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.

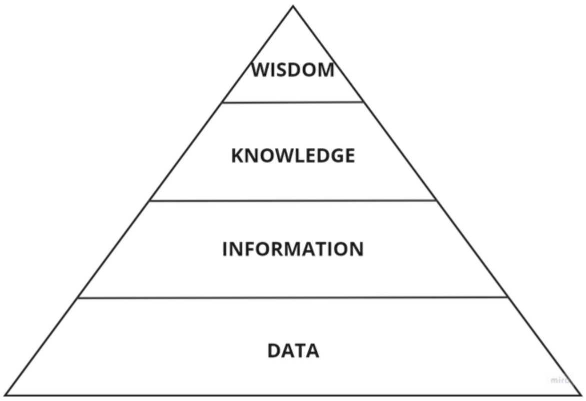


Figure 1.1 DIKW pyramid

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).

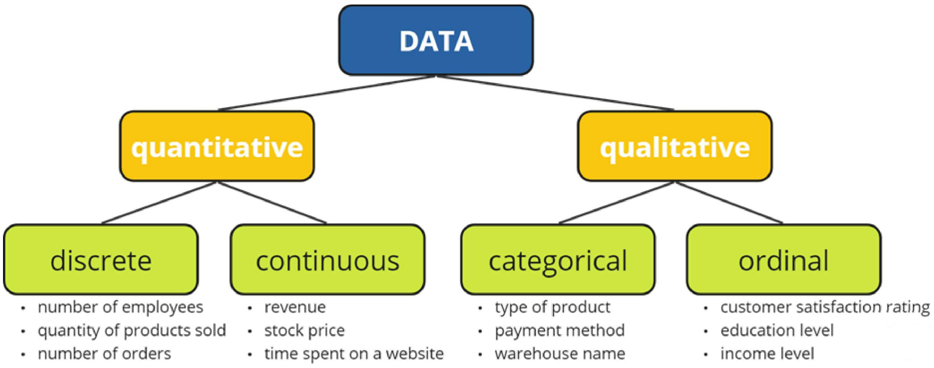


Figure 1.2 Main data types

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.

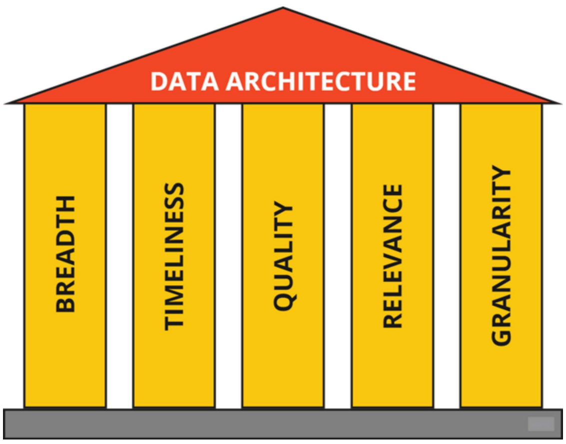


Figure 1.3 Data architecture as a foundation for successful BI

Source: Author, adapted from Howson (2014).

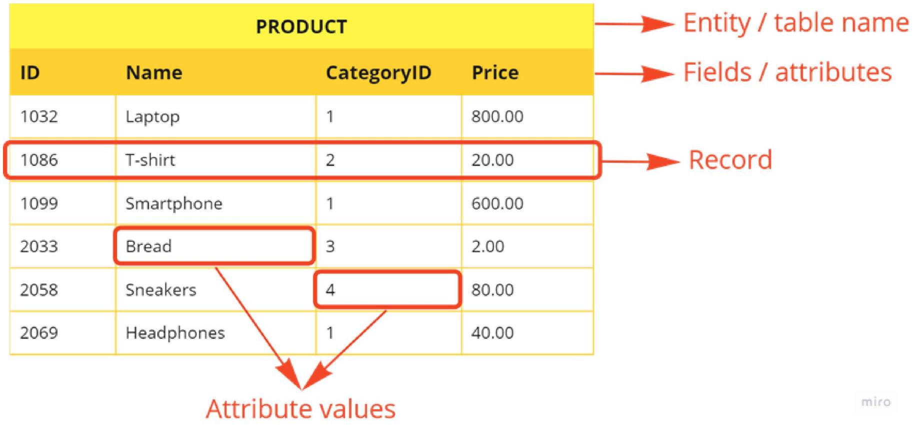
As already mentioned, data can be collected by different sources. This is related to data breadth as one of the pillars of data architecture. It refers to the ability to get multiple data sources which is nowadays, in the era of Big Data, a common way to collect data. On the other hand, combining data 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.



PRODUCT			
ID	Name	CategoryID	Price
1032	Laptop	1	800.00
1086	T-shirt	2	20.00
1099	Smartphone	1	600.00
2033	Bread	3	2.00
2058	Sneakers	4	80.00
2069	Headphones	1	40.00

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.

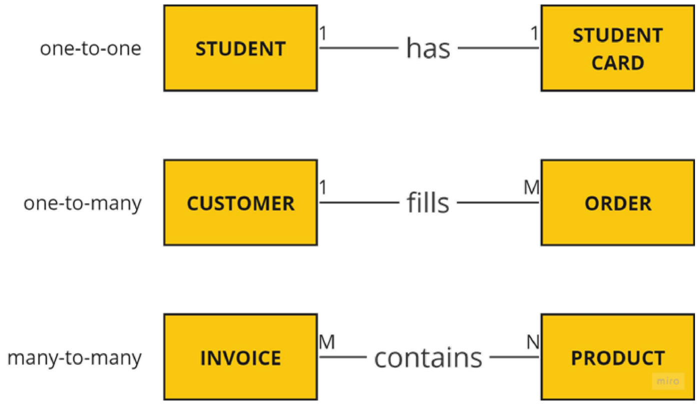


Figure 1.5 Examples of relationships between entities

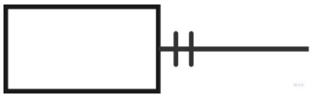
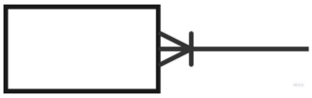
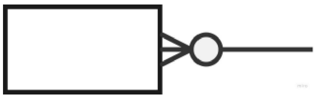
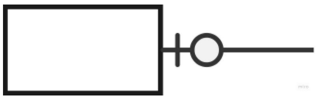
Source: Author.

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.

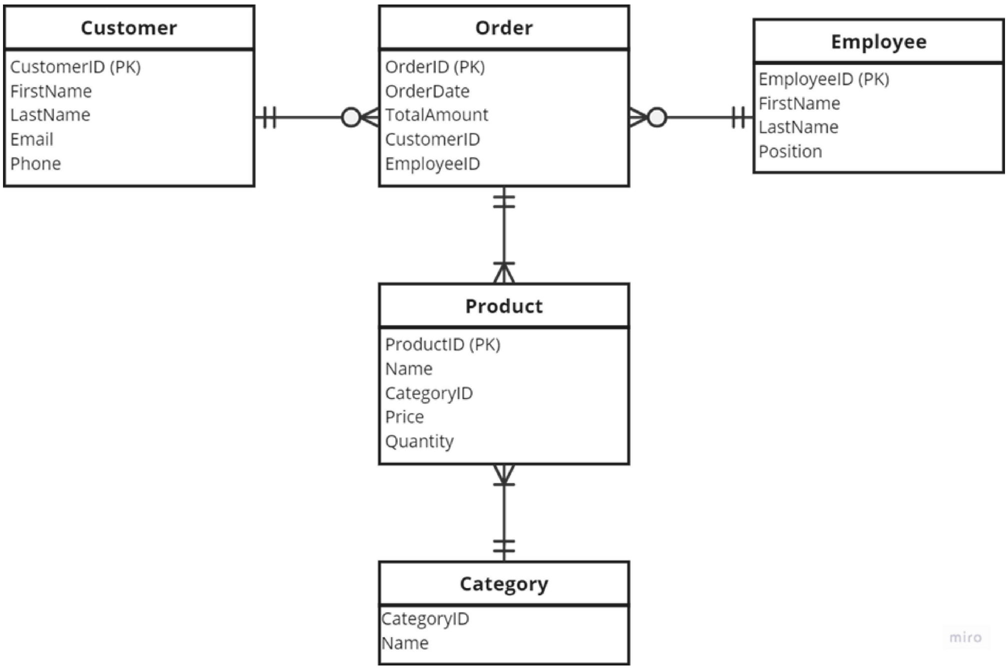


Figure 1.6 Example of sales system ERD

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.