured database system. Databases and file systems frequently contain semi-structured data. Log files, HTML-tagged text, XML files, and JSON data files are

some of the possible formats for storage. (McKinsey Global Institute, 2011; Person & Porway, 2015).

Since **unstructured data** is typically produced by human activity and does not fit into a structured database format, it is entirely unstructured. Text documents, PDFs, blog entries, emails, pictures, and videos are examples of this type of data. (EMC Education Services, 2015; Person & Porway, 2015). According to Sherman (2015), unstructured and semi-structured data must be handled differently from traditional structured data.

That vast volume of structured, and especially unstructured data that is generated, collected and processed at high velocity and complexity is called Big Data. McKinsey Global Institute (2011) defines Big Data as "data whose scale, distribution, diversity, and/or timeliness require the use of new technical architectures and analytics to enable insights that unlock new sources of business value". According to Kitchen and McArdle (2016), in 2001, Doug Laney detailed that Big Data were characterized by three traits (Three V's):

- **volume** (consisting of enormous quantities of data),
- **velocity** (created in real-time),
- **variety** (being structured, semi-structured and unstructured).

According to Howson (2014), the foundation for successful business intelligence is the data architecture (see Figure 1.3) which consists of six important aspects regarding data: breadth, timeliness, quality, relevance and granularity. Data quality is the centre pillar because so much effort goes into ensuring and improving data quality.
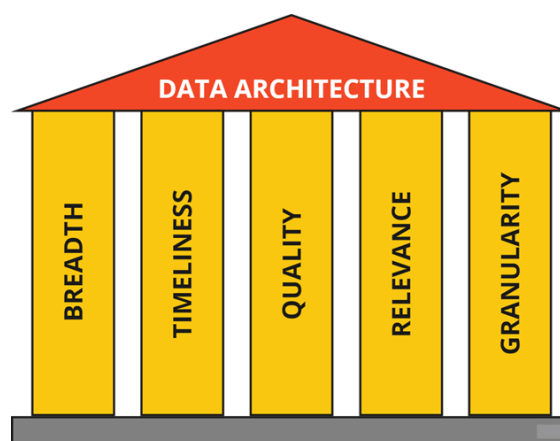
Figure 1.3 Data architecture as a foundation for successful BI

Source: Author, adapted from Howson (2014).

As already mentioned, data can be collected by different sources. This is related to data breadth as one of the pillars of data architecture. It refers to the ability to get multiple data sources which is nowadays, in the era of Big Data, a common way to collect data. On the other hand, combining data from multiple disparate source systems also contributes to data quality problems (Howson, 2014).

In the next sub-chapter the focus will shift from understanding data to harnessing its potential. The relational database management systems (RDBMS) for data storage and retrieval will be explored, along with the visual representation of database structures through Entity-Relationship (ER) diagrams. By connecting our understanding of data sources and data types with data modelling and design, a significant step toward the practical application of Business Intelligence will be taken.

1.3. Data modelling and design

A data model is a formal representation of the data that a business system uses and generates. (Dennis et al., 2018). As already mentioned, structured data is usually stored in a relational database management system (RDBMS). According to Tilley (2020), „a database management system is a collection of tools, features, and interfaces that enables users to add, update, manage, access and analyse data". Some popular RDBMS are Oracle (Oracle), DB2 (IBM) and SQL Server (Microsoft).

In RDBMS, data are organized into tables which contain a collection of records that store information about a particular entity. Tables are represented as two-dimensional structures with vertical columns and horizontal rows. Each column represents a field or attribute of the entity, whereas each row represents a record, which is an instance of the entity (Tilley, 2020). Figure 1.4 shows an example of a table Product.

Figure 1.4 Example of a table in RDBMS

Source: Author.

This table represents a simple product catalogue with 6 products, each having a unique Product ID, Name, Category and Price.

An **attribute** is a particular kind of data about an entity. For example, Customer ID, First Name, Last Name, Address, Postal Code, City, Country, Email Address are the attributes of a Customer entity. A row in a table, or a collection of related fields describing a single instance of an entity (such as a customer), is called a **record**.

Each table in a database must have an attribute which serves as a **primary key**. It is a field (or set of fields) that gives every product in the table a distinct identity. It means that there cannot be two products with the same ID in the table.

The tables in a database are often connected to other tables in the database, i.e. there is a relationship between them. Relationships define how data in one table are related to data in another table.

According to Tilley (2020), three types of relationships can exist between entities: one-to-one, one-to-many and many-to-many. Examples of these relationships are shown in Figure 1.5.
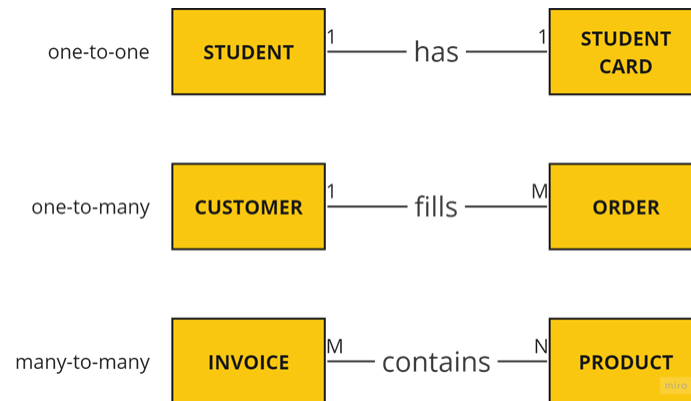
Figure 1.5 Examples of relationships between entities

Source: Author.

The logical structure of a database and the relationships between the tables can be visually represented by the **Entity Relationship Diagram (ERD)**.

There are different notations for creating ERDs, with the most common being the Chen notation and the Martin (Crow's Foot) notation. In this book, a Martin notation will be presented.

According to Martin (Crow's Foot) notation, entity is represented by rectangle. It can be a person, place, event or thing about which data is collected. Attributes are listed as nouns within an entity. Relationships between entities are shown by lines that connect the entities together. Relationships have cardinality which shows how many instances of one entity are associated with an instance of the other (Dennis et al., 2018). In Crow's Foot notation, cardinalities are shown by various symbols. For example, a single bar indicates one, a double bar indicates one and only one, a circle indicates zero and a crow's foot indicates many. Table 1.1 shows various cardinality symbols and their meaning.

Table 1.1 Examples of cardinalities

Symbol	Meaning
	One and only one
	One or many
	Zero or many
	Zero or one

Source: Tilley (2020).

According to Dennis et al. (2018), there are 3 steps in building ERDs:

1. Identify the entities,

2. Add attributes and assign primary keys,

3. Identify relationships.
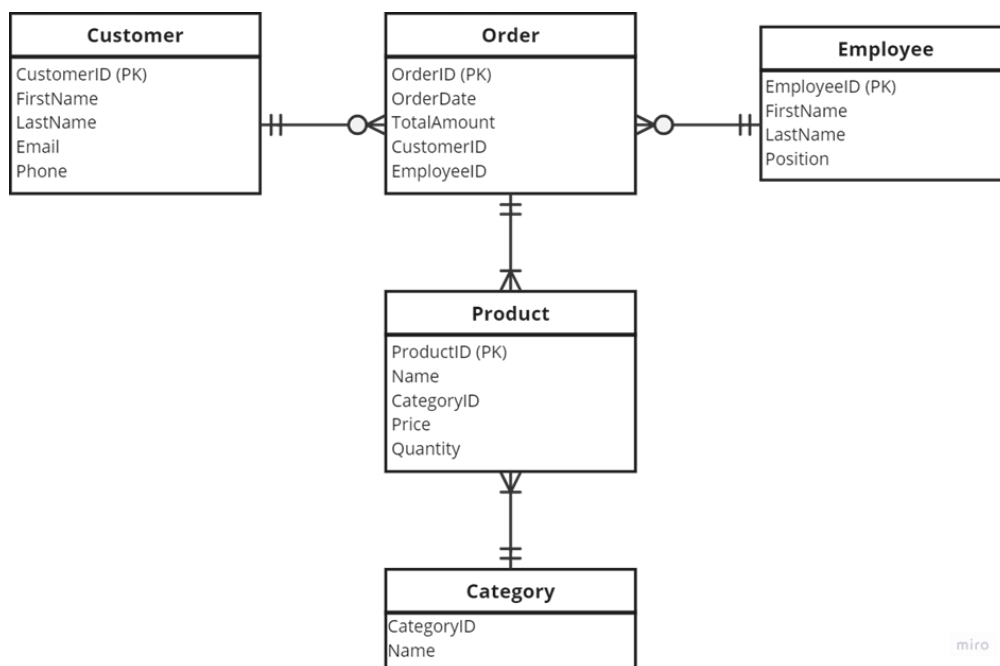
Figure 1.6 Example of sales system ERD

Source: Author.

Figure 1.6 shows part of a sales system ERD. Each entity in this ERD is depicted as a rectangle with a list of its attributes inside. Lines connecting entities are used to show their relationships:

- A Customer can place multiple Orders. Each order has a unique customer attached to it.
- A single order may include more than one product. Every product is linked to a particular purchase.
- An Order is associated with a Customer and an Employee who handled the order.
- A Product is associated with a Category since each product belongs to the single category.

A foundational understanding of data modelling and design, as demonstrated by ER diagrams and RDBMS, is essential for the efficient utilization of data. This comprehension establishes the foundation for shifting towards the pragmatic implementation of data via data-driven decision making, wherein systematically structured data models deliver valuable insights that motivate informed business strategies and decision-making processes.

1.4. Data-driven Decision Making

Data-driven decision-making (DDDM) is a strategic approach that relies on analysing and interpreting data to guide choices and actions. By leveraging insights derived from vast datasets, businesses are able to navigate uncertainty with precision, thereby minimizing risks and maximizing opportunities. Nelson (2022) defines data-driven decision-making as the process of making strategic business decisions that are in line with the aims, objectives, and initiatives of the organization by utilizing facts, metrics, and data. Data-driven decision-making, according to Provost & Fawcett (2013), is the process of making choices that rely more on data analysis than intuition. For instance, a marketer might choose ads only using his extensive industry knowledge and keen sense of what will appeal to consumers. Alternatively, he could base his decision on data analysis showing how customers respond to various advertisements.

In this approach, decisions are not made based on intuition, but rather on hard facts. It involves gathering, analysing , and interpreting data in order to identify patterns, trends, and correlations. Whether it's optimizing operational efficiency, improving customer

experiences, or refining product strategies, data-driven decisions enable businesses to adapt to and thrive in dynamic markets.

By integrating data into the decision-making process, organizations become more adaptable, responsive, and resistant to change, thereby fostering innovation and sustainable growth. A large number of research papers showed that data-driven decision making is associated with increased productivity (e.g. Brynjolfsson & McElheran, 2019; Sala et al., 2022; Colombari et al., 2023). This is the reason why most organizations, especially large ones, are investing in collecting and analysing their data. More than two-thirds of the more than 300 executives surveyed by Bain & Company (2017) say their company invests heavily in data analytics, while more than half anticipate transformational returns on their investments.

There are five steps for making data-driven decisions (Asana, 2022):

1. Understanding company's vision,
2. Finding data sources,
3. Cleaning and organizing data,
4. Performing data analysis,
5. Drawing conclusions.

McKinsey Global Institute (2014) reports that data-driven organizations are 23 times more likely to acquire customers, 6 times as likely to retain customers, and 19 times as likely to be profitable.

The companies with global recognition that make their decisions based on data are Google, Amazon and Netflix.

In order for an organization to realize the complete potential of business intelligence, it is critical to take into account the quality of the data, which is elaborated upon in the following section. This quality ensures the precision and dependability of the conclusions and insights derived from the data.

1.5. Data quality

The degree of accuracy, consistency, reliability, and suitability for a given purpose is referred to as data quality. Data quality is important in the context of business intelligence and data analysis since the conclusions and judgments made from the data mostly depend on its

accuracy and reliability. Poor data quality can lead to incorrect conclusions, flawed strategies, and ultimately, detrimental business outcomes. According to Gartner (2021), every year, poor data quality costs organizations an average $12.9 million.

Data quality can be characterized by the six most commonly used dimensions (Foote, 2022):

- **accuracy:** how accurate are the attribute values in the data?
- **completeness:** is the data complete, without missing information?
- **consistency:** how consistent are the values in and between the databases?
- **timeliness:** how timely is the data?
- **validity:** how data conforms to pre-defined business rules?
- **uniqueness:** is each record uniquely identified, without redundant storage?

According to Sherman (2015), data can be considered high-quality if they have the following characteristics (five Cs of data):

- **clean** – refers to missing items, invalid entries and other similar problems
- **consistent** - uniformity and coherence of data across different sources and within the dataset itself
- **conformed** – it refers to data that adheres to predefined data standards and rules
- **current** - it is essential to use the most current data for decision-making and analysis
- **comprehensive** - it includes all the essential data elements required for the intended decision-making process without omitting critical information.

According to Kenett and Shmueli (2016), almost all data has to be cleaned before it can be used for further analysis. However, the objective determines the degree of cleanliness and the data cleaning strategy. High-quality information for one purpose and low-quality information for another may be found in the same data.

Gartner (2023) has recently identified a set of 12 actions aimed at enhancing data quality. These actions have been classified into four distinct categories, which should be taken into consideration when assessing the integrity of the data:

- focus on the right things in setting strong foundations,
- apply data quality accountability,
- establish "fit for purpose" data quality,
- integrate data quality into corporate culture.

According to Howson (2014), consistent, comprehensive, and accurate data is seen as having a high degree of quality. It is difficult to get good data quality, because organizational and ownership problems have a big impact.

The ability to understand and analyse data is crucial for making informed decisions in the field of business intelligence. Understanding the progression from raw data to actionable insights, encompassing various data sources, types, modelling, and design, is integral to the process of deriving value from information.

REFERENCES

1. Asana (2022). Data-driven decision making: A step-by-step guide [available at: https://asana.com/resources/data-driven-decision-making, access November 5, 2023]

2. Bain & Company (2017). Closing the Results Gap in Advanced Analytics: Lessons from the Front Lines [available at: https://www.bain.com/insights/closing-the-results-gap-in-advanced-analytics-lessons-from-the-front-lines/, access November 5, 2023]

3. Blazquez, D. & Domenech, J. (2018). Big Data sources and methods for social and economic analyses. Technological Forecasting & Social Change, 130, pp. 99-113.

4. Brackett, M. (2015). The Data-Information-Knowledge Cycle. Dataversity [available at: https://www.dataversity.net/the-data-information-knowledge-cycle/, access November 5, 2023]

5. Brynjolfsson, E. & McElheran, K. (2019). Data in Action: Data-Driven Decision Making and Predictive Analytics in U.S. Manufacturing. Rotman School of Management Working Paper No. 3422397 [available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3422397, access November 5, 2023]

6. Chaffey, D. & Wood, S. (2005). Business Information Management: Improving Performance using Information Systems. FT Prentice Hall.

7. Colombari, R., Geuna, A., Helper, S., Martins, R., Paolucci, E., Ricci, R. & Seamans, R. (2023). International Journal of Production Economics, 255.

8. Cotton, R. (2023). The Data-Information-Knowledge-Wisdom Pyramid. Datacamp [available at: https://www.datacamp.com/cheat-sheet/the-data-information-knowledge-wisdom-pyramid, access November 5, 2023]

9. Dalkir, K. (2023). Knowledge Management in Theory and Practice, 4th Edition. MIT Press.

10. Dennis, A., Wixom, B. H. & Roth, R. M. (2018). Systems Analysis and Design, 7th Edition. Wiley.

11. EMC Education Services (2015). Data Science and Big Data Analytics: Discovering, Analysing , Visualizing and Presenting Data. Wiley.

12. Foote, K. D. (2022). Data quality dimensions. Dataversity [available at: https://www.dataversity.net/data-quality-dimensions/, access November 5, 2023]

13. Gartner (2021). How to Improve Your Data Quality [available at: https://www.gartner.com/smarterwithgartner/how-to-improve-your-data-quality, access November 5, 2023]

14. Gartner (2023). Gartner Identifies 12 Actions to Improve Data Quality [available at: https://www.gartner.com/en/newsroom/press-releases/2023-05-22-gartner-identifies-12-actions-to-improve-data-quality, access November 5, 2023]

15. Howson, C. (2014). Successful Business Intelligence: Unlock the Value of BI & Big Data, 2nd Edition. McGraw-Hill Education.

16. Kenett, R. S. & Shmueli, G. (2016). Information Quality: The Potential of Data and Analytics to Generate Knowledge. Wiley.

17. Kitchen, R. & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. Big Data & Society, 3(1).

18. Marr, B. (n.d.). How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read. Bernard Marr & Co. [available at: https://bernardmarr.com/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/, access November 5, 2023]

19. McKinsey Global Institute (2011). Big data: The next frontier for innovation, competition, and productivity. [available at: https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/big%20data%20the%20next%20frontier%20for%20innovation/mgi_big_data_full_report.pdf, access November 5, 2023]

20. McKinsey Global Institute (2014). Five facts: How customer analytics boosts corporate performance [available at: https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/five-facts-how-customer-analytics-boosts-corporate-performance, access November 5, 2023]

21. Nelson, M. (2022). Beyond The Buzzword: What Does Data-Driven Decision-Making Really Mean?. Forbes [available at: https://www.forbes.com/sites/tableau/2022/09/23/beyond-the-buzzword-what-does-data-driven-decision-making-really-mean/?sh=2d35c4eb25d6, access November 5, 2023]

22. Provost, F. & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. Big data, 1(1), pp. 51-59.

23. Rowley, J. (2007). The wisdom hierarchy: representations of the DIKW hierarchy. Journal of Information Science, 33(2), pp. 163–180.

24. Sala, R., Pirola, F., Pezzotta, G. & Cavalieri, S. (2022). Data-Driven Decision Making in Maintenance Service Delivery Process: A Case Study. Applied Sciences, 12(15).

25. Sherman, R. (2015). Business Intelligence Guidebook: From Data Integration to Analytics. Elsevier Inc.

26. Tilley, S. (2020). Systems Analysis and Design, 12th Edition. Cengage.

2. BUSINESS DATA ANALYTICS

Author: Dejan Mirčetić

In the era of digitalization, due to the enormous amount of data generated on a daily basis, traditional knowledge and approaches cannot be used to manage business processes in different areas, so also for manage logistics and supply chains (Nikoličić et al., 2019). Web 2.0, together with Industry 4.0, cloud computing, the Internet of Things (IoT), RFID and other digital technologies have led to generation, storage and transmission of large amounts of data. As the volume and complexity of data increases, so does the complexity and the time required to analyse those data and derive insight from them.

The concept of Big Data was first introduced by Cox and Ellsworth in October 1997, in an ACM digital library article (Tiwari et al., 2018). The study of Big Data and its conceptualization have evolved continuously. Initially, Big Data was characterized by the 3Vs concept, which encompassed **volume**, **velocity**, and **variety**, as discussed in the previous chapter. Subsequently, this characterization expanded to the 5Vs concept, incorporating two additional attributes: **veracity**, and **value** (Nguyen et al., 2018; Tiwari et al., 2018). Volume refers to the magnitude of data generated; the volume of digital data is growing exponentially (Arunachalam et al., 2018). Variety refers to the fact that data can be generated from heterogeneous internal and external sources, in structured, semi-structured, and unstructured formats. Velocity refers to the speed of data generation and delivery, which can be processed in batch, real-time, nearly real-time, or stream- lines. Veracity stresses the importance of data quality because many data sources inherently contain a certain degree of uncertainty and unreliability. Value refers to finding new value contained in the data which can be used for better business planning (Nguyen et al., 2018).

Big Data Analytics (BDA) incorporates two dimensions: **Big Data (BD)** described with the 5Vs concept and **Business Analytics (BA)** which enables to gain insight from data by applying statistics, mathematics, econometrics, simulations, optimizations, or other techniques to help business organizations make better decisions (Wang et al., 2016). Big Data Analytics (BDA) involves the use of advanced analytics techniques to extract valuable knowledge from vast amounts of data with variable types in order to draw conclusion by uncovering hidden patterns

and correlations, trends, and other business valuable information and knowledge, in order to increase business benefits, increase operational efficiency, and explore new market and opportunities (Nguyen et al., 2018; Tiwari et al., 2018). BDA has attracted significant attention in different areas, both academically and business, particularly in logistics and supply chain management.

2.1. BDA in logistics and supply chain management

The supply chains (SCs) represent the network of firms and facilities involved in the transformation of raw materials to final products and distribution of final products to end customers. In SCs, there are physical, financial, and informational flows among different firms. Every day, SCs are becoming more complex, more extended and more global. Therefore, for the successful implementation and management of existing processes in the SC and their continuous alignment with market conditions, modern SC needs highly skilled experts. In order to answer these challenging tasks, SC experts need formal education, which will provide them knowledge and skills from different fields, primarily from logistics, information technology and economics.

The SC is a set of physical elements, their activities and processes through which their interaction takes place. Physical elements, which make up the chain structure, represent a fixed part of the SC. Decisions on the design of the SC structure are made at the strategic level, and at the tactical and operational level, decisions are made on the modalities and rules for the realization of particular logistics processes. Designing a fixed SC and managing an exquisite work together provide a SC management that defines the performance of the chain. Accordingly, the SC management framework consists of three basic elements: (1) SC structures; (2) business processes; (3) and the control components. Each of these elements is directly related to the objectives of the SC, that is, with the degree of fulfilment of the requirements of the end-users, while respecting the critical dimensions of the business that depends on the performance on the market (key performance indicators - KPI). In the modern world competition is no longer between organizations but among SC. Effective SC management has, therefore, become a potentially valuable way of securing competitive advantage and improving organizational performance. SC management is a fact of business, with logistics as a most powerful tool for achieving the ultimate stragegic advantage.

Firms are under heavy pressure to improve SC planning and performance because of factors such as increasing uncertainty and competition. Improving SC performance has become a continuous process that requires an analytical performance measurement system. Considering the number and diversity of logistics processes and SC processes, the resources used for their realization, the parameters that characterize them, as a basis for determining SC performance, a large number of data is used on: geographic, time and quantity determinants of goods, transport means, transport - manipulative assets, warehouse capacities, employees, etc. Data generated through internal operations, as well as transactions with suppliers and customers, can be used to uncover small changes that can make a big impact on an organization with regard to efficiency gains and even cost savings. The other words, the volume of data in every SC is exploding from different data sources, business processes, and IT systems. As the volume and complexity of data increases, so does the complexity and time taken to analyse that data and derive insight from it. Determining, monitoring, and improving logistics and performance SC becomes more complex and involves many processes such as identifying measures, defining targets, planning, communication, monitoring, reporting and feedback. Consequently, conventional approaches cannot be used to make SC decisions and SC management.

In SC management, there is a growing interest in Business Analytics, which is also called **Supply Chain Analytics (SCA)**. SCA is used synonymously with the terms such as 'Big Data Analytics' and 'Business Analytics' within the business and academic communities (Srinivasan and Swink, 2018). SCA refers to the use of data and quantitative tools and techniques to improve operational performance, often indicated by such metrics as order fulfilment and flexibility. Analytics in SCs is not necessarily a new idea since various quantitative techniques and modelling methods have long been used in manufacturing firms. The recent surge of interest in SCA is accompanied by new challenges and opportunities in both business and information technology environments. These challenges include issues arising from managing large amounts of data (e.g. data availability and data quality) and dealing with environmental uncertainties. The properly applied SCA can impact several areas in SC and it can generate significant benefits in logistics performances: improved planning and scheduling; improved responsiveness; improved demand planning; order optimization; optimized inventory management; improved replenishment planning. In recent decades, under the influence of technological development, globalization and increasingly demanding customers, business

paradigms have also changed. Figure 2.1 shows typical periods (with a brief description) in the evolution of logistics, SCM and BDA.
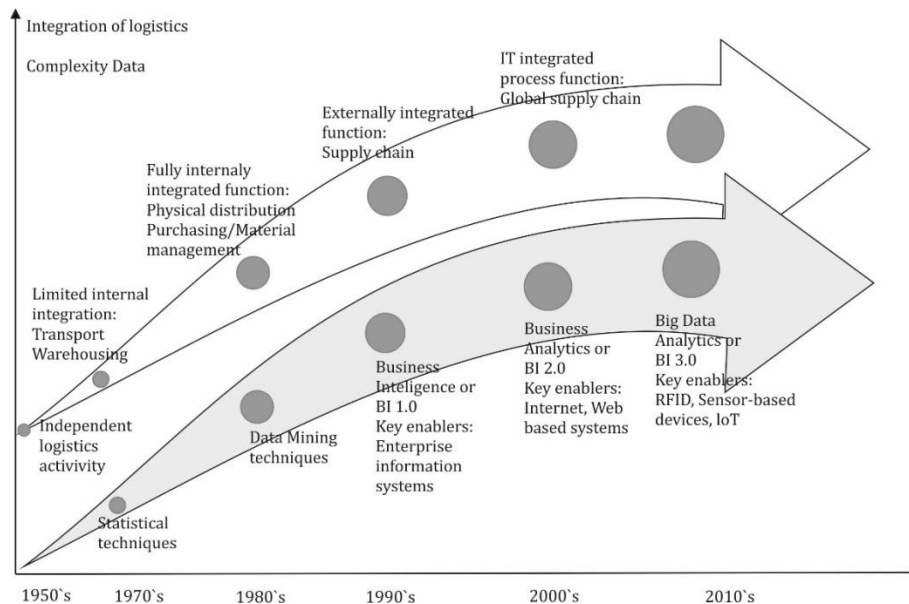
Figure 2.1 Evolution of logistics, SCM, and BDA

Source: Adapted from Arunachalam et al. (2018).

2.2. Tools in business data analytics

Figure 2.2 presents different trends, tools & benefits used in BDA or SCA. All presented analytics techniques can be categorized into three types: descriptive, predictive, and prescriptive. **Descriptive analytics** looks at data and analyse s past events for insight as to how to approach the future. There are looking for the response behind past failures and successes. **Predictive analytics** uses historical data to determine the probable future outcome of an event or the likelihood of a situation occurring. It exploits patterns found in the data to identify future risks and opportunities. **Prescriptive analytics** automatically synthesizes Big Data, business rules, and machine learning to make future predictions. It goes beyond predicting the future by suggesting actions, which needs to be taken in order to achieve desired goals. Also, they are able to demonstrate the implications of each possible decision and act as a decision support tool for SC experts. In the following sub-chapters, we will introduce and describe the various analytical tools used for the BDA in SCM.

Additionally, we will focus on strategies on how to enhance knowledge in BDA for SC experts of the 21st century, via several case studies.
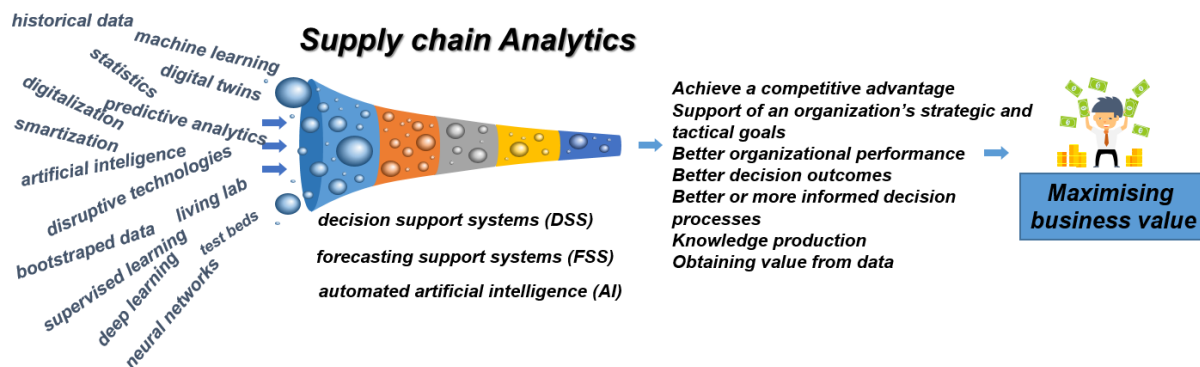
Figure 2.2 Trends, tools & benefits of SCA

Source: Author.

2.2.1. Descriptive analytics

Descriptive analytics provides a summary of descriptive statistics for a given data sample, for example: mean, mode, median, range, histogram and standard deviation. Descriptive analytics describes what happened in the past and derives information from significant amounts of data to answer the question of what is happening. On the basis of real-time information about locations and quantities of goods in the supply chain, managers make decisions at the operational level (e.g. they adjust the schedule of shipments, deploy vehicles, issue orders for restocking products, etc.) (Souza, 2014). It attempts to identify opportunities and problems using online analytical processing system and visualization tools supported by real-time information and reporting technology (e.g. GPS, RFID, transaction bar-code). Common examples of descriptive analytics are reports that provide historical insights regarding the company's production, financials, operations, sales, finance, inventory, and customers (Tiwari et al., 2018).

2.2.2. Predictive analytics

Predictive analytics uses historical data to determine the probable future outcome. Predictive analytics in supply chains derives demand forecasts from past data and answers the questions related to what will be happening or what is likely to happen (Tiwari et al., 2018). It use artificial intelligence, optimisation algorithms and expert systems to predict future behaviours

based on patterns uncovered in the past and the assumption that history will repeat. It exploits patterns found in the data to identify future risks and opportunities and predict the future. This is used to fill in the information that is missing and to explore data patterns using statistics, simulation, and programming.

2.2.3.   Prescriptive analytics

Prescriptive analytics derives decision recommendations based on descriptive and predictive analytics models as well as on mathematical optimization, simulation or multi-criteria decision-making techniques. It goes beyond predicting future outcomes by also suggesting action to benefit from the predictions and showing the decision maker the implications of each decision option. Prescriptive analytics answers the question of what should be happening.

2.3. BDA ecosystem

The main purpose of the BDA ecosystem is to deliver value for the decision-maker. Accordingly, the BDA has a primary goal of providing insight into business processes and leading to the possible answer on how to reduce costs and increase the service level for the final customers. To fulfil its goal, the BDA solutions are usually delivered in the form of Decision Support System (DSS) or Expert System (ES). Therefore, in this and upcoming chapters we will dive into the key pillars of BDA: **business data**, **data mining** and **knowledge discovery** (data analytics, DSS, ES platforms, etc).

Business data

The concept of data is explained in detail in the first chapter of this book. Data is the key factor for conducting any kind of analysis. **Business data** is generated as a result of the execution of processes in a given business environment. In the case of the SC, there are many processes & subprocesses involved in delivering the service to the final customer (Figure 2.3).

Figure 2.3 represents the structure of the SC where in the case of SCA each of the given processes can be observed as a generator of business data. Generated data is different in its importance and influence on the final goals of the company. Accordingly, the business data can be divided into **internally driven** and **externally driven** data. The internally driven data is the data which emerges as a result of company structure, hierarchy and the way the company has decided to operate (for example, manufacturing data, human resource data,

delivery data, accounting data, etc). This data is different for each company and it serves for reporting, analysing and legal reporting to the authorities (accounting data). What is interesting about this data is that the companies are directly controlling and influencing this data and it only has a value for a given company.
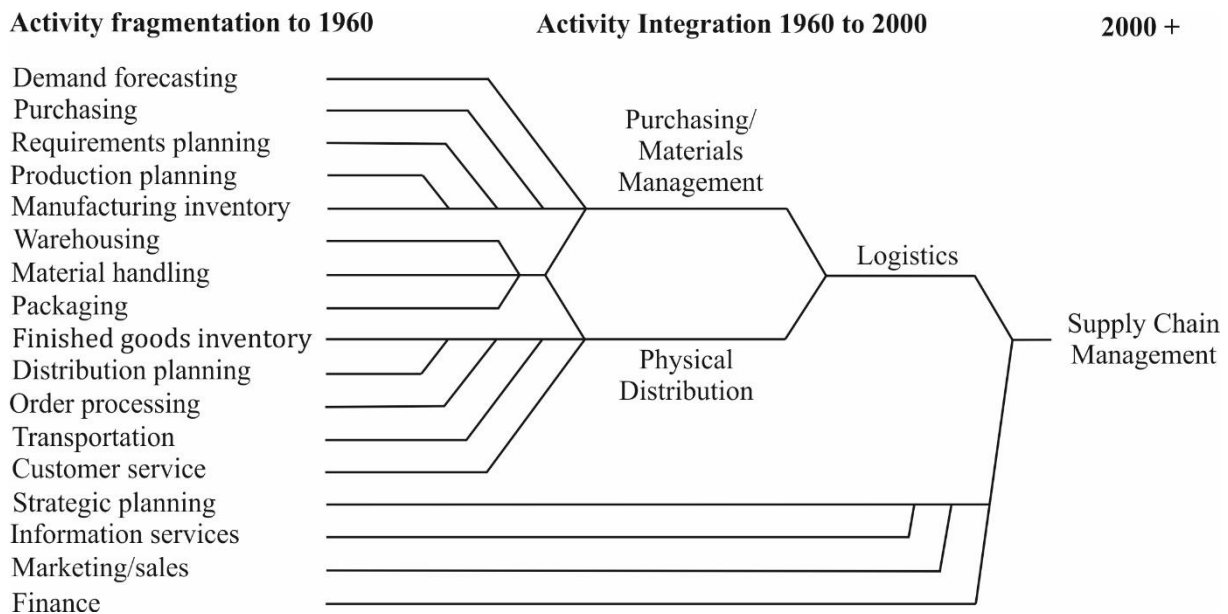
Figure 2.3 Evolution of the logistics and SCs

Source: Hesse & Rodrigue (2004).

External data refers to data generated outside an organisations, which can be public, unstructured, or collected by third-party organisations (private data). For SC analysts, particularly significant is external data shared between companies within supply chains, including market demand data. This kind of externally driven data is important for the companies since it is a result of the market response to the company products and services. The company doesn't have any direct influence on a given data, although companies try via the demand process and its subprocess demand planning to indirectly improve the market response of the customers. Moreover, companies try to model the market engagement to their products via demand planning activities like packaging, promoting the product, making sale promotions, using several distribution channels, etc.

Companies invest a lot of time, money and effort to better understand and model its processes according to the market demand data. This is a very challenging task for several points.

First of all, companies need to establish infrastructure, procedures and contracts with the retailers to track & record the demand data. Usually, companies use sales data from the downstream partner in the SC as a proxy for demand data. In reality, this is not demand data, rather it is the procurement data which can significantly distort the demand data. This is a very common practice since companies don't want to share its data and a large share of companies do not know that this is a bad practice. One of the downsides of this approach because it causes a bullwhip effect among the partners in SC. Another approach is that companies use retailers' point of sale (POS) data as a proxy for demand data (Syntetos et al., 2016). This also has its merits and downsides, since it doesn't take into account out of stock situations on the retail shelves. This approach also needs strong IT infrastructure and contracts with retailers.

The second „problem" with market-generated data is that it doesn't follow the usual statistical generating processes. This is a problem because the majority of the mathematical and statistical methods assume that data is following some statistical generating process. This is very notable in SC, where inventory models assume that demand during the lead time follows the Normal distribution and develop equations for calculating safety stock based on that assumption. According to Mirčetić et al. (2017), Mirčetić et al. (2022) & Mirčetić et al. (2018), 90% of the data from the pool of 97 series in the empirical study of the beer industry, doesn't follow the Normal distribution. Also, inventory models assume that the demand is deterministic and uniformly distributed through all periods, in reality, this is hardly the case. For example in Figure 2.4, the demand for pre-sliced salami in the period of 2015-2022 is presented for Italian manufacturer. The demand shows clear non-deterministic, i.e. stochastic behaviour (with random fluctuations and trend).
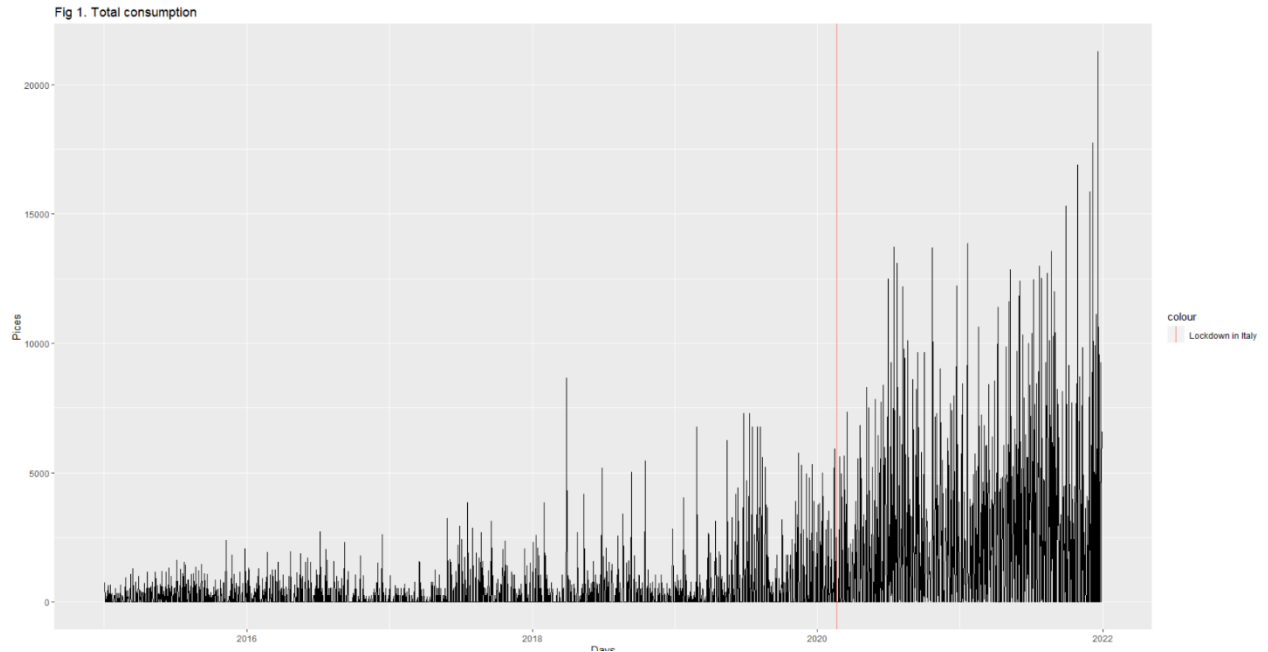
Figure 2.4 The demand for the pre-sliced salami in the period of 2015-2022

Source: Author.

Additionally, the daily demand shows highly volatile behaviour, therefore aggregation of the demand through different time horizons (weekly, monthly, etc) demonstrate a clear galloping demand trend (Figure 2.5).
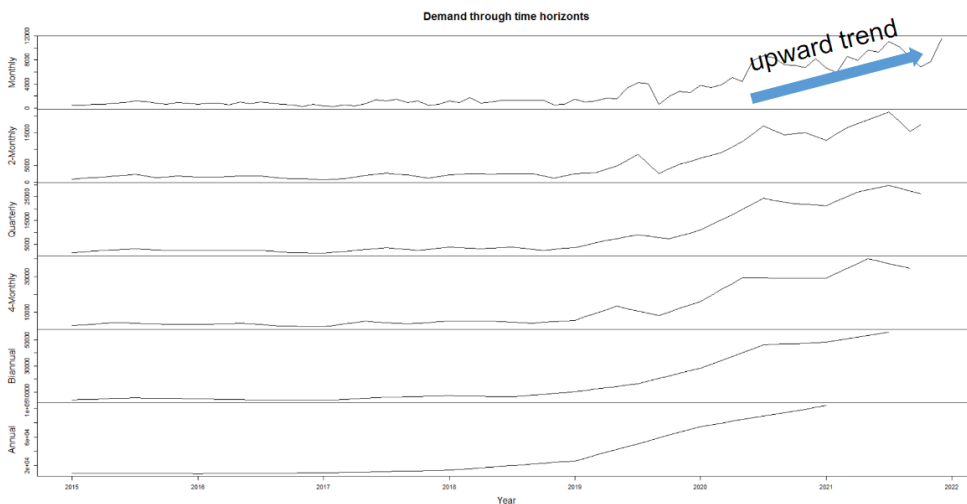
Figure 2.5 Demand aggregation through different time horizons

Source: Author.

Demand aggregation demonstrates a clear new reality from the start of the COVID-19 (an upward trend in the demand consumption process)! These are bulletproof evidence that demand is not deterministic. Regarding the assumption of normality, Figure 2.6 demonstrates a significant discrepancy from the Normal distribution, bearing in mind that it is an extremely right-skewed distribution.

Figure 2.6 demonstrates, that there has been a notable change in the salami demand since the start of the COVID-19 pandemic. The first lockdown in Italy was 2020-02-21 (vertical red line in Figure 2.4), after which demand for presliced salami erupted and reached a historical maximum. The demand was constantly trending, reaching the highest peak on 2021-12-20, with 21280 sliced pieces sold in one day.
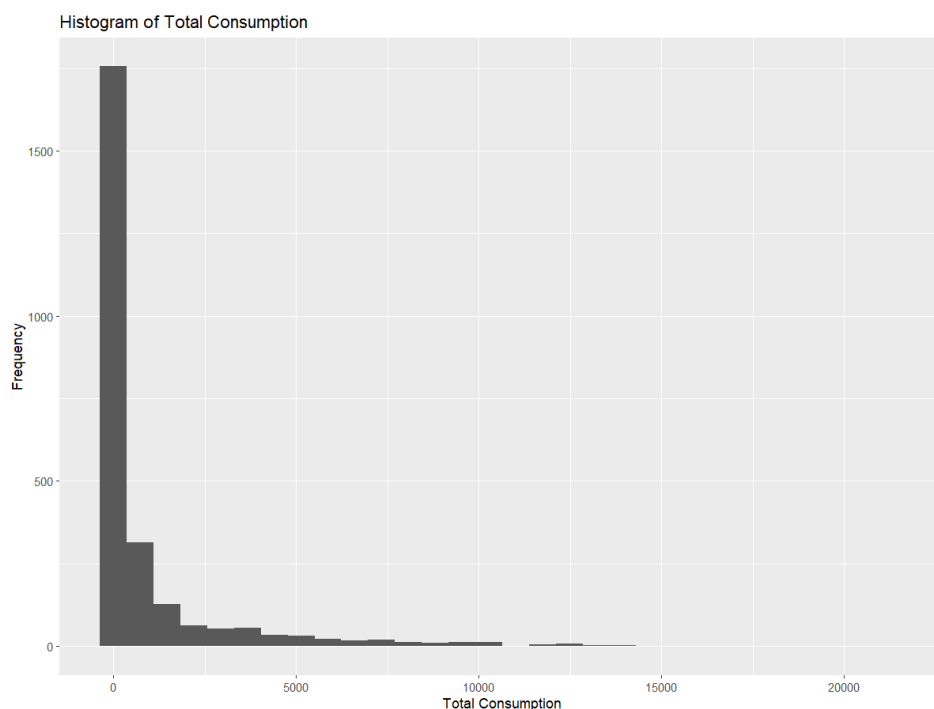
Figure 2.6 Empirical distribution of the demand

Source: Author.

Besides problems with the aforementioned theoretical assumptions about the data, the current situation in a worldwide economy and consequently SCs is posing another question in front of business analysts. These questions emerge as a result of pandemics, war crises, resource shortages, growing inflation, broken worldwide SC, etc. The question that contemporary business analysts need to solve when dealing with the data after the start of the COVID-19 is

how long period and data horizon are now valid for observation and modelling? This is clearly seen in Figure 2.4. If we take a closer inspection of the figure we will notice that before the COVID-19 pandemic and lockdown, consumers never consumed the pre-sliced salami in levels like from the first lockdown. It is noticeable that consumption of a given product erupted. Several questions now emerge:

- Is this just a hype in consumption due to specific conditions during the pandemic;
- Will this trend continue in the future and the company needs to increase its production;
- Does pre-COVID-19 data (from 2015-2020) have any value today and need to be discarded when modelling the demand data for pre-sliced salami?

These are all very hard questions to answer without proper data analytics procedure, which will be presented in the following Chapters.

REFERENCES

1. Arunachalam, D., Kumar, N. & Kawalek, J. P. (2018). Understanding big data analytics capabilities in supply chain management: Unravelling the issues, challenges and implications for practice. Transportation Research Part E, 114, pp. 416-436.

2. Hesse, M. & Rodrigue, J.-P. (2004). The transport geography of logistics and freight distribution. Journal of Transport Geography, 12(3), pp. 171–184.

3. Mircetic, D., Rostami-Tabar, B., Nikolicic, S. & Maslaric, M. (2022). Forecasting hierarchical time series in supply chains: an empirical investigation. International Journal of Production Research, 60(8), pp. 2514-2533.

4. Mirčetić, D. (2018). Unapređenje top-down metodologije za hijerarhijsko prognoziranje logističkih zahteva u lancima snabdevanja (Doctoral dissertation), University of Novi Sad, Serbia.

5. Mirčetić, D., Nikoličić, S., Stojanović, Đ. & Maslarić, M. (2017). Modified top-down approach for hierarchical forecasting in a beverage supply chain. Transportation Research Procedia, 22, pp. 193-202.

6. Mirčetić, D., Ralević, N., Nikoličić, S., Maslarić, M. & Stojanović, Đ. (2016). Expert system models for forecasting forklifts engagement in a warehouse loading operation: A case study. Promet-Traffic&Transportation, 28(4), pp. 393-401.

7. Nikoličić, S., Maslarić, M., Mirčetić, D. & Artene, A. (2019). Towards more efficient logistic solutions: Supply chain analytics. In Proceedings of the 4th Logistics International Conference, Belgrade, Serbia (pp. 23-25).

8. Souza, G. C. (2014). Supply chain analytics. Business Horizons, 57, pp. 595-605.

9. Srinivasan, R. & Swink, M. (2018). An investigation of visibility and flexibility as complements to supply chain analytics: an organizational information processing theory perspective. Prod. Oper. Manag. 27(10), pp. 1849–1867.

10. Syntetos, A. A., Babai, Z., Boylan, J. E., Kolassa, S. & Nikolopoulos, K. (2016). Supply chain forecasting: Theory, practice, their gap and the future. European Journal of Operational Research, 252(1), pp. 1-26.

11. Tiwari, S., Wee, H. M. & Daryanto, Y. (2018). Big data analytics in supply chain management between 2010 and 2016: Insights to industries. Computers & Industrial Engineering, 115, pp. 319-330.

12. Wang, G., Gunasekaran, A., Ngai, E. W. T. Papadopoulos Th. (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. Int. J. Production Economics, 176, pp. 98-110.

13. Zhu, S., Song, J., Hazen, B. T., Lee, K. & Cegielski, C. (2018). How supply chain analytics enables operational supply chain transparency: An organizational information processing theory perspective. International Journal of Physical Distribution & Logistics Management, 48(1), pp. 47-68.

3. DATA MINING AND KNOWLEDGE DISCOVERY

Author: Dejan Mirčetić

As discussed in Chapter 2, the concept of extracting the data and knowledge generation from it is closely related to the **Data Mining techniques**. In contemporary businesses, Data Mining techniques have a high impact on the overall performance of the companies, since operational, tactical and strategic decisions are made based on the input information gained from the Data Mining process. In Chapter 1, it was noted that the world contains over 97 zettabytes of data, with single databases often reaching sizes in the terabytes. In most organisations, data is doubling every two years. This data is stored across various platforms and in different formats, including structured, unstructured and semi-structure data (see Chapter 1). It is estimated that up to 90% of business data exists in an unstructured format. Additionally, a significant portion of this data may contain errors due to inappropriate storage or formatting, or manual errors during data collection. As a result, not all data held by organisations is necessarily accurate or reliable. Nevertheless, this vast amount of data holds valuable strategic information for companies. However, when faced with such a large volume of complex data types, how can they be effectively 'mined' to extract the meaningful insights they contain? The answer lies in Data Mining, which serves to increase revenues and to reduce costs by rapidly and automatically extracting useful knowledge on business insights from massive datasets.

Data Mining emerged from the necessity of efficiently extracting valuable information, requiring techniques that focus on identifying understandable patterns that can be interpreted as useful or interesting knowledge. Thus, **Data Mining is an iterative and interactive process** aimed at discovering valid, novel, useful, and understandable knowledge (patterns, models, rules, etc.) in massive database (Behera et al., 2019). The main objective of Data Mining is to **reveal critical insights** that **support decision-making** within a business organisation.

The upcoming sub-chapters will address how data is 'mined' for Business Analytics and Knowledge Discovery.

3.1. What is Data Mining?

Before defining Data Mining, it's important to place this term in context with the terms it's commonly associated with, connected to, and often mistakenly equated with. Non-experts frequently mix up the terms Data Mining and Big Data technology. However, these are two separate concepts. **Big Data describes** extremely large and complex **datasets** that require specialized software applications for processing. On the other hand, **Data Mining goes a step beyond**, it involves analysing such vast amounts of data to uncover hidden rules and patterns that may not be readily apparent.

Data Mining is a broad term encompassing various analytical techniques, including statistics, artificial intelligence, and machine learning. These methods are used to sift through vast amounts of data stored in an organization's databases or online repositories. The primary goal is to uncover patterns within the dataset. **Business Analytics (BA)** refers to the comprehensive process of utilizing skills, technologies, established practices, and algorithms associated with Data Mining. Hence, **Data Mining commonly serves as the backend of the BA function**, while the frontend of BA function consists of executive reporting metrics and collated information presented in a format that enables managers to make informed business decisions. When using Data Mining, the BA professionals act like a 'data detective' (Lee, 2013), analysing data to better describe and understand an organisation's present and past situation (descriptive analytics), predict future outcomes (predictive analytics) and take effective action (prescriptive analytics).

Data Mining is a core component of the **Knowledge Discovery in Databases (KDD)** process, but it is just one step in the overall process. The Data Mining aspect of the KDD process focuses on using algorithms to extract and identify patterns from data. In the broader KDD process, these mined patterns are evaluated and potentially interpreted to determine which patterns may be considered new "knowledge" (Behera et al., 2019). Defined in this manner, with Data Mining as a backend and Knowledge Discovery as a sort of frontend of BA, they represent, along with business data, its key pillars as outlined in Chapter 2.

Data Mining employees a variety of algorithms to mine huge datasets, identifying patterns that can yield valuable business insights. Data Mining is a tool, not a magical solution. It doesn't passively observe your database and alert you to interesting patterns.

Understanding your business, your data, and analytical methods remains crucial. Data Mining helps business analysts uncover patterns and relationships in data but doesn't determine the value of these patterns for organizations. Therefore, Data Mining does not replace skilled business analysts, instead, it provides them with a powerful new tool to improve their work.

Data Mining involves the computational process of identifying trends, rules, hidden patterns, and other valuable information by analysing large datasets. Data Mining adopt its technique from many research areas, including statistics, machine learning, database systems, visualization, neural networks, etc. It is the process of extracting actionable knowledge from diverse data sources sorted in various formats. Data Mining has become increasingly relevant in recent years due to advancements in data storage technologies (Big Data), Artificial Intelligence (AI), and Robotic Process Automation (RPA).

The **Knowledge Discovery in Databases (KDD)** process involves utilizing the database, including necessary selection, pre-processing, sub-sampling, and transformations, to apply Data Mining algorithms to identify patterns (Behera et al., 2019). It also includes evaluating the results of Data Mining. The common standard for describing the steps of Knowledge Discovery process is CRISP-DM (Cross-Industry Standard Process for Data Mining), which is shown in Figure 3.1.
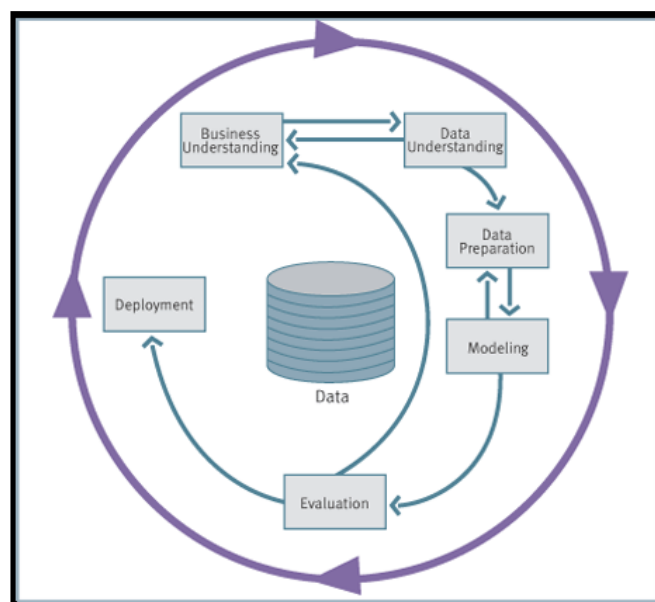
Figure 3.1 Cross-Industry Standard Process for Data Mining (CRISP-DM)

Source: Rahman et al. (2016).

In Figure 3.1, the first phase is business understanding, which involves comprehending the goals to be achieved and conducting detailed fact-finding about resources and assumptions. The second phase, data understanding, explores various descriptive data characteristics. The third phase, data preparation, is the most challenging and time-consuming part of the KDD process, aiming to select relevant data and format it appropriately for analysis. This phase includes activities like data selection, filtering, transformation, and integration. The fourth phase, modelling, entails applying analytical methods and selecting the most suitable algorithms. This phase also involves verifying the model's quality through testing and cross-validation. The fifth phase, evaluation, involves interpreting and assessing the discovered knowledge (Rahman et al., 2016).

3.2. Knowledge Discovery in Logistics and Supply Chain Management

In the context of SC and logistics, Data Mining is different from other general-purpose applications. The reason is related to the already mentioned and discussed diversity of data sources and structure of business data which pose significant challenges for engineers when designing appropriate modelling solutions. The process of generating the knowledge from the data can be summarized in Figure 3.2.

Based on Figure 3.2, the knowledge discovery and generation from the data is highly dependent on the knowledge source and can be divided into the **judgmental** and **statistical**. Both of these directions have their merits, but the procedure of extracting the knowledge from the source is fundamentally different. To apply a more formal approach for Data Mining, i.e., the statistical approach, the key foundation is the existence of quantitative data. In supply chains, there is a large number of sectors, locations, and transactions which generate data flows which can be used for Data Mining and extracting useful feedback. On the other hand, in SC and logistics, there are also a lot of data sources which are not quantitative, and therefore not subjected to formal quantitative procedures. Rather this data is traditionally subjected to expert judgemental panels (Delphi method) and decisions are made based on experts' experience, knowledge and authority.
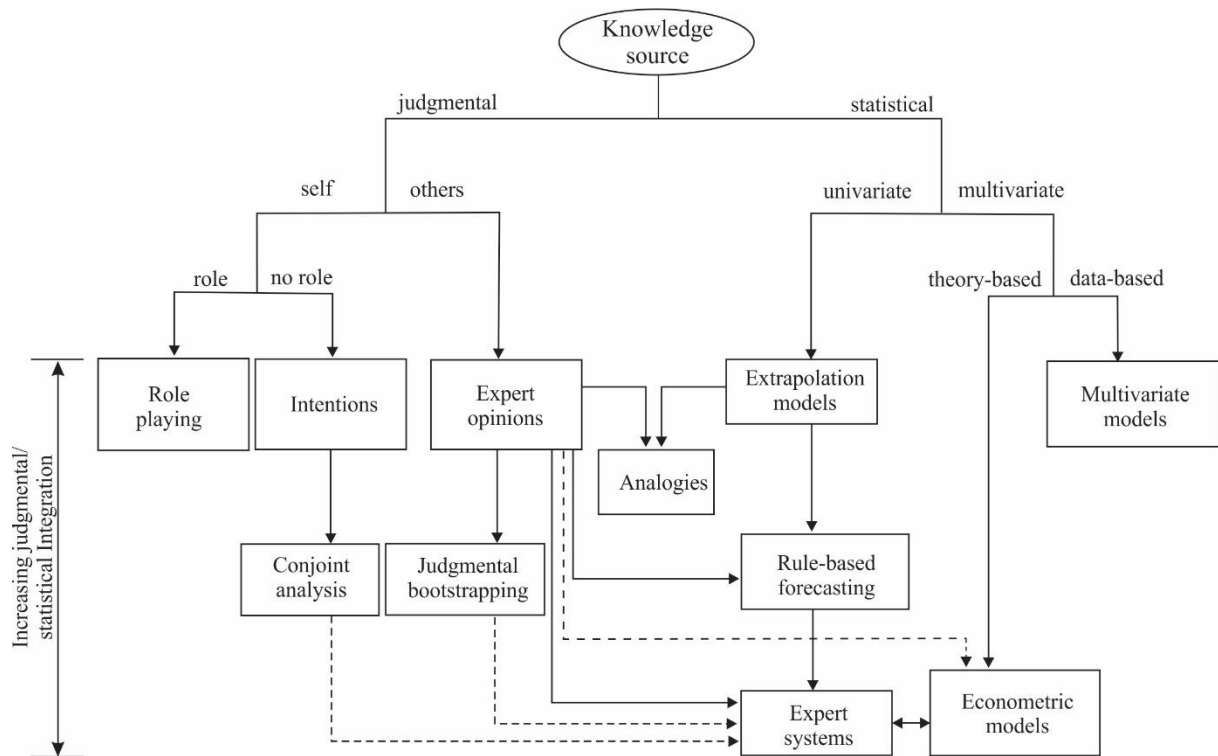
Figure 3.2 Principles of knowledge extraction from the data

Source: Armstrong (2001).

In this book, emphasis will be put on the **quantitative/mathematical techniques**, although we will briefly discuss some of the methodologies for capturing expert knowledge in structured frames, i.e. we will discuss the ES and its applications & examples.

3.3. Delphi's approach to judgmental Knowledge Creation

Valuable insights from seasoned professionals in the supply chain and logistics domain often go unrecognized. These insights should be shared with less experienced professionals in he field. The Delphi method is one approach for capturing and disseminating expert knowledge. According to Steurer (2011), the Delphi method, named after the oracle of Delphi in ancient Greece, which was initially used to consult on various public and personal matters in ancient Greece, in the 1950s, it evolved into a technique where experts replaced the oracle to reach consensus among a group of experts in some field. „Project Delphi", funded by the U.S. Air Force, was the first project using this method to forecast technological developments. Since then, the Delphi method has evolved and improved, finding applications

across various scientific disciplines. The Delphi method was developed to achieve a reliable expert consensus, often serving as a stand-in for empirical evidence when such evidence is lacking. That is, the Delphi technique is an iterative process where experts anonymously provide judgments on a particular issue, aiming to gather consensus and dissent along with their justifications. It is a highly structured group communication process where experts assess uncertain and incomplete knowledge (Naisola-Ruiter, 2022). According to Paivarinta et al. (2011), the Delphi method, among others, is extensively utilized in information systems research. It's employed to choose IS projects, prioritize software development project risks, define IS project requirements, pinpoint key issues in IS management, create a framework for knowledge manipulation activities, comprehend the roles and extents of knowledge management systems in organizations, and examine IS research on offshoring.

The Delphi method has become a standard practice for quantifying the outcomes of group elicitation processes. It is utilized across various disciplines to forecast trends, prioritize research areas, assess potential impacts of different policy choices, establish performance indicators, and develop clinical guidelines, among other applications. **The Delphi techniques is also used in SC and logistics area**. For example, it is highly recommended as an instrument for supply risk identification and assessment, for various kind of evaluation in logistics processes, for determined logistics best practices, for strategic decision-making and policy development, for mapping future SCM practices and for logistics forecasting. The four key characteristics or basic principles of the Delphi method are:

- iterative and multistage process (and data collection as well);
- participant feedback (controlled in some level) with the opportunity for participants to revise their answers;
- statistical determination of group response; and
- certain degree of anonymity.

A typical Delphi process involves presenting a series of questions over multiple rounds. Panelists, selected for their expertise and knowledge, respond anonymously. Each round is followed by feedback on the aggregated responses, allowing participants to see how their answers compare to those of the entire panel. Panelists can then adjust their answers and provide rationales for any changes in subsequent rounds. This iterative process continues until consensus is reached or a predetermined number of rounds is completed.

Steps to conduct the Delphi method

The Delphi method is a structured approach that entails collecting expert insights and opinions to achieve a consensus on a particular topic. The process typically involves four main steps (Figure 3.3).

Figure 3.3 Steps to carry out the Delphi method

Source: Steurer (2011)

Step 1 - Defining the objectives: The initial step involves defining the goals and scope of the Delphi study. This includes identifying the specific questions or topics requiring expert input and outlining the key issues to be discussed. This foundational step ensures the study maintains focus and relevance throughout.

Step 2 – Selection of experts: Choosing the correct group of experts is essential for the success and effectiveness of the Delphi technique. These experts should have pertinent knowledge, skills, and background concerning the subject being studied. The group should be varied to offer a broad spectrum of viewpoints. The number of participants can differ based on the study's size and intricacy, but it's generally advised to include at least 10-15 experts.

Step 3 – Elaboration and launching of questionnaires: This step involves developing questionnaires for gathering input from experts. Typically, the initial questionnaire is open-ended to allow experts to freely share their opinions without influence. In Round 1, experts receive the open-ended questionnaire and independently provide insights, predictions, or suggestions related to the study's objectives. In Round 2, the facilitator summarizes and anonymizes the responses from Round 1 to create a more focused questionnaire for the next round. If it is necessary additional rounds can be conducted to refine opinions based on achieved consensus levels, continuing until a predefined consensus is reached or the facilitator decides to end the process.

Step 4 – Use the results: After completing the Delphi process and achieving a consensus, the results are analyse d and utilized for decision-making, forecasting, policy development, or other purposes outlined in the study's goals. The anonymity of Delphi studies helps ensure that the final results are impartial and reflect the combined expertise of the experts involved.

3.4. Quantitative Data Mining approach for Knowledge Discovery

As demonstrated in Figure 3.2, quantitative **Data Mining is heavily backed on formal mathematical tools, more directly statistical ones**. We could argue that any statistical operation on the data could be regarded as a quantitative analysis. The main purpose of these operations is to extract the real patterns from the data and generate useful insights in the observed process (if the case of analysis in business or supply chains). This is not a straightforward task since there is a significant mismatch between statistical/mathematical theory assumptions and the distributions & patterns which are present in the real business data. The majority of business analyses in practice have a major flow based on that mismatch. It is of vital importance when applying quantitative methods that data fulfils the theoretical mathematic assumptions constrained by the observed model, in order to treat the model results as valid and potentially make decisions based on them.

As discussed in previous section, the Data Mining methods form a core component of the KDD process and are used repeatedly within it. Data Mining is a multidisciplinary technique, with its final analytical methods grounded in mathematics. Statistics, particularly, plays a vital role in data analysis during the data preparation phase, forming the foundation for several data mining methods. Data Mining involves the use of efficient algorithms to uncover expected or believed patterns. As illustrated in Figure 3.4, **Data Mining tasks** can be classified into five categories: clustering, classification, regression, association rules, and generalization. **Clustering** aims to group database objects so that objects within a cluster are similar, while those in different clusters are dissimilar. **Classification** involves learning a function that assigns attribute values to predefined classes. **Regression**, a statistical method, estimates relationships among variables and is commonly used for prediction and forecasting, overlapping significantly with machine learning. **Association rules** are used to describe strong relationships within transaction processes, such as "when A and B then C".

Generalization seeks to express a large amount of data as compactly as possible (Su, 2016). The main techniques used in Data Mining are: classification rules or decision trees, regression, clustering, genetic algorithms, agent-based modelling, etc.
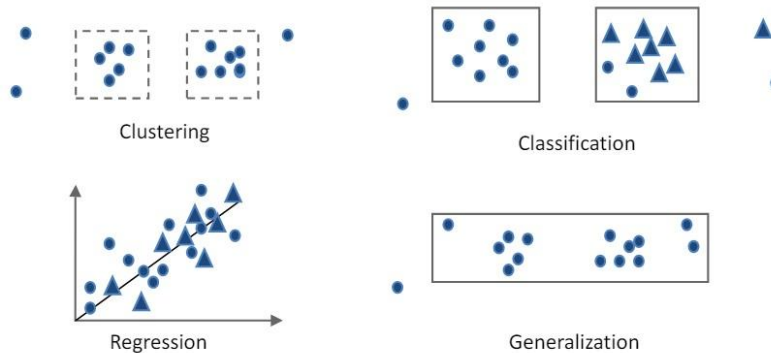
Figure 3.4 Tasks of Data Mining

Source: Su (2016).

The knowledge could be extracted from the processed quantitative data. The **KDD process comprises nine steps** as depicted in Figure 3.5. It is important to note that Data Mining is conducted on transformed data, with non-relevant information already excluded from the original dataset. The patterns discovered through this process are then interpreted and evaluated within a specific context to acquire knowledge that can aid in decision-making.
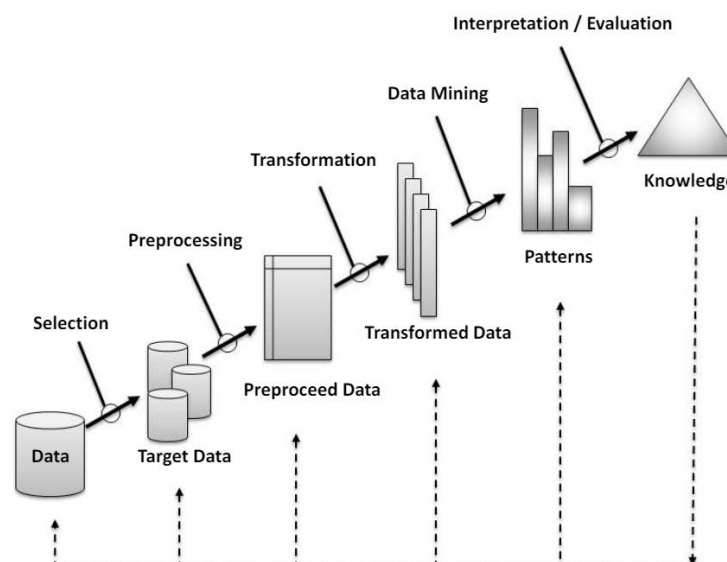
Figure 3.5 Steps that compose the KDD process

Source: Fayyad et al. (1996).

As it previously stated, the common standard for describing the steps of KDD process is CRISP-DM, that represent a leading industrial model. With regard to the current publications and research reviewed by Su (2016), the typical Data Mining techniques and applications in supply chains include:

- **Decision Trees:** Solving suppliers problems that can be reduced to, e.g. for each decision, a set of possible outcomes, together with an assessment of the likelihood of each outcome occurring

- **Regression:** Forecasting and estimating customer demand for a new product

- **Association Rule:** Identifying the cause roots of product fauiler, optimizing the manufacturing capacity and enabling the condition-based maintenance

- **Genetic Algorithm:** Evaluating the improved hypothesis of operating VMI in an uncertain demand environment

- **Clustering Algorithms:** With k-Mean algorithm to categorizing the returned commodities in order to improve the manufacturing processes quality or assigning customers in different segments based on their demographics and purchase behaviours.

- **Multi Agent Data Mining System:** Supporting production planning decisions based on the analysis of historical demand for products

The knowledge extracted by Data Mining is typically stored and presented using **Experts Systems (ES)**. An ES is a sophisticated knowledge system designed to mimic human expertise in various application areas. Olson and Courtney (1992) define ES as „a computer programme within a specific domain, involving a certain amount of Artificial Intelligence to emulate human thinking in order to arrive to the same conclusions as a human expert would". An ES component is ideal to assist a decision-maker in an area where expertise is required (Turban, 1995). Essentially, an ES transfers expertise from an expert (or other source) to the computer. It can either support decision-makers or completely replace them, and it is the most widely applied and commercially successful artificial intelligence technology (Turban et al., 2007). One of the justifications for building an ES is to provide expert knowledge to a large number of users (Kock, 2005). According to Turban et al. (2007), ESs are considered to be part of Decision Support System (DSS), that could be characterized as a computer-based information system that combines models and data in an attempt to solve semi-structured and

unstructured problems with extensive user involvement (Turban et al., 2007; Mirčetić et al., 2016). The next chapter discusses Machine Learning (ML), which refers to computers' ability to learn and represent knowledge from input data. ML can be seen as a bridge between the results produced by Data Mining and the Business Intelligence tools used to present executive reporting metrics in a format that enables managers to make informed business decisions.

REFERENCES

1. Armstrong, J. S. (2001). Introduction. In Armstrong, J. S. (Ed.), Principles of Forecasting: A Handbook for Researchers and Practitioners (pp. 1-12), Springer Nature.

2. Behera, P. C., Dash, C. & Mohapatra, S. (2019). Data Mining and Knowledge Discovery (KDD). International Journal of Research and Analytical Reviews, 6(1), pp. 101-106.

3. Fayyad, U. M., Patetsky-Shapiro, G. & Smith, P. (1996). From Data Mining to Knowledge Discovery in Databases. American Association for Artificial Intelligence, 17, pp. 37-34.

4. Kock, E. D. (2005) Decentralising the Codification of Rules in a Decision Support Expert Knowledge Base (MSc thesis). University of Pretoria; 2005. Available from: http://repository.up.ac.za/handle/2263/22959

5. Lee, P. M. (2013). Use of Data Mining in Business Analytics to Support Business Competitiveness. Review of Business Information Systems, 17(2), pp. 53-58.

6. Mirčetić, D., Ralević, N., Nikoličić, S., Maslarić, M. & Stojanović, Đ. (2016). Expert system models for forecasting forklifts engagement in a warehouse loading operation: A case study. Promet-Traffic&Transportation, 28(4), pp. 393-401.

7. Naisola-Ruiter, V. (2022). The Delphi technique: a tutorial. Research in Hospitality Management, 12(1), pp. 91-97.

8. Olson, D. L., Courtney, J. F. & Courtney, J. F. (1992). Decision support models and expert systems. New York: Macmillan, USA.

9. Paivarinta, T., Pekkola, S. & Moe, C. E. (2011). Grounding Theory from Delphi Studies. In: Proceedings of the 32nd International Conference on Information

Systems (ICIS 2011): Research Methods and Philoshophy, 4-7 December 2011, Shanghai, China.

10. Rahman, F. A., Shamsuddin, S. M., Hassan, S. & Haris, N. A. (2016). A Review of KDD-Data Mining Framework and Its Application in Logistics and Transportation. International Journal of Supply Chain Management, 2(1), pp. 1-9.

11. Steurer, J. (2011). The Delphi method: an efficient procedure to generate knowledge. Skeletal Radiol, 40, pp. 959-961.

12. Su, W. (2016). Knowledge Discovery in Supply Chain Transaction Data by Applying Data Farming (Master thesis). Technical University of Dortmund, Dortmund, DE.

13. Turban, E. (1995). Decision support and expert systems Management support systems. Prentice-Hall, Inc. New York, USA.

14. Turban, E., Rainer, R. K. & Potter, R. E. (2007). Introduction to Information Systems: Supporting and Transforming Business. John Wiley & Sons, Inc. USA.

4. MACHINE LEARNING

Author: Dejan Mirčetić

There are a lot of questions about what is Machine learning (ML). Is it a really process in which machines learn from the external environment by themselves, or it is a formalised process via mathematical algorithms which allow computers to „figure out" rules in the outside world? What tools is ML using? How does the typical data flow look in the ML pipeline? Is it applicable to traditional industries, not just in IT and internet-related industries? Where is the place of ML in the context of business? How to systematically use it for solving business issues? Is there any architecture on how to apply it to SCs?

On these and similar questions, we will try to provide answers in the following chapter, closing with a real case study example of the application of ML algorithms in the food supply chain.

4.1. What is machine learning?

Machine learning is a discipline focused on two interrelated questions: How can one construct computer systems that automatically improve through experience? and What are the fundamental statistical computational-information-theoretic laws that govern all learning systems, including computers, humans, and organizations? The study of machine learning is important both for addressing these fundamental scientific and engineering questions and for the highly practical computer software it has produced and fielded across many applications (Jordan & Mitchell, 2015).

ML arises from this question: could a computer go beyond "what we know how to order it to perform" and learn on its own how to perform a specified task? Could a computer surprise us? Rather than programmers crafting data-processing rules by hand, could a computer automatically learn these rules by looking at data? This question opens the door to a new programming paradigm (Chollet, 2021).

The ML enables a fundamental shift in the programming paradigm (Figure 4.1). In classical programming human programmer inputs rules (programme) and the data that is analysed and processed in agreement to those rules. As a result, the answers are provided at the end.

On the other hand, with ML human programmer inputs the data with answers expected from the data, and outcome the rules.
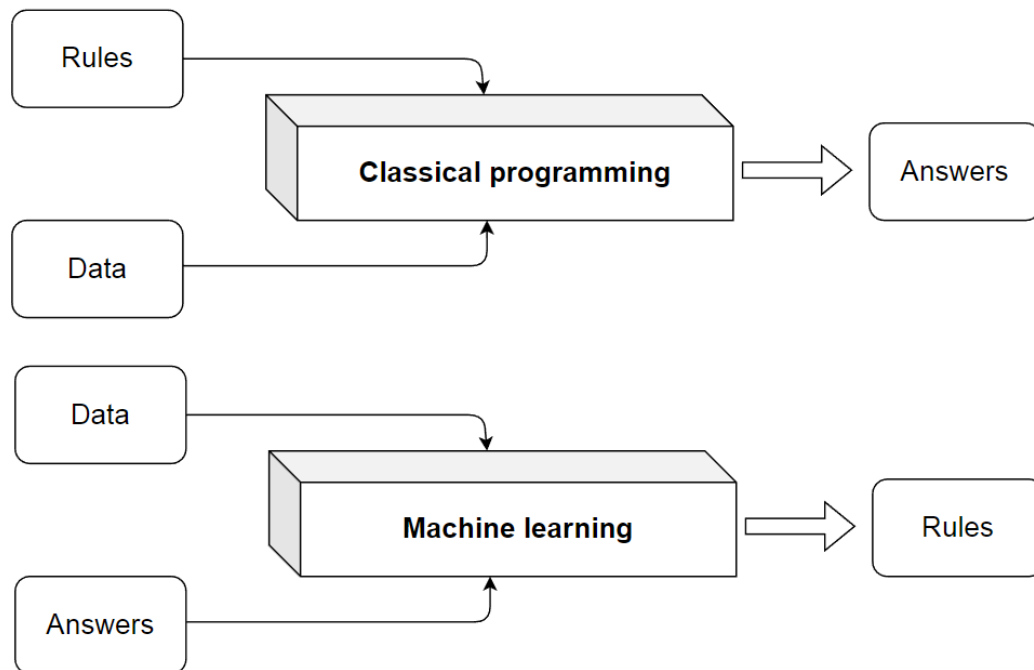
Figure 4.1 The classical programming vs. machine learning system training

Source: Chollet (2021).

Classical programming could be understood as imperative programming since the programmer predefines all the rules and execution of code is performed accordingly, while ML could be understood as declarative programming where we express higher-level goals or describe important constraints, and rely on mathematical algorithms to decide how and/or when to translate that into action.

Today, ML is the foundation of countless important applications, including web search, email anti-spam, speech recognition, product recommendations, and more (Ng, 2017). Many developers of AI systems now recognize that, for many applications, it can be far easier to train a system by showing it examples of desired input-output behaviour than to programme it manually by anticipating the desired response for all possible inputs. The effect of ML has also been felt broadly across computer science and across a range of industries concerned with data-intensive issues, such as consumer services, the diagnosis of faults in complex systems, and the control of logistics chains (Jordan & Mitchell, 2015).

4.2. Foundations and theoretical assumptions of ML

The background of ML lies in mathematics, more specifically in statistics. Therefore ML uses a theory background and algorithms developed in **statistical learning** and there is also debate is the ML a real area of its own or it is just part of statistics. In practice ML algorithms usually lack a certain level of mathematical rigidity and sometimes easily go above some mathematical constraints present in statistics. For example, ML algorithms do not pay a lot of attention to confidence intervals when optimizing the coefficients in parametric algorithms, although this is one of the most important topics in statistics. Generally, there is a **big overlap of ML and statistics** and some of the most notable ML algorithm creators and professors claim that it is just part of statistics (Hastie et al., 2009). Nevertheless, being the area of its own or part of statistics, ML consists of several steps in acquiring knowledge from the data. There is no general consensus about these steps but generally, they can be represented as transformation of different data sources to business intelligence insights.

In a business context, the ML models are useless, without proper support regarding the data pre-processing, data mining and application of the insights to the actual processes. Therefore, creating the ML algorithms without the possibility of updating the model and using its output for an actual decision-making process, does not bring any value to modern companies. Accordingly, in modern-day business analytics, the quantitative ML process is usually part of the business intelligence workflow. More specifically, it is part of business intelligence's important subprocesses (data science and data analytics part of business intelligence). The details about the role of ML in these processes and the actual processes itself of generating the values for the business via ML will be provided in the upcoming subchapter.

4.3. Business intelligence and ML in SCs

Business intelligence, in the context of SCs, is the process of making conclusions about the observed SC processes, based on the modelling of the data from those processes. It is mostly based on statistics, but other mathematical areas come into play: operation research, linear algebra, fuzzy logic (in a case when data is scarce or missing), numerical optimization, metaheuristics, etc. Additionally, new disruptive technologies are also becoming an important

aspect for analysing the data and delivering conclusions: **machine learning, artificial intelligence, digital twins, smartisation, living labs**, etc.

There are no strictly organized procedures on how the business intelligence procedure and ML workflows should be organized, but there are some useful guidelines in the practice and literature which have been proven successful when conducting the analysis. The procedure for conducting the business intelligence is also diverse to the origin of the software that is used for the analysis. For example, Microsoft offers several tools via its channel Microsoft Business Intelligence package, that conduct different tasks: **data ingestion, data storage, data integration, data management, data processing, reporting, data sharing and data science** (Figure 4.2).
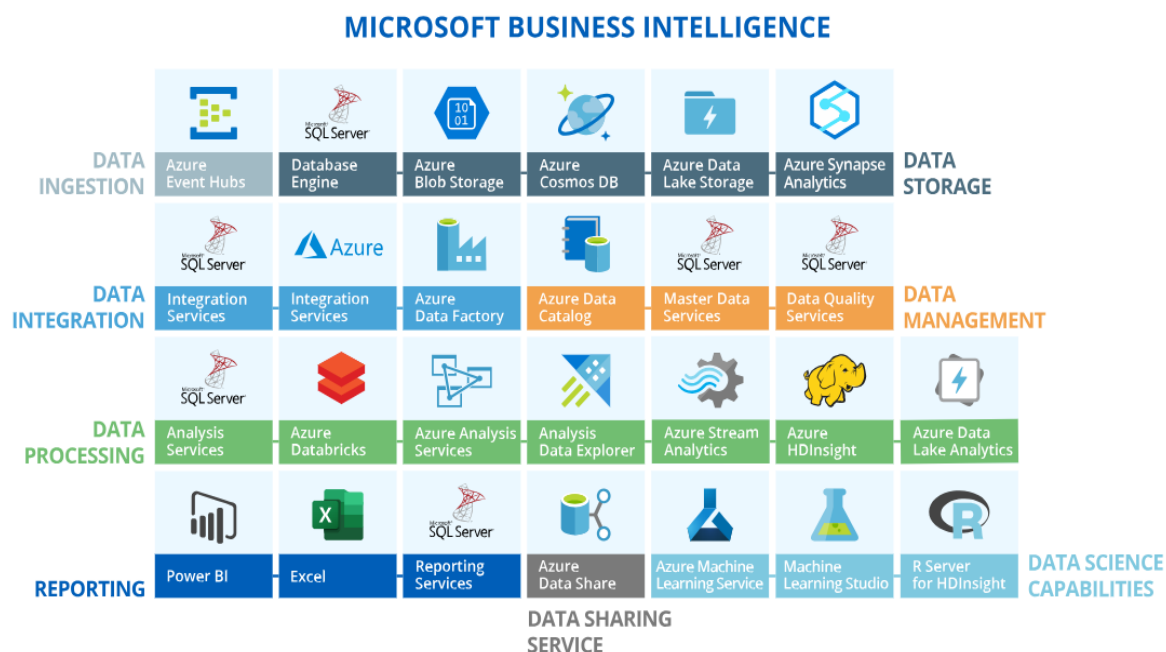
Figure 4.2 Microsoft business intelligence architecture

Source: ScienceSoft (n.d.).

In a given architecture the ML procedures are applied only at the data science level via several tools: Azure ML services, ML studio and R Server for HDInsight. The general procedure of how the data analysis in the context of ML is performed in R Server is presented in Figure 4.3.

Figure 4.3 ML Data analysis steps in R

Source: Wickham et al. (2023).

When dealing with ML, there is usually a **misconception that the majority of time and effort is spent on actually building the ML algorithms**. The reality is totally the opposite, the majority of time is usually spent on wrangling with the data and pre-processing tasks rather than to the modelling process. Sometimes, all the processes before the modelling process are much more ch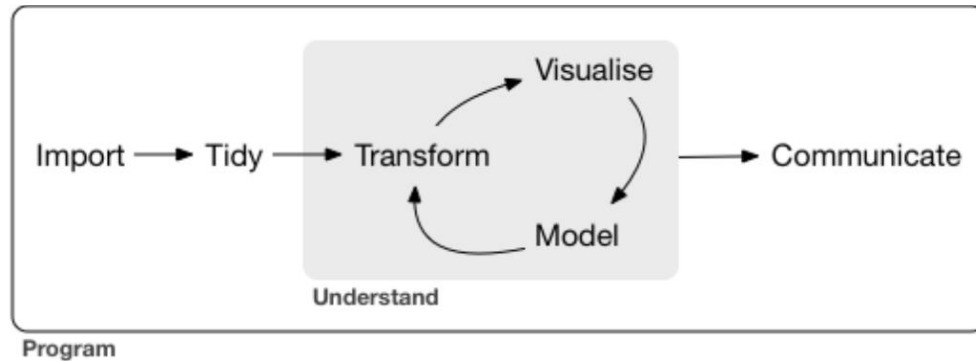allenging and demanding. That is why there is no conses on how these steps need to be performed. Figure 2.9 presents an example of a good approach to transforming the data into business insights and general knowledge. The procedure starts with the import step, which is one of the most important steps in building ML models, since without importing data to the software it is not possible to conduct any kind of analysis. This typically means that you take data stored in a file, database, or web application programming interface (API), and load it into a data frame in R (Wickham et al., 2023). The second step is related to tidying the data which is a procedure unique to R and relates to transforming the data into a specific form for further analysis (each column is variable, and each row is an observation– tibble data frame). The next step is related to the transformation of the data which usually includes narrowing the set of observations to the subsample of interest. Additionally, it may also include creating new variables as combinations of several existing ones or generating summary statistics. Visualization and modelling serve distinct but complementary roles in the realm of data analysis. Visualization is a profoundly human-centric activity, offering insights that may elude more formalized approaches. A well-crafted visualization can reveal unexpected patterns, prompt new inquiries, and even suggest that the original questions may need refinement or different data. In contrast, models provide a mathematical or computational framework for answering precisely formulated questions. They offer scalability and efficiency, making them suitable for handling large datasets. However, models (in which ML are also

included) come with inherent assumptions, and they cannot question or challenge these assumptions. Consequently, models may not have the capacity to surprise or unveil unforeseen insights. The synergy between visualization and modelling is evident in their collaborative role in data analysis. Visualization aids in the initial exploration, encouraging the formulation of precise questions, while models systematically provide answers within the defined parameters. Recognizing the strengths and limitations of each approach is crucial, leading to a more comprehensive and informed data analysis process. The last step represents communication which is vital for the success of data analysis, since if the information is not provided to the decision maker in a right and consistent way, then the whole analytics could be in vain. The key element in data analytics are the ML models, without which, conclusions about the business processes could not be inferred. In order to tackle the specific SC problems, the architecture for general-purpose business intelligence applications (presented in Figure 4.2) needs to be better tuned, as well as the ML models. Accordingly, to transform how supply chains operate via enhancing operational efficiency, improving decision-making, and driving towards the achievement of corporate objectives, the company **Equilibrium AI** developed the AI & ML platform presented in Figure 4.4.
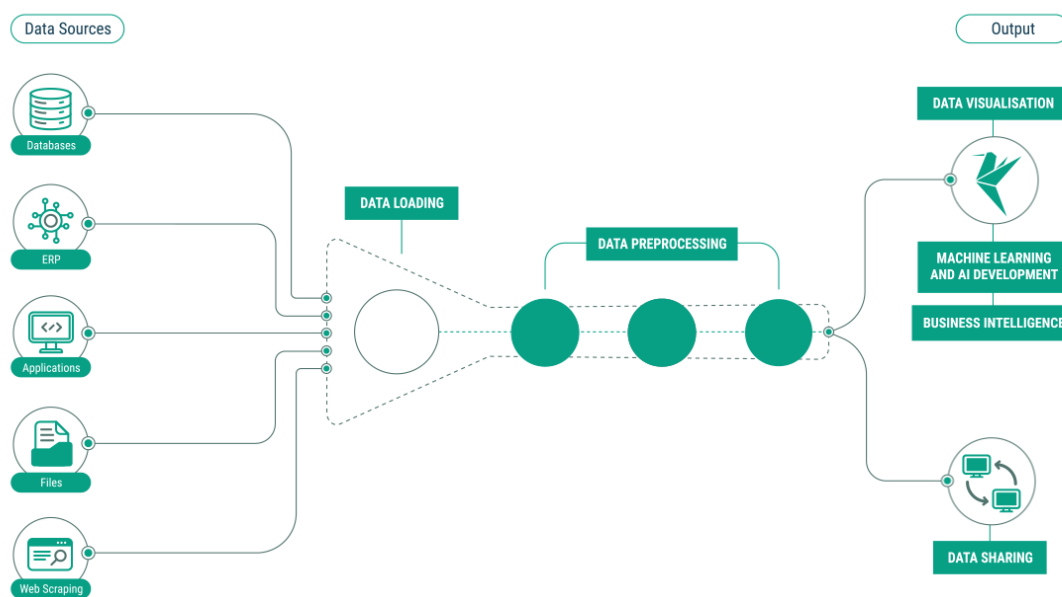
Figure 4.4 The ML data & knowledge pipeline for company Equilibrium AI

Source: Equilibrium AI (n.d.).

The figure represents a good example of everyday practice of how the extraction of knowledge and insights are generated in SC applications. Generally, the process consists of backend and frontend operations in order to create value (business insights) for users. The backend process starts with extracting the data from different data sources usually found in SCs:

- Databases;
- Enterprise resource planning systems (SAP, Navigator, Microsoft Dynamics, etc);
- Applications (web APIs);
- Flat files (csv, xlsx, JSON, etc);
- Web, internet and other online sources.

Each of the data sources has a different structure, protocols and accordingly procedures for how the data is extracted and loaded for cleaning and pre-processing before applying the ML algorithms. Accordingly, this process is performed via data loaders which have pre-programmed code for data mining of different data sources and transitioning the row data to the new database, which is structured and arranged for the application of ML models. Before applying the ML models, there is one additional step called data pre-processing. In this step row data gathered from the companies is checked for the wrong inputs, non-logical values, correct structure of inputs, outliers, double entries, NA, NaN, etc. The procedure continues with merging the outside data with company data. This data is usually related to external factors which can potentially influence the observed SC business process, for example, weather data, consumer price index, average income in a given region, specific demographic characteristics in a given area, gas prices, pandemic outbreaks, social network comments about company products, etc. This is very important because it holistically collects all the possible factors (internal and external) which are possibly influencing on a given business process, which increases the chance that the ML models will find the right signal in the data and be able to make the right conclusion and rules what are the root cause reasons why is business process behaving as observed.

After merging internal and external data, pre-processing consists of signal detection, removing the noise from data, feature engineering and randomly dividing the data on train and test (sometimes on validation data if the neural network model are developed). Data which comes out of the pre-processing step is cleaned and structured for the application of ML models.

The output part consists of data visualization, ML & AI development and data sharing. Sometimes, this data is shared without applying the ML models to other platforms which conduct different kinds of analysis (just reporting to stakeholders or government agencies). The visualization process is performed via the frontend part of the platform which is user-centric and allows users to make requests on what data, how and in what settings they want to see the observed SC data (for example Figure 4.5).

Figure 4.5 The typical visualization part of ML platform in SCs

Source: Author.

Conversely, the ML part of processing the data is hidden from the eyes of the users and it is not easy to understand. That is why the ML models are sometimes regarded **as black box** models in which there is no clear understanding of how exactly the machine connected the observed input with the observed output. This is one of the obstacles which is preventing the broader usage of ML models in practice, especially ones that are complex to interpret (Rostami-Tabar & Mircetic, 2023). Accordingly, we could divide the ML models into those with high interpretability-low flexibility and low interpretability-higher flexibility (Figure 4.6). In general, as the flexibility of an ML method increases, usually accuracy of the ML model increases and interpretability decreases (Mirčetić et al., 2016).
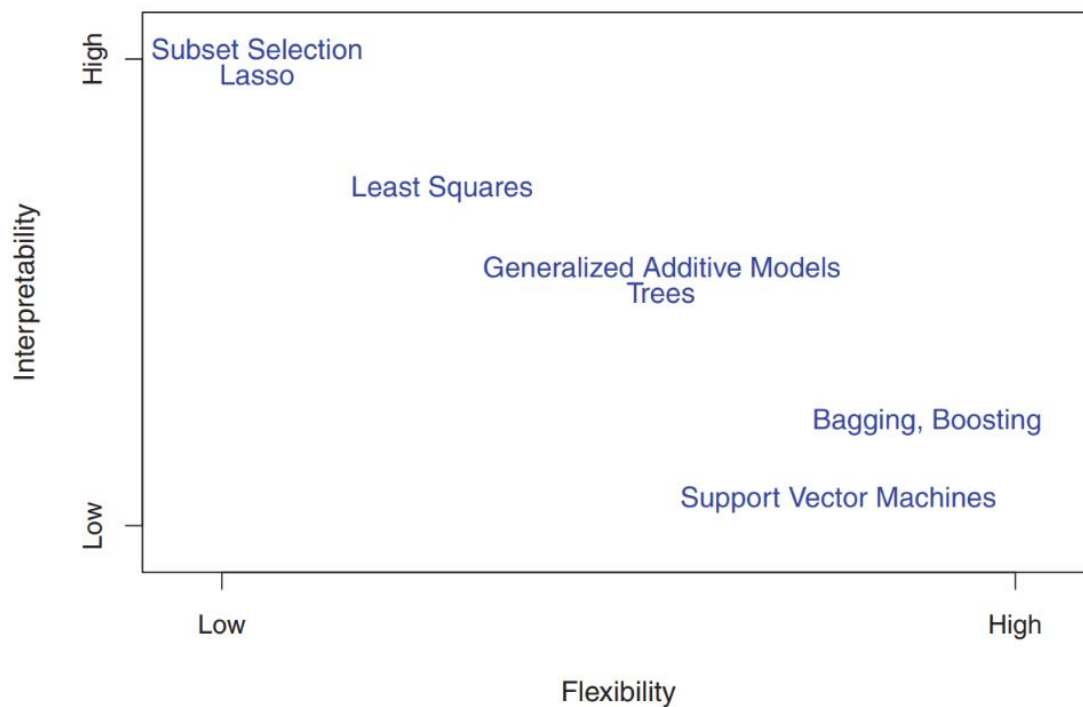
Figure 4.6 A representation of the trade-off between flexibility and interpretability, using different ML methods

Source: Hastie et al. (2009).

ML and SC business data

If the users better understand the visualization and graphics like Figure 4.5, why do we need ML at all, and could we skip modelling the data with ML and just make informative graphics? Unfortunately no. Maybe the main reason why we need ML models is because it is not possible in all situations to have easily readable and detectable patterns in the data seen via graphics (like in Figure 4.5). The more common situation is that graphics cannot usually reveal the mystery of what is happening in the observed SC business data and we need stronger tools in the form of ML algorithms to dig deeper into data and search for **data-generating rules** (Figure 4.7).

Figure 4.7 Statistical characteristics of products in the food supply chain (summarized for all products)

Source: Author.

It is very hard to make easy conclusions from Figure 4.7 and derive business rules about the data-generating process. To find patterns in the data from the figure, we have developed an ML model which could be used to summarize the characteristics and detect important signals in data. Accordingly, the developed ML model for a food supply chain is presented in Equation 1. The basic driver and the backbone of this ML model is the autoregressive integrated moving average model, with the general following form:

$$y_t^i = c + (\phi_1 y_{t-1}^i + ... + \phi_p y_{t-p}^i) + (\theta_1 e_{t-1} + ... + \theta_q e_{t-q}) + e_t \; ; \tag{1}$$

$$y_t - y_{t-1} = c + \phi_1(y_{t-1} - y_{t-2}) + ... + \phi_p(y_{t-p} - y_{t-p-1}) + (\theta_1 B e_t + ... + \underbrace{\theta_q e_{t-q}}_{\theta_q B^q e_t} + e_t) \; ;$$

$$y_t - B y_t = c + \phi_1(y_{t-1} - B y_{t-1}) + ... + \phi_p(y_{t-p} - B y_{t-p}) + (e_t(1 + \theta_1 B + ... + \theta_q B^q)) \; ;$$

$$(1-B)y_t = c + \phi_1(1-B)(y_{t-1}) + ... + \phi_p(1-B)y_{t-p} + e_t(1 + \theta_1 B + ... + \theta_q B^q) \; ;$$

$$(1-B)y_t = c + \phi_1(1-B)B y_t + ... + \phi_p(1-B)B^p y_t + e_t(1 + \theta_1 B + ... + \theta_q B^q) \; ;$$

$$\underbrace{(1-B)^d y_t}_{\substack{differencing \\ d\_\deg ree}} \cdot \underbrace{(1 - \phi_1 B - ... - \phi_p B^p)}_{AR(p)} = c + \underbrace{e_t(1 + \theta_1 B + ... + \theta_q B^q)}_{MA(q)} .$$

The ML model clearly demonstrates its low interpretability and black box characteristics. It is hard for an average business user to understand the connections between input and output data. Moreover, for the average business user when confronted with the presented model, the question emerges! What is Equation (1)? We could argue that Equation (1) represents the rules from Figure 4.1, generated by ML data & knowledge pipeline, which reveal the mystery about data-generating processes in a given SC business setting.

At first look, the developed ML model in Equation (1) doesn't seem to improve our understanding of the data. We are still confused as with Figure 4.7, but the ML model has a crucial advantage over the figure. In essence, the ML model is a **mathematical formula** which may not be easily understandable to a human user but is completely understandable to a computer, which **can be programmed to use a given formula** and make **business decisions** based on discovered rules.

REFERENCES

1. Chollet, F. (2021). Deep learning with Python. Simon and Schuster.

2. Equilibrium AI (n.d.). Equilibrium AI Data Pipeline [available: https://eqains.com/, access: January 23, 2024]

3. Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). The elements of statistical learning: Data mining, inference, and prediction (Vol. 2). Springer.

4. Jordan, M. I. & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), pp. 255-260.

5. Mirčetić, D., Ralević, N., Nikoličić, S., Maslarić, M. & Stojanović, Đ. (2016). Expert system models for forecasting forklifts engagement in a warehouse loading operation: A case study. Promet-Traffic & Transportation, 28(4), pp. 393-401.

6. Ng, A. (2017). Machine learning yearning. [available: http://www.mlyearning.org/, access: January 23, 2024]

7. Rostami-Tabar, B. & Mircetic, D. (2023). Exploring the association between time series features and forecasting by temporal aggregation using machine learning. Neurocomputing, 548, 126376.

8. ScienceSoft (n.d.). Microsoft Business Intelligence to Drive Robust Analytics and Insightful Reporting [available: https://www.scnsoft.com/services/business-intelligence/microsoft, access: January 23, 2024]

9. Wickham, H., Çetinkaya-Rundel, M. & Grolemund, G. (2023). R for data science. O'Reilly Media, Inc.

5. BUSINESS PROCESS MANAGEMENT AND PROCESS MINING

Author: Dario Šebalj

To stay competitive in today's business environment, effective management and ongoing process improvement are critical. This chapter examines Business Process Management (BPM) and Process Mining, two important parts of business intelligence that help companies analyse , optimize, and enhance their operational processes.

BPM provides an organized and structured method for identifying, designing, executing, monitoring, and improving business processes while aligning them with the organization's strategic objectives. Process mining, on the other hand, is a tool for identifying and enhancing actual processes through the extraction of knowledge from event logs found in modern business information systems. The combination of Business Process Management and Process Mining enables an objective, data-driven method for understanding and improving business processes.

By leveraging these methodologies, organizations can find hidden inefficiencies and problems, adapt to changing market demands, and improve their performance and customer satisfaction. The fundamental concepts, methodologies, tools, and real-world applications of BPM and process mining will be covered in this chapter.

5.1. Business process

Every organization, regardless of size or sector, is a complex system of interconnected processes. These processes are the structured activities undertaken to accomplish a specific organizational goal. For instance, in a manufacturing company, key processes might include product design, raw materials procurement, manufacturing, quality control, and distribution. In a service-oriented business like a bank, processes include account opening, loan processing, customer service, and compliance checks. Organizations use processes on a daily basis, and these processes can be as varied as the organizations themselves. In a hospital, processes

range from patient admission to medical treatment and discharge. In an educational institution, they encompass student enrolment, course delivery, and examination administration. Each process is a sequence of steps, involving various departments and personnel, and often supported by technology.

According to Dumas et al. (2018), every business process consists of several events and activities. **Events** correspond to things which do not have a duration and happen atomically (e.g. 'Order received'). On the other side, **activities** are tasks or operations that are interconnected and whose execution fulfils the goal of the business process (e.g. 'Pay the invoice'). A typical process, aside from events and activities, includes **decisions**, which indicate a stage at which the process decides which direction it will go in the future. For example, in the sales process, one decision point could be when the salesperson checks whether the product is in stock. If a product is in stock, the process moves on to the next activity. If there is no product in stock, the process proceeds in a different way (e.g. by informing the customer that the order cannot be fulfilled). The important parts of a process are actors/participants and objects. **Actors** include people, organizations or software systems that perform the process activities, while **objects** are equipment, materials, paper documents (physical objects), electronic documents and records (informational objects).

Dumas et al. (2018) state that the execution of a process results in one or more **outcomes**. An outcome should, in theory, benefit all parties involved in the process (*positive outcome*). Sometimes this value is only partially or never reached (*negative outcome*).

Von Scheel et al. (2015) define **business process** as „a collection of tasks and activities (business operations and actions) consisting of employees, materials, machines, systems, and methods that are being structured in such a way as to design, create, and deliver a product or a service to the consumer".

Understanding a process is just the beginning. The true problem, and opportunity, is to manage these processes in a systematic and planned manner. This brings us to the following chapter: Business Process Management (BPM). In this section, we will look at the approaches and frameworks that allow organizations to not only manage but execute their processes. BPM is more than just process recording and analysis; it is a comprehensive method to developing, implementing, monitoring, and constantly improving business processes.

5.2.   Business Process Management

In scientific and professional literature, we can find different definitions of Business Process Management. Gartner (n.d.) defines BPM as „a discipline that uses various methods to discover, model, analyse , measure, improve and optimize business processes". According to Camunda (n.d.), BPM is „a systemic approach for capturing, designing, executing, documenting, measuring, monitoring, and controlling both automated and non-automated processes to meet the objectives and business strategies of a company". Swenson and Rosing (2015) proposed some wider and maybe most precise definition: „Business process management (BPM) is a discipline involving any combination of modelling, automation, execution, control, measurement, and optimization of business activity flows in applicable combination to support enterprise goals, spanning organizational and system boundaries, and involving employees, customers, and partners within and beyond the enterprise boundaries".

According to Freund and Rücker (2012), the new BPM projects often include one of these scenarios:

1. Process improvement using information technology, (IT)
2. Documentation of current processes,
3. Introduction of entirely new processes.

Dumas et al. (2018) see BPM as a continuous cycle comprising the following phases:

- **Process identification** - A business problem is given in this step. Processes that are important to the problem being solved are identified, defined, and linked. The result of process identification is a new or improved process architecture. This architecture shows all of an organization's processes and how they connect to each other. It is used to choose which process or set of processes to handle for the rest of the lifecycle.

- **Process discovery (As-is process modelling)** - This is where the current state of all the important processes is documented, usually in the form of one or more "as-is" process models.

- **Process analysis -** During this step, problems with the current As-is process are identified, documented, and, if possible, measured using performance indicators. A structured list of issues is the outcome of this step. These issues are ranked in order of possible impact and estimated effort needed to fix them.

- **Process redesign (To-be process modelling)** – This phase's objective is to find process modifications that will enable the company to meet its performance targets while also addressing the issues found in the preceding phase. This phase usually results in a To-be process model.

- **Process implementation** - The adjustments needed to transfer the As-is process to the To-be process are planned and carried out during this phase. Automation and organizational change management are the two aspects of process implementation. The term "organizational change management" describes the collection of actions necessary to change the way of working of all participants involved in the process. The creation and implementation of IT systems (or improved versions of current IT systems) to support the future process is referred to as process automation.

- **Process monitoring** – after the implementation of the redesigned process, relevant data is gathered and analyse d to assess the performance of the process. Corrective action is initiated after bottlenecks, recurring errors, or deviations from the intended behaviour are identified.

This cycle must be repeated continuously because new problems might arise in the same or some other processes. This BPM lifecycle is shown in Figure 5.1.
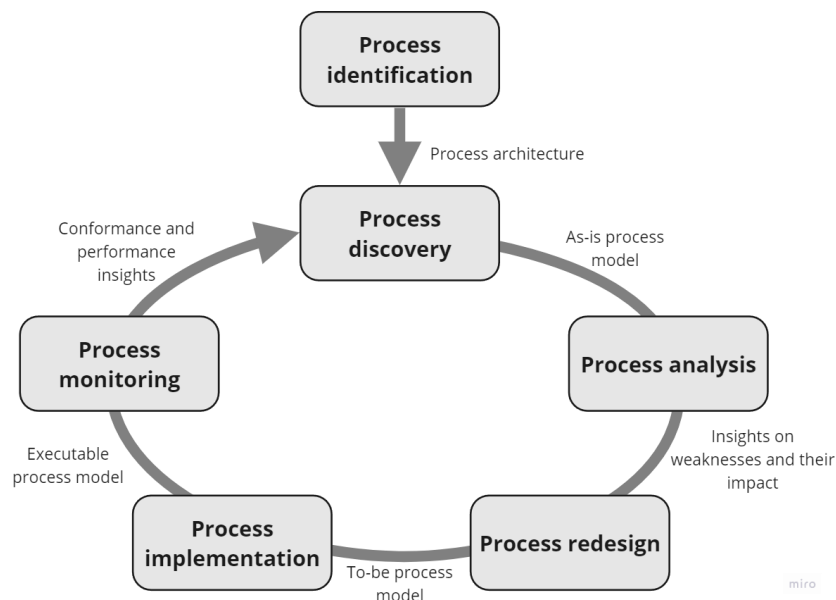
Figure 5.1 BPM lifecycle

Source: Dumas et al. (2018).

Freund and Rücker (2012) list several roles which are involved in BPM projects:

- **Process owner** – person who has strategic responsibility for the processes. He has budget authority, and he is often a member of the first or second tier of management. For example, the process owner can be the company's CEO.

- **Process manager** – person who has operational responsibility for the processes. He is often a low- or middle-level manager. For example, the sales manager could be the process manager.

- **Process participant** – person who works with the process and creates value (e.g. salesperson).

- **Process analyst** – person who understands BPM in general and BPMN in particular, and he is a centre of every BPM project.

BPM helps businesses match their processes with their overall goals, become more efficient, and adapt to changing environments. In the next section, the methods and tools that are used to make accurate models of business processes will be presented.

More than merely drawing diagrams, business process modelling aims to capture the core processes in a way that makes them easier to understand, communicate, and analyse . Stakeholders may use it to visualize complex processes, see inefficiencies and bottlenecks, and conceptualize improvements and innovations.

In the next section, the most popular modelling method BPMN (Business Process Model and Notation) will be presented. It will be discussed how this tool can be used to effectively document business processes.

5.3.   Business Process Modelling

In order to provide standardized, graphical notation for documenting, designing and analysing business processes, the **Business Process Model and Notation (BPMN)** was introduced. According to Lucidchart (n.d.), the Business Process Management Initiative (BPMI) created the Business Process Modelling Notation, which has undergone numerous changes. The initiative was taken over by the Object Management Group (OMG) after that group merged with it in 2005. OMG released BPMN 2.0 and changed the method's name to Business Process Model

and Notation. With a wider range of symbols and notations for Business Process Diagrams, it established a more comprehensive standard for business process modelling.

These four element categories are represented by BPMN (Lucidchart, n.d.; Freund and Rücker, 2012):

- **flow objects:** events, tasks (activities), and gateways,
- **connecting objects:** sequence flow, message flow and association,
- **participants:** pool and lanes,
- **artifacts:** data objects, data store and annotations.

5.3.1. Events

Aagesen and Krogstie (2015) define events as something that happens in a process. There are three types of events in BPMN: start, intermediate and end events. Start event is a trigger for the beginning of the process. Intermediate events occur during the business process and often mark some milestones or waiting in the process. End events mark the end of a business process. They are represented by circles.

Figure 5.2 Start, intermediate and end event notations

Source: Author.

According to Dumas et al. (2018), the event should be named as [object] + [verb in past participle]. Here are some examples of how to name the events: „Invoice sent", „Order submitted", „Products received".

Table 5.1 shows different types of start, intermediate and end events (OMG, 2006).

Table 5.1 Event types

Type	Description	Symbol
Start event		
None	The type of event is not displayed.	○
Message	A message arrives from a participant and triggers the start of the process.	✉
Timer	The process is triggered at the specific time (e.g. every Monday at 9am).	🕐
Conditional	The event is triggered when some condition is met (e.g. when inventory level is lower than 500 pieces).	▤
Intermediate event		
None	The type of event is not displayed.	◎
Message	A message arrives from a participant and triggers the event. This causes the process to continue if it was waiting for the message.	✉
Timer	It can act as a delay mechanism. For instance, if the process is awaiting the delivery of a product.	🕐
End event		
None	The type of event is not displayed.	○
Message	A message is sent to a participant at the end of the process.	✉
Error	An error should be generated at the end of the process.	Ⓝ
Terminate	All activities in the process should be immediately ended.	◉

Source: OMG (2006).

### 5.3.2.	Tasks (activities)

Tasks are something that is carried out during a process, activities performed by a person or system. It is represented by a rectangle with rounded corners.

In BPMN, there is a special subset of regular task called sub-process. It is represented by a rectangle with a '+' sign at the bottom. It serves to represent the process within the process. In this way, the complexity of the main process, i.e. the process in focus, is reduced.

Figure 5.3 Task and Sub-process notations

Source: Author.

The task should be named as [verb in imperative] + [object] (Dumas et al., 2018). Here are a few tasks as examples: „Send the invoice" or „Submit the order".

### 5.3.3.	Gateways

Gateways are locations where processes split or merge. They are represented by the diamond shape. There are three most common types of gateways: XOR (exclusive) gateway, OR (inclusive) gateway, and AND (parallel) gateway.

Figure 5.4 OR, XOR and AND gateway notations

Source: Author.

According to von Rosing et al. (2015), **OR gateway**, when splitting, allows one or more branches to be activated, based on conditions. Before merging, all active incoming branches must be completed in order to continue the flow. An example of an XOR gateway is shown on Figure 5.6.
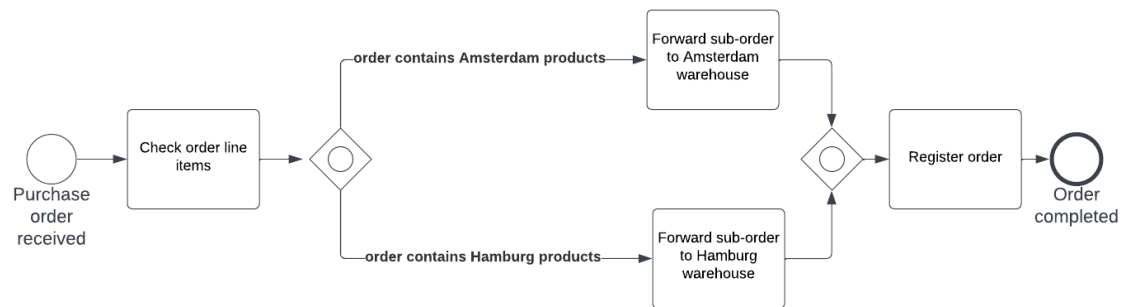
Figure 5.5 An example of the use of OR gateway

Source: Dumas et al. (2018).

In this example, a company has warehouses in Amsterdam and Hamburg, where it keeps different products. Upon receipt of an order, it is split among these warehouses: a sub-order is sent to Amsterdam if certain products are kept there, and a sub-order is sent to Hamburg if certain products are kept there. The procedure then ends when the order is registered (Dumas et al., 2018). We can see that the process can go in both directions (if ordered products are kept in both warehouses) or just in one direction (if ordered products are kept just in one warehouse).

XOR gateway, when splitting, routes sequence flow to only one of the outgoing branches, based on conditions. When merging, it awaits one incoming branch to complete before continuing the flow (von Rosing et al., 2015).

AND gateway is used to execute two or more tasks that do not have any order dependencies on each other and can be executed simultaneously (Dumas et al., 2018). When merging, it awaits all the in branches to complete before continuing the flow (von Rosing et al., 2015). An example of using XOR and AND gateways is shown in Figure 5.6.
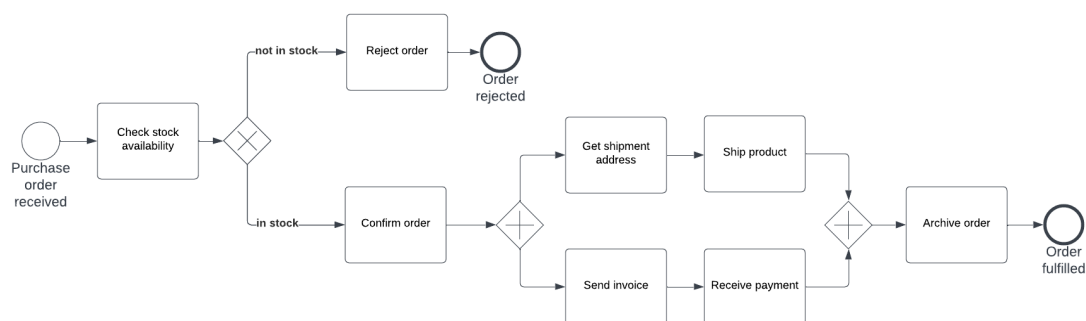
Figure 5.6 An example of the use of XOR and AND gateways
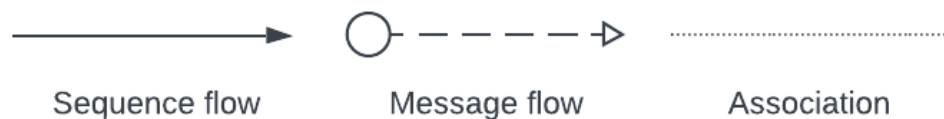
Source: Dumas et al. (2018).

In this example, upon receipt of an order, a salesperson checks stock availability. There is just one and only one possible path – whether the products are in stock or not. On the other hand, it does not matter whether the "Send invoice" or "Get shipment address" activity is carried out first. But only after both sets of activities ("Get shipment address" – "Ship product" and "Send invoice" – "Receive payment") are executed, the order can be archived.

5.3.4. Connecting objects

In BPMN, there are three types of connecting objects: sequence flow, message flow and association.

According to von Rosing et al. (2015), a **sequence flow** shows the order in which tasks will be completed in a process. It is represented by a solid line with a solid arrowhead. A **message flow** is represented by a dashed line. There is a circle on one side of the line and a white arrowhead on the other. It is used to represent the message flow between the process pools. An **association** is used to connect text with flow objects. It is represented by a dotted line.

Figure 5.7 Sequence flow, message flow and association

Source: von Rosing et al. (2015).

5.3.5. Participants

BPMN provides two elements to model process participants: pools and lanes. According to Dumas et al. (2018), **pools** are used to model a whole organization, and a **lane** to model a department or business unit. For example, a pool can be "Company X", and lanes "Sales Department", "Warehouse" and "Accounting". By using pools and lanes, it can be easily seen which participant is doing which activity.
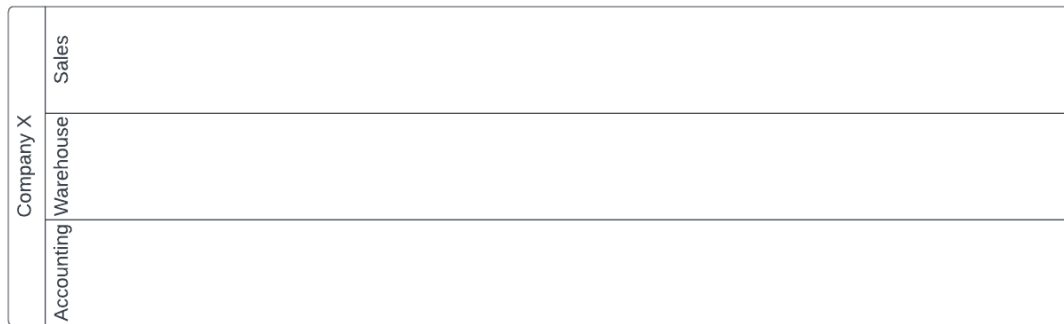
Figure 5.8 Pool and lanes

Source: Author.

5.3.6. Artifacts

There are different types of artifacts: data objects, data store and annotations. **Data objects** represent the data that is required to perform certain tasks (data as input) or is a result of the task execution (data as output). For example, an "Order" document is created after the "Create order" task is executed. On the other hand, the "Send invoice" task requires an invoice as input in order to execute this task. Dumas et al. (2018) states that data objects can be physical objects carrying information (e.g. paper invoice) or electronic objects (e.g. email or an invoice in PDF).

Figure 5.9 Data objects

Source: Author.

According to Dumas et al. (2018), **data store** is a place which contains data objects, e.g. database for electronic objects or a filling cabinet for physical ones. Data stores can be used by process activities to extract and store data objects. For example, "Check raw materials availability" task looks up the supplier's catalogue .
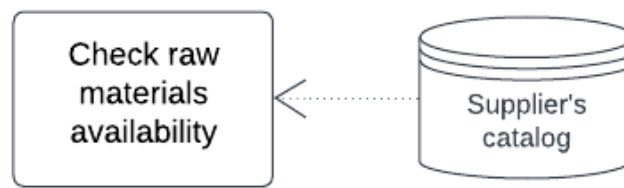
Figure 5.10 Data store

Source: Dumas et al. (2018).

Annotations are a mechanism for a modeler to provide additional text information for the reader of the BPMN diagram (von Rosing et al., 2015).

Figure 5.11 Annotation

Source: Author.

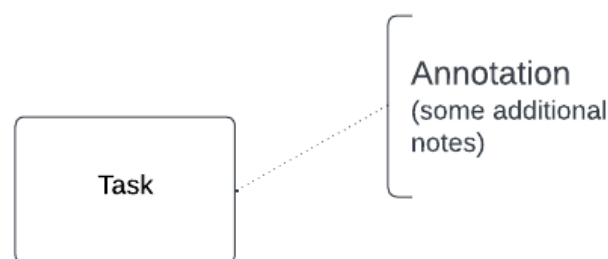BPM has been recognized as an important framework for organizations aiming to optimize their operations and align their processes with strategic objectives. This foundational knowledge is essential for the next topic.

5.4.   Process Mining

Process Mining stands at the intersection of data mining and process modelling. It represents an innovative approach to understanding and enhancing business processes. In contrast to the theoretical and methodological focus of BPM, Process Mining explores the actual (real) data generated by business processes. It uses data from various information systems to provide an objective, real-time view of process execution.

Figure 5.11 shows the difference between BPM and Process Mining. In traditional Business Process Management, a process model is developed first. Then, people and IT systems perform tasks and activities in accordance with this model. In Process Mining, historical data from the IT systems is used to create a process model. This model shows the actual, real processes.
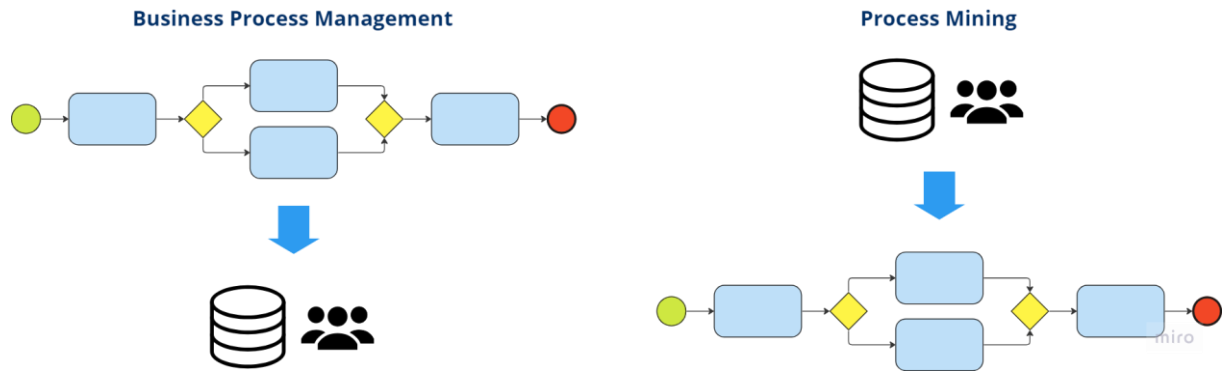
Business Process Management

Process Mining

Figure 5.12 Business Process Management vs. Process Mining

Source: Author.

IEEE (2012) defines **Process Mining** as "techniques, tools, and methods to discover, monitor and improve real processes (i.e., not assumed processes) by extracting knowledge from event logs commonly available in today's (information) systems". An event log is a digital record of the events that have been executed within an information system.

In order to perform a process mining analysis, the event log must contain a case ID, an activity name and a timestamp. A **case** (process instance) is the entity being handled by the process that is analyse d (e.g. customer orders, insurance claims, etc.), an **activity** is a well-defined step in the process (IEEE, 2012) and the **timestamp** is the date and time at which the activity is performed.

Table 5.2 shows an example of an event log. In this example, there are two cases (1001 and 1002), each consisting of a series of events for handling customer inquiries.

Table 5.2 Example of an event log

Case ID	Activity Name	Timestamp	Resource
1001	Call Received	2023-15-12 09:00	Agent A
1001	Issue Identified	2023-15-12 09:15	Agent A
1002	Call Received	2023-15-12 10:17	Agent C
1001	Escalation	2023-15-12 10:20	Agent A
1002	Information Given	2023-15-12 10:26	Agent C
1002	Call Concluded	2023-15-12 10:28	Agent C
1001	Tech Support Call	2023-15-12 11:43	Agent B
1001	Issue Resolved	2023-15-12 11:59	Agent B

Source: Author.

After extracting the data (event logs) from the information systems (e.g. as a CSV or XLS file), the data is imported into special process mining software. Nowadays, there is a wide range of process mining software. The most popular are ProM, Fluxicon Disco, ARIS Process Mining, Celonis etc. Based on the imported data, the process mining software discovers a process model. This model can then be analyse d to determine whether there are some bottlenecks, problems or opportunities for improvement.

According to van der Aalst (2018), Process Mining is applicable to all kinds of operational processes (organizations and systems). Analysing hospital treatment procedures, enhancing customer service procedures in a multinational company, comprehending booking site users' browsing habits, assessing baggage handling system malfunctions, and refining X-ray machine user interfaces are a few examples of applications.

Reil et al. (2021) analyse d a successful implementation of process mining in the practical fields of supply chain management. They stated that in 2020, the Swedish-Swiss energy and automation technology group ABB faced challenges like connecting over 40 ERP systems and managing terabytes of process data. The implementation of process mining in their production processes enabled ABB to gain insights into their global business network performance and move towards a fully digitized supply chain. Benefits included reduced inventory costs, boosted sales processes, improved productivity, on-time deliveries, optimized equipment usage, and increased capacity. The incoming logistics procedures of the automotive supply chain, which are vulnerable to bottlenecks that can cause large revenue losses, benefited greatly from this strategy. Process mining proved useful in efficiently resolving these problems.

BPM's structured way of managing and improving processes makes it possible for businesses to adapt to changing customer needs and operational problems. Process Mining, on the other hand, offers deep insights into actual process performance, highlighting areas of improvement. The integration of BPM and Process Mining is not just a strategic advantage but knowing how to use them and knowing how they work will be important for businesses to be ready for the future.

REFERENCES

1. Aagesen, G. & Krogstie, J. (2015). BPMN 2.0. for Modelling Business Processes. In vom Brocke, J., Rosemann, M. (Eds.). Handbook on Business Process Management 1, 2nd edition. Heidelberg: Springer, pp. 219-250.

2. Camunda (n.d.). What is Business Process Management? [available at: https://camunda.com/glossary/business-process-management-bpm/, access December 28, 2023]

3. Dumas, M., La Rosa, M., Mendling, J. & Reijers, H. A. (2018). Fundamentals of Business Process Management, 2nd Edition. Springer.

4. Freund, J. & Rücker, B. (2012). Real-Life BPMN: Using BPMN 2.0 to Analyse , Improve, and Automate Processes in Your Company. Camunda.

5. Gartner (n.d.). Business Process Management (BPM) [available at: https://www.gartner.com/en/information-technology/glossary/business-process-management-bpm, access December 28, 2023]

6. IEEE (2012). Process Mining Manifesto. IEEE Task Force on Process Mining [available at: https://www.tf-pm.org/upload/1580737631545.pdf, access December 28, 2023]

7. Lucidchart (n.d.). What is Business Process Modelling Notation [available at: https://www.lucidchart.com/pages/bpmn, access December 28, 2023]

8. OMG (2006). Business Process Modelling Notation Specification. Object Management Group.

9. Reil, T., Groher, E. & Siegfried, P. (2021). Process Mining in Supply Chain Management. Supply Chain Management Journal, 12(2), pp. 1-13.

10. Swenson, K. D. & von Rosing, M. (2015). Phase 4: What Is Business Process Management?. In von Rosing, M., Scheer, A.-W., von Scheel, H. (Eds.). The complete business process handbook. Waltham: Morgan Kaufmann, pp. 79-88.

11. van der Aalst, W. (2018). Foreword: Process Mining Book. Fluxicon [available at: https://fluxicon.com/book/read/foreword/, access December 28, 2023]

12. von Rosing, M., Scheer, A.-W. & von Scheel, H. (2015). The BPM Way of Modelling. In von Rosing, M., Scheer, A.-W. and von Scheel, H. (Eds.). The complete business process handbook. Waltham: Morgan Kaufmann, pp. 431-457.

13. Von Scheel, H., von Rosing, M., Fonesca, M., Hove, M. & Foldager, U. (2015). Phase 1: Process Concept Evolution. In von Rosing, M., Scheer, A.-W. and von Scheel, H. (Eds.). The complete business process handbook. Waltham: Morgan Kaufmann, pp. 1-9.

6. INFORMATION SYSTEMS IN LOGISTICS

Author: Dario Šebalj

The integration of technology and management systems plays an important role in improving efficiency, accuracy and strategic decision making. Three essential systems for streamlining business operations in logistics are Enterprise Resource Planning (ERP) systems, Warehouse Management Systems (WMS) and Transportation Management Systems (TMS).

ERP systems form the backbone of a company, integrating various departments (such as accounting, procurement, sales, production etc.) and processes into a unified system. WMS, on the other hand, focuses on optimizing warehouse operations, ensuring effective inventory management and storage optimization. Lastly, TMS is dedicated to the planning, execution and optimization of the transport of goods. This system is critical in reducing transportation costs and improving logistics efficiency.

This chapter not only provides an overview of each system but also explores how integration can lead to a more cohesive and intelligent business environment.

6.1.   Enterprise Resource Planning (ERP) Systems

In the beginning companies were divided into different departments, depending on the functions they performed. Thus, there was a department of production, procurement, sales, finance, etc. Each department operated in isolation in such a way that it had its own data collection and analysis system. Those systems were not connected to each other. Today, organizations are considered as one system, and all the departments are its sub-systems (Leon, 2014). They all share the same, centralized database.

The existence of independent information systems for each department led to inefficiencies, data inconsistencies and redundancy and challenges in decision-making. The shift towards an integrated system was an important approach to organizational management since organizations are seen as a single, unified system. The critical component of this integration

is the centralized database which serves as a core part of the organization, ensuring that all departments have access to consistent, **real-time data**. This leads to better communication, coordination and collaboration between departments.

Figure 6.1 shows the difference between the traditional approach where departments are independent and each department has its own database and the modern approach where departments share one central database.
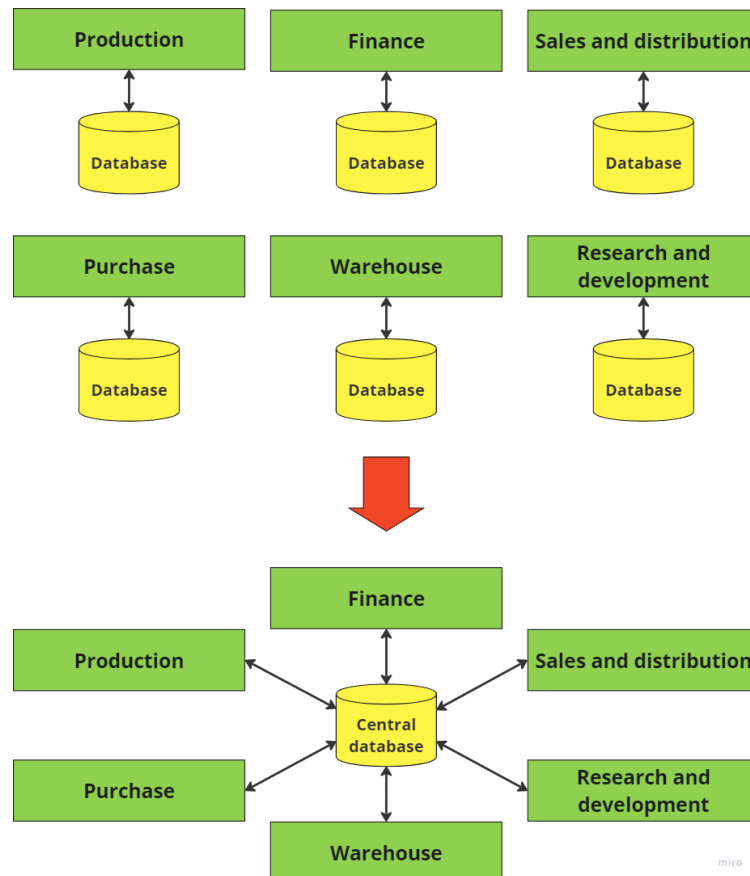
Figure 6.1 Difference between independent departments and departments which share the same central database

Source: Author, according to Leon (2014).

According to Bradford (2015), **enterprise resource planning (ERP)** systems are business systems that combine and organize data from various departments within the organization to create a single, comprehensive system that serves the needs of the whole enterprise. ERP systems integrate and coordinate processes and functions that were previously fragmented and supported by different legacy systems, or older, standalone business systems,

in a seamless manner, improving all aspects of critical operations, including purchasing, accounting, manufacturing, and sales.

In other words, the ERP system is a complex, modular software solution that integrates all of the company's business functions, helps in business process management and shares a single database for the whole system.

An enterprise resource planning (ERP) system is considered as a cross-functional system that automates and integrates the essential business operations of an organization in order to maximize effectiveness and efficiency (Mahmood et al., 2019).

Bradford (2015) states that companies can implement a module or modules of ERP software without having to buy and implement the complete package because most of it is flexible enough.

According to Bradford (2015), ERP systems are often thought of as "back-office" systems since they integrate "back office" functions like order fulfilment, purchasing, accounting, and finance. ERP systems are now more than just back-office systems; they include front office, customer-facing, and supply chain-facilitating modules.

ERP systems have many **advantages**, such as (Bradford, 2015; Paredes Hernandez, 2023):

- Improved transparency and insights – data from every department can be accessed by executive-level employees,
- Real-time access to information – data is available in real-time to all users in all departments,
- Reduction of operational costs – through lower inventory costs, production costs or purchasing costs,
- Single interface through all modules – modules in ERP system look the same and provide the same way of functioning,
- Scalability – cloud-based ERP systems enable the use of additional computing resources in case of company and data growth,
- Improved customer service - a new system such as ERP software can enable more personalized and expedited customer service because it centralizes all customer data.

Some of the **disadvantages** of ERP systems are (Bradford, 2015; Paredes Hernandez, 2023; Oracle, n.d.a):

- Complex and time-consuming implementation – the implementation of an ERP system can take between a few months and several years, depending on the size of the company,

- Price – ERP systems are often very expensive, especially the popular ones: SAP and Microsoft Dynamics NAV,

- Change management - it will take a lot of time and effort to make sure that every important employee is adequately trained on how to use the new system.

The main reason why companies implement ERP systems is to support their company's growth. Also, the implementation of an ERP is the goal of a considerable number of businesses in order to enhance their productivity and processes (Software Path, 2022).

Implementation of ERP systems is very complex and the most ERP projects fail. According to Saunders (2022), about 80% of ERP projects fail. 25% of ERP projects were cancelled or delayed, and another 55% have missed the stakeholders' expectations for the project.

Mahmood et al. (2019) conducted research in which they identified the most critical issues/challenges faced by organizations when implementing the ERP:

1. **Top management support** – the support, strategic direction and active involvement of top management are essential for the successful implementation and management of ERP systems,

2. **Change management** - resistance, particularly from middle managers accustomed to traditional methods, poses significant challenges to adopting new ERP systems,

3. **Training and development** - the complexity of ERP systems requires extensive and ongoing employee training, with insufficient training leading to potential ERP failures and often representing hidden costs for organizations.

4. **Effective communication** - clear and continuous communication and coordination among various departmental users is vital for successful ERP implementation and organizational change management,

5. **System integration** - involves the complex task of integrating various ERP modules with existing business applications and legacy systems within the organization,

a process essential for optimizing business processes and improving efficiency, yet often costly and complex.

Another equally crucial aspect of the ERP system implementation is the financial investment required for the implementation and maintenance of ERP systems. This forthcoming sub-chapter will explore the various components of ERP system costs, encompassing both the initial investment and the ongoing operational expenses.

6.1.1.   Costs of ERP systems

Enterprise Resource Planning (ERP) systems have become an integral part of modern business operations, offering a range of benefits from improved efficiency to enhanced data integration. However, the implementation of such systems comes with significant costs that organizations need to consider carefully.

ERP systems have traditionally been used by companies that sell tangible goods. These comprehensive software solutions were designed to serve large multinational organizations. As a result, their implementation was exceedingly costly and complex. ERP modules like purchasing, sales, and logistics are the foundation for the financial reporting processes, and automating these across a global organization could create significant returns on investment (Berry, 2021).

The total cost of implementing an ERP system includes expenses related to software licensing, hardware requirements, implementation, maintenance, consulting, formal and informal training and customization. These costs are usually referred to as **total cost of ownership (TCO)**. It can vary significantly depending on the scope of implementation, the complexity of the software, and the chosen ERP vendor. For mid-sized organizations, the investment in packaged ERP software alone can amount to a few million dollars (Leon, 2014; Tilley, 2020).

In addition to software costs, the implementation of ERP systems often requires substantial investments in IT infrastructure. This includes servers, storage systems, network components, and possibly upgrading existing components that are nearing the end of their life cycle (Bradford, 2015). While cloud computing can reduce some of these hardware costs as the ERP software runs on the vendor's servers, the initial investment in infrastructure remains a significant component of the overall cost.

The hidden costs associated with ERP implementations, such as consultancy fees, also play a significant role in the total expenditure. These costs include the fees for external consultants who are familiar with the ERP system but may not have in-depth knowledge of the organization's specific business processes (Leon, 2014).

Nearly 80% of total costs occur after the purchase of the hardware and software (Tilley, 2020).

The cost of ERP software is influenced by various factors. For example (Hale, 2019; Wood, 2023):

- **Deployment method** – ERP systems can be deployed on a cloud, on premise or as a combination of the two.
- **Number of users** – ERP systems with a lower number of users could cost less.
- **Applications required** – a number of modules can range from core modules to some specific modules.
- **Customization level** – any additional upgrades of the initial software increase the price of the ERP system.
- **User training and support** – usually, but not always, implementation fees include a year of customer support. Real-time, consistent support may come at an additional expense.
- **Hardware upgrades** – companies may need to buy extra hardware (e.g. servers, storage, network infrastructure) during implementation in order to support their new ERP.

The average budget per user for an ERP project, according to a Software Path (2022) report, is $9,000. However, this cost varies based on the size of the business and the number of users. According to Hale (2019), maintenance costs may amount to 10% to 20% of the initial license fee.

6.1.2.   ERP systems trends

In recent decades organizations have spent millions of dollars in implementing ERP systems (Ruivo et al., 2020). The ERP software revenue is growing 8% year over year to a market

value of $44 billion in 2023 (Haranas, 2023) and it is projected to reach $62 billion by 2028 (Statista, 2023).

Nowadays, there is a large number of ERP software vendors. According to Davidson (2023) top ERP software vendors are Microsoft, SAP, Oracle, Sage, Epicor and Infor.

When it comes to purchasing ERP software, manufacturing is the industry with the highest representation (27%). At 20%, construction is in the second place. Together, distribution and transportation, which are included in the supply chain industry's broader definition, account for 16% (Wood, 2023).

According to Statista (2023), the **requirement for customization** is one of the primary customer preferences in the ERP software market. Software that can be customized to meet the unique demands and specifications of businesses is essential. The need for flexible and scalable cloud-based ERP solutions has increased as a result. Customers also want software that is simple to use and seamlessly integrates with other systems.

The future and trends of Enterprise Resource Planning (ERP) systems are shaped by evolving business needs and technological advancements. As of 2023, several key trends are prominent in the field of ERP (Luther, 2023):

- **Cloud ERP** - Cloud-based ERP solutions are becoming increasingly popular due to their simpler deployment, lower costs, elasticity, and the ability to accommodate business growth. The pandemic has accelerated the shift from on-premises software to cloud ERP, as these systems allow employees to work remotely with ease. According to Wood (2023), 42% of companies used Cloud-based ERP in 2022 (compared to 2013, when this percentage was only 4%). Typically, cloud ERP is provided as software as a service (SaaS), which means users must pay a monthly, quarterly, or annual fee for ongoing access.

- **Two-Tier ERP** - The two-tier ERP approach is gaining traction. This strategy uses a primary ERP system at the corporate level, while subsidiaries and divisions operate using a different, often cloud-based, ERP solution. Larger companies might keep using their main ERP system for financials and other core processes, while smaller business units would look for solutions tailored to their specific requirements.

- **Digital transformation** - ERP systems are playing a crucial role in the digital transformation of businesses. By integrating digital technology into all business functions, ERP systems are boosting revenue, competitiveness, and improving customer service and communication.

- **Integration with other technologies** - Modern ERP systems are increasingly integrated with other technologies, such as IoT and social media, to enhance core processes and provide greater visibility and a better customer experience.

- **Personalization** - ERP systems are evolving to offer more personalized experiences to customers, supported by AI-based assistive and conversational user interfaces like chatbots. This trend is facilitated by cloud ERP platforms designed for easier configuration and industry-specific solutions.

- **AI-Powered insights and improvements** - AI and machine learning are being embedded into ERP systems, providing valuable business insights by analysing operational and customer data. This integration helps in optimizing a range of business processes and improving personalisation.

- **Predictive analytics** - The use of predictive analytics in ERP systems is on the rise, focusing on analysing data to predict future trends and outcomes, which aids in better decision-making and strategic planning.

- **Mobile ERP** - Mobile ERP is becoming more common, offering on-the-go access to critical business data and facilitating remote work. Mobile ERP apps with user-friendly interfaces help employees to complete tasks efficiently, irrespective of their location.

These trends indicate a significant shift in ERP systems towards more adaptable, personalized, and integrated solutions that align with modern business practices and technological advancements.

ERP systems manage basic supply chain functions like inventory control and order fulfilment, but they are typically very basic. Their main objective was to assist with financial processes. The inventory management module of an ERP was not very good at managing the labour in the warehouse, but it might be pretty good at tracking inventory valuation for the enterprise balance sheet. As a result, the market saw the emergence of best-of-breed logistics applications that could complement the ERP and close gaps. Warehouse management systems

(WMS) and transportation management systems (TMS) emerged as the two main categories of logistics applications (Berry, 2021).

6.2.   Warehouse Management Systems

From the moment materials or goods enter a distribution or fulfilment centre until they leave, a **warehouse management system (WMS)** enables companies to monitor and manage warehouse operations. The primary objective of a WMS is to facilitate the efficient and economical movement of materials and goods through warehouses. Picking, receiving, put away, and inventory tracking are just a few of the many tasks that a WMS performs to facilitate these movements. WMS software systems provide real-time visibility into a company's entire inventory, both in transit and warehouses, and are a crucial part of supply chain management (O'Donnell, 2020).

According to SAP (n.d.a), a warehouse management system optimizes various warehouse activities. It streamlines the receiving and put-away process using RFID technology and integrates with other software for efficient item handling. In inventory management, WMS provides real-time visibility and supports advanced analytics for better stock control. For order picking, packing, and fulfilment, it guides efficient storage, retrieval, and packing, employing technologies like RF scanning and robotics to optimize order processing. Shipping processes are enhanced by integrating with logistics software, ensuring timely and accurate deliveries. WMS also aids in labour management, offering insights into labour costs and productivity, and supports efficient task management. Additionally, it facilitates yard and dock management, improving loading efficiency, and supports cross-docking for perishable goods. Finally, WMS provides valuable warehouse metrics and analytics, enabling better decision-making and process optimisation.

SAP (n.d.a) lists 5 benefits of a WMS:

1. **Improved operational efficiency** - WMS systems enhance efficiency and handling volumes by automating and streamlining warehouse processes from inbound receipts to outbound deliveries.

2. **Reduced waste and costs** - WMS helps in minimizing waste, especially for date-restricted or perishable stock, and optimizes warehouse space utilisation.

3. **Real-time inventory visibility** - It offers real-time insight into inventory movement, aiding in accurate demand forecasts and improved traceability.

4. **Improved labour management** - WMS aids in forecasting labour needs and optimising task assignments based on various factors, thereby enhancing employee morale.

5. **Better customer and supplier relationships** - WMS leads to improved order fulfilment and faster deliveries, increasing customer satisfaction and enhancing supplier relations.

Warehouse management system development is still being influenced by technological advancements. For example, those are (Scullin, 2023):

- **Automated picking tools** - technologies like voice automated order picking, robotic order picking, and pick-to-light systems, coupled with sophisticated barcoding,

- **Automatic Guided Vehicles (AGVs)** - enhance storage and retrieval processes, crucial for tasks like pallet and rack storage, container management, and automating the receiving process,

- **Internet of Things (IoT)** - By integrating IoT, various automated and manual elements are controlled within a unified network, enhancing inventory control, labour planning, and customer experience through faster fulfilment rates,

- **Augmented (AR) and virtual reality (VR)** - AR technology, through devices like smart glasses, provides real-time overlays of instructions or information in a warehouse environment, aiding in tasks like route navigation and bin location without the use of hands. VR is utilized for training and safety purposes, such as training lift truck operators and improving delivery routes.

Berry (2021) states that the WMS market is very mature, and there are many well-known software companies that offer a wide range of features to help with even the most complicated warehouse tasks. A lot of the top WMS providers now offer delivery models in the cloud. 40-50% of new WMS customers now choose cloud delivery over on-premises deployments. Some of the popular WMS vendors are (Gartner, n.d.): SAP Extended Warehouse Management (EVM), Oracle Warehouse Management (WMS Cloud), Microsoft Dynamics 365 Supply Chain, Manhattan WMS and Infor WMS.

## 6.3.   Transportation Management Systems

A Transportation Management System (TMS) is a crucial software in logistics that optimizes the movement of goods across various modes of transport. As part of a broader Supply Chain Management system, TMS optimizes load and delivery routes, tracks freight, and automates tasks like trade compliance and freight billing. This system not only ensures timely delivery but also reduces costs, thereby benefiting both businesses and customers. It offers comprehensive visibility into transportation operations, aids in compliance, and simplifies the shipping process across land, air, or sea (SAP, n.d.b; Oracle, n.d.b).

According to Berry (2021), there are several ways that a TMS can lower the cost of transportation. The shipping department can save a lot of time and effort by automating the process of booking and tracking shipments. Routing guide capabilities ensure that shipping clerks choose the method of transport that costs the least for each shipment. A lot of TMS products can optimise less-than-truckload shipments into full-truckload shipments, which are much cheaper.

Some of the benefits of TMS are (SAP, n.d.b; Inbound Logistics, 2023):

- **cost savings** – TMS significantly reduces both administrative and shipping costs, optimizing load and freight management,
- **real-time visibility** – provides critical insights into the transportation process, enhancing route efficiency and tracking,
- **greater customer satisfaction** – ensures on-time delivery and improves customer experience through better tracking and billing processes,
- **improved efficiency** – TMS enhances the overall efficiency of transportation operations,
- **enhanced decision making** – offers valuable data for informed decision-making, improving strategic planning in transportation management.
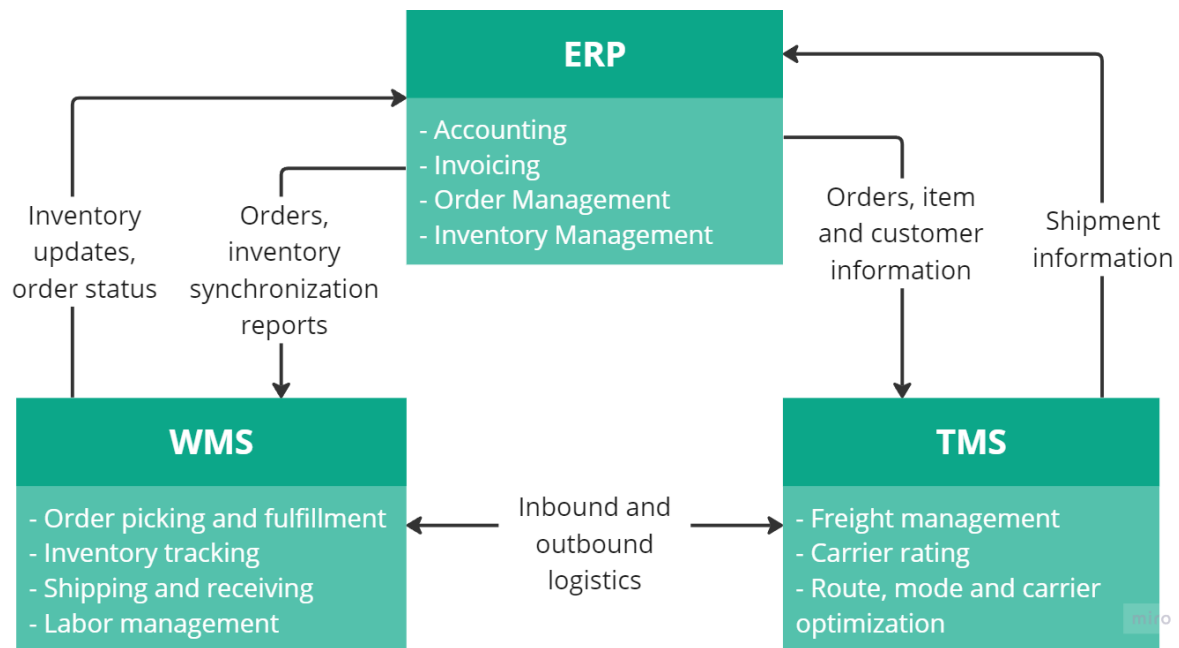
Figure 6.2 The connection between ERP, WMS and TMS

Source: Essex (2020).

Figure 6.2 shows the connection between ERP, WMS and TMS systems. According to Essex (2020), an ERP system manages accounting, invoicing, order, and inventory management. The WMS assists with fulfilment, shipping, and receiving tasks in a warehouse, such as picking and storing goods, and updates the ERP system's inventory module with real-time data from barcode and RFID scans. The ERP system provides order details to the TMS for shipment preparation and execution. The TMS returns shipment details to the ERP for accounting and order management, and potentially updates customer relationship management (CRM) modules for customer updates on order status.

In this chapter, the important role of ERP, WMS and TMS systems for logistics was described. These systems, critical in modern logistics, collectively enhance efficiency, ensure precise inventory management, and optimize transportation processes. The integration of ERP, WMS, and TMS is not just a technological advancement but a strategic necessity, driving businesses towards greater efficiency, accuracy, and customer satisfaction in the field of logistics and supply chain management.

REFERENCES

1.   Berry, J. (2021). Logistics in the Cloud-Powered Workplace. In Sullivan, M. & Kern, J. (Eds.). The Digital Transformation of Logistics. Piscataway: IEEE Press.

2.   Bradford, M. (2015). Modern ERP: Select, Implement, and Use Today's Advanced Business Systems, 3rd Edition. Lulu.com

3.   Davidson, R. (2023). Top 6 ERP Software Vendors. SoftwareConnect [available at: https://softwareconnect.com/erp/top-vendors/, access January 15, 2024]

4.   Esex, D. (2020). Transportation management system (TMS). TechTarget [available at: https://www.techtarget.com/searcherp/definition/transportation-management-system-TMS, access January 15, 2024]

5.   Gartner (n.d.). Warehouse Management Systems Reviews and Ratings [available at: https://www.gartner.com/reviews/market/warehouse-management-systems, access January 17, 2024]

6.   Hale, Z. (2019). What Factors Determine the Cost of ERP Software?. Software Advice [available at: https://www.softwareadvice.com/resources/erp-software-pricing/, access January 15, 2024]

7.   Haranas, M. (2023). Oracle, Microsoft, SAP, Workday Lead Cloud ERP Market: Gartner. CRN [available at: https://www.crn.com/news/cloud/oracle-microsoft-sap-workday-lead-cloud-erp-market-gartner, access January 17, 2024]

8.   Inbound Logistics (2023). Transportation Management System: Meaning, Importance,   and   Benefits   [available   at: https://www.inboundlogistics.com/articles/transportation-management-system/, access January 17, 2024]

9.   Leon, A. (2014). ERP demystified, 3rd edition. McGraw Hill Education.

10.   Luther, D. (2023). 8 ERP Trends for 2023 & The Future of ERP. Oracle Netsuite [available   at: https://www.netsuite.com/portal/resource/articles/erp/erp-trends.shtml, access January 15, 2024]

11.   Mahmood, F., Khan, A. Z. & Bokhari, R. H. (2019). ERP issues and challenges: a research synthesis. Kybernetes, 49(3), pp. 629–659.

12. O'Donnell, J. (2020). Warehouse management system (WMS). TechTarget [available at: https://www.techtarget.com/searcherp/definition/warehouse-management-system-WMS, access January 15, 2024]

13. Oracle (n.d.a). What are the benefits of an ERP system? [available at: https://www.oracle.com/hk/erp/what-is-erp/erp-benefits/, access January 20, 2024]

14. Oracle (n.d.b). What Is a Transportation Management System? [available at: https://www.oracle.com/scm/logistics/transportation-management/what-is-transportation-management-system/, access January 15, 2024]

15. Paredes Hernandez, J. (2023). The advantages and disadvantages of ERP systems. IBM [available at: https://ibm.com/blog/enterprise-resource-planning-advantages-disadvantages/, access January 16, 2024]

16. Ruivo, P., Johansson, B., Sarker, S. & Oliveira, T. (2020). The relationship between ERP capabilities, use, and value. Computers in Industry, 117, 103209.

17. SAP (n.d.a). What is a warehouse management system (WMS)? [available at: https://www.sap.com/products/scm/extended-warehouse-management/what-is-a-wms.html, access January 15, 2024]

18. SAP (n.d.b). What is a transportation management system (TMS)? [available at: https://www.sap.com/products/scm/transportation-logistics/what-is-a-tms.html, access January 15, 2024]

19. Saunders, P. (2022). Are ERP Projects Really The Stuff Of Nightmares?. Forbes [available at: https://www.forbes.com/sites/sap/2022/06/28/are-erp-projects-really-the-stuff-of-nightmares/, access January 20, 2024]

20. Scullin, Ch. (2023). 7 Smart Warehouse Technologies to Implement Today. Camcode [available at: https://www.camcode.com/blog/smart-warehouse-technologies/, access January 10, 2024]

21. Software Path (2022). What 1,384 ERP projects tell us about selecting ERP (2022 ERP report) [available at: https://softwarepath.com/guides/erp-report, access January 15, 2024]

22. Statista (2023). Enterprise Resource Planning Software – Worldwide [available at: https://www.statista.com/outlook/tmo/software/enterprise-software/enterprise-resource-planning-software/worldwide, access January 15, 2024]

23. Tilley, S. (2020). Systems Analysis and Design, 12th Edition. Boston: Cengage Learning.

24. Wood, L. (2023). How Much Does ERP Cost?. SoftwareConnect [available at: https://softwareconnect.com/erp/pricing/, access January 22, 2024]

7. E-LOGISTICS

Author: Michał Adamczak

This chapter is devoted to the most important issues related to e-logistics. The chapter not only defines this concept but also presents it in a broader context of the possibilities offered by data analysis to optimize logistics processes. The chapter includes topics such as:

- the context in which e-logistics operates, including the concept of e-business,
- basic definitions of e-logistics,
- development of e-logistics,
- modern e-logistics technologies and tools,
- practical e-logistics solutions.

7.1. Introduction

The development of digital technologies has a long history. Therefore, it cannot be said that digital solutions or sharing information in supply chains is a modern solution. On the contrary, from the perspective of time and the point of view of professionally active people, the digital aspect is already a mature solution that has become a permanent part of the course of logistics processes. In other words, we can no longer imagine, let alone operate in logistics without a parallel flow of digitally recorded information.

The modern economy is called the post-industrial economy or the digital economy. However, this does not mean that the flow of materials has been completely stopped or abandoned. It is the flow of materials that is key to economic turnover and consumption. This will continue to be the case as long as people's needs are met by material goods. Of course, some people's needs can be met by digital content, but in the foreseeable future, it will not be possible to meet all people's needs with digital goods. Thus, it seems that the coexistence of material and digital flows will constitute an inseparable tandem for the next decades. The term digital economy is intended to emphasize the role and scope of material and information flows. As will

be shown in this chapter, digital flow is becoming increasingly important for building conditions that improve the efficiency of material flow and thus improve the competitive position of specific enterprises and entire supply chains.

E-logistics is therefore a solution that fits into the mainstream of the modern economy. It is both a response to the requirements of the modern economy and a solution that provides new opportunities for doing business.

7.2. E-business

E-logistics is a solution operating within the broader concept of e-business. E-business can be loosely defined as a business process that uses the Internet or other electronic medium as a channel to complete business transactions (Jayashankar et al., 2003). Within e-business, we can distinguish such detailed activities as: e-commerce, e-advertising, e-marketing, electronic banking, electronic auctions etc. In these types of business activity, the adjective "electronic" indicates that these activities are carried out strictly in electronic (digital) form using the Internet, mobile connection etc. (Skitsko, 2016). The location of e-logistics within e-business and other concepts using data transmission via the Internet is presented in Figure 7.1.
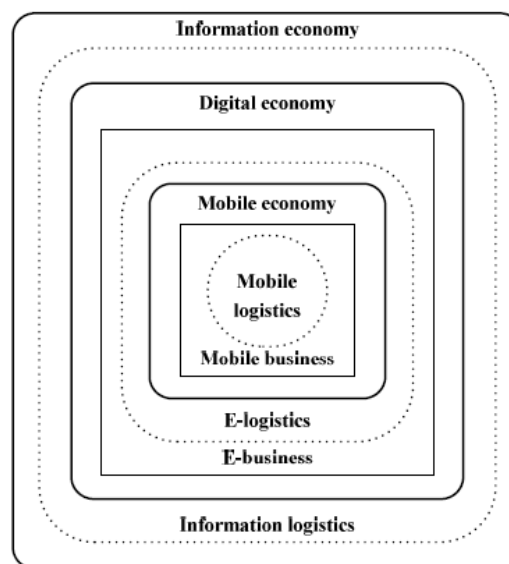
Figure 7.1 E-logistics in concept of e-business

Source: Skitsko (2016)

Within e-business, there are several basic models of communication between market participants (Shemet, 2012):

- B2B (business-to-business). In this model, there is interaction among the companies (enterprises, legal entities) intending to obtain various benefits.

- B2C (business-to-customer). In this model, the company interacts with its end consumer.

- C2C (customer-to-customer). In this model, people (physical persons) interact with each other with the help of various communication means and technology.

- C2B (customer-to-business). In this model opinions or ideas of the end consumers expressed by various means, in particular, on various Internet forums, social media, email etc. considerably influence the products making (their characteristics, features, price etc.) by the producer.

- B2G (business-to-government). In this model, the company interacts with the state administrative organs.

- C2G (customer-to-government). In this model, there is interaction between a person and state administrative organs.

- G2B (government-to-business), G2C (government-to-customer). In these models state administrative organs provide the companies (enterprises) and physical persons with information services via Internet.

The impact of the development of data processing technologies and the Internet on supply chains can be distinguished in three areas (Jayashankar et al., 2003):

- development of systems supporting enterprise management (ERP) and material flow planning (APS);

- development of systems supporting the business decision-making process that operates in real-time;

- sharing information between enterprises.

All of the areas mentioned above also occur in the context of implementing logistics processes, which became the basis for the creation of the e-logistics concept.

7.3. Definition of e-logistics

At the outset, it is difficult to provide a single definition of e-Logistics. This is because it is a concept very closely linked to the technical possibilities of acquiring, collecting, processing and transmitting data and information. Therefore, the definitions of this concept have changed over time and will probably continue to change.

„E-Logistics is a dynamic set of communication, computing, and collaborative technologies that transform key logistical processes to be customer-centric, by sharing data, knowledge and information with the supply chain partners." (Wang et al., 2004)

Another interesting definition, although narrow in scope, is presented by a team of authors Quirk, Forder and Bentley. *„E-Logistics is using Internet-based technologies for supporting the acquisition of material, warehousing, transportation and enables distribution through routing optimization with inventory tracking"* (Quirk et al., 2003)

Both of the above definitions focus on the aspect of data that accompanies the flow of materials in supply chains. The task of e-logistics is, therefore, to track the flow of materials to better control it and provide information about this flow in real-time to all its stakeholders, which in turn will enable the synchronization of this flow in the supply chain (Mangiaracina et al., 2015).

According to another definition, e-logistics are logistic processes implementing the flow of products purchased in electronic sales channels (Erceg & Damoska Sekuloska, 2019). An illustration of this way of understanding e-logistics is presented in Figure 7.2.

Figure 7.2 E-logistics

Source: Moroz et al. (2014)

By comparing both approaches to defining e-logistics, the basic characteristics of traditional logistics and e-logistics supporting the flow of materials in e-commerce were compared. The results of the comparison are presented in Table 1.

Table 7.1 Basic differences between traditional and e-logistics

Scope	Traditional logistics	E-logistics
Shipment type	Bulk	Parcel
Customer	Strategic	Unknown
Customer service	Reactive, Rigid	Responsive, flexible
Distribution model	Supply-driven push	Demand-driven pull
Inventory / Order flow	Un-directional	Bidirectional
Destinations	Concentrated	Highly dispersed
Demand	Stable consistent	Highly seasonal, fragmented
Orders	Predictable	Variable

Source: Song & Hou (2004)

To sum up the above definitions, their similarities should be pointed out. Logistics activities carried out for the needs of material flow are very similar to each other, regardless of whether they concern traditional flow or those carried out within e-commerce. When describing e-logistics, it should be noted that in both approaches, this concept is related to the flow of data describing the material flow. The basic functions of e-logistics are the same for both areas (Skitsko, 2016):

- formation of an information environment in which interact the participants of the logistics chain of goods supply;
- definition of characteristics of electronic information flows;
- formation of requirements and needs to the companies which provide information and communication services and corresponding connections;
- organization of the use of international standards of product identification;
- maintenance of correct and reliable operation, development of the information system of the company;
- collection, analysis, storage, transformation and organization of information transfer in electronic form;
- selection of the necessary data for management decision-making.

The implementation of these functions would not be possible without digital technologies that enable the collection, collection and analysis of data. The description of the most important of them, which had the greatest impact on the development of e-logistics, is presented in the next subsection.

7.4.   Development of e-logistics

Based on the presented definitions, it can be clearly stated that the beginning of e-logistics dates back to the times when the first IT systems supporting the management of material flow, material requirement planning (MRP) and distribution resource planning (DRP) systems were created. These systems began to develop in the 1960s. They were the first solutions for the parallel flow of materials and digitally recorded information. The following years saw the dynamic development of these systems, which led to the creation of enterprise resource planning (ERP). In parallel, systems dedicated to individual logistics functions were developed:

transport management systems (TMS) and warehouse management systems (WMS) (Wang, 2016). More details on IT systems can be found in Chapter 6 of this handbook.

The development of ERP systems, and especially the concentration of data and the multidimensionality of this data, allowed the creation of decision support systems (DSS) (Turbanet al., 2002). The development of the Internet and the possibility of exchanging data between the systems of individual enterprises initiated the development of ERPII class systems allowing for the integration of data between partners in supply chains (Møller, 2005). Data exchange between partners is possible thanks to the Electronic data interchange (EDI) solution (Huang, et al., 2008).

Another milestone in the development of e-logistics was the creation of electronic marketplaces (EM). The creation of platforms connecting enterprises directly with customers (and other configurations presented in the subsection on e-business) allowed for the creation of new business models and thus requirements for logistics (Wang, et al., 2007).

In parallel with the development of EM, systems for collecting and analysing large data sets were developed, allowing for the implementation of computing processes in the cloud. The development of big data collection technology and the possibility of analysing it and sharing analytical tools and analysis results remotely via the Internet has provided completely new possibilities, especially in the area of information supply to DSS and, consequently, the possibility of optimizing logistics processes, especially in such areas as: forecasting, inventory management, transport management and human resources management (Waller & Fawcett, 2013). To sum up the development of digital technologies used in e-logistics, we can use the observation of the authors Merali, Papadopoulos and Nadkarni (2012), who presented four-step changes in ICTs since the 1960s, which had a major influence on the e-logistics development (Merali et al., 2012):

- connectivity (between people, applications, and devices);
- capacity for distributed storage and processing of data;
- reach and range of information transmission;
- rate (speed and volume) of information transmission.

Undoubtedly, the above-mentioned steps in the development of ICT technologies have influenced the possibilities of the practical application of digital solutions in logistics processes.

These changes also clearly present the direction in which digital technologies are developing. The technologies that are currently used in e-logistics are described in more detail in the next subchapter.

7.5.   Modern technologies supporting e-logistics

The development of Industry 4.0 and Logistics 4.0 provides additional opportunities to expand the solutions and services offered within e-logistics. Among the main technologies supporting e-logistics are currently:

- Blockchain;
- Internet of Things and sensors (IoT);
- Generative Artificial Intelligence (AI);

Blockchain is a distributed database system between all participants in the same network. This system records and stores data in the form of linked blocks forming a collection of records. They are permanent and therefore cannot be deleted. It is important to know that there is no possibility of making updates or any modifications. However, it is possible to add or read a recording (Dutta et al., 2020).

Blockchain technology makes it possible to track different transactions along the whole supply chain in a secure and traceable manner. The documented transactions and data are irrevocably stored in the blockchain and cannot be used or read without consensus. Every time a consignment is being transported or handled, the transaction can be documented, creating a permanent history from the manufacturer to the trader or consumer (Aritua et al., 2021).

The Internet of Things (IoT) allows non-computer devices to communicate with each other. The concept is based on a wide range of technologies, from communication protocols through sensors collecting data, infrastructure enabling data transmission to systems analysing the collected data (Minerva, 2015). IoT solutions are often combined with RFID (radio-frequency identification) sensors, giving the possibility of not only local identification of goods or cargo but also transferring this data to any user. IoT solutions can be created in two variants (Idrissi et al., 2022):

- Internet-centric – the main element of the system are services offered in cloud computing and the system objects are data providers;

- Object-centric – a solution in which the central point of the network is object that can be controlled using messages transmitted over the Internet.

IoT solutions are widely used in logistics. The IoT makes it possible to trace various information to control the quality of the goods such as light, humidity, temperatures, vibrations, shocks, etc. (Dash et al., 2019). For example, at Maersk, a container carrier wishes to market a service that requires additional insurance on the whole journey. The transport conditions (vibration, temperature, humidity, magnetism, position, etc.) can be monitored in the instrumented container. This information can be also uploaded to the Blockchain to trigger partial payments during shipment AI is the simulation of human intelligence processes by machines and computer systems. Knowledge generation by artificial intelligence is carried out in three steps (Samoili et al., 2020):

- learning - the acquisition of information and its rules of use;
- reasoning - the use of rules to conclude;
- self-correction.

The AI application allows the system to give precise indications to each operator on each order. The system can do this through history-based learning. This helps achieve maximum efficiency, especially in picking-intensive warehouses, such as e-commerce (Dash et al., 2019).

> The presented technologies do not constitute a closed catalogue of solutions used within e-logistics. The cooperation of these technologies within the acquisition, collection and processing of data in order to create information supporting effective managerial decisions is particularly important.

7.6.   E-logistics in practice

Regardless of how e-logistics is defined, these solutions function in virtually every aspect of logistics activity, regardless of the function or material flow phase. According to the literature review presented earlier, attention should be focused on the connection between suppliers and recipients. This is where data exchange and connecting entities to improve

the efficiency of material flow seem particularly important. This is currently possible thanks to the generally accessible Internet and automatic data acquisition. Practical e-logistics solutions are offered by virtually all logistics operators, especially those operating on the global market. An excellent example of solutions used within e-logistics is those offered by Dachser. This European logistics operator provides its customers with a direct connection to transport and warehouse management systems, thanks to which customers have uninterrupted real-time access to data on the implementation of this operator's logistics processes. The functions offered by Dachser (n.d.) within e-logistics include:

- product and service analysis - this tool allows you to quickly determine the optimal or desired delivery times for shipments within Europe;
- online ordering - automatic import of data to orders saves time. The address import function from the ERP system complements address management. This functionality also allows you to send documents, save information on dangerous goods, as well as send future orders and use your own barcodes;
- control of all transport costs - allows you to quickly obtain information on the transport price without having to submit extensive inquiries;
- inventory tracking - allows you to track processes taking place in warehouses - from checking the status of order receipts to batch monitoring. This functionality allows you to immediately determine shortages and inventory levels;
- current information on the status of the shipment and its location - the Track & Trace function allows you to create an individual link for each shipment, which will inform you about the current status of the shipment. This link can then be forwarded to customers or partners;
- online invoice management - online access to all shipment data. Data is available in PDF files, Excel tables and CSV files. We can also send this data digitally via the EDI centre.;
- electronic pallet accounting - manages loading equipment that requires tracking, i.e. euro pallets and racks.

Another global logistics operator that uses e-logistics solutions on a large scale is DHL. In addition to the very similar functions presented above for another operator, DHL also uses solutions from the field of machine learning, augmented reality and artificial intelligence to

a large extent. Augmented reality is used to optimize the warehouse infrastructure and logistics operations conducted there. Machine learning and artificial intelligence are used to increase the efficiency of the business and increase the organization's resilience by focusing the actions taken on proactive instead of reactive actions. Taking proactive actions is possible thanks to the analysis of large data sets and searching for relationships between causes and effects in them. It is therefore possible to predict the formation of future phenomena based on past events. Such actions also have an impact on increasing the value of services directed to DHL customers and increase their competitive position (DHL, 2017). This shows that a logistics operator can offer not only classic logistics services in the form of transport, storage or order handling but also advanced services in the field of data analysis and recommending solutions resulting from these analyses. E-logistics solutions are therefore becoming a source of competitive advantage, and the services resulting from them are a natural element of cooperation between the links of the supply chain.

7.7. Summary

The solutions operating within e-logistics are as diverse as the definitions of this concept. Two main trends in defining this concept can be distinguished. In a broader sense, e-logistics are all kinds of digital solutions that accompany the flow of materials. In a narrower sense, e-logistics is defined as the implementation of logistics processes accompanying e-commerce. Of course, both approaches are not mutually exclusive. The presented history of development, the presented expansion of the scope in which e-logistics functions and the expected directions of development clearly show that regardless of the way in which this concept is defined, it will be the object of interest of both business practitioners and researchers.

Although, as noted in the introduction to this chapter, material flow will not be replaced by information flow, information flow largely determines the efficiency of material flow. Supporting information processes implemented within e-logistics with methods and data analysis tools seems to be particularly important in this respect. Modern technical solutions allow for the collection of large sets of data and the search for relationships between these data in order to prepare information useful in making managerial decisions.

Detailed solutions in the field of data analysis, data preparation for making managerial decisions have been discussed in the remaining chapters of this manual. They present not only

business analytics concepts but also ERP systems that allow for data collection, BI tools that allow for data analysis and visualization, as well as modern issues related to the use of machine learning in data analysis and data security.

> The lack of a clear context for defining the concept of e-logistics is caused by the rapid development of the subject and the blurring of the boundaries between individual solutions supporting the implementation of information flow.

REFERENCES

1. Aritua, B., Wagener, C., Wagener, N. & Adamczak, M. (2021). Blockchain solutions for international logistics networks along the new silk road between Europe and Asia, Logistics, 5(3), pp. 1-14.

2. Dachser (n.d.). eLogistics: Internetowy portal do zarządzania logistyką [available at: dachser.pl/pl/elogistics-116, access April 07, 2024]

3. Dash, R., McMurtrey, M., Rebman, C. & Kar U.K. (2019). Application of Artificial Intelligence in Automation of Supply Chain Management, Journal of Strategic Innovation and Sustainability, West Palm Beach, 14(3), pp. 43-53.

4. DHL 2017. The 21st Century Spice Trade: A Guide to the Cross-Border E-Commerce Opportunity [available at: http://www.dhl.com/content/dam/downloads/g0/press/publication/g0_dhl_express_cross_border_ecommerce_21st_century_spice_trade.pdf, access June 23, 2018].

5. Dutta, P., Choi, T.M., Somani, S. & Butala, R. (2020). Blockchain technology in supply chain operations: applications, challenges and research opportunities. Transp Res Part E: Logist Transp Rev,142(102067).

6. Erceg, A. & Damoska Sekuloska, J. (2019). E-logistics and e-SCM: how to increase competitiveness. LogForum, 15(1), pp. 155-169.

7. Huang, Z., Janz, B. & Frolick, M. (2008). A comprehensive examination of Internet-EDI adoption. Information Systems Management, 25(3), pp. 273-286.

8. Idrissi, Z. K., Lachgar, M. & Hrimech, H. (2022). Blockchain, IoT and AI revolution within transport and logistics, 2022 14th International Colloquium of Logistics and Supply Chain Management (LOGISTIQUA), EL JADIDA, Morocco, 25-27 May 2022.

9. Swaminathan, J. M. & Tayur, S. R. (2003). Models for Supply Chains in E-Business. Management Science, 49(10), pp. 1387-1406.

10. Mangiaracina, R., Marchet, G., Perotti, S. & Tumino, A. (2015). A review of the environmental implications of B2C e-commerce: a logistics perspective. International Journal of Physical Distribution & Logistics Management, 45(6), pp. 565-591.

11. Merali, Y., Papadopoulos, T. & Nadkarni, T. (2012). Information systems strategy: past, present, future? The Journal of Strategic Information Systems, 21(2), pp. 125–153.

12. Minerva, R., Biru A. & Rotondi, D. (2015). Towards a definition of the Internet of Things (IoT), IEEE.

13. Moroz, M., Nicu, C., Pawel, I. D. D., Polkowski, Z. (2014). The transformation of logistics into e-logistics with the example of electronic freight exchange, Zeszyty Naukowe Dolnośląskiej Wyższej Szkoły Przedsiębiorczości i Techniki. Studia z Nauk Technicznych, 3, pp. 111-128.

14. Møller, C. (2005). ERP II: a conceptual framework for next-generation enterprise systems?, Journal of Enterprise Information Management, 18(4), pp. 483–497.

15. Samoili, S., Cobo, M.L., Gomez, E., De Prato, G. Martinez-Plumed, F. & Delipetrev, B. (2020). Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence, Joint Research Centre, Luxembourg: Publications Office of the European Union, pp. 1-97.

16. Shemet A. D. (2012). Forms of E-commerce and its place in the system of digital economy, Science and Transport Progress. Bulletin of Dnipropetrovsk National University of Railway Transport, Dnipropetrovsk, Ukraine, 41, pp. 311-315.

17. Skitsko V. I. (2016). E-logistics and m-logistics in information economy. LogForum, 12(1), pp. 7-16.

18. Song, Y. & Hou, H., (2004). On traditional M. F and Modern M. F, Journal of Beijing Jiaotong University (Social Sciences Edition), 3(1), pp. 10-16.

19. Quirk, A., Forder, J. & Bentley, D. (2003). Electronic Commerce and the Law, 2nd edition, John Wiley & Sons Ltd., USA.

20. Turban, E., McLean, E. & Wetherbe, J. (2002). Information Technology for Management: Transforming business in the digital economy, John Wiley & Sons, New York.

21. Wang, J., Yang, D., Guo, D. & Huo Y., (2004). Taking Advantage of E-Logistics to Strengthen the Competitive Advantage of Enterprises in China [in:] Proceedings of The Fourth International Conference on Electronic Business, Bejing, pp. 185-189.

22. Wang, Y., Potter, A. & Naim, M. M. (2007). Electronic marketplaces for tailored logistics, Industrial Management and Data Systems, 107 (8), pp. 1170–1187.

23. Wang, Y. (2016). E-logistics: an introduction, in Wang Y.I. & Pettit S., E-Logistics: Managing Your Digital Supply Chains for Competitive Advantage,Kogan Page, pp. 3-31.

24. Waller, M. A. & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management, Journal of Business Logistics, 34(2), pp. 77–84.

8. GIS IN LOGISTICS

Author: Dario Šebalj

Geographic Information Systems (GIS) have revolutionized the logistics industry by providing powerful tools for spatial analysis and decision-making. As businesses increasingly operate in a globalized environment, the ability to visualize and analyse geographical data is essential for optimizing supply chains, managing transportation networks, and enhancing overall efficiency. GIS technology enables logistics professionals to map routes, track shipments, and analyse spatial patterns, leading to more informed decisions and improved resource allocation. This chapter explores the integration of GIS in logistics, highlighting its applications, benefits, and future potential. By understanding how GIS can be leveraged in logistics, businesses can gain a competitive edge, reduce costs, and enhance customer satisfaction.

8.1. Geographic Information Systems (GIS)

A Geographic Information System (GIS) is a computer-based tool that integrates, stores, analyse s, and visualizes geographic data. It connects spatial data with descriptive information to help users understand and interpret spatial relationships, patterns, and trends. GIS is used across various industries for mapping, analysis, and decision-making, providing valuable insights into the spatial dimensions of data (Jonker, 2023; GisGeography, 2024a; Esri, n.d.a; National Geographic, n.d.).

According to Esri (n.d.b) and GisGeography (2024b), the history of Geographic Information Systems (GIS) dates back to the early 1960s when the first computerized GIS was developed by **Roger Tomlinson**, often referred to as the "father of GIS". This initial system was created for the Canada Land Inventory to assist in land-use management and resource planning. Throughout the 1970s and 1980s, advancements in computer technology, remote sensing, and spatial analysis led to the development of more sophisticated GIS software. In 1969, Esri (Environmental Systems Research Institute) was founded and became a pivotal player in the GIS industry, introducing the ArcGIS platform, which significantly enhanced the capabilities and accessibility of GIS technology. By the 1990s, GIS technology had evolved to include a wider range of applications, from urban planning to environmental management.

The integration of GIS with GPS (Global Positioning Systems) and the advent of the Internet further expanded its usage. Today, GIS is an integral tool in various sectors, including transportation, logistics, agriculture, and public safety, providing critical insights and aiding in decision-making processes.
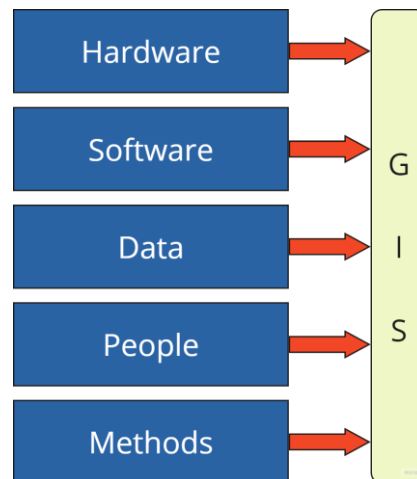
Figure 8.1 Components of GIS

Source: Author, adapted from Kishore and Rautray (n.d.).

Figure 8.1 shows the five essential components of a Geographic Information System (GIS), according to Kishore and Rautray (n.d.):

- **Hardware:** The physical devices used to run GIS software and store data, such as computers, servers, GPS devices, and other peripherals.
- **Software:** The programmes and applications that perform GIS functions, enabling users to analyse and visualize spatial data.
- **Data:** The spatial and non-spatial information that GIS systems analyse , including maps, satellite imagery, and tabular data.
- **Methods:** The techniques and procedures used to analyse GIS data, such as algorithms and statistical models.
- **People:** The professionals and users who operate and manage GIS technology, from data analysts to decision-makers.

Geographic information systems have a wide range of applications across various industries, making them indispensable tools for spatial data analysis and decision-making. In business intelligence, GIS is utilized for market analysis, site selection, and logistics optimization, helping

companies to make data-driven decisions based on geographic trends (Longley et al., 2015). Environmental management leverages GIS for natural resource management, environmental monitoring, and disaster response, enabling more effective conservation efforts and emergency planning (Goodchild et al., 2018). By integrating and analysing spatial data, GIS improves decision-making processes through precise geographic insights and visualizations, enabling organizations to identify patterns and relationships that are not immediately apparent in traditional data formats (Longley et al., 2015).

One of the core components of GIS technology is the concept of layers. According to Esri (n.d.c), a layer is a slice of the geographic reality in a particular area. Each layer in a GIS corresponds to a specific type of data, such as roads, land use, elevation, water bodies, or population density. Figure 8.2 shows the example of different kinds of data on one map (streets, buildings and vegetation), each corresponding to one layer.
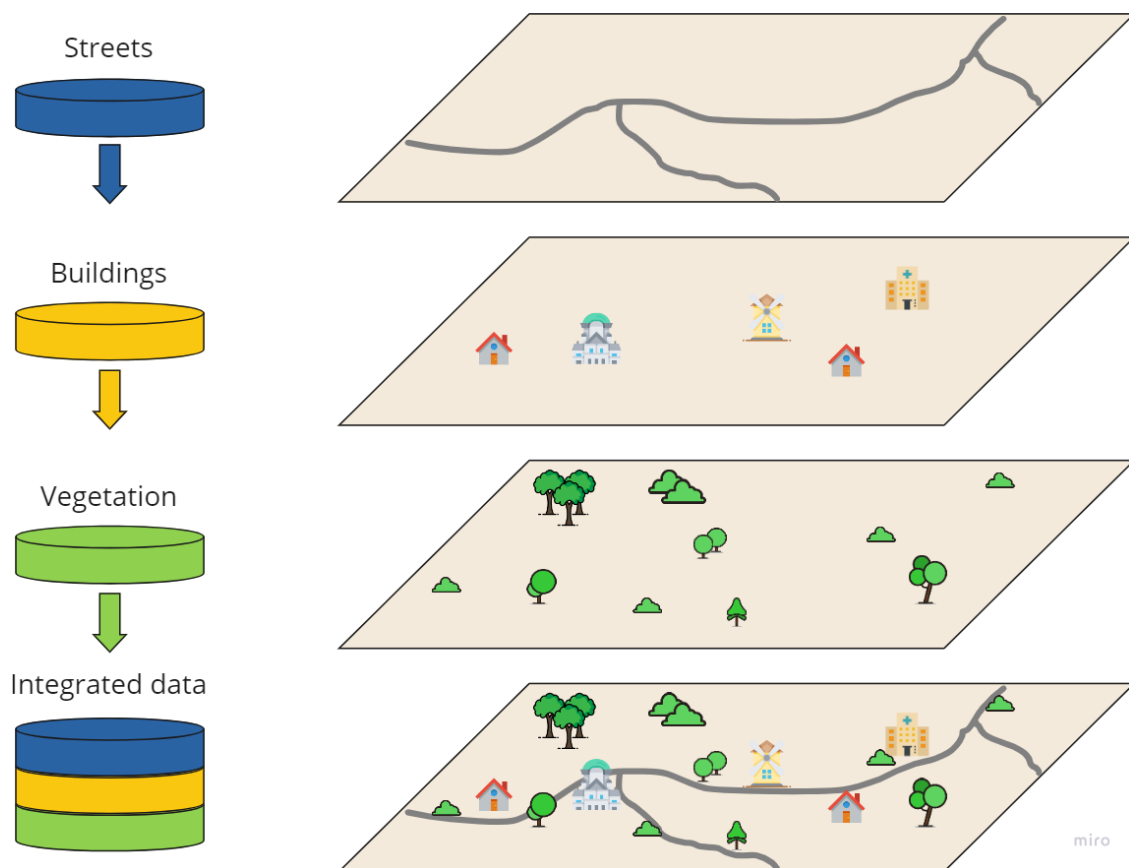
Figure 8.2 GIS layers

Source: Author, adapted from National Geographic (n.d.).

Geographic information systems rely on a variety of data types to represent, analyse , and visualize geographical information. GIS data can be broadly categorized into two main types: raster and vector data.

According to Dempsey (2024), the predominant form of GIS data is **vector data**. Points, lines and polygons used to represent geographic data are examples of vector data. In a vector representation, all lines are captured as points connected by precisely straight lines (Longley et al., 2015). Point data represent discrete data points or specific locations, like schools, city names or points of interest. Line data represent linear features like roads and rivers and polygons are used for area features such as lakes, administrative boundaries, and forests (Dempsey, 2024).

Raster data is a grid-based data structure composed of pixels or cells, each with an associated attribute. The most common sources of raster data are satellite imagery, aerial imagery, remotely sensed data, and data with shaded relief and topography (Dempsey, 2024; Longley et al., 2015).
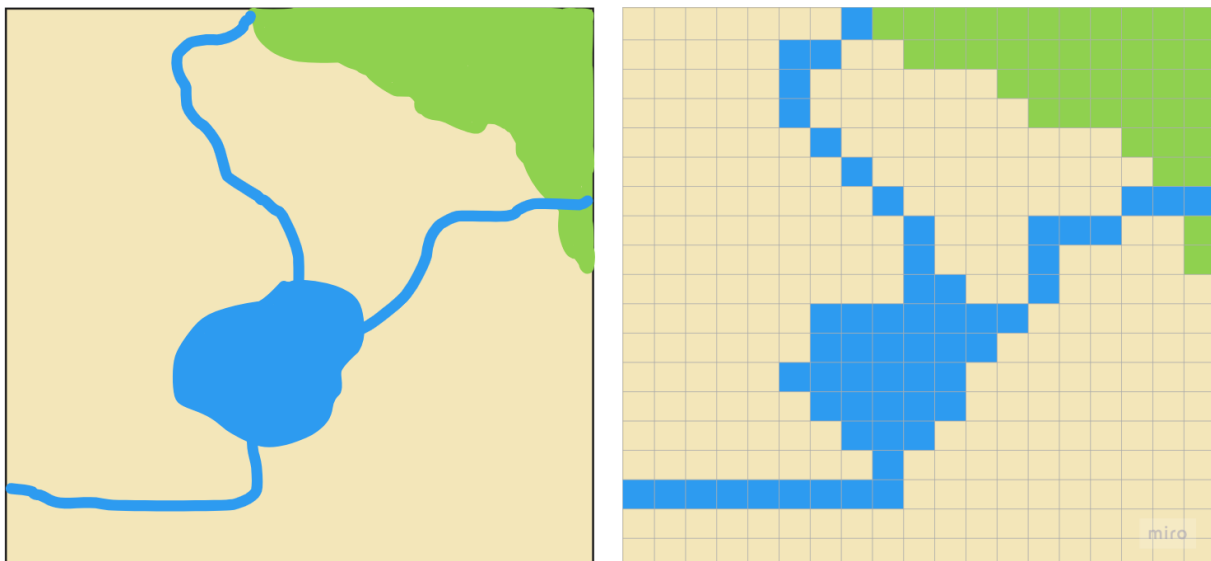
Figure 8.3 Vector (left) and raster (right) data

Source: Author.

Figure 8.3 shows two map representations using vector (on the left) and raster (on the right) data. Vector data include polygons (lake and forest) and lines (rivers), and raster data include a grid where each cell represents one colour (blue, yellow or green).

It is important to understand the different kinds of GIS data in order to use them effectively in business intelligence. Vector data is ideal for precise mapping and analysis of discrete geographical features, while raster data is great for representing continuous data and large-scale environmental data.

8.2. GIS in logistics

Geographic Information Systems (GIS) have fundamentally transformed the logistics sector, providing tools that enable more efficient, cost-effective, and strategic decision-making processes. The integration of GIS in logistics allows the visualization, analysis, and interpretation of spatial data, which is crucial for optimizing routes, managing supply chains, and enhancing overall operational efficiency.

To address logistical challenges, GIS technology combines state-of-the-art data management and analytical techniques with the science of geography. Logistics professionals can see patterns, relationships, and trends that are not visible in traditional data formats by using it to make it easier to overlay different data sets on a map. According to Esri (2017), strategic planning and operational optimization benefit greatly from this spatial perspective.

One of the primary applications of GIS in logistics is route optimization. By analysing spatial data, logistics companies can determine the most efficient routes for delivery, reducing travel time, fuel consumption, and overall operational costs. For instance, GIS can account for traffic patterns, road conditions, speed limits to optimize routing in real time (Ramzan, 2023). This capability not only improves efficiency but also enhances customer satisfaction by ensuring timely deliveries.

Sureshkumar et al. (2017) conducted a study that highlights the numerous advantages of GIS and emphasizes its transformative potential in route optimization for traffic management. GIS enables real-time data application for dynamic traffic adjustments and comprehensive spatial analysis by facilitating the integration of various data types, including GPS and satellite imagery. Because of this integration, there are major time and cost savings due to shorter travel distances and less fuel used. Decision-making processes are improved by the spatial visualization capabilities of GIS, which reveal patterns and trends that are hidden in conventional data formats. All things considered, the study shows that GIS-based route

optimization not only lowers environmental impact and increases operational efficiency, but it also offers a strong framework for dealing with intricate urban traffic issues.

The application of Geographic Information Systems (GIS) in optimizing municipal solid waste (MSW) collection routes has proven to be highly effective in enhancing operational efficiency and reducing costs. Singh and Behera (2018) demonstrated that the integration of GIS and the network analyst tool in ArcGIS significantly reduced haul distances by an average of 27.78%, highlighting substantial improvements in waste management logistics in Kanpur, India. Similarly, Nguyen-Trong et al. (2016) utilized a combined approach of GIS, equation-based optimization, and agent-based modelling to dynamically optimize waste collection routes in Hagiang City, Vietnam, achieving a cost reduction of 11.3%. These studies underscore the transformative potential of GIS in addressing the complexities of urban waste management, particularly through the integration of real-time data and advanced modelling techniques. By leveraging GIS for spatial analysis and route optimization, municipalities can achieve more sustainable and efficient waste management practices, thereby enhancing overall service delivery and reducing environmental impact.

Hemidat et al. (2017) conducted a study which aims to improve the efficiency of municipal solid waste (MSW) collection in the several Jordan cities by using GIS techniques. The researchers developed optimized waste collection scenarios using the ArcGIS Network Analyst tool, aiming to reduce operational costs, vehicle operating times, and environmental impacts. The optimized scenarios showed notable savings compared to the current state (S0). Specifically, Scenario S1 resulted in cost savings of 15%, 6%, and 11% for Irbid, Karak, and Mafraq, respectively. Scenario S2 demonstrated cost savings of 13%, 3%, and 6% for the same cities. The combined scenario (S3) yielded the highest savings, with 23%, 8%, and 13% reductions in total costs. These findings highlight the substantial impact of GIS-based route optimization on reducing operational costs, vehicle operating times, and environmental impacts by minimizing fuel consumption and emissions.

GIS-based analytics significantly enhance blood supply chain management by providing real-time visibility and facilitating better decision-making. The integration of GIS with data mining and other analytic techniques allows for efficient tracking, management, and optimization of blood resources, leading to improved operational efficiency and reduced wastage (Delen et al., 2011).

Also, GIS plays a vital role in urban infrastructure planning and management by providing a robust platform for integrating and analysing spatial data. The use of GIS in this context enables more informed decision-making, leading to optimized infrastructure investments and enhanced service delivery. The study conducted by Irizarry et al. (2013) highlights the effectiveness of GIS in managing urban infrastructure and improving operational efficiencies.

The use of GIS in route optimization across different domains, such as municipal solid waste management, blood supply chain management, and urban infrastructure planning, has demonstrated substantial benefits. GIS enhances operational efficiency by integrating spatial data with advanced analytical tools, facilitating real-time decision-making and optimizing resource utilization. Studies have shown significant cost reductions and improved service delivery through GIS-based route optimization, underscoring its critical role in managing complex logistical operations. By leveraging GIS technology, organizations can achieve sustainable practices, reduce environmental impacts, and enhance overall operational effectiveness.

8.3.   Future trends in GIS

Geographic information systems are undergoing significant transformations driven by technological advancements and increasing demands for spatial data analysis. This subchapter will explore the future trends in GIS, focusing on emerging technologies, cloud computing, the integration of big data and the role of artificial intelligence (AI) and machine learning (ML).

The future of GIS is shaped by several key trends and innovations that are transforming how we collect, analyse , and utilize spatial data. A significant trend is the integration of advanced technologies such as cloud computing, AI, machine learning (ML), and drone-based data collection. These technologies enhance the efficiency and capabilities of GIS, allowing for real-time data processing and more sophisticated spatial analyses. **Cloud computing** is revolutionizing GIS by providing scalable and accessible platforms for storing and processing large datasets. This shift enables organizations to leverage vast amounts of geospatial data without the need for significant on-premises infrastructure. The adoption of GIS as a service is growing, allowing users to access powerful GIS tools and data analysis capabilities through

cloud platforms. This trend is making GIS more accessible and cost-effective, particularly for smaller organizations and industries with limited resources. **AI and ML** are playing important roles in automating and enhancing spatial data analysis. These technologies can identify patterns, make predictions, and provide insights from complex datasets that would be challenging to analyse manually. For instance, AI algorithms can process satellite imagery to detect changes in land use, while ML models can predict traffic patterns based on historical data. The integration of AI and ML with GIS is enabling more accurate and timely decision-making across various sectors, from urban planning to disaster management. The advancements in **drone technology** are also significant trends in GIS. Drones equipped with high-resolution cameras and sensors are increasingly used for data collection in hard-to-reach areas. These tools provide real-time, high-accuracy data that can be integrated into GIS for detailed mapping and analysis. This trend is particularly beneficial for environmental monitoring, infrastructure inspection, and agricultural management. Another emerging trend is the use of **augmented reality** (AR) and **virtual reality** (VR) in GIS. These technologies offer new ways to visualize and interact with spatial data, providing immersive experiences that can enhance understanding and decision-making. For example, AR can overlay geospatial data onto real-world views, helping users visualize underground utilities or navigate complex environments. VR can create detailed simulations of urban landscapes, allowing planners to explore different scenarios and their potential impacts. **Real-time data analysis** is becoming increasingly important in GIS applications. The ability to process and analyse data as it is collected enables more responsive and dynamic decision-making. This capability is enhanced by the integration of GIS with the Internet of Things (IoT), where data from connected devices can be continuously monitored and analyse d. Real-time GIS is being used in applications such as traffic management, emergency response, and environmental monitoring, where timely information is critical. The expansion of GIS applications into new industries and sectors is also noteworthy. GIS is now being used in fields such as healthcare, where it helps track disease outbreaks and optimize healthcare delivery. In retail, GIS analyse s customer demographics and optimizes store locations. Technology is also crucial in smart city initiatives, providing the spatial intelligence needed to manage urban infrastructure and resources efficiently (Kerski, 2022; MGISS, 2023).

As can be seen, the integration of GIS in logistics has revolutionized the industry by enhancing operational efficiency, reducing costs, and improving customer satisfaction. As GIS technology

continues to evolve, its applications in logistics will expand, offering even more sophisticated tools for addressing complex challenges. By leveraging these advancements, logistics companies can maintain a competitive edge and adapt to the dynamic demands of the global market.

REFERENCES

1. Delen, D. & Erraguntla, M. (2011). Better management of blood supply-chain with GIS-based analytics. Annals of Operations Research, 185, 181-193.

2. Dempsey, C. (2024). Types of GIS Data Explored: Vector and Raster. Geography Realm [available at: https://www.geographyrealm.com/geodatabases-explored-vector-and-raster-data/, access June 9, 2024]

3. Esri (2017). The ArcGIS Book: 10 Big Ideas about Applying The Science of Where, 2nd Edition. Esri Press.

4. Esri (n.d.a). What is GIS? [available at: https://www.esri.com/en-us/what-is-gis/overview, access May 27, 2024]

5. Esri (n.d.b). History of GIS? [available at: https://www.esri.com/en-us/what-is-gis/history-of-gis, access May 27, 2024]

6. Esri (n.d.c). Layer. GIS dictionary [available at: https://support.esri.com/en-us/gis-dictionary/layer, access June 8, 2024]

7. GisGeography (2024a). The Remarkable History of GIS [available at: https://gisgeography.com/history-of-gis/, access May 27, 2024]

8. GisGeography (2024b). What is GIS? Geographic Information Systems [available at: https://gisgeography.com/what-is-gis/, access May 27, 2024]

9. Goodchild, M. F., Steyaert, L. T., Parks, B. O., Johnston, C., Maidment, D., Crane, M. & Glendinning, S. (2018). GIS and Environmental Modelling: Progress and Research Issues, 4th Edition. John Wiley & Sons.

10. Hemidat, S., Oelgemöller, D., Nassour, A., Nelles, M. (2017). Evaluation of Key Indicators of Waste Collection Using GIS Techniques as a Planning and Control Tool for Route Optimization. Waste and Biomass Valorization, 8, 1533-1554.

11. Hguyen-Trong, K., Nguyen-Thi-Ngoc, A., Nguyen-Ngoc, D. & Dinh-Thi-Hai, V. (2017). Optimization of municipal solid waste transportation by integrating GIS

analysis, equation-based, and agent-based model. Waste Management, 59, pp. 14-22.

12. Irizarry, J., Karan, E. P. & Jalaei, F. (2013). Integrating BIM and GIS to improve the visual monitoring of construction supply chain management. Automation in Construction, 31, pp. 241–254.

13. Jonker, A. (2023). What is a geographic information system (GIS)? IBM [available at: https://www.ibm.com/topics/geographic-information-system, access May 27, 2024]

14. Kerski, J. (2022). 5 Trends in GIS and How to Successfully Navigate Them. Esri [available at: https://community.esri.com/t5/esri-young-professionals-network-blog/5-trends-in-gis-and-how-to-successfully-navigate/ba-p/1169616, access June 9, 2024

15. Kishore, P. & Rautray, S. (n.d.). The five essential components of GIS. Infosys BPM [available at: https://www.infosysbpm.com/blogs/geospatial-data-services/gis-five-essential-components.html, access June 8, 2024]

16. Longley, P. A., Goodchild, M. F., Maguire, D. J. & Rhind, D. W. (2015). Geographic Information Science and Systems, 4th edition. John Wiley & Sons.

17. MGISS (2023). The Future of Gis: Trends and Innovations in Geospatial Technology [available at: https://mgiss.co.uk/the-future-of-gis-trends-and-innovations-in-geospatial-technology/, access June 9, 2024]

18. National Geographic (n.d.). GIS (Geographic Information System) [available at: https://education.nationalgeographic.org/resource/geographic-information-system-gis/, access May 27, 2024]

19. Ramzan, H. (2023). Optimizing Route Planning with GIS: A Comprehensive Approach for GIS Engineers. Medium [available at: https://medium.com/@hadiaramzan.2199/optimizing-route-planning-with-gis-a-comprehensive-approach-for-gis-engineers-f12d94dd7a16, access June 8, 2024]

20. Singh, S. & Behera, S. N. (2018). Development of GIS-Based Optimization Method for Selection of Transportation Routes in Municipal Solid Waste Management. Advances in Waste Management, pp. 319–331.

21. Sureshkumar, M., Supraja, S. & Bhavani Sowmya, R. (2017). GIS Based Route Optimization for Effective Traffic Management. International Journal of Engineering Research And Management, 4(3), pp. 62-65.

9. DATA VISUALISATION METHODS

Author: Dario Šebalj

In today's data-driven world, the ability to efficiently translate complex datasets into clear, intuitive visualizations is a necessity for organizations that want to use their data effectively. Data visualization goes beyond a purely aesthetic representation; it is a fundamental component of business intelligence that helps decision makers identify trends, outliers and patterns hidden in raw data. This chapter introduces different types of visualizations, from simple charts such as bar charts and line charts to more complicated graphical representations such as heat maps and bullet graphs. Each type of visualization serves different purposes and is suitable for different data sets, so it is critical for data analysts to select the appropriate visual to effectively convey the intended message.

Data visualization is the process of transforming information into a visual context, such as a map or chart, and is used to make data easier for the human mind to understand and draw conclusions from. The main goal of data visualization is to facilitate the identification of patterns, trends and outliers in large data sets. Common types of data visualization include charts, tables, maps and dashboards (Brush, 2022; GeeksForGeeks, 2024).

Due to the growing popularity of big data and data analytics projects, visualization is now more important than ever. Companies are increasingly using machine learning to collect huge amounts of data, which can be difficult and slow to process, understand and explain. This can be sped up with the help of visualization, which also makes the information easier to understand for stakeholders and business owners (Brush, 2022).

Before choosing a visualization method, it is important to understand the context of the visualisation.

9.1. Understanding of the situation context

Nussbaumer Knaflic (2015) states that understanding and contextualizing is the first and most important step before engaging with data visualization techniques and storytelling methods. Understanding the audience is the key aspect of context. Nussbaumer Knaflic

emphasizes the importance of knowing who the audience is, their level of expertise, and what is important to them. This understanding ensures that data visualization and storytelling are tailored to the audience's needs and preferences, making the information more relevant and engaging.

According to IBM (n.d.), general background information helps the audience understand the significance of a particular data point. For example, if a company's email open rate is below average, we should show how the open rate compares to the industry as a whole to illustrate that there is a problem with this marketing channel for the company. The audience needs to understand how the current performance compares to a specific target, benchmark or other key performance indicators (KPIs) in order to be motivated to take action.

There are three important questions which need to be answered (Nussbaumer Knaflic, 2015; IBM, n.d.):

- **Who:** This involves identifying the audience and understanding their perspective in order to know how the story should be tailored. This ensures that the visualization is aimed directly at the target audience, making it more effective and engaging. For example, while a quarterly annual report may only require a high-level summary data, a financial analyst may need a detailed trend analysis over several years. Deciding on the complexity, level of detail and insights to emphasize depends on who will be looking at the visualization.

- **What:** This refers to the key message or insight to be communicated to the audience. It is about being clear about the action or decision that the data visualization is intended to influence. The context defines the purpose of the visualization. Is it to persuade, inform, explore or confirm? Each purpose can lead to different decisions about the type of visualization and the data points highlighted. For example, a persuasive visualization designed to gain support for a new initiative may focus on different data than a visualization designed to simply inform about past performance.

- **How:** This is about choosing the most appropriate and effective means of communicating the story or insight, taking into account the medium, format and visualization techniques that will best resonate with the target audience. Certain types of datasets also require special visualization. For example, scatter plots are good for showing the relationship between two variables, and line graphs are a good way

to show time series data. The visuals should help the audience to understand the main message. An incorrect arrangement of charts and data can have the opposite effect and confuse rather than enlighten the audience.

When it comes to visualization and data analysis, a distinction is made between explorative and explanatory visualization. Exploratory visualization motivates the user to investigate the data or topic more thoroughly in order to make their own discoveries. Explanatory visualisation brings the results to the forefront, conveying the author's hypothesis or argument to the reader (Schwabish, 2021).

Having explored the importance of understanding the situational context in which data visualisations are used, it is clear that this foundational knowledge determines the way information is best communicated and perceived by an audience.

The next critical step is to effectively engage the audience. The next subchapter describes strategies for engaging and maintaining the viewer's interest. This includes selecting elements that enhance the visual appeal and readability of the data visualization and ensure that key insights do not go unnoticed. By using techniques that draw the viewer's eye and highlight important data, visualizations can be more than just informative — they can be captivating and compelling.

9.2. Methods to attract attention

When designing data visualizations, it is very important to capture and direct the audience's attention. The interplay between the mechanics of vision and the principles of visual perception determines how effectively a visualization conveys the intended message. Understanding how the human eye perceives visual elements is the first step in creating compelling visualizations.

Approximately 70% of the sense receptors in our bodies are dedicated to vision (Few, 2012).

The eye is pre-attentively attracted to certain visual attributes, i.e. these attributes are processed quickly and automatically in the visual system without conscious effort. Attributes

such as colour, size, shape and orientation can be used to highlight critical data points or areas within a visualization and immediately attract the viewer's attention.

According to Schwabish (2021), **Gestalt theory** describes how people typically group visual elements. The word *Gestalt* means *pattern* (Cairo, 2013). It was developed by German psychologists in the early 20th century. When it comes to creating diagrams and other visualizations, the six principles of Gestalt theory are particularly helpful.

The principle of **proximity** states that our perception groups objects together when they are in close proximity to each other (e.g. Figure 9.1).
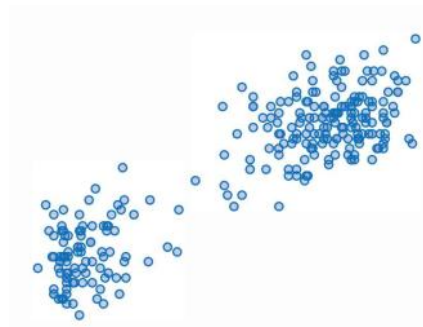
Figure 9.1 Proximity as Gestalt theory principle

Source: Schwabish (2021).

The principle of **similarity** says that the human brain categorizes objects based on their shared attributes such as colour, shape, or direction (e.g. Figure 9.2).
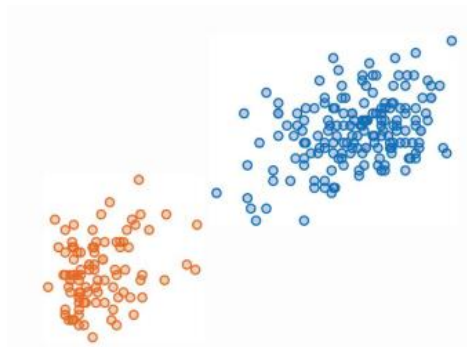
Figure 9.2 Similarity as Gestalt theory principle

Source: Schwabish (2021).

According to the principle of **enclosure**, bounded objects are perceived as a group (e.g. Figure 9.3).
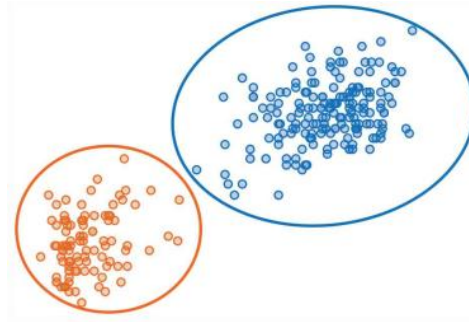
Figure 9.3 Enclosure as Gestalt theory principle

Source: Schwabish (2021).

According to the principle of closure, our brain tends to ignore gaps and fill in missing information in order to form a complete structure. When analysing a line chart that contains missing data, we tend to mentally fill in the gaps using the simplest approach (e.g. Figure 9.4).

Figure 9.4 Closure as Gestalt theory principle

Source: Schwabish (2021).

The principle of **continuity** suggests that elements aligned in a straight line or a smooth curve are perceived by the viewer as more related than elements that do not lie on the line or curve (e.g. Figure 9.5).
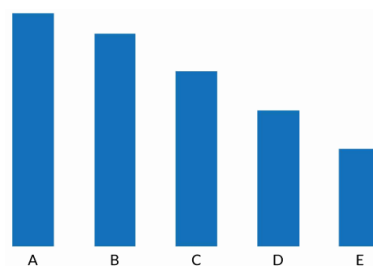
Figure 9.5 Continuity as Gestalt theory principle

Source: Schwabish (2021).

Based on the principle of **connection**, our perception categorizes objects that are connected to each other as belonging to the same group (e.g. Figure 9.6).
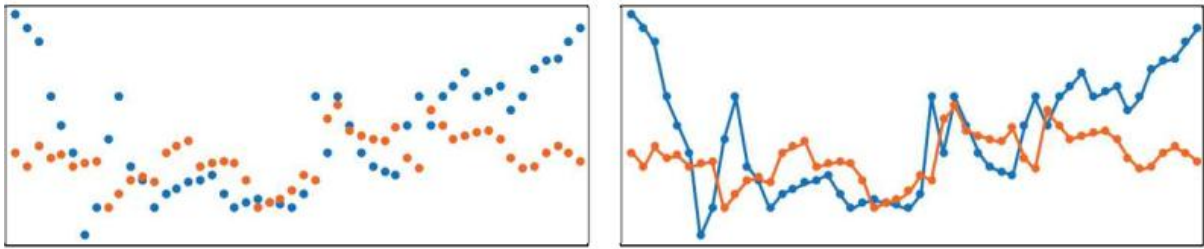
Figure 9.6 Connection as Gestalt theory principle

Source: Schwabish (2021).

There is a very important concept for data visualization, a subset of Gestalt theory called pre-attentive processing. Schwabish (2021) explains that pre-attentive features draw our attention to a specific area of a graph or image.

These features refer to the visual qualities that the human visual system can perceive in the early stages of visual processing without conscious attention, and which are usually measured in milliseconds. One of the pre-attentive attributes used in this book is colour (blue) and weight (bold text). The next very popular example is finding a specific number in the number matrix (e.g. Wexler et al., 2017). Figure 9.7 shows the number matrix without (on the left side) and with (on the right side) pre-attentive attributes.

Figure 9.7 Using pre-attentive attributes in data visualization

Source: Wexler et al. (2017).

From the left matrix, it takes a long time to guess how many 9s are there. But just one change in the matrix makes a big difference. Only the colour was changed – 9s are red and all the other numbers are light grey. Colour (in this case, hue) is one of the several pre-attentive attributes. Figure 9.8 shows examples of some pre-attentive attributes often used in data visualization.
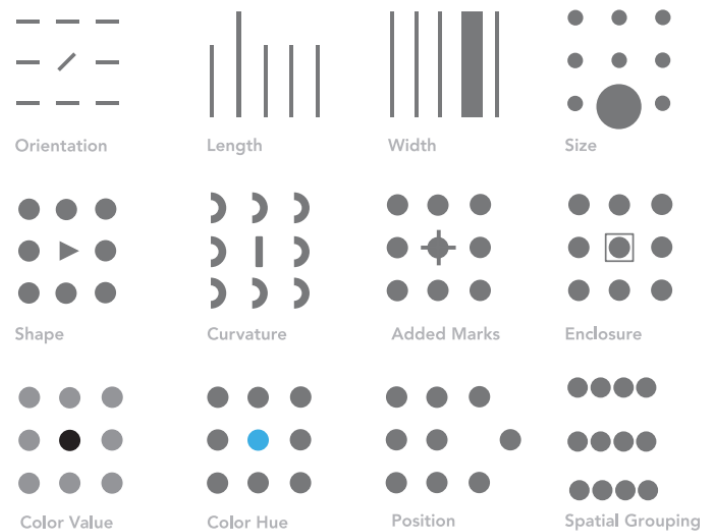
Figure 9.8 Types of pre-attentive attributes used in data visualization

Source: Wexler et al. (2017).

Pre-attentive attributes enable viewers to recognize patterns, outliers, or important data points almost instantly. By employing strategies that guide the viewer's eye to the most significant information, data visualizations can significantly enhance the communication and understanding of complex datasets.

Next subchapter will examine how different types of data and the insights they are intended to provide influence the choice of visualization methods. From simple bar charts to more complex heat maps or bullet graphs, choosing the right visualization method is essential in ensuring that the data not only captures attention but also communicates the intended message effectively and accurately.

9.3. Choosing the right visualization method

The first step in selecting the appropriate method is a thorough understanding of the data. What are the key messages we want to convey? What type of data are we dealing with? Are we working with time series data, geographical information or hierarchical structures? The type of data used can have a significant impact on the chosen visualization method. Choropleth maps, for example, are best suited for displaying geographical data, while line charts may be more appropriate for time series data.

The background context, and expectations of the audience also play an important role in the choice of visualization method. A technical audience may appreciate detailed and complex visualizations such as heat maps or network diagrams. A general audience, on the other hand, may find simpler charts, such as bar charts or line charts, more accessible and engaging.

Interactivity is another important aspect to consider. Interactive visuals, such as dynamic dashboards, allow users to explore different levels of data by filtering, zooming and selecting specific elements. This interactivity can lead to deeper insights as users can tailor the visualization to their specific questions.

A visualization method does not always have to be charts. It can also be a table or even simple text. As Few (2012) states, the purpose of tables and graphs is to effectively convey important information and provide the reader with important, meaningful and useful insights.

In the next few subchapters, the most popular visualization methods will be briefly explained.

9.3.1. Simple text

Nussbaumer Knaflic (2015) suggests using simple text when there is only one or two numbers to share (Figure 9.9).

23%
sales growth compared
to 2022

Figure 9.9 Simple text in visualization

Source: Author.

According to Schwabish (2021), this simple text is often referred to as BAN (Big Ass Numbers). They are typically used to draw attention to key metrics or performance indicators and give the viewer immediate access to important information. By highlighting areas that require attention or action, BANs typically aid in decision making by helping users focus their attention on important aspects of the data (Tay, 2024). Although BANs are simple, they can be enhanced with subtle visual elements such as colour coding or icons to indicate performance against

targets or changes over time. For example, a red downward arrow next to a sales number can immediately indicate a decline, while a green upward arrow signals growth.

9.3.2. Table

Tables are an essential part of data visualization because they provide a structured and clear way to present numerical data. Tables are incredibly useful when it comes to presenting detailed information with precision and clarity, even though they may not have the same visual impact as charts or maps. According to Schwabish (2020), in most cases, they are not intended to provide a quick visual representation of data. Instead, tables are useful when the exact values of the data or estimates need to be shown. While they are not the ideal option for presenting a lot of data or in a small space, a well-designed table can help the reader find specific numbers as well as spot trends and outliers. Few (2012) notes that tables are useful for reference and one-to-one comparisons due to their simple structure and the fact that the quantitative values are expressed as text that we can immediately understand without having to translate it.

Tables should be formatted as follows (Schwabish, 2020; Nussbaumer Knaflic, 2015):

- remove all borders around the table,
- lighten the grid lines as much as possible or remove them completely,
- clearly separate the header from the body of the table,
- align the text in the table and header to the left, and the numbers to the right,
- use an appropriate level of data detail (e.g. use numbers with one decimal place if this is sufficient to understand the data).

9.3.3. Bar chart

A bar chart is ideal for displaying numerical values by groups or categories (e.g. if we want to display the number of employees by department). It can be aligned vertically or horizontally. Horizontal alignment (as in Figure 9.10) is recommended if the category names are too long or if there are too many categories. The bar chart in Figure 9.10. shows sales (quantitative data) per region (qualitative data).
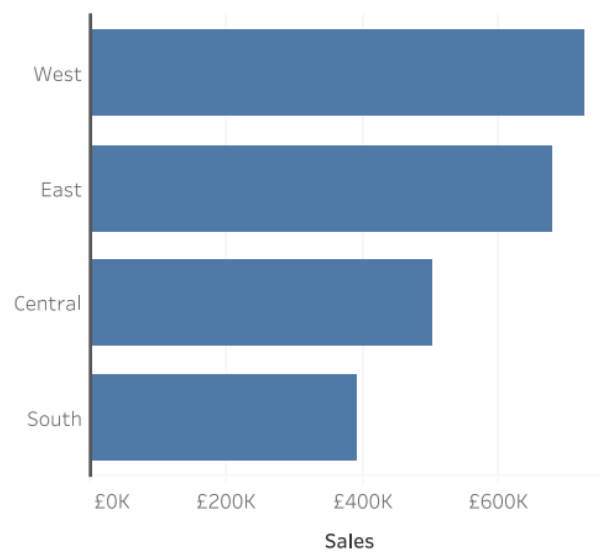
Figure 9.10 Bar chart in visualization

Source: Wexler et al. (2017).

According to Few (2013), the most effective way to represent measures related to discrete items on a nominal or ordinal scale is a bar chart. It is easy to compare individual values by simply comparing the height of the bars. The axis of a bar chart must start at zero. If the axis starts at a value other than zero, this can overemphasize the difference between the bars and distort our perception of the values in the bar chart, which is based on the length of the bars (Schwabish, 2021).

9.3.4.   Line chart

A line chart is used to represent the change of a quantitative value, which lies on the y-axis, over time, which is positioned on the horizontal x-axis. Yi (n.d.a) suggests that a line chart should not contain more than five lines. Also, it is not necessary to include a zero baseline for a line chart. It is acceptable to extend the range of the vertical axis to the point where the value changes are most informative if a zero line is neither understandable nor helpful.
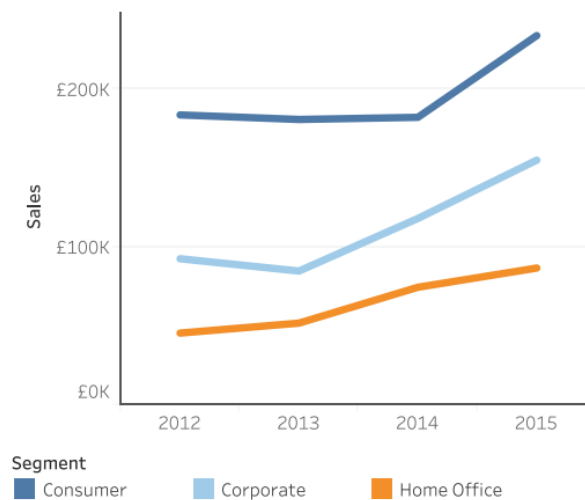
Figure 9.11 Line chart in visualization

Source: Wexler et al. (2017).

The line chart in Figure 9.11 shows the sales (quantitative data) over a period of 4 years and is also broken down by segment.

An area chart (Figure 9.12), which is a variant of the line chart, adds shade between the line and a zero baseline (Yi, n.d.a).
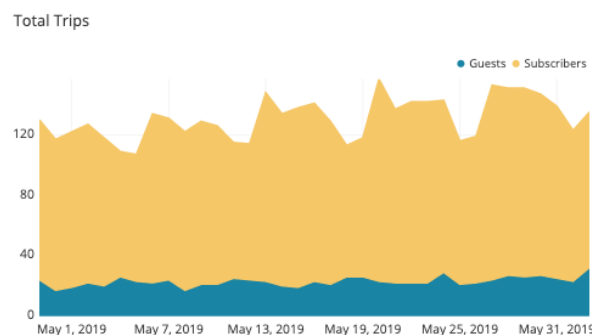
Figure 9.12 Area chart in visualization

Source: Yi (n.d.)

The area chart can be seen as a cross between a line chart and a bar chart, as the values can be interpreted not only by their vertical positions but also by the area shaded between each point and the baseline (Yi, n.d.).

9.3.5. Scatterplot

A scatterplot is used when we want to see if there is a relationship between two quantitative variables. According to The Data Visualisation Catalogue ue (n.d.), the patterns seen on a scatterplot can be used to interpret the nature of the correlation. These are: positive (values increase together), negative (one value decreases while the other increases) or zero (no correlation).
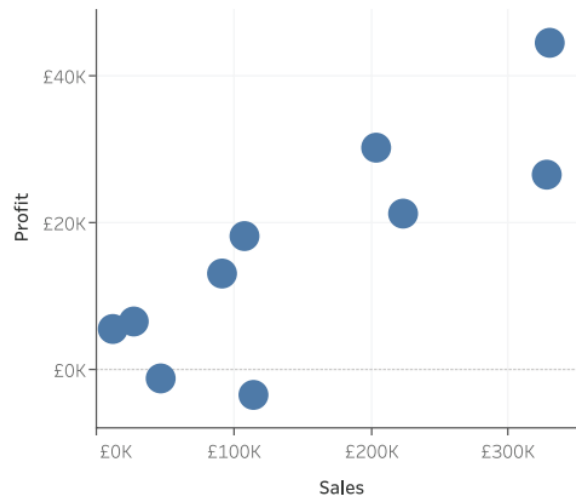
Figure 9.13 Scatterplot in visualization

Source: Wexler et al. (2017).

The scatterplot in Figure 9.13 shows the relationship between profit and sales (both quantitative variables).

According to Yi (n.d.), it is very important to mention that in a scatter plot, just because we see a relationship between two variables, it does not mean that changes in one variable cause changes in the other. This leads to the widely used phrase in statistics: "correlation does not imply causation."

9.3.6. Choropleth map

A choropleth map uses differences in shading or colouring within predefined areas to indicate the values or categories in those areas (Wexler et al., 2017). According to Schwabish (2021), the colour palette on the choropleth map is easy to understand, smaller values correspond to lighter colours and larger values to darker colours.
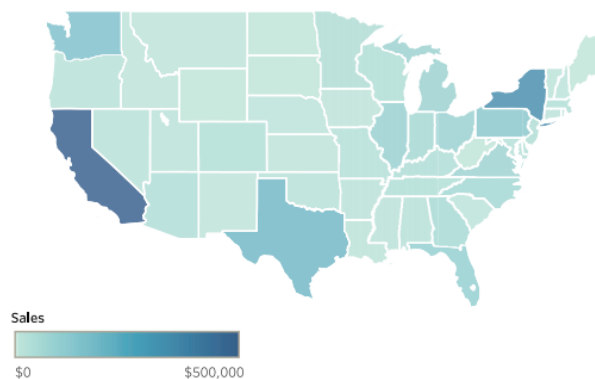
Figure 9.14 Choropleth map in visualization

Source: Wexler et al. (2017).

The choropleth map in Figure 9.14 shows the sum of sales in different states in USA.

9.3.7.   Heatmap

A heatmap is a visualization of data in tabular format, where coloured cells represent the relative magnitude of the numbers (Nussbaumer Knaflic, 2015).

Since colour is a key element of this type of chart, you need to make sure that the colour palette you choose matches the data. The most common type of colour is a sequential colour, where darker colours correlate with higher values and lighter colours correlate with lower values, or vice versa (Yi, n.d.b).

Figure 9.15 Heatmap in visualization

Source: Author, adapted from Nussbaumer Knaflic (2015).

The heatmap in Figure 9.15 shows different values of some quantitative data (e.g. sales) by category (in rows) and region (in columns).

9.3.8.   Bullet graph

A bullet graph is invented by Stephen Few in 2005 (Few, 2013). It is basically a bar chart with a single black horizontal bar representing an actual value, an additional (vertical) marker for a target value (that we want to achieve) and shaded areas in the background representing a scale of success (e.g. poor, good, excellent).
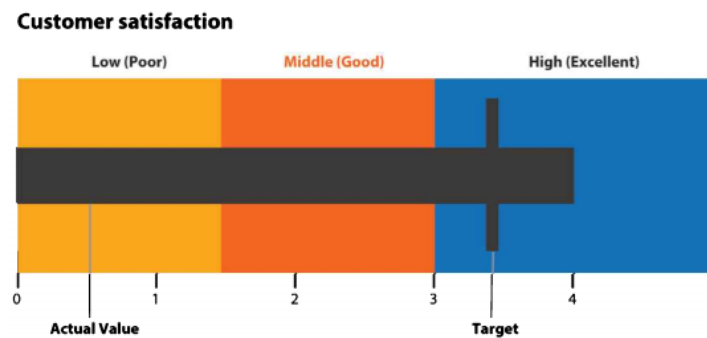
Figure 9.16 Bullet graph in visualization

Source: Schwabish (2021).

The bullet graph in Figure 9.16 shows that we want to achieve a customer satisfaction rating of 3.4 (out of 5). Our current satisfaction rating is 4, which is above the target value. In the background there are three areas of customer satisfaction – low (poor), middle (good) and high (excellent).

Choosing the right visualization method is very important for effective communication, but the design principles that guide these techniques also play a critical role in the clarity and effectiveness of the data presentation. In the next section, we will look at the importance of layout, typography, colour schemes, and the strategic use of space, which are critical to making visualizations not only aesthetically pleasing, but also easy to understand and interpret.

9.4. The guidelines for good visualization design

Effective visualization design is about enhancing the viewer's ability to understand and interact with data. This involves a careful balance between aesthetic elements and functionality, with the choice of colour, font and layout playing a crucial role in conveying information clearly and efficiently. Simplicity should also be a guiding principle. But according to Cairo (2013),

graphics should not simplify messages. They should clarify them, highlight trends, uncover patterns and reveal realities that were not previously visible.

A common pitfall is overcomplicating a visual with too many elements that confuse rather than clarify. The goal is to make the data accessible and understandable to the target audience and ensure that the visualization serves its purpose, which is to inform and support decision-making.

The clarity and effectiveness of data visualization can be improved by removing unnecessary distractions. Nussbaumer Knaflic (2015) emphasizes that any element that does not add value or directly support the understanding of the data is considered **clutter**. This includes unnecessary grid lines, excessive colours, irrelevant data points, and overly complex chart decorations. These elements can distract attention from the key messages the data is intended to convey. She recommends techniques such as simplifying colour schemes, minimizing text and using white space. By using white space strategically, you can create a visual hierarchy that highlights key data points and makes the overall presentation clearer and easier to understand. In addition, the effective use of white space can help create a balanced layout that is less cluttered and more organized.

Colour is a very important part of any data visualization. It serves not only to attract attention, but also to organize information and convey meaning effectively. When used correctly, colour can greatly enhance the clarity and impact of a visualization. There are several suggestions for the effective use of colour in visualizations (Few, 2012; Cairo, 2013; Few, 2013; Nussbaumer Knaflic, 2015; Wexler et al, 2017; Schwabish, 2021; Lidwell, 2023; Interaction Design Foundation, n.d.):

- **Choose appropriate colour combination:** using the colour wheel, it is possible to create visualizations that are visually balanced and pleasing to the eye. There are several common colour combinations (Figure 9.17):
    - **monochromatic** – one colour in different shades,
    - **analogous** – three colours next to each other on the colour wheel. These colours are pleasing to the eye and create harmonious design,
    - **complementary** – two opposite colours on the colour wheel. These colours are contrasting colours and should be used to emphasize something (e.g. increase - green/decline - red).

- o **triadic** – three equally distant colours on the colour wheel. These colours are dynamic and attract attention.

In addition, warmer colours should be used for foreground elements and cooler colours for background elements.
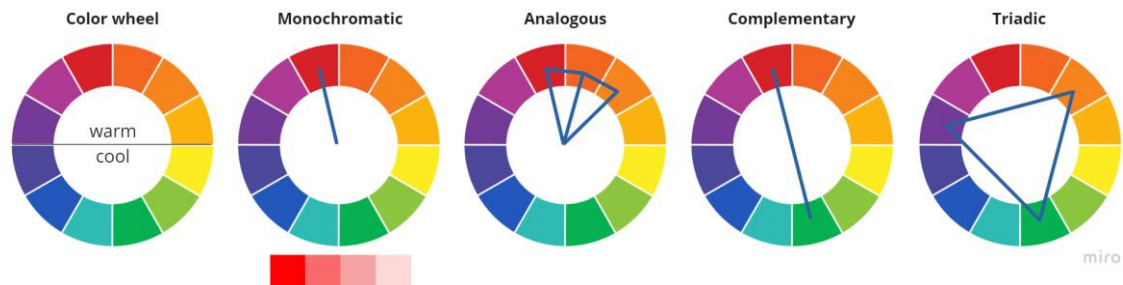
Figure 9.17 Colour combinations

Source: Author, adapted from Lidwell (2023).

- **Choose appropriate colour schemes:** the choice of colour scheme depends on the type of data to be visualized. A **sequential** colour scheme is the use of one colour from light to dark and is ideal for displaying numerical data that progresses from a low to a high value (e.g. sales by state). A **diverging** colour scheme is useful for highlighting values above or below a midpoint (e.g. profit/loss). A **categorical** colour scheme is best for categorical data where the colours need to distinguish different groups without implying an order or value (e.g. product categories). Figure 9.18 shows different colour schemes used in visualizations.
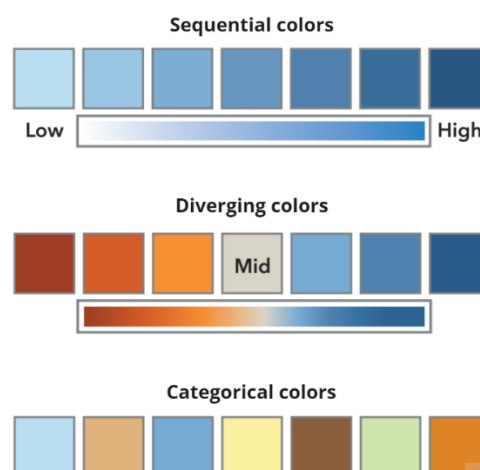
Figure 9.18 Colour schemes

Source: Wexler et al. (2017).

- **Use colour sparingly:** excessive use of colour can cause confusion and make a chart harder to understand. The colour palette should be limited to what the human eye can quickly distinguish between, about five different colours.

- **Consider colour blindness:** about 8% of men and 0.5% of women are colour blind. Avoid colour combinations that are difficult for colour-blind users to distinguish, such as red and green. Instead of these colours, the better combination is orange and blue.

- **Colours should be consistent:** consistent use of colour across multiple visualizations allows the viewer to easily understand and compare data. Once a colour scheme has been established for specific data types or categories, it should be maintained in all associated visualizations.

- **Use colour to highlight important data:** colour can be a strong indicator of where to look. Using a bright or contrasting colour can draw attention to key data points or findings, while more neutral colours can be used for less critical information. Some authors suggest that creating a clear, understandable visualization should start with a grey colour. All data elements in the chart (e.g. bars in the bar chart or lines in the line chart) should be grey. Then add labels and colour only for the elements you want to highlight.

- **Keep it simple:** in the field of data visualization, the KISS principle, an acronym for "Keep It Simple, Stupid", is very relevant and helpful. It refers to the use of simple charts, as complex charts or overly detailed visuals can overwhelm users and make it difficult to recognize the key messages or data points. It also means that you should avoid visual clutter and reduce unnecessary visual components such as overly bright colours, fonts, and grid lines. When applying the KISS principle, the focus should be on the data itself and not on decorative or overly complex design elements.

These principles are crucial to ensuring that data visualizations achieve their primary goal of communicating complex information in a way that is accessible and understandable to all audiences.

It is clear from this chapter that effective data visualization is a critical component in the process of data-driven decision making. Based on an understanding of situational context, this chapter has highlighted the importance of tailoring visualizations to the specific needs of the target audience. Through a detailed examination of different visualization methods

and design principles, it is shown that thoughtful visual representation of data is important to enable better understanding and communication of complex information.

REFERENCES

1. Brush, K. (2022). Data visualization. TechTarget [available at: https://www.techtarget.com/searchbusinessanalytics/definition/data-visualization, access April 15, 2024]

2. Cairo, A. (2013). The functional art: An introduction to information graphics and visualization. New Riders.

3. Data Visualisation Catalogue ue (n.d.). Scatterplot [available at: https://datavizcatalogue ue.com/methods/scatterplot.html, access April 17, 2024]

4. Few, S. (2012). Show Me the Numbers. Analytics Press.

5. Few, S. (2013). Information dashboard design: Displaying data for at-a-glance monitoring. Analytics Press.

6. GeeksForGeeks (2024). What is Data Visualization and Why is It Important? [available at: https://www.geeksforgeeks.org/data-visualization-and-its-importance/, access April 15, 2024]

7. IBM (n.d.). What is data visualization? [available at: https://www.ibm.com/topics/data-visualization, access April 15, 2024]

8. Interaction Design Foundation (n.d.). Colour Theory [available at: https://www.interaction-design.org/literature/topics/colour-theory, access April 18, 2024]

9. Lidwell, W., Holden, K. & Butler, J. (2023). Universal Principles of Design, 3rd Edition. Quarto Publishing Group USA.

10. Nussbaumer Knaflic, C. (2015). Storytelling with data: A data visualization guide for business professionals. Wiley.

11. Schwabish (2020). Ten guidelines for better tables. Journal of Benefit-Cost Analysis, 11(2), pp. 151-178.

12. Schwabish (2021). Better data visualization: A guide for scholars, researchers and wonks. Columbia university press.

13. Tay, J. (2024). Effective use of BANs. Medium [available at: https://medium.com/@e0373084/eye-catching-bans-88d29632e4fa, access April 12, 2024]

14. Wexler, S., Shaffer, J. & Cotgreave, A. (2017). The big book of dashboards: Visualizing your data using real-world business scenarios. Wiley.

15. Yi, M. (n.d.a). A complete guide to line charts. Attlassian [available at: https://www.atlassian.com/data/charts/line-chart-complete-guide, access April 17, 2024]

16. Yi, M. (n.d.b). A complete guide to heatmaps. Attlassian [available at: https://www.atlassian.com/data/charts/heatmap-complete-guide, access April 17, 2024]

10. DATA ETHICS AND INFORMATION SECURITY

Author: Dario Šebalj

In the era of digital transformation, the ethical handling and security of data have emerged as major issues for individuals and organizations. As vast amounts of personal and sensitive information are collected and processed daily, ensuring that this data is managed responsibly and securely is very important.

This chapter examines the principles of data ethics, highlighting the moral considerations and best practices for data handling, and explores the various threats to information security. By understanding and addressing these issues, we can protect privacy, maintain trust, and foster a safer digital environment.

10.1.   The importance of data ethics

Data ethics refers to the moral principles and practices guiding the collection, processing, sharing, and utilization of data to ensure respect for individuals' rights, societal well-being, and trust. It encompasses transparency, accountability, fairness, and privacy, ensuring that data practices are aligned with ethical standards and legal frameworks to prevent harm and promote responsible innovation (Cognizant, n.d.; Gov.uk, 2020; Knight, 2021; McKinsey, 2022; Cepelak, 2023).

In today's digital landscape, the ethical handling of data is crucial for maintaining trust and securing a competitive edge. The McKinsey (2022) article on data ethics highlights the importance of integrating ethical considerations into data management practices. It points out three common mistakes: assuming data ethics are irrelevant, relying solely on legal and compliance teams for oversight, and prioritizing short-term financial gains over ethical practices. To address these issues, they recommend several strategies. First, companies should establish clear, company-specific guidelines for data ethics. These guidelines serve as a foundation for ethical data management and help in setting

a standard across the organization. Second, forming diverse teams to handle data-related issues ensures a range of perspectives and reduces the risk of biased decision-making. Third, involving senior leadership as champions of data ethics initiatives is crucial for driving these practices throughout the organization.

Figure 10.1 5C of Data Ethics

Source: Author, adapted from Atlan (2023).

Figure 10.1 shows 5C of Data Ethics, outlined by Atlan (2023), which represents essential principles for ethical data handling:

- **Consent:** Obtain informed, voluntary consent from individuals before collecting their data, ensuring transparency about data usage.
- **Collection:** Only collect data necessary for specific purposes, avoiding excessive data collection.
- **Control:** Allow individuals to access, review, and update their data, ensuring they have control over its use.
- **Confidentiality:** Protect data from unauthorized access and breaches through robust security measures.
- **Compliance:** Adhere to legal and regulatory requirements, conducting regular audits to ensure ongoing compliance.

Similar to Atlan's principles, Cote (2021) identifies five core principles of data ethics that are essential for business professionals to uphold:

- **Ownership** emphasizes that individuals retain ownership over their personal information. It is both unlawful and unethical to collect personal data without explicit consent. Companies must obtain consent through clear agreements or digital privacy policies, ensuring that users are aware and agree to data collection practices.

- **Transparency** involves clear communication regarding how data will be collected, stored, and used. Businesses must inform individuals about the methods and purposes of data collection. This transparency builds trust and allows users to make informed decisions about their data. Deceptive practices or withholding information about data usage are both unethical and unlawful.

- **Privacy** focuses on the responsibility of businesses to protect the privacy of personal information. Even with consent, personal data should not be made publicly available without the individual's explicit permission. Companies must implement robust security measures to safeguard personal information from unauthorized access or breaches.

- **Intention** pertains to the ethical motivations behind data collection and usage. Data should be collected and used for purposes that are beneficial and not harmful to individuals or society. Ethical data practices involve using data to enhance user experiences and improve services without exploiting or causing harm.

- **Outcome** considers the broader impacts of data usage on individuals and society. Businesses must evaluate the potential consequences of their data practices and strive to avoid negative outcomes. This principle emphasizes the need for ethical foresight and responsibility in data-driven decision-making.

Guzman & Dyer (2020) emphasize that ethical challenges in data practices are not straightforward and often lack clear-cut solutions. They stated that there is a discrepancy between ethical expectations online and offline. Many individuals perceive a form of exceptionalism in online spaces, where traditional ethical rules do not seem to apply. This mindset can lead to the justification of actions online that would be deemed unethical offline. Authors are proposing an ethical approach that bridges both realms, emphasizing that ethical principles should remain consistent regardless of the medium.

The paper on data ethics published by Basl et al. (2021) explores the complex process of moving from abstract ethical principles to concrete, implementable promises in the context of big data and artificial intelligence (AI). They concluded that it is difficult and important

to make this shift in order to guarantee that ethical behaviour is not just theoretical but also practical and significant.

According to O'Reilly (2018), four anonymized case studies have been developed by Princeton's Centre for Information Technology Policy and Centre for Human Values to encourage ethical discourse. One of the case studies explores the ethical dilemmas posed by an automated healthcare app designed to assist adult-onset diabetes patients using AI. It underscores the need to balance technological benefits with ethical principles such as autonomy, fairness, and accountability. Addressing these ethical challenges is crucial for the responsible integration of AI in healthcare, ensuring that it serves the best interests of all patients. There are some key issues that need to be addressed:

- **Paternalism:** The app aims to encourage healthier behaviours among patients by nudging them towards better choices. While this can improve health outcomes, it raises ethical questions about autonomy and paternalism. Is it ethical for the app to influence patient behaviour, or should patients have complete autonomy over their health decisions?

- **Consent and transparency:** The app collects sensitive health data to function effectively. Ensuring informed consent and transparency about data collection, usage, and sharing is crucial. Patients must be fully aware of what data is being collected, how it will be used, and who will have access to it.

- **Data privacy and security:** Handling sensitive health data requires stringent privacy and security measures. The case study emphasizes the need for robust data protection protocols to safeguard patient information from breaches and unauthorized access.

- **Responsibility and accountability:** Determining who is responsible for the app's decisions and actions is another critical aspect. If the app makes an incorrect recommendation that adversely affects a patient's health, identifying the responsible party (developers, healthcare providers, or the app itself) is complex but necessary for accountability.

Modern data management is fundamentally based on data ethics, which guarantees fair, transparent, accountable, and privacy-respecting data practices. Organizations can foster responsible innovation, avoid legal pitfalls, and increase trust by upholding ethical standards. It is not only a best practice but also a requirement for sustainable and responsible growth to

incorporate strong ethical frameworks into data management practices as data becomes more and more essential to operations and decision-making.

Another important aspect is information security since data ethics and information security are intrinsically linked. Ensuring ethical data practices lays the foundation for robust information security measures. Protecting data from unauthorized access, breaches, and other security threats not only preserves privacy and confidentiality but also upholds the ethical principles discussed in this chapter.

10.2.   Fundamentals of information security

Information security refers to the comprehensive set of practices and principles aimed at safeguarding information and information systems from unauthorized access, use, disclosure, disruption, modification, or destruction. It ensures the confidentiality, integrity, and availability of data through the implementation of protective measures, policies, and technologies. These measures include access control, encryption, disaster recovery, and compliance with legal and regulatory standards to mitigate risks and protect against potential threats (Fruhlinger, 2020; CISCO, n.d., NIST, n.d.).

Information security is now essential to an organization's credibility and integrity in the digital era. Its importance is highlighted by the growing dependence on digital data and the rise in cyberthreats that compromise sensitive data. First and foremost, information security protects sensitive data from unauthorized access, breaches, and theft. This includes personal information, financial data, intellectual property, and confidential business communications. As cyberattacks become more sophisticated, the risk of data breaches escalates, potentially leading to severe financial losses and reputational damage. For instance, the 2017 Equifax breach exposed the personal information of 147 million people, resulting in a settlement of up to $425 million (Federal Trade Commission, 2022). Such incidents highlight the dire consequences of inadequate information security measures.

Furthermore, information security is critical for retaining consumer trust and confidence. In an age when data privacy is crucial, customers are becoming more concerned about how their information is handled. A strong information security architecture ensures that consumers' data is secure, which fosters loyalty and trust. According to an IBM survey, **75%**

of customers would not buy from a firm that they do not trust to preserve their data (PR Newswire, 2018). Thus, information security is both a technological need and a strategic business imperative.

Information security is also critical for mitigating operational disturbances. Cyberattacks, such as ransomware, may disrupt corporate operations by preventing users from accessing essential systems until a ransom is paid. The 2021 Colonial Pipeline ransomware attack, which led to fuel shortages across the eastern United States, exemplifies the disruptive potential of such threats (Kerner, 2022). By implementing strong security measures, organizations can safeguard their operational continuity and resilience against such disruptions.

According to Kim and Solomon (2018), information is considered secure if it meets three major tenets:

- **confidentiality:** sensitive information is accessed only by authorized individuals,
- **integrity:** information can only be altered by those with permission,
- **availability:** information and resources are accessible to authorized users whenever needed.

Those tenets are often called CIA triad, as shown in Figure 10.2.
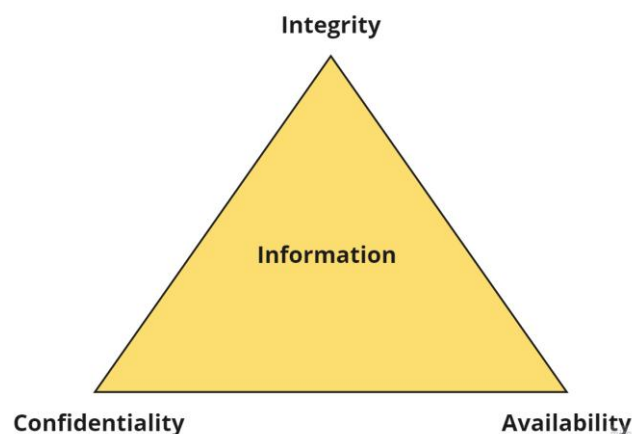
Figure 10.2 The CIA triad

Source: Author, adapted from Kim and Solomon (2018).

In information security, the concepts of risk, threat, and vulnerability are central to understanding and managing security. **Risk** is the likelihood that something bad will happen to an asset. In information security, risk is the probability of a harmful event affecting

the confidentiality, integrity, or availability of information. For example, a company might assess the risk of a cyber-attack on its network by considering both the likelihood of such an attack and the potential damage it could cause. A threat is any action that has the potential to cause harm to information systems or data. **Threats** can be intentional, such as cyber-attacks by hackers, or unintentional, such as natural disasters or human error. **Vulnerability** refers to weaknesses or gaps in a system's defence that can be exploited by threats to cause harm. These can be flaws in software, hardware, organizational processes, or human behaviour (Kim and Solomon, 2018).

To effectively protect information systems, organizations need to understand how risks, threats, and vulnerabilities interact. For instance, a vulnerability in software (such as a security flaw) might be exploited by a threat actor (such as a hacker), leading to a data breach, which constitutes a risk to the organization. By identifying and mitigating vulnerabilities, organizations can reduce the risk of threats causing significant damage.

10.2.1. Data breaches

Data breaches have become a significant threat in the digital age, affecting organizations across various sectors. These breaches compromise sensitive information, leading to financial losses, reputational damage, and legal consequences. Understanding the magnitude and impact of data breaches is crucial for developing effective security strategies to protect against such incidents. According to Fortinet (n.d.), a **data breach** "is an event that results in confidential, private, protected, or sensitive information being exposed to a person not authorized to access it". These breaches can happen in various ways (Kaspersky, n.d.a):

1. **Accidental insider:** An employee unintentionally accesses sensitive information without proper authorization. For example, using a colleague's computer and viewing confidential files.
2. **Malicious insider:** An individual with authorized access intentionally misuses data for harmful purposes. This could involve stealing or leaking sensitive information.
3. **Lost or stolen devices:** Unencrypted and unsecured devices such as laptops or external drives containing sensitive data are lost or stolen, making the information vulnerable to unauthorized access.

4. **Malicious outside criminals:** Hackers use various methods to breach systems, including phishing attacks, brute force attacks, and malware. These cybercriminals exploit vulnerabilities in software, networks, and user behaviour to gain access to sensitive data.

According to Kaspersky (n.d.a), common methods used in data breaches include **phishing**, where cybercriminals impersonate trusted entities to trick individuals into divulging sensitive information. Another method is **brute force attacks**, where hackers use software to repeatedly guess passwords, exploiting weak or reused credentials to gain unauthorized access to accounts. Additionally, **malware**, such as spyware, is used to infiltrate systems and steal data undetected. These methods will be explained later in the chapter.

Human error accounts for **95%** of all data breaches (Cybernews, 2022).

The 21st century has witnessed some of the most severe data breaches, highlighting the vulnerabilities in digital systems and the critical need for robust security measures. According to Hill and Swinhoe (2022) and ESET (n.d.), there are some of the most notable data breaches:

- **Yahoo (2013-2014):** Yahoo experienced one of the largest data breaches in history, with all three billion of its user accounts compromised in 2013. This breach exposed names, email addresses, dates of birth, and security questions and answers. Another breach in 2014 affected 500 million accounts, further highlighting the company's security vulnerabilities.

- **Marriott International (2018):** Marriott announced a breach affecting approximately 500 million guests. Hackers had access to the Starwood guest reservation database, exposing personal information such as names, addresses, phone numbers, email addresses, and passport numbers. This breach was a result of unauthorized access dating back to 2014. The Information Commissioner's Office (ICO), a UK data regulatory authority, ultimately fined the corporation £18.4 million in 2020 for failing to protect the privacy of its customers' personal information.

- **Adult Friend Finder (2016):** The Adult Friend Finder breach exposed the personal information of 412 million accounts, including names, email addresses, and passwords, many of which were poorly encrypted. This incident raised significant concerns about the security practices of online platforms handling sensitive personal information.

- **MySpace (2013):** The MySpace data breach, which occurred in 2013, resulted in the exposure of over 360 million user accounts. The data included names, email addresses, and passwords. Hackers later sold this information on the dark web, which highlighted the vulnerabilities in the security systems of social media platforms during that time.

- **LinkedIn (2021):** In 2021, personal data of 700 million LinkedIn users was posted on a dark web forum. The hacker used data scraping techniques via LinkedIn's API to obtain email addresses, phone numbers, and other personal details. Although not a traditional hack, this incident raised serious concerns about data privacy and the potential misuse of scraped data for social engineering attacks.

- **Equifax (2017):** This breach exposed the personal data of nearly 148 million Americans, 15.2 million Brits, and 19,000 Canadians. Hackers exploited a vulnerability in the Apache Struts web application framework that Equifax had failed to patch. The stolen data included social security numbers, birth dates, and addresses, leading to an estimated $1.7 billion in costs for Equifax.

- **eBay (2014):** eBay disclosed a breach that affected 145 million users. The attack originated from compromised employee login credentials, leading to the exposure of names, email addresses, physical addresses, phone numbers, and encrypted passwords. This breach highlighted vulnerabilities in employee access controls and the importance of strong authentication measures.

- **Target (2013):** A breach affecting over 41 million payment card accounts and the contact information of more than 60 million customers. Cybercriminals accessed customer data, including names, phone numbers, email addresses, credit and debit card numbers, and encrypted PINs. Target faced substantial legal and settlement costs, including a $10 million class-action lawsuit and an $18.5 million multistate settlement.

These breaches demonstrate the far-reaching consequences of cyberattacks and the critical need for strong cybersecurity practices. By learning from these high-profile cases, organizations can better protect their data, improve their security protocols, and minimize

the risk of future breaches. Data breaches are often the result of a range of underlying threats. Understanding these threats is critical to building successful information security measures. Cyber threats may emerge from a variety of sources, including malicious insiders, cybercriminals, and even state-sponsored actors. In addition, vulnerabilities in systems and networks may be exploited to get unauthorized access to sensitive information.

10.2.2. Information security threats

Security threat is a malicious act that attempts to corrupt or steal data, compromise an organization's systems, or compromise the company as a whole (TechTarget, 2024). There are many different information security threats that seriously jeopardize data availability, integrity, and confidentiality.

According to Kim and Solomon (2018) and Grubb (2021), **malware** (malicious software) is designed to infiltrate, damage, or disable computers and networks. Common types of malware include viruses, worms, trojans, ransomware, and spyware. A **virus** is a type of malware that attaches itself to a legitimate programme or file and spreads to other programmes and files when the infected software is executed. Viruses can corrupt or delete data, disrupt system operations, and spread to other systems through email attachments, network connections, or removable media. Unlike viruses, **worms** are standalone malware that can self-replicate and spread independently across networks without needing to attach to a host programme. Worms exploit vulnerabilities in operating systems or applications to propagate, often causing network congestion and overloading systems by consuming bandwidth and resources. A **trojan**, or trojan horse, is malware disguised as legitimate software. Users are tricked into installing it, believing it is a harmless or useful programme. **Ransomware** is a type of malware that encrypts the victim's data, rendering it inaccessible until a ransom is paid to the attacker. Ransomware attacks can be devastating, leading to significant data loss and operational disruption if the ransom is not paid or backups are not available. **Spyware** is malware designed to gather information about a person or organization without their knowledge. It can collect various types of data, such as keystrokes, browsing habits, and personal information, and transmit this data to a third party.

Another type of threat is **phishing**. Kosinski (2024) explains that phishing attacks involve fraudulent emails, texts, calls, or websites designed to deceive individuals into disclosing

personal information or downloading malware. These attacks exploit human error and trust, making them highly effective. To combat phishing, organizations must use advanced threat detection tools and provide robust employee training to recognize and respond to these scams effectively.

Phishing is the leading cause of data breaches, accounting for 16% and costing organizations an average of **$4.76 million** per breach (Kosinski, 2024).

There are four main types of phishing (Forbes, 2024):

- **Email phishing:** using email to steal sensitive information. Attackers may target large audiences by assuming the identity of reputable organizations.
- **Spear phishing:** sending individualized emails, texts, or phone calls with the intent of gaining access to computer systems or sensitive information. When using this technique, attackers usually use data from open databases, social media, or past breaches to bolster their legitimacy.
- **Whaling:** it focuses on high-ranking or senior personnel, including finance officers and chief executives. Attackers create very convincing, highly tailored communications to get sensitive data and information from a business.
- **Vishing:** making phone calls or leaving voicemails under the guise of a reliable source. The goal is to get bank accounts, take advantage of personal information, and steal money.

Insider threats are security risks that originate from within the organization. They can be employees, contractors, or business partners who have access to the organization's systems and data. These threats can be particularly dangerous because insiders often have legitimate access to sensitive information and systems, making their malicious activities harder to detect (TechTarget, 2024).

Another type of threat are **Distributed Denial-of-Service** (DDoS) attacks. They aim to disrupt the normal traffic of a targeted server, service, or network by overwhelming it with a flood of internet traffic. This is achieved by using multiple compromised computer systems as sources of attack traffic. When these devices, often distributed globally, simultaneously

send numerous requests to the target, they consume its available bandwidth and resources, leading to service outages and preventing legitimate users from accessing the service (TechTarget, 2024).

Internet security threats are closely tied to the actions of hackers, who exploit vulnerabilities in systems for various malicious purposes. According to Grubb (2021), hackers are often categorized based on their intentions and methods. Two primary categories are white hat hackers and black hat hackers. **White hat hackers**, also known as ethical hackers, use their skills for defensive purposes. They work to protect organizations from cyber threats by identifying and fixing security vulnerabilities before malicious hackers can exploit them. **Black hat hackers**, in contrast, engage in illegal activities with malicious intent. They exploit security vulnerabilities for personal gain, which can include stealing data, spreading malware, or causing disruptions.

One crucial defence against information security threats is the use of strong passwords. Strong passwords, which should be complex and unique for each account, significantly reduce the risk of unauthorized access.

Table 10.1 shows the time it takes a hacker to brute force a password, according to research conducted by Hive Systems (2024).

Table 10.1 Time it takes a hacker to brute force a password in 2024

Number of characters	Numbers only	Lowercase letters	Upper and lowercase letters	Numbers, upper and lowercase letters	Numbers, upper and lowercase letters, symbols
4	Instantly	Instantly	3 secs	6 secs	9 secs
5	Instantly	4 secs	2 mins	6 mins	10 mins
6	Instantly	2 mins	2 hours	6 hours	12 hours
7	4 secs	50 mins	4 days	2 weeks	1 month
8	37 secs	22 hours	8 months	3 years	7 years
9	6 min	3 weeks	33 years	161 years	479 years
10	1 hour	2 years	1k years	9k years	33k years
11	10 hours	44 years	89k years	618k years	2m years
12	4 days	1k years	4m years	38m years	164m years
13	1 month	29k years	241m years	2bn years	11bn years
14	1 year	766k years	12bn years	147bn years	805bn years
15	12 years	19m years	652 bn years	9tn years	56tn years
16	119 years	517m years	33tn years	566tn years	3qd years
17	1k years	13bn years	1qd years	35qd years	276qd years
18	11k years	350bn years	91qd years	2qn years	19qn years

Source: Author, adapted from Hive Systems (2014).

Understanding the various types of information security threats, such as phishing attacks, malware, and DDoS attacks, highlights the critical need for robust cybersecurity measures. These threats pose significant risks to personal data, financial information, and organizational integrity. Because of this, it becomes essential to adopt comprehensive security strategies. The following sub-chapter presents suggestions for maintaining strong internet security, including practical tips that people and institutions can use to protect their digital resources.

10.2.3.   Information security recommendations

Large number of security risks, such phishing and malware, can be greatly minimized by understanding them and putting them into practice. There are several most important recommendations for ensuring information security (Rubenking & Duffy, 2023; NSW Government, n.d.; Kaspersky, n.d.b):

- **Use strong passwords:** create complex passwords combining letters, numbers, and symbols for each account. Avoid using easily guessable information like birthdays. Use a password manager to store and manage your passwords securely.

- If possible, **enable multi-factor authentication (MFA):** add an additional layer of security by requiring two or more verification methods to access your accounts, such as a password and a one-time code sent to your phone.

- **Keep software updated:** regularly update your operating systems, browsers, and applications to patch security vulnerabilities. Enable automatic updates whenever possible to ensure you are always protected against the latest threats.

- **Be aware of phishing scams:** do not click on links or download attachments from unknown or suspicious emails. Verify the sender's information and look for signs of phishing, such as misspellings or urgent requests for personal information.

- **Use secure connections:** ensure your internet connection is secure by using Virtual Private Networks (VPNs) and avoiding public Wi-Fi for sensitive activities like online banking. Check for "https://" in the URL, indicating a secure connection.

- **Backup data regularly:** regularly back up your data to external drives or cloud storage services. This practice ensures you can recover your information in case of hardware failure, theft, or a ransomware attack.

- **Install antivirus software:** use reputable security software to detect, prevent, and remove malware. Keep your antivirus software updated and perform regular scans to ensure your system is clean.

- **Monitor accounts regularly:** frequently check your financial and online accounts for any unauthorized activities. Set up alerts for unusual transactions and report any suspicious behaviour immediately to your service provider.

By understanding the ethical implications of data handling and recognizing the security threats that exist, individuals and organizations can develop effective strategies to protect sensitive information. From establishing ethical guidelines and using strong, unique passwords to implementing advanced security measures and staying informed about potential threats, these practices collectively ensure the integrity, confidentiality, and availability of data. By prioritizing data ethics and robust security protocols, a safer, more trustworthy digital environment can be created.

REFERENCES

1.   Atlan (2023). Data Ethics Unveiled: Principles & Frameworks Explored [available at: https://atlan.com/data-ethics-101/, access May 17, 2024]

2.   Basl, J., Sandler, R. & Tiell, S. (2021). Getting from commitment to content in AI and data ethics: Justice and explainability. Atlantic Council [available at: https://www.atlanticcouncil.org/in-depth-research-reports/report/specifying-normative-content/, access May 17, 2024]

3.   Cepelak, C. (2023). What is Data Ethics? Datacamp [available at: https://www.datacamp.com/blog/introduction-to-data-ethics, access May 14, 2024]

4.   CISCO (n.d.). What Is Information Security? [available at: https://www.cisco.com/c/en/us/products/security/what-is-information-security-infosec.html, access May 20, 2024]

5.   Cognizant (n.d.). Data ethics [available at: https://www.cognizant.com/us/en/glossary/data-ethics, access May 14, 2024]

6.   Cote (2021). 5 Princples of Data Ethics for Business. Harvard Business School Online [available at: https://online.hbs.edu/blog/post/data-ethics, access May 17, 2024]

7.   Cybernews (2022). World Economic Forum finds that 95% of cybersecurity incidents occur due to human error [available at: https://cybernews.com/editorial/world-economic-forum-finds-that-95-of-cybersecurity-incidents-occur-due-to-human-error/, access May 21, 2024]

8.   ESET (n.d.). 5 scary data breaches that shook the world [available at: https://www.eset.com/in/about/newsroom/corporate-blog/corporate-blog/eset-5-scary-data-breaches-that-shook-the-world/, access May 21, 2024]

9.   Federal Trade Commission (2022). Equifax Data Breach Settlement [available at: https://www.ftc.gov/enforcement/refunds/equifax-data-breach-settlement, access May 20, 2024]

10.   Forbes (2024). Cybersecurity Stats: Facts And Figures You Should Know [available at: https://www.forbes.com/advisor/education/it-and-tech/cybersecurity-statistics/, access May 24, 2024]

11. Fortinet (n.d.). What Is A Data Breach? [available at: https://www.fortinet.com/resources/cyberglossary/data-breach, access May 21, 2024]

12. Fruhlinger, J. (2020). What is information security? Definition, principles, and jobs. CSO [available at: https://www.csoonline.com/article/568841/what-is-information-security-definition-principles-and-jobs.html, access May 20, 2024]

13. Gov.uk (2020). Data Ethics Framework: glossary and methodology [available at: https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework-glossary-and-methodology, access May 14, 2024]

14. Grubb, S. (2021). How Cybersecurity Really Works: A Hands-on Guide for Total Beginners. No starch press.

15. Guzman, L. & Dyer, S. (2020). Ten questions we're asking about ethics, data, and open source research. Amnesty International [available at: https://citizenevidence.org/2020/11/10/ethics-data-open-source/, access May 17, 2024]

16. Hill, M. & Swinhoe, D. (2022). The 15 biggest data breaches of the 21st century. CSO Online [available at: https://www.csoonline.com/article/534628/the-biggest-data-breaches-of-the-21st-century.html, access May 21, 2024]

17. Hive Systems (2024). Are Your Passwords in the Green? [available at: https://www.hivesystems.com/blog/are-your-passwords-in-the-green?utm_source=tabletext, access May 24, 2024]

18. Kaspersky (n.d.a). How Data Breaches Happen & How to Prevent Data Leaks [available at: https://www.kaspersky.com/resource-centre/definitions/data-breach, access May 21, 2024]

19. Kaspersky (n.d.b). Top 15 internet safety rules and what not to do online [available at: https://www.kaspersky.com/resource-centre/preemptive-safety/top-10-preemptive-safety-rules-and-what-not-to-do-online, access May 25, 2024]

20. Kerner, S. M. (2022). Colonial Pipeline hack explained: Everything you need to know. TechTarget [available at: https://www.techtarget.com/whatis/feature/Colonial-Pipeline-hack-explained-Everything-you-need-to-know, access May 20, 2024]

21. Kim, D. & Solomon, M. G. (2018). Fundamentals of Information Systems Security, 3rd Edition. Jones & Bartlett Learning.

22. Knight, M. (2021). What Is Data Ethics?. Dataversity [available at: https://www.dataversity.net/what-are-data-ethics/, access May 14, 2024]

23. Kosinski, M. (2024). What is a phishing attack? IBM [available at: https://www.ibm.com/topics/phishing, access May 24, 2024]

24. McKinsey (2022). Data ethics: What it means and what it takes [available at: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/data-ethics-what-it-means-and-what-it-takes, access May 14, 2024]

25. National Institute of Standards and Technology (NIST) (n.d.). Information security [available at: https://csrc.nist.gov/glossary/term/information_security, access May 20, 2024]

26. NSW Government (n.d.). 10 Tips for Cyber Security [available at: https://www.digital.nsw.gov.au/sites/default/files/2022-09/top-10-cyber-security-tips.pdf, access May 25, 2024]

27. O'Reilly (2018). Case studies in data ethics [available at: https://www.oreilly.com/content/case-studies-in-data-ethics/, access May 17, 2024]

28. PR Newswire (2018). New Survey Finds Deep Consumer Anxiety over Data Privacy and Security [available at: https://www.prnewswire.com/news-releases/new-survey-finds-deep-consumer-anxiety-over-data-privacy-and-security-300630067.html, access May 20, 2024]

29. Rubenking, N. J. & Duffy, J. (2023). 12 Simple Things You Can Do to Be More Secure Online. PC mag [available at: https://www.pcmag.com/how-to/12-simple-things-you-can-do-to-be-more-secure-online, access May 25, 2024]

30. TechTarget (2024). Top 10 types of information security threats for IT teams [available at: https://www.techtarget.com/searchsecurity/feature/Top-10-types-of-information-security-threats-for-IT-teams, access May 24, 2024]

LIST OF TABELS

LIST OF FIGURES

Dario Šebalj, PhD

Assistant Professor at the Faculty of Economics and Business in Osijek, Department of Quantitative Methods and Informatics. He is a lecturer in a field of business process management, ICT project management, business analysis and e-logistics. He has published more than 20 scientific papers in journals and international conference proceedings. He was an executive editor of the journal Ekonomski vjesnik from 2020 to 2023, and secretary of the Alumni EFOS association from 2017 to 2019. He worked as a researcher on two Erasmus+ projects as well as a Croatian Science Foundation project.

ORCID: 0000-0002-8295-7847

Dejan Mirčetić, PhD

Research Associate, Institute for Artificial Intelligence Research and Development of Serbia Assistant Professor, Faculty of Technical Sciences, University of Novi Sad. Dr Dejan Mirčetić earned his PhD in the field of demand forecasting and supply chain analytics and is the author of more than 60 scientific papers, with research primarily focused on solving real-world business and industrial challenges. At the Institute for Artificial Intelligence of Serbia, Dr Mirčetić is responsible for the design and development of AI solutions for business and industry. He currently leads the ISIOP project, part of the AIPlan4EU platform within the Horizon 2020 research and innovation programme. In addition to his work in supply chain management, Dr Mirčetić makes significant contributions to the application of artificial intelligence in healthcare and the food industry, focusing on innovative approaches to process optimisation and efficiency enhancement.

ORCID: 0000-0002-1602-7908

Michał Adamczak, PhD Eng.

Research and teaching employee at the Poznan School of Logistics. Head of the Department of Logistics at this university. Member of the Board of the Polish Logistics Association. Specializes in management, inventory, supply chain management and production management. In his work, he uses logistics data analysis tools, statistical analysis, modelling and process simulations, including the Ms Excel spreadsheet. Author of open and closed training programmes on the above-mentioned topics. Implementer of several dozen consulting projects for commercial and manufacturing companies. Lead scientist in many projects implemented under the ERASMUS+ programme. Author of over 100 scientific publications in recognized domestic and foreign journals. Participant of many international conferences.

ORCID: 0000-0003-4183-7264

BUSINESS ANALYTICS SKILLS FOR THE FUTURE-PROOF SUPPLY CHAINS