UNIVERSITY OF DEBRECEN

BUSINESS INFORMATICS

EDITOR: PÉTER LENGYEL

AUTHORS: ISTVÁN FÜZESI
TAMÁS KOVÁCS
PÉTER LENGYEL
ÁDÁM PÉNTEK
RÓBERT SZILÁGYI
VIKTOR TAKÁCS
LÁSZLÓ VÁRALLYAI

Authors:

ISTVÁN FÜZESI (CHAPTER 7)
TAMÁS KOVÁCS (CHAPTER 1)
PÉTER LENGYEL (CHAPTER 3)
ÁDÁM PÉNTEK (CHAPTER 2)
RÓBERT SZILÁGYI (CHAPTER 4)
VIKTOR TAKÁCS (CHAPTER 5)
LÁSZLÓ VÁRALLYAI (CHAPTER 6)

Professional readers:

ISTVÁN FÜZESI (CHAPTER 1)
TAMÁS KOVÁCS (CHAPTER 2)
PÉTER LENGYEL (CHAPTER 4)
ÁDÁM PÉNTEK (CHAPTER 3)
RÓBERT SZILÁGYI (CHAPTER 5)
VIKTOR TAKÁCS (CHAPTER 6)
LÁSZLÓ VÁRALLYAI (CHAPTER 7)

ISBN 978-963-615-180-5
ISBN (PDF) 978-963-615-181-2

Kiadta: a Debreceni Egyetemi Kiadó,
az 1795-ben alapított Magyar Könyvkiadók
és Könyvterjesztők Egyesülésének a tagja
dupress.unideb.hu
Felelős kiadó: Karácsony Gyöngyi
Készült a Debreceni Egyetemi Kiadó nyomdájában,
2024-ben

BUSINESS INFORMATICS

(Academic Lecture notes)

Editor:
Péter Lengyel

Debreceni Egyetemi Kiadó
Debrecen University Press
2024

Table of Contents

CHAPTER 1: INTRODUCTION TO BUSINESS INFORMATICS

(Tamás Kovács)

1. Introduction

IT as a discipline is relatively young, but its roots go back to antiquity. The Ancient Greeks and Romans already used mechanical calculators, such as the abacus and the mechanical computer of ancient Cyrene. In the Middle Ages, scientists such as Al-Khwarizmi further developed computing techniques and introduced the decimal number system. The development of modern computer science began in the 17th century with the work of Blaise Pascal and Gottfried Wilhelm Leibniz, who developed mechanical calculators capable of more complex operations. In the 19th century, Charles Babbage created the analytical calculating machine, the first programmable computer. The 20th century brought the real revolution in computing. In 1937, John Atanasoff and Clifford Berry created the first electronic digital computer, the ABC. During the Second World War, Alan Turing used a computer called Colossus to decipher the Enigma code, which was instrumental in winning the war. After the war, electronic computers developed rapidly. The 1950s saw the first transistor computers, which were much smaller and faster than their predecessors. In the 1960s, integrated circuits were developed, allowing computers to be miniaturised and mass-produced. In the 1970s, the first microprocessors appeared, enabling the development of personal computers (PCs). The spread of the IBM PC in the 1980s brought a real revolution in computing. In the 1990s, Internet-connected computers spread into a global network. In the 21st century, the rapid development of information technology continues. The rise of artificial intelligence, big data, cloud computing and mobile phones is revolutionising computing. This discipline is constantly evolving and permeates every aspect of modern life. Business Informatics can, therefore, be seen as an application area.

Business Informatics is an interdisciplinary field that has evolved from the fusion of economics, information technology and management. It aims to effectively use modern information tools and technologies to optimise and increase the efficiency of economic and business processes. Business Informatics includes business intelligence, database management, decision support systems, project management, IT strategy planning and other related areas. Business IT professionals play a key role in modern organisations, building bridges between business and IT. They can identify business needs and design and implement effective IT solutions. It is a dynamic field that is constantly adapting to technological advances and changing business needs. Business IT professionals are in high demand in the job market and have excellent career prospects.

By learning Business Informatics:

- You can learn to use IT tools effectively to optimise business processes.
- You will be able to identify and analyse business needs.
- Learn to design and implement effective IT solutions.
- You will have excellent career prospects in the labour market.

2. Basic concepts

Information systems (IS): these are software applications, databases, hardware and networks that work together to collect, store, process, analyse and disseminate information within an organisation.

Data management involves effective organisation, storage, provision and retrieval of data to support informed decision-making.

Business Process Management (BPM): focuses on mapping, analysing and improving the flow of activities within a business to optimise efficiency and effectiveness.

Decision Support Systems (DSS): These computer-based tools provide managers with information, analysis and modelling capabilities to help them make data-driven decisions.

Enterprise Resource Planning (ERP): a software package integrating core business functions such as accounting, human resources, inventory management and manufacturing into a single system.

Project management: involves the tools and techniques for planning, implementing, monitoring and controlling projects to achieve defined objectives within budget and time constraints.

By understanding these basic concepts, you will understand how business informatics uses technology to improve business operations and achieve organisational goals. These concepts are just a starting point, and business IT encompasses a wide range of technologies and applications that are constantly evolving as the technological world develops.

2.1. Data, information, knowledge

Informatics deals with information and its processing. In the professional community of informatics, three core concepts are used: data, information and knowledge. The relationship between these concepts is illustrated in the following figure.

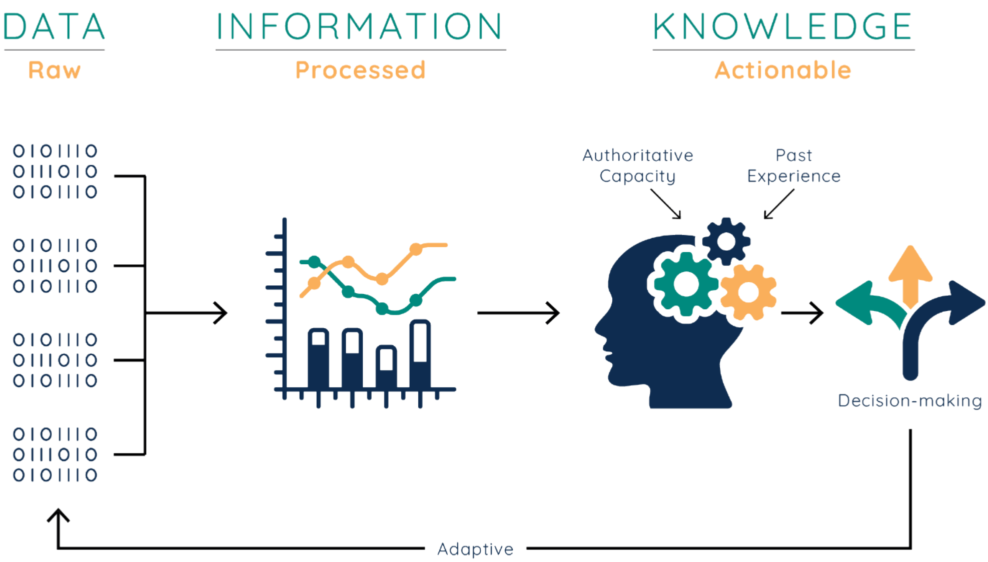


Figure 1.1. The flow from data to information to knowledge

Source: https://internetofwater.org/wp-content/uploads/2018/12/01-Moving-from-Data-to-Action.pdf [Accessed: 30.05.2024]

The first interpretation is that the impulses from our environment that affect the receiver are data: everything we see, hear and feel (whether by smell, touch or emotion). Data becomes information by being given meaning by the receiver beyond perception. To use another example (often cited in the literature on the relationship between the two concepts), if we see an inscription in an unknown language, we perceive its characters (data). Still, we do not know what the inscription means, so it has no information content for us. So, in this approach, information is a subset of data. According to the third approach, we are not aware of the stimuli in the environment that are not important to us anyway, and consequently, there is no need to categorise them. What remains are the stimuli in our environment that we take into account – that is, that are important to us for some reason, that has meaning: this is information. In this approach, the characteristic of information is usually emphasised that it is acquired by adding to the 'knowledge' of the perceiver – information is, therefore, value (and perception). Data is nothing more than the stored (recorded or recordable) information form. Despite their apparent contradictions, the three approaches have one thing in common: they all see information as something meaningful, important and valuable to the receiver. Finally, knowledge, and when we put information into context, we take a systems approach to the data, using our relevant skills and prior knowledge. We use this to make many decisions at work and in our lives. In terms of processing, data is primarily raw, and information is processed, while knowledge is a form of actionable, actionable and supportive knowledge based on our past experiences.

2.2. Definitions of IT in business

Information technology is the set of tools and methods that deal with producing, storing, processing and transmitting information. Suppose this can be reduced to a specific processing process. In that case, i.e. the range of information that can be processed, the set of methods used to process it, the way it is stored, etc., then it can be called an information technology system (IT system).

Sometimes, the information is processed at the point of origin, but this is not (nowadays) the case: information (usually) needs to be transferred from one system to another, and the processes involved are collectively called communication. The communication process involves at least two parties: the one who communicates the information (the sender) and the one who receives the information (the receiver). Communication can only be successful if the information communicated by the sender is presented to the receiver with the same meaning – and this can only be ensured if both parties to the communication process follow specific (predefined) rules. These sets of rules are called protocols, which are a set of specifications for the communication process. If we want to emphasise the interconnectedness of the ICT devices involved in a communication process, i.e. the fact that they do not transmit and receive information between each other randomly but do so with a purpose, we talk about communication networks.

Some information is passed from one IT device to another by providing communication protocols. These devices then do something with the information. What exactly is obviously determined by the person interpreting the information? In this case, the emphasis is on automation. An algorithm is defined as a finite sequence of elementary actions (which cannot be further subdivided into operational parts) defined by rules to solve a task. If the following conditions are satisfied for an algorithm, then the algorithm is automatable:

- has a clearly defined initial activity (i.e. it has exactly one 'start')
- deterministic and ordered (after each activity is completed, the next activity to be performed can be clearly defined)
- it has at least one activity which is not followed by another activity (it has an end)

Information technology devices that can store and automatically execute information and the algorithms that operate on them are called computers. The algorithms that a computer can interpret are called programs, and the information stored and processed on the computer is called data – these two components together form the software layer of the computer. In addition, the computer is made up of the set of "physical" (sometimes called "tangible") elements (electronic devices, wires, screws, cases, etc.) that make up the device, called hardware.

2.3. Neumann principles

Of course, the development of computers did not happen overnight. Still, one of the milestones in this long process was a few theses by János Neumann, in which he formulated what he thought was necessary for an electrical device to function independently. These principles are called Neumann's, and the devices that implement them are called computers.

The Neumann Principles (a set of requirements for (electronic) computers):

1) ***Fully electronic operation*** (to understand this principle, we must bear in mind that in earlier computing devices, some mechanical structure: gears, gear

chains, sliding rails, etc., were involved in the operations), whose elements are logic circuits (the basic logic operations: AND, OR, NOT).

2) ***Use of the binary number system*** to perform operations. This is essentially a "necessary" feature of electronic operation: the physical creation of biphasic states in an electronic system results in the easy discriminability of information, whether it is the creation and removal of magnetisation caused by an electric current or the easy and quick exchange of double reflective properties, and other physical implementations.

3) ***Stored program principle:*** data and programs are stored in the same way (and in the same place) in the system: this is memory, which consists of parts (compartments) of the same size and distinguishable by serial numbers. Once the data and the program have been entered into the system, the control unit can access the contents of the memory without external (human) intervention and perform the operations it has been designed to perform on the stored data: its operation is automatic.

4) ***Sequential instruction execution:*** the computer only executes a single operation at a time – one of the major drawbacks of Neumann machines, which would obviously have significantly higher performance if they were running in parallel. Computers as we know them can do several things at once, which is apparently only 'virtual' parallelism.

5) The computer must have the following ***functional units***:

- control unit (CPU): system control tasks, interpreting instructions and execution;
- operation processing unit (ALU): perform elementary mathematical and logical operations independently;
- logic and logic operations;
- storage unit (memory): see memory management. Principle 3;
- input/output (i.e. communication) units (I/O) devices: using which a person using a computer can communicate with the computer.

The Neumann principles, in addition to their unchanged definition of the foundations of computer design and operation for more than half a century (even if in today's modern computers, the above structural elements are not necessarily separated: the CPU and ALU together form the processor, etc.), have a fundamental influence on the scale of computing. The first computer based on

Neumann principles was the EDVAC (Electronic Discrete Variable Automatic Calculator) (1944–1948).

As an outlook, our accelerated world has led to the experimental emergence of so-called quantum computers, which are now replacing traditional architecture. These computers calculate in binary, with a bit being either 1 or 0. The qubits in quantum computers are based on the laws of the quantum world so that they can be 1, 0 or both. Just as the performance of a conventional processor is primarily determined by whether it is 8, 16, 32 or 64 bits, the number of qubits it handles is a fundamental question for the performance of a quantum computer. There are currently 53 known qubit systems, which Google calls Sycamore.

3. Objectives of business informatics

3.1. The main tasks

Business IT professionals collect data from various sources, including transactional systems, sensors, and social media. This data is then processed and stored in databases or data warehouses. They use various analytical techniques to extract insights from the data, including statistical analysis, machine learning, and data mining. To help users quickly and easily understand the data and its insights, they visualize the data using charts, graphs, and maps. Additionally, they produce reports on the analysis results, providing managers with valuable information about business performance and potential trends. Furthermore, they develop Decision Support Systems (DSS), computer programs that help managers make better decisions by integrating data, analysis, and decision support tools.

3.2. The objectives

Improving business efficiency can help businesses by improving business processes, reducing costs and increasing productivity.
Increasing revenues can help businesses develop new products and services, enter new markets and better serve their existing customers.

Reduce risks: it can help businesses identify and manage risks by providing insight into market trends, competitor activity and operational efficiency.

Facilitate better decision-making: it can help managers make better decisions by providing accurate and timely information on business performance.

Gaining a competitive advantage: it can help businesses gain a competitive advantage by making them more efficient and effective.

It is a powerful tool that can help businesses achieve their goals. Investing in business IT solutions can give businesses a significant return on investment (ROI).

3.3. Tasks and applications of business informatics

Reporting tools: these tools help users create reports from data, which they can visualise with charts, graphs and maps.

Data visualisation tools: these tools help users visualise data in an easy-to-understand visual format.

Data analysis tools: These tools help users gain insights from data by applying statistical analysis and machine learning techniques.

Data mining tools: These tools help users discover hidden patterns and trends in data.

Decision Support Systems (DSS): These systems help managers make better decisions by integrating data, analytics and decision support tools.

Many software solutions are available on the market, and choosing the right software depends on the business's specific needs. Choosing software that is easy to use, flexible and scalable as your business grows is important. Business IT is a great tool that can help businesses achieve their goals.

4. IT risks and opportunities

Business information (BI) is the branch of information science that focuses on improving business processes and systems through the effective use of information technology (IT). However, IT implementation is not without risks, and it is crucial that these risks are managed to maximise opportunities.

IT risks:

Data security risks: The systems store sensitive data that can be stolen or manipulated by hackers or other malicious actors. Data security risks can cause severe financial losses and reputational damage to businesses.

Data quality risks: poor quality data can lead to misleading analytical results, which can lead to poor decision-making. Data stored in systems must be accurate, complete and up-to-date.

System failures: hardware and software failures can lead to downtime and data loss. Business IT systems must be reliable and available to businesses.

Human error: human error, such as incorrect data entry or poor configurations, can have serious consequences. It is important to design and use IT systems in a way that minimises the risk of human error.

Legal and compliance risks. Non-compliance can lead to financial penalties and legal consequences.

IT opportunities:

Improving efficiency: can help businesses by improving business processes, reducing costs and increasing productivity. It can help businesses develop new products and services, enter new markets and better serve existing customers.

Reduce risks: This can help businesses identify and manage risks by providing insight into market trends, competitor activity, and operational efficiency.

Facilitate better decision-making: This can help managers make better decisions by providing them with accurate and timely information on business performance.

Gaining a competitive advantage: Businesses can gain a competitive advantage by making them more efficient and effective.

Managing IT risks in business:

Implement data security measures – Data security measures such as firewalls, encryption and access controls can help protect data from unauthorised access.

Data quality controls can help ensure that data stored in business IT systems is accurate, complete and up-to-date.

5. Skills and competencies required for jobs in the business

The IT sector is growing fast, and businesses are constantly looking for talented IT professionals. The skills and competencies required for a successful IT career are varied, but the most important include technical knowledge, problem-solving skills, communication and teamwork. This chapter reviews the key skills and competencies required for IT business positions, including technical and soft skills. In addition, some IT jobs may require additional specific skills and competencies. For example, a web developer must be familiar with web development languages and frameworks, while a data scientist will need statistical and machine-learning skills. The following are the most common skills required to fill a job in business IT.

Technical skills:

- **Security**: understanding cyber security threats and solutions is key to protecting business data and systems.
- **Programming**: while not all IT roles require coding, some level of programming skills is beneficial, especially for tasks such as automation and customisation. Some jobs require at least a working knowledge of a programming language such as Python, Java, C++ or JavaScript.
- **Systems and networks**: Managing and troubleshooting computer systems, networks and devices is essential. This includes knowledge of different operating systems (Windows, Linux, Mac).
- **Cloud computing**: Cloud technology is now widely used by businesses. Expertise in cloud platforms such as AWS, Azure or GCP is valuable.
- **Data analytics**: Many IT jobs involve working with data. The ability to analyse, interpret and visualise data is a sought-after skill.

Soft Skills:

- **Communication**: excellent written and verbal communication is essential to convey technical information to technical and non-technical audiences.
- **Problem-solving**: IT professionals routinely troubleshoot and solve complex problems. Strong analytical and logical thinking is key.
- **Organisation and time management**: prioritising tasks, managing deadlines and staying organised is essential in a fast-paced IT environment.

In addition, some business IT jobs may require specific skills, depending on the field. Business analysts, for example, may need business acumen, while IT security professionals may need a thorough understanding of security protocols and tools. In addition to having mixed knowledge, some roles may require a more generalist approach: e.g., DataAnalysts may already have a working knowledge of Python programming language, alongside strong business and analytical skills.

Business IT professionals can work in both research and commerce. In business, there are different uses, which vary depending on professional experience.

The most commonly employed areas are:

- Management consultancy
- IT consultancy
- IT account manager
- Systems analysis and organisation
- Business analyst
- IT project manager
- IT auditor
- Solution Designer
- Enterprise Architect
- Information Technology Management

There are many ways to develop, but the most important ones include continuous training, obtaining certificates, gaining practical experience, networking and self-learning:

- **Continuous training**: a range of courses, online courses, and boot camps are available to help you learn the latest technologies. These programmes are flexible and fit different learning styles and schedules.
- **Earn certifications**: industry-recognised certifications demonstrate expertise and make your CV more attractive to prospective employers. There are many certifications in various IT fields, so everyone can find the right one for them.
- **Gaining experience**: practical experience is essential to reinforce the theoretical knowledge you have learned. There are many ways to gain experience, including volunteering, participating in side projects, or contributing to open-source projects.
- **Networking**: networking with other IT professionals is an excellent way to learn, develop your career and keep up with the latest trends. There

are many industry events and conferences where you can meet other professionals, expand your network and build valuable contacts.

- **Self-learning**: Many online resources, blogs and professional journals help you stay current.

  Continuous learning and development is essential for IT professionals to succeed. By applying these strategies, they can expand their knowledge, acquire new skills and remain competitive.

Key-works, tasks and job opportunities

Business Information Technology (BIT) is a dynamic field that offers many exciting job opportunities for professionals with different skills and experience. Business IT professionals collect, analyse and visualise data to provide businesses with valuable insights that can lead to better decision-making and more efficient operations.

The most common job opportunities for business IT professionals

Unsurprisingly, due to the traineeship orientation, jobs have been created in each company that are linked or partially linked to the tasks. These range from creating the simplest queries to building executive dashboards, complex system design, and data mining. Thus, it can be said that the field is constantly evolving and very diverse.

Data Analyst: Data Analysts focus on collecting, analysing and visualising data to provide business insights.

Report Writer: Report writers are responsible for producing reports on the analysis results.

Business Intelligence (BI) Developer: BI developers develop and maintain business IT software and systems.

Data Scientist: Data Scientists use advanced statistical and machine-learning techniques to extract insights from data.

Data Architect: Data architects are responsible for designing and developing databases and data warehouses.

REFERENCES

Bach, M. P., Praničević, D. G., Šebalj, D., & Pihir, I. (2018). Employment of business informatics graduates: Preliminary results. In Central European Conference on Information and Intelligent Systems (pp. 55–60). Faculty of Organization and Informatics Varazdin.

Beynon-Davies, P. (2019). Business information systems. Bloomsbury Publishing. ISBN 978-1352009998

Helfert, M. (2008). Business informatics: An engineering perspective on information systems. Journal of Information Technology Education: Research, 7(1), 223–245. ISSN 1539-3585

Helfert, M. (2010). Business informatics. In Encyclopedia of Library and Information Sciences (Vol. 1, pp. 687–692). DOI: 10.1081/E-ELIS3-120043478

Herdon, M., Bakó, M., Lengyel, P., Nagyné, P. I., Rózsa, T., Szilágyi, R., & Várallyai, L. (2011). Üzleti informatika jegyzet. Debreceni Egyetem AGTC GVK. ISBN 978-963-318-158-2

Pedersen, M. K., & Larsen, M. H. (2000). The efficiency opportunity impact of information systems in an organizational economics framework of informatics. Copenhagen Business School.

Stair, R. M., & Reynolds, G. (2008). Fundamentals of business information systems. Thomson Learning. ISBN 978-0324406271

Wigand, R. T., Mertens, P., Bodendorf, F., König, W., & Schumann, M. (2003). Introduction to business information systems. Springer Science & Business Media. ISBN 978-3540001584

Zając, D., Nycz, M., & Pólkowski, Z. (2018). The graduate profile in the field of informatics and business informatics in the context of labour market needs.

CHAPTER 2: DATABASE SYSTEMS

(Ádám Péntek)

1. Introduction to databases

A database is a structured collection of data organized to facilitate the efficient storage, retrieval and management of information. It is a central repository for storing and managing data, allowing users to store, retrieve and update information as required.

The database is made up of tables organized in rows and columns. Each row represents a record or entry, while each column represents a particular attribute or field within that record. These tables are linked through relationships, forming a relational database model that helps maintain data integrity and consistency.

Databases are used in various applications and industries, including business, finance, healthcare and education. Their role in supporting the functioning of organizations is indispensable. They store a wide range of information, including customer data, data related to financial records, inventory data, employee information, and data essential for production and other operations.

Because data is now a core business asset, in addition to storage, databases provide features such as data security, data integrity, concurrency control, and backup and recovery procedures to ensure the reliability and availability of the information they hold.

1.1. The database of related concepts

1. Data structure: data are organized in a structured format. It typically consists of tables, rows and columns. This format allows efficient storage and retrieval of data and the creation of relationships between individuals.

2. Data integrity: integrity rules and constraints are enforced to ensure that data remain accurate, consistent and reliable. Integrity constraints, such as primary keys, foreign keys and unique constraints, help prevent data inconsistencies and ensure that data meet defined criteria.
3. Data independence: changing the structure (schema) of the database does not necessarily require changing the applications that use the data. The separation between the logical and physical aspects of the database allows for greater flexibility and adaptability.
4. Data security: a critical issue for databases. Sensitive and confidential information is stored in them, so procedures are needed to protect against data breaches, such as user authentication, authorization, encryption, protection against unauthorized access, etc.
5. Data consistency: data must comply with predefined rules and restrictions.
6. Queries: databases provide powerful tools that allow users to retrieve specific information from the database using structured query language (SQL) or other query languages. Queries can filter, sort, and aggregate data based on specific criteria, allowing users to extract meaningful insights from the database.
7. Data recovery: tools that can efficiently recover databases in the event of data loss due to hardware failures, software errors or other unforeseen events. Backup and recovery strategies ensure that data can be restored consistently during a disaster or system failure.

1.2. Types of databases

Relational databases: relational databases are the most widely used type of database management system (DBMS). Data is organized in tables of rows and columns, where each row represents a record, and each column represents an attribute. Relational databases use a structured query language (SQL) to manipulate and query data. Examples of relational databases include MySQL, PostgreSQL, Oracle Database, Microsoft SQL Server and MsAccess, which is used in classroom exercises.

NoSQL databases. Unlike relational databases, NoSQL databases do not use a table schema and support flexible data models such as key-value pairs, document-oriented, column-oriented and graph databases. NoSQL databases are highly scalable and can efficiently handle different data types. Examples of NoSQL databases include MongoDB, Cassandra, Couchbase and Redis.

Document databases. Each document contains key-value pairs and can have a variable schema for dynamic data structures. Document databases suit applications with variable data requirements and complex data models. Examples of document databases are MongoDB, Couchbase, CouchDB and RavenDB.

2. Database Management System (DBMS)

2.1. Definition of DBMS

A database management system (DBMS) is a software application that facilitates database creation, management and administration. It acts as an intermediary between users and the underlying database and provides an interface for users to interact with the data stored in the database.

In principle, a DBMS offers several vital functions:
Data definition: the DBMS allows users to define the database structure, including creating tables, defining data types, defining relationships between tables, and enforcing integrity constraints. This ensures meaningful and consistent organization and structuring of data.

Data manipulation: the DBMS allows users to perform various operations on the data stored in the database. These include inserting new records, updating existing records, deleting obsolete records and querying the database to retrieve specific information. Database management systems usually support a query language, such as SQL (Structured Query Language), which allows users to interact with the database using standardized commands.

Data security and access control: DBMS provide mechanisms to control access to the database and ensure data security. It allows administrators to define user roles and privileges, restrict access to sensitive data, and enforce authentication and authorization policies to prevent unauthorized access and data breaches.

Data integrity and consistency: the DBMS applies integrity constraints and enforcement policies to maintain data integrity and consistency. It ensures that data stored in the database conforms to predefined rules and constraints, preventing inconsistencies, errors and data corruption. DBMS also supports transactions, allowing users to perform multiple database operations as a single atomic unit, ensuring that data remains consistent even during errors or interruptions.

Data recovery and backup: the DBMS provides data recovery and backup mechanisms to protect against data loss caused by hardware failures, software errors or other unforeseen events. It allows administrators to perform regular database backups, perform incremental backups, and implement disaster recovery plans to ensure that data can be restored consistently during a disaster or system failure.

2.2. Database management systems

Different database management systems (DBMS) are tailored to specific data storage and management requirements. These include relational database management systems covering 95% of all systems.

A relational DBMS (RDBMS) is the most common type of DBMS, which organizes data into tables with rows and columns. Each table represents an entity, and the relationships between entities are created using keys. RDBMS follows the relational model and uses the SQL (Structured Query Language) language to manipulate and query data. Examples of RDBMS include MySQL, PostgreSQL, Oracle Database, SQL Server and SQLite.

An object-oriented DBMS (OODBMS) stores data as objects consisting of attributes and methods. It is well-suited for applications with complex data structures and relationships. OODBMS supports encapsulation, inheritance and polymorphism, making it suitable for modelling real entities and their behaviour. Examples of OODBMS include db4o and ObjectDB.

NoSQL (Not Only SQL) DBMS is a non-relational database management system designed to handle large amounts of unstructured and semi-structured data. NoSQL databases offer flexibility, scalability and performance advantages over traditional RDBMS, especially for web-scale applications and big data analytics. NoSQL databases can be categorized into different types, including document-oriented, key-value, column-family and graph databases. Examples of NoSQL databases include MongoDB, Cassandra, Redis, Couchbase and Neo4j.

3. Relational database management system (RDBMS)

A relational database management system (RDBMS) is a software system that enables the creation, maintenance and management of relational databases. RDBMS is based on a relational model of data, which organizes data into tables of rows and columns. The RDBMS provides a structured and efficient mechanism for storing, retrieving and managing data in different applications and domains. Its tabular structure characterizes it. The RDBMS organizes data into tables, each representing an entity or relationship. Tables consist of rows (also known as tuples) and columns (also known as attributes), where each row represents a single record or instance of the entity, and each column represents a particular attribute or characteristic of the entity. Relationships are established between entities to simulate the "mini world" we want to store and how it works. Primary keys uniquely identify each record in a table, while foreign keys create relationships between tables by referencing the primary keys of related tables. The relationships enable data integrity and consistency in the database.

Data accuracy is one of the most critical requirements for databases. The RDBMS provides various constraints to ensure data integrity and consistency. These constraints include primary key constraints, unique constraints, foreign key constraints, control constraints and non-zero constraints. These ensure that data remains accurate, reliable and consistent. ACID properties (Atomicity, Consistency, Isolation, Durability) guarantee that database transactions are executed reliably, consistently, and in isolation and that changes made to the database are persistent and permanent. A consistent, easy-to-learn language has been developed to structure the database and facilitate data management. The RDBMS uses SQL to define, query and manipulate relational databases. SQL provides a comprehensive set of commands for creating tables, inserting, updating and deleting data, querying data with SELECT statements, and performing data manipulation and transaction control operations.

3.1. Tables, records and fields

In relational databases, tables, records, and fields are essential components that organize and store data in a structured way. Tables are the primary organizational units of a relational database. They represent collections of related data units or objects within a database schema. Tables store and manage data logically and structured, allowing efficient retrieval, management, and information analysis. Tables can contain multiple records, each representing a unique instance of the data unit stored in the table.

A record, a tuple or row, is a single instance or entry within a table. It represents a complete data set associated with a particular entity or object. The records consist of the values of each field defined in the table, which together represent the attributes and properties of the entity. Each record in a table contains values corresponding to each field or attribute defined for that table. For example, in a table storing employee data, each record may represent a single employee, with fields such as employee ID, name, class, and salary. Fields, also known as columns or attributes, are the unique data elements or properties that define the table's structure. Each field represents specific information about the entities or objects stored in the table. Fields define the characteristics and properties of the data stored in the table, such as data type, length, format, and constraints. Fields define the structure and properties of the data stored in a table, including data type, size, and constraints. Common examples of fields are names, dates, numbers, texts, and balloon values.

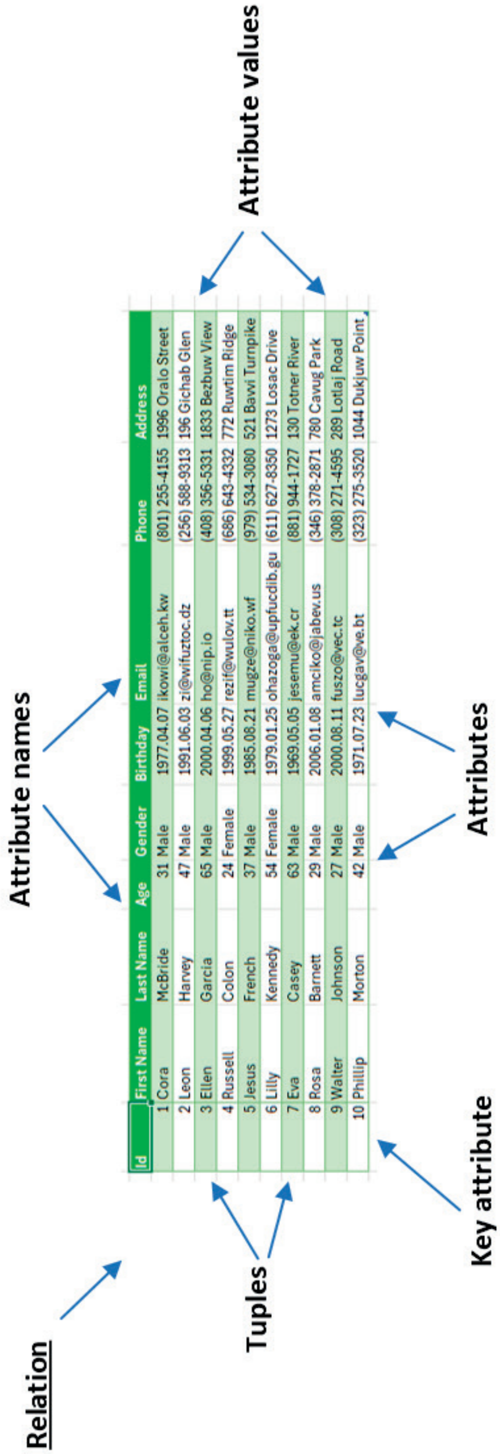
3.2. Primary and foreign keys

Primary and foreign keys create links between tables and ensure data integrity. A primary key is a column or combination of table columns uniquely identifying each record or row. They must have unique values for each record and cannot contain NULL values. Defining a primary key ensures data integrity by preventing duplicate records and providing a unique identifier for each record in the table. Primary keys are defined when the table is created or modified.

A foreign key is a column or group of columns in a table that establishes a relationship with another table's primary key or unique key. Foreign keys represent relationships between tables, allowing data to be linked and referenced across multiple tables. They ensure referential integrity by ensuring that the values of the foreign key column(s) match the values of the corresponding primary key or unique key column(s) of the related table. If a foreign key is defined, it helps maintain data consistency by preventing operations such as deleting or updating records that violate referential integrity constraints.

3.3. Normalization

Normalization is a technique used in database design to optimize data storage, reduce redundancy and improve data integrity. Normalization eliminates data anomalies such as insertion, update, and deletion and ensures the database remains flexible, scalable, and maintainable. It does this by structuring the relational database in such a way as to minimize the redundancy and dependency of the data stored in it. This involves breaking down larger tables into smaller, related tables and defining their relationships. Normalization helps to improve data integrity, reduce data redundancy and simplify database maintenance. However, it can also increase the complexity of database queries and result in performance degradation for certain types of queries.



Id	First Name	Last Name	Age	Gender	Birthday	Email	Phone	Address
1	Cora	McBride	31	Male	1977.04.07	ikowi@alceh.kw	(801) 255-4155	1996 Oralo Street
2	Leon	Harvey	47	Male	1991.06.03	zi@wifuztoc.dz	(256) 588-9313	196 Gichab Glen
3	Ellen	Garcia	65	Male	2000.04.06	ho@nip.io	(408) 356-5331	1833 Bezbuw View
4	Russell	Colon	24	Female	1999.05.27	rezif@wulov.tt	(686) 643-4332	772 Ruwtim Ridge
5	Jesus	French	37	Male	1985.08.21	mugze@niko.wf	(979) 534-3080	521 Bawi Turnpike
6	Lilly	Kennedy	54	Female	1979.01.25	ohazoga@upfucdib.gu	(611) 627-8350	1273 Losac Drive
7	Eva	Casey	63	Male	1969.05.05	jesemu@ek.cr	(881) 944-1727	130 Totner River
8	Rosa	Barnett	29	Male	2006.01.08	amciko@jabev.us	(346) 378-2871	780 Cavug Park
9	Walter	Johnson	27	Male	2000.08.11	fuszo@vec.tc	(308) 271-4595	289 Lotlaj Road
10	Phillip	Morton	42	Male	1971.07.23	lucgav@ve.bt	(323) 275-3520	1044 Dukjuw Point

Figure 2.1. Relation – Row (Tuple, Record, Key, Values)

Source: own editing

The steps of the normalization are:

1. First normal form (1NF): ensures that each table has a primary key and that each column contains atomic (indivisible) values.
2. Second normal form (2NF): Eliminates partial dependencies by ensuring that all non-key attributes are functionally fully dependent on the full primary key.
3. Third Normal Form (3NF): Eliminates transitive dependencies by ensuring that all non-key attributes are functionally dependent only on the primary key and not on other non-key attributes.

4. Structured Query Language (SQL)

SQL, or Structured Query Language, is a powerful programming language for managing and manipulating relational databases. It is the standard language for interacting with relational database management systems (RDBMS) such as MySQL, PostgreSQL, SQL Server, Oracle Database and SQLite. SQL provides a rich set of commands and syntax for performing various database operations, including querying, inserting, updating, deleting and manipulating schema. It is a versatile and widely used language that is indispensable for database administrators, developers, analysts and other professionals involved in managing and administrating relational databases. Thanks to its intuitive syntax and comprehensive command set, SQL allows users to interact efficiently with databases and easily perform complex data operations. Learning SQL is essential for anyone working with relational databases, and it is a valuable skill in data management and analysis.

The statements are grouped into 5 categories:
Data Query Language (DQL): contains a set of commands for querying and retrieving data from databases. The SELECT statement is the primary DQL command to retrieve data from one or more tables based on specified criteria. SQL queries can filter, sort and aggregate data to extract meaningful information from the database.

Data Definition Language (DDL): provides commands to define and manage the structure of databases and database objects. DDL commands include CREATE, ALTER and DROP, which are used to create, modify and delete database schemas, tables, indexes, views and other objects. DDL commands allow users to define the structure of the database, specify data types, set constraints, and create relationships between tables.

Data manipulation language (DML): contains commands for manipulating data within tables. DML commands include INSERT, UPDATE and DELETE, which are used to add new records, modify existing records and remove records from tables. DML commands allow users to modify data stored in the database and maintain data integrity and consistency.

Data Control Language (DCL): contains commands for managing user access and permissions. DCL commands include GRANT and REVOKE, which are used to grant or revoke privileges and permissions for users and roles. DCL commands allow administrators to control access to data, enforce security policies, and ensure data privacy and confidentiality.

Transaction Control Language (TCL): contains commands for managing database transactions. TCL commands include COMMIT, ROLLBACK and SAVEPOINT, which are used to manage transactions' atomicity, consistency, isolation and durability (ACID) properties. TCL commands allow users to control transaction boundaries, commit or rollback changes, and maintain data integrity in a multi-user environment.

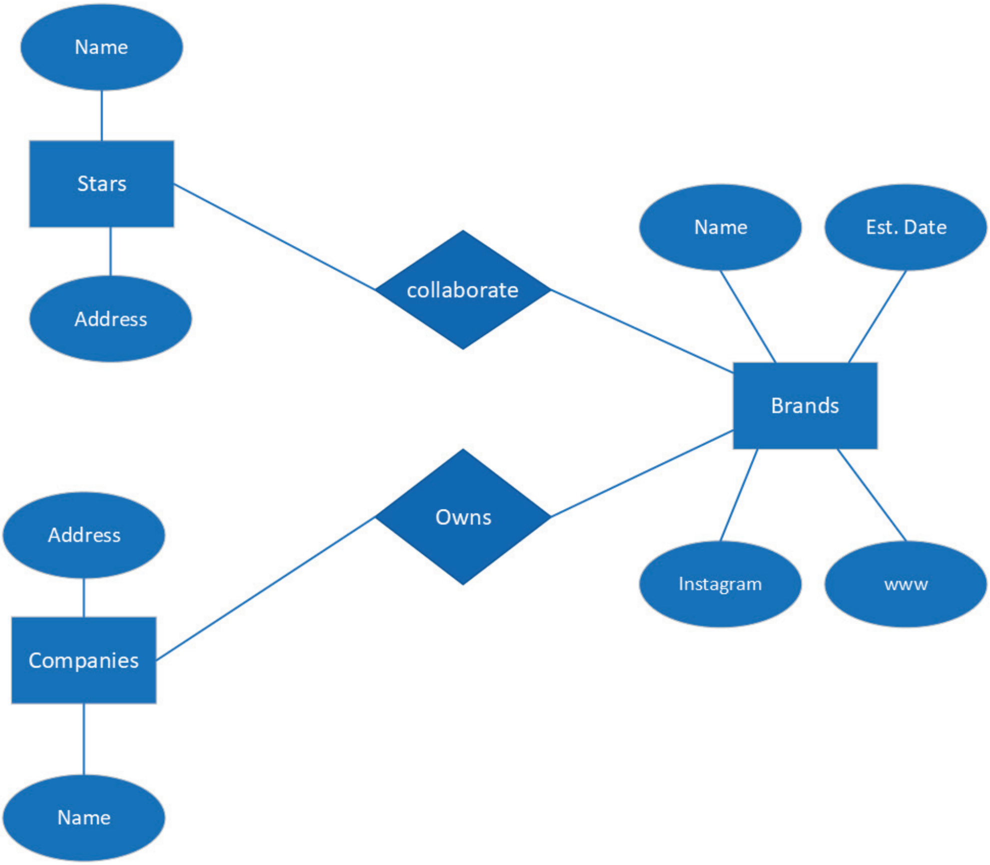


Figure 2.2. ER diagram that represents a simple database about brands. The entity sets are Brands, Stars, and Companies

Source: own editing

5. Conceptual Data Modeling

Conceptual data modelling is a key database design phase that focuses on understanding and representing an organization's core business concepts and relationships. It involves creating a high-level, abstract representation of data requirements and structures, independent of any specific database management system (DBMS) or implementation details. The primary objective of conceptual data modelling is to define the key entities, attributes and relationships relevant to the organization's business processes and data requirements. It also serves as a communication tool between stakeholders, including business analysts, data engineers and end users, to ensure a common understanding of data elements and their relationships. Modelling is an iterative process that involves continuous refinement and validation of the model based on stakeholder feedback. To gather requirements and refine the model, you can use interviews, workshops, and documentation reviews.

Entity-relationship (ER) modelling: Entity-relationship modelling is the most commonly used approach in conceptual data modelling. It represents key entities (objects or concepts) within an organization and the relationships between them. Entities represent an organisation's key objects or concepts, such as customers, products, orders, employees, etc. Relationships define the relationships and interactions between entities, such as "is-a", "has-a" or "belongs-to" relationships. Attributes describe properties or characteristics of entities. They provide additional details and information about the entities and are typically represented as columns in a database table. Attributes can be simple (atomic) or complex (composed of several sub-attributes), and their data type can be text, numeric, date, etc. The nature and extent of relationships between entities in the conceptual data model are determined by cardinality and multiplicity. Cardinality describes the number of instances of an entity that can be associated with another entity. It is expressed in terms such as 'one to one', 'one to many' or 'many to many'. Multiplicity specifies an entity's minimum and maximum number of occurrences within a relationship. It is often represented graphically using entity-relationship diagrams (ERDs) for ease of understanding. ERDs provide a visual representation of entities, relationships, attributes, and their cardinality and multiplicity limits. Boxes represent entities, lines represent relationships, and diamonds represent relationship types. Attributes are listed in boxes representing entities.

5.1. Logical Data Modeling

Logical data modelling is an essential phase of the database design process that focuses on transforming the conceptual data model into a more detailed and structured representation that reflects the logical organization of data in a relational database management system (RDBMS). The logical data model builds on the conceptual data model developed in the conceptual data modelling phase. While the conceptual data model focuses on the definition of high-level entities, attributes and relationships, the logical data model adds additional details, including data types, constraints and normalization. As with conceptual data modelling, logical data modelling often uses entity-relationship diagrams (ERDs) to represent entities, relationships and attributes. However, in the logical data model, entities are transformed into database tables, attributes become columns of tables, and foreign keys represent relationships. Normalization is one of the most important aspects of logical data modelling, which aims to organize data efficiently and reduce data redundancy. In which large tables are broken down into smaller, related tables to eliminate data anomalies and ensure data integrity. Normal forms such as the first normal form (1NF), the second normal form (2NF), the third normal form (3NF) and the Boyce-Codd normal form (BCNF) are used to ensure proper normalization of each table.

Logical data modelling defines the data types and constraints for each column in the table. Data types, such as integer, character, date, etc., define the data type stored in each column. Constraints enforce rules and conditions on the data stored in the database, such as primary key constraints, foreign key constraints, unique constraints, and control constraints. In addition, the process involves defining each table's primary and foreign keys to create relationships between tables. If necessary, indexes can be defined to optimize query performance. Enabling faster data retrieval based on specific columns.

6. Database security

Database security is a critical aspect of information security that focuses on protecting the data stored in databases from unauthorized access, disclosure, modification or destruction. Since databases contain sensitive and valuable information, their protection is essential to maintain the data's confidentiality, integrity and availability. Access control mechanisms regulate who can access the database and what operations can be performed on the data. Role-based access control (RBAC) allocates rights and privileges to users based on their organizational roles and responsibilities. User authentication mechanisms such as passwords, biometrics and multi-factor authentication verify the identity of users accessing the database.

Encryption techniques protect data stored in databases from unauthorized access by converting it into ciphertext using encryption algorithms. Transparent Data Encryption (TDE) encrypts data at the storage level, ensuring that data remains encrypted while stored on disk. Column-level encryption selectively encrypts sensitive columns of database tables, allowing finer control over data protection.

Control mechanisms monitor and record database activity, including login attempts, data modifications and access control changes. Database audit logs record who accessed the database, what operations were performed and when they were performed. Real-time monitoring tools analyze database activity in real-time and detect suspicious or anomalous behaviour that may indicate security incidents.

Data masking techniques hide sensitive information in databases by replacing actual data with fictitious or obfuscated data. Reduction techniques selectively hide or anonymize sensitive data in query results, reports or application interfaces, preventing unauthorized disclosure.

Patch management includes the application of software patches, updates and security patches to the database management system (DBMS) and related software components. Regular patching helps mitigate security vulnerabilities and protect against known exploits and malicious attacks.

Backup and recovery mechanisms protect databases from data loss and corruption caused by hardware failures, software errors or malicious attacks. Regular backups ensure that data can be restored to its previous state during a disaster or security incident.

Database activity monitoring solutions analyze database traffic and user activity to identify potential security threats, policy violations and data breaches. DAM solutions provide real-time alerts, reports, and forensic analytics so organizations can immediately detect and respond to security incidents.

Database security measures must be consistent with industry regulations, data protection laws and organizational security policies. Compliance frameworks such as GDPR, HIPAA, PCI DSS and SOX have requirements for protecting sensitive data, controlling access and maintaining audit trails.

REFERENCES

Ardeleanu, S. (2016). Relational database programming. ISBN 978-1-484220-79-5.

Churcher, C. (2012). Beginning database design: From novice to professional. ISBN 978-1-4302-4210-9.

Date, C. J. (2019). Database design and relational theory. ISBN 978-1484255391.

Gajdos, S. (2016). Adatbázisok. ISBN 978-963-313-195-4.

Garcia-Molina, H., Ullman, J. D., & Widom, J. (2009). Database systems: The complete book. ISBN 0-13-606701-6.

Kovács, L. (2004). Adatbázisok tervezésének és kezelésének módszertana. Budapest: ComputerBooks. ISBN 963618321X.

Negi, M. (2019). Fundamentals of database management system: Learn essential concepts of database systems. ISBN 978-9388176620.

Taylor, A. G. (2000). Database development for dummies (1st ed.). ISBN 0-7645-0752-9.

CHAPTER 3: INFORMATION SYSTEMS

(Péter Lengyel)

1. Introduction to Information Systems

1.1. Definition and basic concepts of information systems

In today's digital age, Information Systems (IS) are the backbone of almost every organization, driving operational excellence, fostering innovation, and enabling informed decision-making. Understanding the fundamental concepts of information systems is crucial for students aspiring to excel in the modern business environment.

What is an Information System?

An information system is an integrated ensemble of people, processes, data, and technology that supports an organisation's operations, management, and decision-making functions. Unlike mere data, the output of an information system is information that has been processed to provide value to its users.

Key Concepts in Information Systems:

- **Interdisciplinary Nature**: Information Systems is a field that intersects business, technology, and management. It emphasizes not only the technological components but also their application in solving business problems and supporting strategic goals.
- **Socio-technical Systems**: This concept highlights that an information system comprises social and technical elements. It includes the technology (hardware, software, databases, and networks), the people (users and IT professionals), and the processes that make the system functional and useful.

- **Data vs. Information**: It's vital to differentiate between these two concepts. Data represents raw facts and figures, while information is data that has been processed and organized to add value for its users. This transformation is at the heart of what information systems do.

Components of Information Systems:

- People: From IT professionals who develop and manage these systems to end-users who interact with them, people are crucial.
- Processes: These are the methods and procedures that outline how business activities are conducted using the system.
- Technology: This includes:
  - Hardware: Physical devices and equipment.
  - Software: The programs and applications.
  - Databases: Organized collections of data.
  - Networks: The infrastructure that supports communication between system components.

Objectives of Information Systems:

- Enhance Operational Efficiency: Streamlining business processes to increase productivity and reduce costs.
- Foster Innovation: Developing new products, services, and business models through technological capabilities.
- Build Customer and Supplier Intimacy: Strengthening relationships through efficient and responsive operations.
- Support Informed Decision Making: Providing timely and relevant information to enable better business decisions.
- Gain Competitive Advantage: Achieving superior performance in the marketplace through distinctive capabilities.
- Ensure Survival: Meeting basic regulatory requirements and adapting to the changing business environment.

Discussion Points and Activities:

- Case Study Analysis: Review how a specific company leveraged information systems to solve a business problem or gain a competitive advantage.
- Group Discussion: What are some everyday examples of information systems you interact with? Discuss their components and what makes them effective or ineffective.
- Research Assignment: Investigate a failed information system project. Identify the factors that led to its failure and propose solutions.

Understanding the definition and basic concepts of information systems is the first step toward appreciating their strategic role in today's businesses. As we delve deeper into each component and explore various information systems, we will understand how these systems are designed, implemented, and managed to achieve organizational goals.

1.2. Historical development of information systems

The evolution of information systems (IS) is a fascinating journey that mirrors the advancements in technology and the changing needs of businesses. Information systems have undergone significant transformations from simple manual systems to today's complex digital solutions. This section explores the key milestones in developing information systems and their impact on businesses and society.

The Genesis and Evolution:

- **Pre-computer era** (Before the 1950s): Early information systems were manual and paper-based. Businesses use ledgers, filing systems, and manual processes to collect, process, and store information. The main challenge was the time-consuming nature of processing large volumes of data.
- **The advent of computers** (1950s–1960s): The introduction of mainframe computers marked the beginning of the digital information system era. These early computers were large and expensive and primarily used by big corporations to process vast data. Languages like COBOL and FORTRAN emerged, facilitating business and scientific computing.
- **Personal computers and databases** (1970s–1980s): The invention of personal computers (PCs) and the development of database management systems (DBMS) democratized computing. Information storage and retrieval became more efficient, paving the way for more sophisticated information systems.
- **The Internet and World Wide Web** (1990s): The Internet revolutionized information systems by enabling unprecedented connectivity and access to information. The development of the World Wide Web and browsers in the early 1990s made sharing and accessing information globally easier.
- **Enterprise Resource Planning (ERP) Systems** (Late 1990s): ERP systems integrated all facets of an enterprise into a unified information

system. Companies could manage their operations through a single system, from production to human resources.

- **The age of Big Data and Cloud Computing** (2000s-Present): The explosion of digital data and the advent of cloud computing have transformed information systems. Big data analytics, cloud-based services, and mobile computing have enabled more agile, scalable, and data-driven information systems.

Impact on Businesses and Society:

- **Efficiency and productivity**: The evolution of information systems has significantly increased businesses' efficiency and productivity by automating processes and enabling faster decision-making.
- **Globalization**: Information systems have made it easier for businesses to operate globally, connecting markets and people worldwide.
- **Innovation**: The advancements in information systems have fostered innovation, leading to new products, services, and business models.

1.3. The role of information systems in modern enterprises

In the digital landscape of modern business, information systems (IS) have emerged as critical tools for achieving strategic goals, enhancing operational efficiency, and fostering innovation. As enterprises face the complexities of globalization and technological evolution, the adept use of IS is a key determinant of success.

Strategic and Operational Impact

Information systems enable businesses to carve out competitive advantages by providing insights that inform decision-making and improve operational efficiencies. By automating routine tasks, IS increases productivity and reduces errors, contributing to operational excellence. Additionally, real-time data monitoring allows for swift adjustments to business strategies in response to market changes.

The strategic deployment of IS in supply chain management and customer relationship management (CRM) has revolutionized these areas. By streamlining processes and enhancing communication, businesses can optimize costs and build stronger relationships with customers and suppliers.

Innovation and Growth

IS are foundational to data-driven innovation, opening new avenues for exploring business models, products, and services. They support the creation and delivery of digital products, facilitating traditional business model transformation and market expansion. Moreover, IS plays a crucial role in enabling global operations, breaking down geographical barriers and fostering international growth.

Challenges and Adaptation

Implementing and managing information systems comes with challenges, including data security and privacy concerns, keeping pace with technological advancements, and aligning IS with business strategies. Overcoming these obstacles requires effective leadership, a culture of continuous learning, and an openness to change.

The role of information systems in modern enterprises is multifaceted, driving strategic initiatives, operational efficiencies, and innovation. As technology evolves, the importance of IS in shaping the future of business in the digital era continues to grow. Organisations that successfully integrate and leverage these systems will find themselves well-positioned to thrive in an increasingly competitive and complex business environment.

2. Fundamental components of Information Systems

Information systems are integral to the modern business landscape, enabling organisations to operate efficiently, make informed decisions, and compete in a global market. At the core of any information system are five fundamental components: hardware, software, data, people, and processes. Understanding these components' roles and interplay is crucial for effectively leveraging IS.

2.1. Hardware and software

The physical devices and equipment constitute any information system's technological backbone. Hardware includes computers, servers, data centres, switches, routers, and other devices essential for operations. The choice of hardware impacts the system's performance, scalability, and reliability. Software refers to the programs and applications that run on the hardware, instructing it on performing tasks. This includes operating systems, enterprise

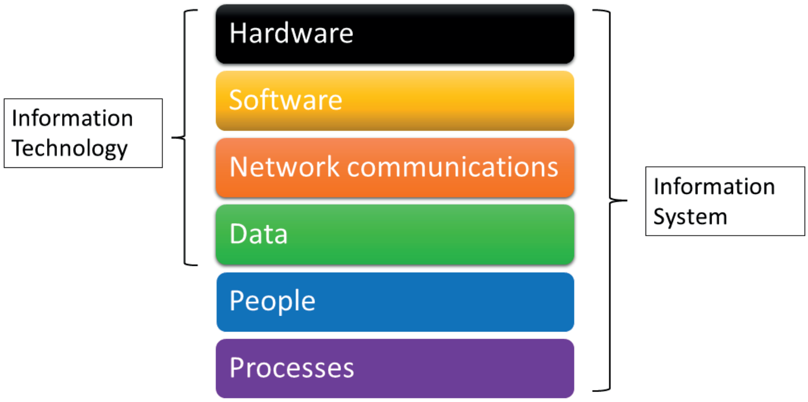


Figure 3.1. Fundamental Components of Information Systems

Source: Pham & Desai-Naik, 2024

software, application software, and databases. The software determines the functionalities available in an information system, from data processing and analysis to user interface design and interaction.

2.2. Data and information

In information systems, the terms 'data' and 'information' are foundational yet distinct concepts that play pivotal roles. Understanding the nuances between data and information is essential for anyone looking to master information systems, as it lays the groundwork for grasping how these systems transform raw data into actionable insights. Data refers to raw facts and figures that, on their own, may not carry significant meaning to the untrained eye. These can be numbers, characters, images, or other measurements collected through observation. In information systems, data is the raw input that needs to be processed and analyzed to become functional. For example, the daily sales figures across different regions, the number of clicks on a website, or the temperatures recorded by a weather station all constitute data. Data is indispensable because it is the fundamental building block of information systems used to produce information.

Transforming data into information

In information systems, the transition from data to information is a critical process that involves several steps, including data collection, storage, processing, and analysis. Modern information systems utilize sophisticated algorithms and computing power to handle vast amounts of data and convert them into information efficiently. This capability allows organizations to:

- Make strategic decisions based on trends and insights derived from data analysis.
- Improve operational efficiency by identifying bottlenecks or inefficiencies in processes.
- Enhance customer satisfaction through personalized services based on customer data analysis.

The distinction between data and information is crucial in studying information systems. While data represents the raw inputs collected from various sources, information stands for processing data that has been structured or analyzed to support decision-making. The ability of information systems to effectively transform data into information is what empowers organizations

2.3. Networks and communication

Communication technologies encompass the hardware and software enabling network connectivity and the methods and tools facilitating user interaction. Email, instant messaging, video conferencing, and collaborative platforms are all examples of communication technologies integrated into information systems. These technologies enhance collaboration, streamline workflows, and enable real-time decision-making by ensuring relevant information is readily accessible to all stakeholders. With the increasing reliance on networks and communication technologies, ensuring the security and reliability of these systems has become paramount. Encryption, firewalls, and secure protocols are essential to protect data integrity and confidentiality. Additionally, redundancy, failover systems, and disaster recovery plans are crucial to maintaining system availability and reliability.

Networks and communication technologies are indispensable to modern information systems, facilitating the efficient and secure exchange of data and information. As the backbone of information systems, they enable organizations to operate in an increasingly interconnected world, driving collaboration, innovation, and strategic decision-making. Understanding the principles of networks and communication is fundamental for anyone designing, implementing, and managing information systems, underscoring their role in today's digital landscape.

2.4. People and processes

Information systems are not solely about technology but fundamentally human systems embedded within organizations. The success of these systems hinges not just on sophisticated software and hardware but equally on the people who design, implement, use, and manage them, as well as the processes they follow to achieve organizational goals. This subsection explores the interaction between human elements and organizational procedures within information systems.

The role of people:

- **Users**: At the heart of any information system are its users. They range from front-line staff interacting with the system daily to management using it for strategic decision-making. Their needs, skills, and satisfaction are critical for the system's effectiveness.
- **IT Professionals**: This group includes system analysts, developers, network administrators, and specialists who design, implement, and maintain information systems. Their expertise is crucial in ensuring the systems meet organizational needs and are secure, reliable, and scalable.
- **Executives and managers**: Leadership is pivotal in setting the vision for leveraging information systems to achieve strategic objectives. They are responsible for aligning IT investments with business goals and fostering a culture that embraces technological advancement.

Processes refer to the sequences of tasks or activities performed to achieve a specific outcome. In the context of information systems, processes can be manual or automated, and they dictate how tasks are carried out within the system, impacting efficiency, accuracy, and productivity. Business processes are the core activities that drive an organization's operations and strategic goals. Information systems are often designed to streamline and automate business processes, enhancing efficiency and enabling better decision-making. IT processes encompass the methodologies and practices IT professionals use to plan, develop, implement, and manage information systems. This includes software development lifecycles, project management practices, and IT service management.

3. Types and applications of Information Systems

This section explores the diverse types of information systems (IS) and their applications within organizations. These systems vary widely in their functions, ranging from operations and management to strategic decision support, each tailored to fulfil specific organizational needs. Understanding the variety and applications of these systems provides insights into how they support business processes, decision-making, and strategic goals.

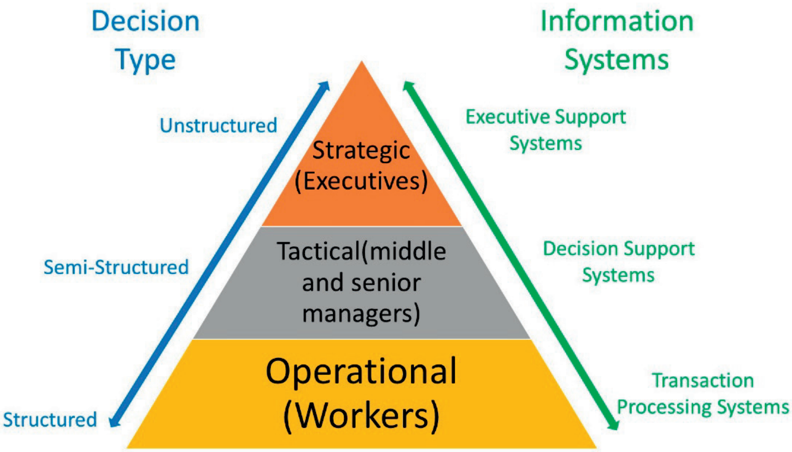


Figure 3.2. A three-level pyramid model of different types of Information Systems

Source: Pham & Desai-Naik, 2024

Transaction Processing Systems (TPS)

Definition: TPS are the backbone of daily operations in many organizations, designed to handle large volumes of routine, repetitive transactions such as billing, payroll, or inventory management.

Application: They are critical in sectors like retail, banking, and manufacturing, ensuring efficiency and accuracy in transaction processing. By automating these tasks, TPS save time and reduces errors, facilitating smooth operational processes.

Management Information Systems (MIS)

Definition: MIS provide managers with reports and access to the organization's current performance and historical records, primarily focusing on internal sources of information.

Application: MIS are used in decision-making processes regarding planning, controlling, and managing departments or divisions. They help managers analyze organizational performance, spot trends, and make informed decisions.

Decision Support Systems (DSS)
Definition: DSSs are interactive, computer-based systems that help managers make decisions by using data and models to solve unstructured or semi-structured problems.
Application: DSS use them in situations requiring judgment, evaluation, and insights for decision-making, such as forecasting and business planning. They enable managers to perform “what-if” analysis, considering various scenarios before making decisions.

Customer Relationship Management (CRM) Systems
Definition: CRM systems help businesses manage and maintain customer relationships, track engagements and sales, and deliver actionable data.
Application: They are crucial for sales, marketing, and customer service departments, enabling the organization to better understand and serve its customers, thus enhancing customer satisfaction and loyalty.

Supply Chain Management (SCM) Systems
Definition: SCM systems manage the flow of goods, data, and finances related to a product or service from procurement to delivery.
Application: They are vital for coordinating and integrating these flows both within and among companies. SCM systems improve the visibility of demand, supply, and logistics, ensuring efficient operations across the supply chain.

Enterprise Resource Planning (ERP) Systems
Definition: ERP systems integrate core business processes in one comprehensive information system, facilitating information flow across the organization.
Application: They streamline and standardize processes, providing a holistic view of the business from financials and HR to service delivery, thereby enhancing operational efficiency and strategic planning.

Knowledge Management Systems (KMS)
Definition: KMS support the creation, capture, storage, and dissemination of information and knowledge within an organization.
Application: These systems are essential for promoting knowledge sharing, innovation, and organisational cultural change. They enable employees to

access organizational knowledge bases, best practices, and lessons learned, fostering a collaborative learning environment.

The variety and application of information systems within organizations highlight their integral role in supporting and enhancing business operations, decision-making, and strategic initiatives. By effectively deploying these systems, organizations can achieve greater efficiency, improved decision-making capabilities, and enhanced competitiveness in their respective industries. Understanding these types offers valuable insights into how information technology can be leveraged to meet specific organizational needs and objectives.

4. Security and ethical issues in Information Systems

Security and ethical issues are increasingly at the forefront of information system design and implementation. As these systems become more integral to organizational operations and society, the implications of security breaches and unethical use of information technology can be profound. Addressing these concerns is not just about protecting data; it's about safeguarding trust and ensuring the responsible use of technology.

4.1. Principles of data protection and security

Data protection and security principles form the bedrock for responsibly managing, storing, and transmitting data, crucial for maintaining trust and compliance in the digital world.

Core principles Include:

- Data Minimization: Collect only essential data.
- Purpose Limitation: Use data solely for its intended purpose.
- Data Accuracy: Ensure information is accurate and updated.
- Storage Limitation: Retain data no longer than necessary.
- Integrity and Confidentiality: Protect data from unauthorized access through encryption and access controls.
- Accountability: Organizations must demonstrate compliance with data protection policies.
- Transparency: Be clear about the nature and use of collected data.
- Secure Data Transmissions: Encrypt data during transmission to prevent interception.

Adopting a comprehensive approach that includes regular security assessments, implementing layered security measures, and training employees on best practices is vital. Staying updated with legal and regulatory requirements, such as GDPR and CCPA, is also crucial for compliance. In essence, adhering to these principles guards against data breaches and builds trust with stakeholders, ensuring a secure digital environment.

4.2. Ethical and legal issues in information systems

Ethical and legal issues in information systems are pivotal concerns as technology increasingly intersects with daily life and business operations. These issues revolve around privacy, data accuracy, intellectual property rights, and access to technology.

Ethical Considerations:

- Privacy: Protecting individuals' personal information against unauthorized access and misuse.
- Accuracy: Ensuring information is correct and up-to-date to prevent harm.
- Intellectual Property: Respecting copyrights and ownership of digital content.
- Access: Bridging the digital divide and ensuring equitable access to technology.

Legal standards govern information system practices, addressing privacy (GDPR in Europe, CCPA in California), cybersecurity, and copyright laws. Compliance is mandatory, with regulations evolving alongside technological advancements.

Navigating ethical and legal landscapes in information systems demands a balance between leveraging technology for benefit while safeguarding rights and fostering inclusivity. Compliance with laws and ethical codes is essential for responsible operation and trust-building.

4.3. Modern challenges and solutions in data protection

Modern challenges in data protection are driven by rapid technological advancements, increasing cyber threats, and the complexity of global data flows. Solutions to these challenges require a multifaceted approach, blending technological, legal, and organizational strategies to safeguard data effectively.

Challenges:

- Sophisticated Cyberattacks: Hackers employ advanced techniques to exploit vulnerabilities, necessitating equally sophisticated defence mechanisms.
- Big Data and IoT: The explosion of data from various sources, including IoT devices, complicates data management and protection.
- Cloud Computing: While offering scalability and efficiency, cloud environments present unique security concerns, especially regarding data sovereignty and third-party risks.
- Compliance with Multiple Regulations: Organizations operating globally must navigate a labyrinth of data protection regulations, making compliance increasingly complex.

Solutions:

- Advanced Security Technologies: Utilizing AI and machine learning can help predict and mitigate cyber threats more effectively.
- Privacy by Design: Integrating data protection measures from the initial design phase of products and services.
- Enhanced Encryption: Applying robust encryption standards for data at rest and in transit.
- Regular Audits and Training: Conduct security audits and train employees on data protection best practices.
- Data Protection Officers (DPOs): Appointing DPOs to oversee data protection strategies and ensure regulatory compliance.

Implementing these solutions requires a commitment to continuous improvement and adaptation to evolving threats and regulatory landscapes. By proactively addressing these modern challenges, organizations can protect their data assets and maintain trust with stakeholders.

5. Trends in Information Systems

The landscape of information systems is continuously evolving, shaped by technological advancements, changing organizational needs, and societal shifts. Several future trends are set to redefine how information systems are developed, deployed, and utilized, offering new opportunities and challenges.

Artificial Intelligence and Machine Learning: AI and machine learning are increasingly integrated into information systems, enhancing decision-making, automating routine tasks, and providing personalized user experiences. These technologies are becoming more sophisticated, enabling systems to learn from data, predict outcomes, and make intelligent decisions in real-time.

Internet of Things (IoT): The expansion of IoT devices is transforming information systems, enabling the collection and analysis of data from a myriad of connected devices. This trend facilitates more responsive and context-aware systems, impacting everything from smart homes and cities to advanced manufacturing processes.

Blockchain Technology: Originally known for underpinning cryptocurrencies, blockchain is gaining traction in various sectors for its ability to provide secure, transparent, and tamper-proof record-keeping. Its applications in information systems range from enhancing data security and integrity to facilitating smart contracts and decentralized applications.

Cloud Computing and Edge Computing: The cloud continues to be a significant trend, offering scalability, flexibility, and cost-efficiency. Edge computing, which processes data closer to where it is generated, is increasingly important, especially for IoT and mobile applications, reducing latency and bandwidth use.

Cybersecurity Mesh: As cyber threats become more sophisticated, a more integrated approach to cybersecurity is emerging. Cybersecurity mesh is a flexible, modular architecture connecting disparate security services for a more cohesive and responsive defence strategy.

Privacy-Enhancing Computation: In response to growing data privacy concerns, privacy-enhancing computation techniques that allow data to be processed securely without compromising privacy are being developed. This trend is critical for maintaining user trust and complying with stringent data protection regulations.

Quantum Computing: Although still in its early stages, quantum computing has the potential to revolutionize information systems by solving complex problems much faster than traditional computers. Its impact could be game-changing in cryptography, drug discovery, and climate modelling.

The future of information systems is poised at the intersection of innovation and complexity. These trends highlight the field's dynamic nature, emphasizing the need for ongoing learning and adaptation. As these technologies continue to develop, they promise more efficient, secure, and intelligent information systems that can address both current challenges and future opportunities, reshaping the digital landscape.

In concluding our exploration of information systems, we've traversed from foundational concepts to the brink of tomorrow's innovations, uncovering the integral role these systems play in the tapestry of modern life. Our journey illuminated the complexities of designing, implementing, and securing these systems while forecasting emerging technologies' transformative potential.

As we've seen, information systems are much more than mere conduits for data; they are the engines of innovation, the backbone of enterprises, and the facilitators of global connectivity. As we stand on the cusp of a future rich with possibilities—from AI-driven analytics to quantum computing breakthroughs—the promise of information systems to enhance, innovate, and transform our world is unmistakable.

Yet, this journey also underscores the importance of ethical stewardship and vigilant security in an era where data is both a valuable asset and a vulnerable target. The delicate balance between leveraging technology for advancement and safeguarding our digital and societal fabric demands careful attention.

As we look ahead, the information systems landscape will undoubtedly continue to evolve, shaped by the ingenuity of those who navigate its challenges and harness its potential. The journey through the world of information systems is not solitary. Still, a shared voyage that calls on us to contribute to a future where technology amplifies our capabilities, enriches our lives and fosters a more connected, informed, and ethical world.

In essence, the horizon of information systems is vast and promising, filled with opportunities for innovation, growth, and profound societal impact. As we move forward, let us do so with curiosity, responsibility, and an unwavering commitment to using technology as a force for good, steering towards a future where information systems continue transforming and elevating our collective human experience.

REFERENCES

Acquisti, A., Dinev, T., & Keil, M. (2019). Editorial: Special issue on cyber security, privacy and ethics of information systems. Information Systems Frontiers. https://doi.org/10.1007/S10796-019-09971-5

Behara, R. S., Gundersen, D. E., & Capozzoli, E. A. (1995). Trends in information systems outsourcing. Journal of Purchasing and Materials Management, 31(2), 45–53. https://doi.org/10.1111/J.1745-493X.1995.TB00202.X

Ceccarelli, A., Bondavalli, A., Froemel, B., Hoeftberger, O., & Kopetz, H. (2016). Basic Concepts on Systems of Systems. In Springer. https://doi.org/10.1007/978-3-319-47590-5_1

Dantas dos Santos, W. (2019). The entropic and symbolic components of information. BioSystems, 177, 63–71. https://doi.org/10.1016/J.BIOSYSTEMS.2019.05.003

Evangelista Silva, S., Ribeiro, F. N., Camarda, R. F., & Amorim, V. J. P. (2018). Lifecycle Information Systems: The concept, principles, and an approach. *iSys*. https://doi.org/10.5753/ISYS.2018.370

Galinkin, E. (2022). Towards a Responsible AI Development Lifecycle: Lessons From Information Security. arXiv.org. https://doi.org/10.48550/arXiv.2203.02958

Kapur, N. K. (2023). Information Lifecycle Management. Forschung zur Digitalisierung der Wirtschaft. https://doi.org/10.1007/978-3-658-40304-1_3

Milchev, G., & Miltchev, R. (2018). Development of Information Technologies, Planned Obsolescence and Modification of the Life-Cycle of the CAD/CAM/CAE Systems. European Journal of Sustainable Development, 7(3), 217. https://doi.org/10.14207/EJSD.2018.V7N3P217

Nold, E. G. (1997). Trends in health information systems technology. American Journal of Health-System Pharmacy, 54(3), 269–274. https://doi.org/10.1093/AJHP/54.3.269

Safronov, V. V., Barabanov, V. F., Nuzhnyy, A. M., & Grebennikova, N. I. (2021). Integrated information system of lifecycle support of networks and electrical network equipment. *Journal of Physics: Conference Series*. https://doi.org/10.1088/1757-899X/1035/1/012037

Saha, S., Das, A., Kumar, A., Biswas, D., & Saha, S. (2019). Ethical hacking: Redefining security in information system. In Book Title. https://doi.org/10.1007/978-981-15-0361-0_16

Taherdoost, H. (2022). An overview of trends in information systems: Emerging technologies that transform the information technology industry. Cloud Computing and Data Science, 4(1). https://doi.org/10.37256/ccds.4120231653

Pham, L. – H. T., & Desai-Naik, T. (2024). Using information systems for competitive advantage. In Information systems for business 2e. Workforce LibreTexts. Retrieved from https://workforce.libretexts.org/Courses/Evergreen_Valley_College/Information_Systems_for_Business_2e/07%3A_Leveraging_Information_Systems_for_Strategic_Advantage/7.04%3A_Using_Information_Systems_for_Competitive_Advantage [Accessed: 30.05.2024]

Zelinska, O., Potapova, N., & Yemelianova, A. O. (2023). Information system for maintaining the register of clients of the bank. Visnik Khmelnytskyi National University, 2307–5732. https://doi.org/10.31891/2307-5732-2023-317-1-94-99

Zemmouchi-Ghomari, L. (2021). Basic Concepts of Information Systems. In InTechOpen. https://doi.org/10.5772/INTECHOPEN.97644

CHAPTER 4: BUSINESS ANALYSIS

(Róbert Szilágyi)

1. Data, information, and knowledge

Data are facts of the world. For example, financial transactions, age, temperature, number of steps from my house to my office are simply numbers. The information appears when we work with those numbers, and we can find value and meaning. The information can help us to make informed decisions.

We can talk about knowledge when the data and the information turn into rules to assist the decisions. We can't store knowledge because it implies a theoretical or practical understanding of a subject. However, using predictive analytics, we can simulate intelligent behaviour and provide a good approximation.

An example of data information:
How much? – not detailed question
20 – a numerical data
What is 20? – question to understand the data
What was the question? – another question to find the answer to the original question

The nature of data
We can find data in all the situations of the world around us, in all the structured or unstructured, in continuous or discrete conditions, in weather records, and in stock market logs. Data can be seen as the essential raw material of any kind of human activity.

As shown in the following figure, we can see **Data** in two distinct ways: **Categorical** and **Numerical**.

Categorical data are values or observations that can be sorted into groups or categories. There are two types of categorical values: nominal and ordinal. A nominal variable has no intrinsic ordering to its categories. For example, housing is a categorical variable having two categories (own and rent). An ordinal variable has an established ordering. For example, age as a variable with three orderly categories (young, adult, and elder).

Numerical data are values or observations that can be measured. There are two kinds of numerical values: discrete and continuous. Discrete data are values or observations that can be counted and are distinct and separate. For example, the number of lines in a code. Continuous data are values or observations that may take on any value within a finite or infinite interval. For example, an economic time series such as historic gold prices.

1.1. The data analysis process

Data analysis helps us to make this possible through exploring the past and creating predictive models.

The data analysis process is composed of the following steps:

- The statement of problem
- Obtain your data
- Clean the data
- Normalize the data
- Transform the data
- Exploratory statistics
- Exploratory visualization
- Predictive modeling
- Validate your model
- Visualize and interpret your results
- Deploy your solution

All these activities can be grouped as shown in the following figure:

Figure 4.1. Data analysis process activities

Source: Cuesta, 2013

The problem

The problem definition starts with high-level questions such as how to track differences in behaviour between groups of customers or what's going to be the gold price in the next month. Understanding the objectives and requirements from a domain perspective is the key to a successful data analysis project.

Types of data analysis questions are listed as follows:

- Inferential
- Predictive
- Descriptive
- Exploratory
- Causal
- Correlational

Data preparation

Data preparation is about obtaining, cleaning, normalising, and transforming the data into an optimal dataset, trying to avoid any possible data quality issues such as invalid, ambiguous, out-of-range, or missing values. This process can take a lot of your time. Analyzing data that has not been carefully prepared can lead you to highly misleading results.

The characteristics of good data are listed as follows:

- Complete
- Coherent
- Unambiguous
- Countable
- Correct
- Standardized
- Non-redundant

Data analysis is the process in which raw data is ordered and organized, which is used in methods that help to explain the past and predict the future. Data analysis is not about the numbers. It is about making/asking questions, developing explanations, and testing hypotheses. **Data Analysis** is a multidisciplinary field which combines **Computer Science**, **Artificial Intelligence and Machine Learning**, **Statistics and Mathematics**, and **Knowledge Domain** as shown in the following figure.

What about big data?

Big data is a term used when the data exceeds the processing capacity of a typical database. We need big data analytics when the data grows quickly, and we need to uncover hidden patterns, unknown correlations, and other useful information.

There are three main features of big data:

- **Volume:** Large amounts of data (GB, TB, PB)
- **Variety:** Different types of structured, unstructured, and multi-structured data (e.g. Database, photo, web, video, mobile, social, unstructured)
- **Velocity:** Needs to be analyzed quickly (periodic, near real-time, real-time)

Big data is an opportunity for any company to gain advantages from data aggregation, data exhaust, and metadata. This makes big data a useful business analytic tool, but there is a common misunderstanding about what big data is. **Apache Hadoop** is the most popular implementation for solving large-scale distributed data storage, analysis, and retrieval tasks. The **NoSQL** and **massively parallel processing** (**MPP**) data stores are also useable for large-scale data storage.

One of the main challenges for big data is how to store, protect, backup, organize, and catalogue the data on a petabyte scale. Another main challenge of big data is the concept of data ubiquity.

Figure 4.2. The algorithm vs data

Source: Cuesta, 2013

2. Quantitative versus qualitative data analysis

Quantitative and qualitative analysis can be defined as follows:

- **Quantitative data:** It is numerical measurements expressed in terms of numbers.
- **Qualitative data:** It is categorical measurements expressed in terms of natural language descriptions.

Quantitative analytics involves the analysis of numerical data. The type of the analysis will depend on the level of measurement.

There are four kinds of measurements:

- **Nominal:** Data has no logical order and is used as classification data.
- **Ordinal:** Data has a logical order, and differences between values are not constant.
- **Interval:** Data is continuous and depends on logical order. The data has standardized differences between values but does not include zero.
- **Ratio:** Data is continuous with logical order and regular interval differences between values, which may include zero.

Qualitative analysis can explore the complexity and meaning of social phenomena. Data for qualitative study may include written texts (for example, documents or email) and/or audible and visual data (for example, digital images or sounds).

Exploratory vs. explanatory analysis

Exploratory analysis (EDA), Describe the characteristics of the data using descriptive statistics and data visualization techniques. Identify patterns, trends, and relationships within the data using appropriate statistical methods. Highlight any anomalies or outliers that require further investigation (VanderPlas, 2020).

Explanatory analysis is what happens when you have something specific you want to show to an audience.

The goals of exploratory data analysis are listed as follows:

- Detection of data errors
- Checking of assumptions
- Finding hidden patterns (such as tendency)
- Preliminary selection of appropriate models
- Determining relationships between the variables

Data Validation, Storytelling and visualisation

Data quality in the ETL process

The idea of the ETL (Extract Transform Load) process is well-known. The accuracy and relevance of data analysis rely on the ability to bring high data quality using the ETL process. The ETL process is complex and time-consuming but important and unavoidable.

Data Validation

In the validation process, we refer to good-quality data characteristics to assess the quality of the dataset we use. Using characteristic, several steps of data validation are performed. The most important steps are the following:

Data Accuracy: In assessing the accuracy of data and initial data exploration was conducted.

Data Completeness: In assessing the data completeness we conducted data sorting to search for missing data.

Data Consistency: To assess the data consistency, data matching and cleaning were conducted using. Only after the data validation process can we start the discussion of the story of our data.

3. Data storytelling

Most researchers and scholars did not agree on the definition of data storytelling is, many scholars have centered on how the use of visual charts and narratives can support communication and insights. Jeffrey Heer used the terms narrative and visualisation to refer to Data storytelling with a chain of events coupled with visual charts.

Storytelling – Who, What, and How

When it comes to explanatory analysis, there are a few things to think about and be extremely clear on before visualizing any data or creating content. First, To whom **(Who)** are you communicating? It is important to have a good understanding of who your audience is and how they perceive you. Second, **What** do you want your audience to know or do? You should be clear **how** you want your audience to act and take into account how you will communicate to them and the overall tone that you want to set for your communication.

Only after you can concisely answer these first two questions are you ready to move forward with the third: How can you use data to help make your point?

Who – The audience

The more specific you can be about who your audience is, the better position you will be in for successful communication. Sometimes, this means creating different communications for different audiences. Identifying the decision maker is one way of narrowing your audience.

What – Action
What do you need your audience to know or do? This is the point where you think through how to make what you communicate relevant for your audience and form a clear understanding of why they should care about what you say.

How – Mechanism
How will you communicate to your audience? The method you will use to communicate to your audience has implications on a number of factors, including the amount of control you will have over how the audience takes in the information and the level of detail that needs to be explicit. We can think of the communication mechanism along a continuum, with the live presentation at the left and a written document.

Data Visualisation Techniques
Various techniques and methods have been used to present insights using data visualisation from standard to more sophisticated and complex techniques that deal with complex data sets. Simple visualisation techniques include but are not limited to charts, tables, graphs, maps, infographics, and dashboards. The data visualisation aims to expose something new about the underlying patterns and relationships within the data. The visualization needs to look good and meaningful to help organizations make better decisions. Visualization is an easy way to jump into a complex dataset (small or big) to describe and explore the data efficiently.

Situating diagrams in graphic communication
A major problem with visual methods in the natural and social sciences is the different terminology used to describe diagrams. Table 4.1. summarizes the differences and similarities between three categories of graphic communication used as qualitative data collection methods: Charts, Tables, and Lists. Although the categories may overlap, they are discussed as distinct and mutually exclusive categories.

Table 4.1. Differences in diagrams and tables and lists (adapted from Umoquit, 2013)

	Diagrams	**Tables and lists**
Purpose	Ability to simplify complex ideas	Highly structured ideas and variables
Abstraction	Some level of abstraction	Concrete
Structure (internal)	More direct in terms of representation; structure inherent where pre-agreed notation is used	Most direct in terms of representation
Notation	May use pre-set or agreed conventions or notations	Uses pre-set or agreed conventions or notations
Spatial Meaning	Spatial arrangements of signs usually carry meaning	Spatial arrangements of signs carry meaning
Verbal/visual spectrum	Composite of both visual and verbal	Verbal signs dominate over visual

Main visualization types

Amounts – Bars, Vertical Bars, Dots, Groped Bars, Vertical Grouped Bars, Stacked Bars, Heatmap

Distributions – Histogram, Density Plot, Cumulative Density, Boxplots, Violins, Strip Charts, Sina Plots, Stacked Histograms, Overlapping Densities

Proportions – Pie Chart, Bars, Vertical Bars, Stacked Bars, Multiple Pie Charts, Grouped Bars, Stacked Bars, Stacked Densities, Mosaic Plot, Treemap, Parallel Sets

Relationships – Scatterplot, Bubble Chart, Paired Scatterplot, Slopegraph, Density Contours, 2D Bins, Hex Bins, Correlogram, Line Graph, Connected Scatterplot, Smooth Line Graph

Geospatial data – Map, Choropleth, Cartogram, Cartogram Heatmap

Uncertainty – Error Bars, 2D Error Bars, Graded Error Bars, Confidence Strips, Eyes, Half-Eyes, Quantile Dot Plot, Confidence Band, Graded Confidence Band, Fitted Draws

Improving the visualisation understandability

Reducing clutter in visualizations is absolutely crucial for effective communication and understanding.

Cognitive Overload: Our brains have limited processing power for visual information. When cluttered with unnecessary elements, a visualisation overwhelms viewers and makes it difficult to focus on the key message. Clutter competes for attention with the important data, hindering comprehension and potentially leading to misinterpretations.
Clarity and Focus: A clean visualization guides the viewer's eye towards the most important information. Removing distractions ensures viewers understand the intended message quickly and accurately. Clear visuals are like well-written stories, engaging the audience and leaving a lasting impression.
Aesthetics and Engagement: Clutter can make visualizations appear messy and unprofessional, reducing their impact and appeal. A clean, well-designed visualization is informative and aesthetically pleasing, creating a positive user experience and fostering engagement.
Accessibility: Visualizations with excessive clutter can be challenging for people with visual impairments or cognitive disabilities. By simplifying the design and reducing extraneous elements, you make your visualizations more accessible to a wider audience.
Efficiency and Speed: Clutter-free visualizations are easier to create and edit, saving time and resources. Additionally, viewers can process information faster from uncluttered designs, leading to quicker decision-making and improved workflow.

Here are some tips for reducing clutter in your visualizations:

- **Prioritize**: Identify the key message you want to convey and focus on elements that support it. Eliminate anything that doesn't contribute to understanding.
- **Simplify design**: Use clean lines, and appropriate color palettes, and avoid excessive decorative elements.
- **Use data efficiently**: Choose the right visualization type for your data and avoid cramming too much information into a single chart.
- **Leverage white space**: Blank space around and between elements allows the data to breathe and improves visual hierarchy.
- **Pay attention to labels and legends**: Keep them clear, concise, and well-placed.
- **Test and iterate**: Get feedback from your target audience and refine your visualization based on their input.
- **Remove chart border**: Chart borders are usually unnecessary. Instead, think about using white space to differentiate the visual from other elements on the page as needed.

- **Remove gridlines**: If you think it will be helpful for your audience to trace their finger from the data to the axis, or you feel that your data will be more effectively processed, you can leave the gridlines. But make them thin and use a light colour like grey.
- **Remove data markers**: Every single element adds cognitive load on the part of your audience.

Highlighting in visualisation

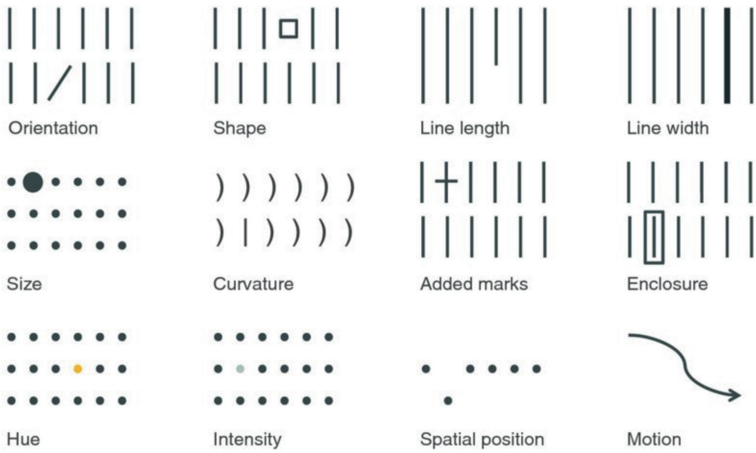


Figure 4.3. Visualisation highlighting possibilities

Source: Nussbaumer, 2015

As we can see in Figure 4.3. there are plenty of possibilities to highlight the relevant information. The combination of the mentioned solutions is also interesting.

4. Machine Learning and Business Intelligence

This chapter, among many sources, processes the article by Szilágyi and Tóth (2024). Machine learning is a subfield of computer science concerned with building algorithms that, to be useful, rely on a collection of examples of some phenomenon. These examples can come from nature, by humans or generated by another algorithm. Machine learning can also be defined as the process of solving a practical problem by gathering a dataset and algorithmically building a statistical model based on that dataset. That statistical model is assumed to be used somehow to solve the practical problem.

ML has a large amount of algorithms generally split into three groups, given how the algorithm is training:

- Supervised learning
- Unsupervised learning
- Reinforcement learning

Data-driven research has been energized by the growth of Big Data analytics and developments in machine learning. Therefore, the machine learning method is one of the possible approaches for creating a Big Data decision support tool. Tsolaki et al. (2022) grouped data mining and machine learning methods by application area.

AI visualisation is one of the distinctive features of Power BI. Key-influencer visualisation, decomposition tree, and anomaly detection are the three different types of AI visualisation in Power BI. In this research, we will use decomposition trees.

Implementing machine learning techniques often requires a thorough understanding of mathematics and computer science; however, ML is difficult to use due to the high level of expertise required. A self-service framework ML can, therefore, be used by analysts with less technical knowledge. Numerous research papers have addressed the self-service deployment of machine learning for business intelligence using Microsoft Power BI.

Decomposition Tree Visual: Self-Service Decision Tree

The concept of a decision tree is used by decomposition tree visualisation to represent data in multiple dimensions. In the visualisation, the data is summarised in a graph, allowing the user to dive into the data dimensions in any order.

The decomposition tree in Power BI allows the user to visualise data across multiple dimensions. It automatically aggregates the data and allows the user to drill down into the dimensions in any order. In addition, it is a visualisation with artificial intelligence, so the user can prompt it to find the next dimension and break it down according to the user's criteria. Thank you for this tool. It is possible to perform root cause analysis and ad hoc investigations with it.

The gradient boosting algorithm uses statistical regression and classification as primary methods for predictive data mining. The idea behind the gradient boosting procedure is to train a system using a mathematical model to maximize a differentiable loss function from a set of random "inputs" and "outputs". The procedure automatically aggregates data and allows drilling down into selected dimensions in any order. It is an artificial intelligence (AI) visualisation, which makes it a valuable tool for ad hoc investigations. There is an "AI Splits" or "AutoAI" to figure out where to look next in the data.

4.1. Data-Driven Business Analysis

In today's data-rich world, businesses are generating and collecting information at an unprecedented rate. But data alone isn't enough. **Turning that data into actionable insights and driving informed decisions** is where data-driven business analysis comes in.

What is data-driven business analysis?
It's an approach to business analysis that **leverages data analytics and visualization techniques** to understand business problems, assess potential solutions, and measure their effectiveness. It's about:

- **Asking the right questions**: Identifying key business issues and opportunities that can be addressed with data.
- **Collecting and cleansing data**: Gathering relevant data from various sources and ensuring its accuracy and completeness.
- **Applying analytical techniques**: Using statistical analysis, data mining, and machine learning to uncover patterns, trends, and relationships within the data.
- **Visualizing insights**: Creating clear and compelling visualizations to communicate data-driven findings to stakeholders.
- **Recommending informed solutions**: Drawing conclusions and proposing data-backed solutions to improve business performance.
- **Key Performance Indicators** (KPIs) are the main instruments of Business Performance Management. KPIs are the measures that are translated to both the strategy and the business process. These measures are often designed for an industry sector with assumptions about business processes in organizations.

Benefits of data-driven business analysis:
- **Improved decision-making:** Data-driven insights lead to more informed decisions, reducing the risk of guesswork and gut feeling.
- **Enhanced efficiency and cost savings:** Identifying inefficiencies and optimizing processes based on data leads to improved resource utilization and cost reduction.
- **Better customer understanding:** Analyzing customer data helps you better understand their needs, preferences, and behavior, allowing for targeted marketing and improved customer experience.

- **Increased innovation:** Data can reveal hidden patterns and opportunities, paving the way for innovative solutions and competitive advantage.
- **Stronger communication and collaboration:** Data-driven insights provide a common ground for discussion and collaboration across different teams within an organization.

Getting started with data-driven business analysis:

- **Develop your data literacy:** Understand basic data analysis concepts and techniques.
- **Identify your business problems:** What are the key challenges facing your organization?
- **Choose the right tools:** There are various data analysis and visualization tools available depending on your needs and budget.
- **Focus on high-quality data:** Ensure data accuracy and completeness for reliable insights.
- **Collaborate with stakeholders:** Involve data scientists, business leaders, and other stakeholders in the process.
- **Tell a compelling story:** Use data visualizations and storytelling to communicate insights effectively.

Remember, data-driven business analysis is an ongoing journey. By continuously learning, adapting, and integrating data into your analysis, you can unlock its full potential to drive meaningful improvements in your business.

4.2. Large Language Models in business analysis

Before discussing this topic, some key concepts about LLM should be clarified.
Chatbot: a computer program designed to simulate a conversation with human users, especially over the Internet.
Generative Model: a model that creates new data, not only classifies, but predicts based on the input data.
Generative Pre-Trained Transformer (GPT): is a machine-learning model that uses unsupervised and supervised learning techniques to understand and generate human-like language.
Language Model: a type of artificial intelligence model that is trained to generate text that resembles human language.
Natural Language Processing (NLP): NLP is an area of artificial intelligence that involves the use of algorithms to analyze human language, such as text and speech, to extract meaning and useful information.

Neural Network: a machine learning model consisting of connected processing nodes trained on data to perform a specific task based on their connections.
Supervised Fine-Tuning: a machine learning technique in which a pre-trained model is further trained on a smaller, labelled data set to improve performance on a specific task.
Transfer Learning: the ability of tools like ChatGPT to use the knowledge gained from one task to improve performance in another related task.
Unsupervised Pre-Training: a machine learning technique in which a model is trained on a large dataset without labelled examples, allowing it to learn the underlying structure and patterns in the data.

Security challenges of GPT-4

Although GPT-4 shows increased performance in areas such as reasoning, knowledge retention, and encoding compared to earlier models such as GPT-2 (Radford et al., 2019) and GPT-3, during its development based on the analyses carried out, specific risks can be observed.
Hallucination: GPT-4 is prone to "hallucinating," i.e., "producing meaningless or untrue content from certain sources." This can be particularly harmful as users increasingly trust the information provided by the model.
Harmful content, biased content: Language models can create different types of harmful content. These may be content that may violate development guidelines and may cause harm on an individual or social level. Language models can amplify biases and perpetuate stereotypes.
Disinformation and (user) influencing activities: GPT-4 can create believable, realistic, and targeted content, including news articles, messages, conversations, and emails.
Data protection: GPT-4 learned from several previously established and publicly available data sources that may contain publicly available personal data. Because of the above, the model can know people who are highly present on public Internet interfaces.
Cyber security: In the case of vulnerability discovery and exploitation, external cyber security experts tested whether GPT-4 can help discover, assess, and exploit vulnerabilities in computer systems. They found that GPT-4 can explain certain vulnerabilities, but it no longer performs well when exploiting known vulnerabilities.

Large Language Models in business

Large Language Models (LLM) are already making waves in various industries, and business analysis is no exception. Here's how LLMs are transforming the role of business analysts:

Improved Communication and Collaboration:

- **Generating Reports and Summaries:** LLMs can automatically generate reports and summaries of complex data sets, making information easily digestible for different stakeholders. This improves communication and collaboration between analysts and other business units.
- **Personalized Insights:** LLMs can tailor reports and recommendations based on individual needs and preferences. This ensures that stakeholders receive the information most relevant to them.

Additional Benefits:

- **Reduced Costs:** Automating tasks and improving efficiency through LLMs can lead to business cost savings.
- **Increased Productivity:** Analysts can focus on higher-level tasks and strategic thinking, increasing productivity and value creation.
- **Democratization of Data Insights:** LLMs can make data analysis more accessible to non-technical users, fostering a data-driven culture within the organization.

Overall, LLMs are powerful tools that can significantly enhance the role of business analysts. By leveraging their strengths while addressing their limitations, businesses can gain a competitive edge through data-driven decision-making and insightful business analysis.

4.3. What new skills may be needed in businesses?

Prompt design and prompt engineering: Prompts are input instructions or questions that we give to AI systems to generate responses. Prompts are critical because they determine what results you get. For GPT or other language models to provide accurate and useful answers, it is important to create appropriate prompts. Prompt engineering is the process of creating questions or instructions that help the model achieve the desired results. In prompt engineering, users must learn how to create efficient and accurate prompts so that the model provides the desired results. This is key to the application of AI, as good questions can be critical to success.

Knowing and using MI art: GPT and similar models can also be used in creative fields, such as art and design. DALL-E, for example, is a model that generates images based on textual descriptions. Proficiency in AI art enables the creation of new and exciting works.

Acquiring programming basics: Although it is not necessary to be a professional programmer when using GPT and other AI models, basic programming knowledge can be an advantage. These skills allow us to customize and integrate the models into our applications or projects. For example, if a company is developing its GPT-based chatbot, programming fundamentals can help fine-tune how the chatbot works.

API (Programming Interface) use: APIs allow applications to communicate with other services and systems. There are a growing number of APIs in the field of AI that allow developers and companies to easily access AI systems and integrate them into their applications. For example, an e-commerce company can use APIs to offer customized recommendations and a personalized shopping experience to its customers, increasing purchase conversion.

These skills and approaches can help companies take advantage of the opportunities offered by AI and large language models. AI is an increasingly widespread technology, and those who properly prepare and apply it can gain a competitive advantage in the age of digitization and automation.

ICT related suggested skills for employees:

- **Marketing department:** May require proficiency in digital marketing tools, social media management platforms, and content creation software.
- **Sales department:** Needs skills like customer relationship management (CRM) systems, email marketing platforms, and data visualization tools.
- **Non-technical roles:** Even basic computer literacy, effective communication, and problem-solving skills are crucial for navigating daily tasks and utilizing various software efficiently.

Additionally, also related skills that companies have to consider:

- **Emerging technologies:** Skills related to artificial intelligence, blockchain, and cloud-based solutions are increasingly sought after.
- **Data literacy:** The ability to understand, analyze, and interpret data is becoming essential across all job roles.
- **Soft skills:** Communication, collaboration, and problem-solving are crucial for leveraging technology effectively.

REFERENCES

Aspin, A. (2021). Miscellaneous visual styles. In Pro Power BI theme creation (pp. 213–230). Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-7068-4_11

Aspin, A. (2022). Third-party visuals. In Pro Power BI dashboard creation (pp. 185-202). Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-8227-4_10

Azamfirei, R., Kudchadkar, S. R., & Fackler, J. (2023). Large language models and the perils of their hallucinations. Critical Care, 27(1). https://doi.org/10.1186/s13054-023-04393-x

Bishop, C. M. (1994). Neural networks and their applications. Review of Scientific Instruments, 65(6), 1803-1832. https://doi.org/10.1063/1.1144830

Burkov, A. (2019). The hundred-page machine learning book. ISBN 978-1999579500

Cole, N. K. (2015). Storytelling with data: A data visualization guide for business professionals. John Wiley & Sons. Retrieved from ProQuest Ebook Central.

Cuesta, H. (2013). Practical data analysis. Packt Publishing. Retrieved from ProQuest Ebook Central.

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Wirtz, J., & Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. International Journal of Information Management, 71. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Ehrenmueller-Jensen, M. (2020). Adding smart visualizations. In Self-service AI with Power BI desktop (pp. 89-106). Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-6231-3_5

Fleischhacker, A. J. (2023). A business analyst's introduction to business analytics: Coding in R for fast, easy, and visual Bayesian inference. ISBN 979-8857923412

Forbes. (2023). The current state of LLMs in business intelligence and what needs to change. Retrieved from https://www.forbes.com/sites/forrester/2023/04/28/the-current-state-of-llms-in-business-intelligence-and-what-needs-to-change/?sh=3bf4d32247d8] Accessed 2024.02.05.]

Fortino, A. (2020). Data visualization for business decisions: A laboratory manual. Mercury Learning & Information. Retrieved from ProQuest Ebook Central.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189-1232. http://www.jstor.org/stable/2699986

Gualo, F., Rodriguez, M., Verdugo, J., & Caballero, I., Piattini, M. (2021). Data quality certification using ISO/IEC 25012: Industrial experiences. Journal of Systems and Software, 176, 110938. https://doi.org/10.1016/j.jss.2021.110938

Harvard Business Review (HBR). (2024). Why becoming a data-driven organization is so hard. Retrieved from https://hbr.org/2022/02/why-becoming-a-data-driven-organization-is-so-hard

Heer, J., Bostock, M., & Ogievetsky, V. (2010). A tour through the visualization zoo. Communications of the ACM, 53(6), 59–67. https://doi.org/10.1145/1743546.1743567

King, M. R. (2022). The future of AI in medicine: A perspective from a chatbot. Annals of Biomedical Engineering. https://doi.org/10.1007/s10439-022-03121-w

Knaflic, C. N. (2015). Storytelling with data: A data visualization guide for business professionals. Wiley. ISBN 978-1-119-00225-3

Lee, B., Riche, N. H., Isenberg, P., & Carpendale, S. (2015). More than telling a story: Transforming data into visually shared stories. IEEE Computer Graphics and Applications, 35(5), 84–90. https://doi.org/10.1109/mcg.2015.99

Lee, C., Panda, P., Srinivasan, G., & Roy, K. (2018). Training deep spiking convolutional neural networks with STDP-based unsupervised pre-training followed by supervised fine-tuning. Frontiers in Neuroscience, 12, 435. https://doi.org/10.3389/fn

Lo, L. S. (2023). The CLEAR path: A framework for enhancing information literacy through prompt engineering. Journal of Academic Librarianship, 49(4). https://doi.org/10.1016/j.acalib.2023.102720

MacNeil, S., Tran, A., Mogil, D., Bernstein, S., Ross, E., & Huang, Z. (2022). Generating diverse code explanations using the GPT-3 large language model. Proceedings of the ACM Conference on International Computing Education Research, 2, 37–39.

Manning, C., & Schutze, H. (1999). Foundations of statistical natural language processing. MIT Press.

Marr, B. (2015). Big data: Using SMART big data, analytics and metrics to make better decisions and improve performance. Newark: John Wiley & Sons, Incorporated. [Accessed: 07.09.2023]. ProQuest Ebook Central.

McKinsey. (2024). The data-driven enterprise of 2025. Retrieved from https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025 [Accessed: 2024.02.05.]

Medium.com. (2023a). Prompt engineering for AI language models. Retrieved from https://medium.com/@nirajranasinghe/prompt-engineering-for-ai-language-models-f6d226603c34 [Accessed: 28.09.2023.]

Medium.com. (2023b). The ChatGPT list of lists: A collection of 1500 useful, mind-blowing, and strange use cases. Retrieved from https://medium.com/mlearning-ai/the-chatgpt-list-of-lists-a-collection-of-1500-useful-mind-blowing-and-strange-use-cases-8b14c35eb [Accessed: 28.09.2023.]

Medium.com. (2023c). 5 best ways to use ChatGPT API. Retrieved from https://divakersingh29.medium.com/5-best-way-to-use-chatgpt-api-c6e0a8356a34 [Accessed: 28.09.2023.]

Murgia, E., Abbasiantaeb, Z., Aliannejadi, M., Huibers, T., Landoni, M., & Pera, M. S. (2023). ChatGPT in the classroom: A preliminary exploration on the feasibility of adapting ChatGPT to support children's information discovery. UMAP 2023 – Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization, 22–27. https://doi.org/10.1145/3563359.3597399

OpenAI. (2023). GPT-4 observed safety challenges. Retrieved from https://cdn.openai.com/papers/gpt-4-system-card.pdf [Accessed: 21.09.2023.]

Patriarca, R., Di Gravio, G., Cioponea, R., & Licu, A. (2022). Democratizing business intelligence and machine learning for air traffic management safety. Safety Science, 146, 105530. https://doi.org/10.1016/j.ssci.2021.105530

Paul, J., Ueno, A., & Dennis, C. (2023). ChatGPT and consumers: Benefits, pitfalls and future research agenda. International Journal of Consumer Studies, 47(4), 1213-1225. https://doi.org/10.1111/ijcs.12928

Pavlik, J. V. (2023). Collaborating with ChatGPT: Considering the implications of generative artificial intelligence for journalism and media education. Journalism and Mass Communication Educator. https://doi.org/10.1177/10776958221149577

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners.

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. Retrieved from https://www.cs.ubc.ca/~amuham01/LING530/papers/radford2018improving.pdf

Roubtsova, E., & Michell, V. (2014). KPIs and their properties defined with the EXTREME method. In Shishkov, B. (Ed.), Business modeling and software design. BMSD 2013. Lecture notes in business information processing (Vol. 173, pp. 88–102). Springer, Cham. https://doi.org/10.1007/978-3-319-06671-4_7

Szilágyi, R., & Tóth, M. (2024). Use of LLM for SMEs, opportunities and challenges. Journal of Agricultural Informatics, 14(2). https://doi.org/10.17700/jai.2023.14.2.703

Tsolaki, K., Vafeiadis, T., Nizamis, A., Ioannidis, D., & Tzovaras, D. (2022). Utilizing machine learning on freight transportation and logistics applications: A review. ICT Express. https://doi.org/10.1016/j.icte.2022.02.001

Turing. (2024). How LLMs are changing the face of business analytics. Retrieved from https://www.turing.com/resources/how-llms-are-changing-the-face-of-business-analytics [Accessed: 2024.02.05.]

Umoquit, M., Tso, P., Varga-Atkins, T., O'Brien, M., & Wheeldon, J. (2013). Diagrammatic elicitation: Defining the use of diagrams in data collection The Qualitative Report, 18(30), 1–12. https://doi.org/10.46743/2160-3715/2013.1487

Vassiliadis, P., Simitsis, A., & Skiadopoulos, S. (2002). Conceptual modeling for ETL processes. In Proceedings of the 5th ACM international workshop on Data Warehousing and OLAP (DOLAP '02) (pp. 14-21). https://doi.org/10.1145/583890.583893

Wilke, C. (2019). Fundamentals of data visualization: A primer on making informative and compelling figures. O'Reilly. ISBN 978-1492031086

Yang, J., Chen, Y.-L., Por, L. Y., & Ku, C. S. (2023). A systematic literature review of information security in chatbots. Applied Sciences (Switzerland), 13(11). https://doi.org/10.3390/app13117228

CHAPTER 5: BUSINESS PROCESS MANAGEMENT (BPM)

(Viktor Takács)

The chapter is a review of Dumas et al. (2013), extended with a summary of ARIS eEPC.

An organisation's business process is a collection of interrelated activities to achieve specific goals. It defines the sequence of steps required to transform inputs (resources) into outputs (products or services) that deliver customer value.

Effective business process management (BPM) involves:

- Identifying and documenting existing processes.
- Analyzing processes for inefficiencies and bottlenecks.
- Designing and implementing improvements.
- Monitoring and measuring process performance.

BPM is by no means the only discipline concerned with improving organisations' operational performance. Below, we briefly introduce some related disciplines and identify key relations and differences between these disciplines and BPM.

Total Quality Management (TQM) is an approach that historically preceded and inspired BPM. TQM focuses on continuously improving and sustaining the quality of products and, by extension, services. In this way, it is similar to BPM in its emphasis on the necessity of ongoing improvement efforts. However, where TQM emphasises the products and services themselves, the view behind BPM is that the quality of products and services can best be achieved by focusing on improving the processes that create these products and services. It should be admitted that this view is controversial, as contemporary TQM adepts would rather see BPM as one of the various practices that are commonly found within a TQM program. Not so much a theoretical distinction but an empirical one is that application of TQM is primarily found in manufacturing domains—where the products are tangible—while BPM is more oriented to service organizations.

Operations Management is a field concerned with managing a firm or organisation's physical and technical functions, particularly those relating to production and manufacturing. Probability theory, queuing theory, decision analysis, mathematical modelling, and simulation are all important techniques for optimizing the efficiency of operations from this perspective. What is different between operations management and BPM is that operations management is generally concerned with controlling an existing process without necessarily changing it, while BPM is often concerned with making changes to an existing process to improve it.

Lean is a management discipline that originates from the manufacturing industry, particularly the engineering philosophy of Toyota. One of the main principles of Lean is the elimination of waste, i.e. activities that do not add value to the customer. The customer orientation of Lean is like that of BPM, and many of the principles behind Lean have been absorbed by BPM. In that sense, BPM can be seen as a more encompassing discipline than Lean. Another difference is that BPM emphasises using information technology as a tool to improve business processes and make them more consistent and repeatable.

Six Sigma is another set of practices originating from manufacturing, engineering, and production practices at Motorola. The main characteristic of Six Sigma is its focus on the minimization of defects (errors). Six Sigma places a strong emphasis on measuring the output of processes or activities, especially in terms of quality. Six Sigma encourages managers to systematically compare the effects of improvement initiatives on the outputs. In practice, Six Sigma is not necessarily applied alone but in conjunction with other approaches. A popular approach is to blend the philosophy of Lean with the techniques of Six Sigma, leading to an approach known as Lean Six Sigma. Nowadays, many of the techniques of Six Sigma are commonly applied in BPM as well.

In summary, we can say that BPM inherits from the continuous improvement philosophy of TQM, embraces the principles and techniques of operations management, Lean and Six Sigma, and combines them with the capabilities offered by modern information technology in order to optimally align business processes with the performance objectives of an organization.

1. What is a Business Process?

We define a business process as a collection of interrelated events, activities, and decision points that involve multiple actors and objects, and that together lead to an outcome that is valuable to the organization and its customers. Business processes planned in operative level, when activities are simple and can be seen as one single unit of work, we call them tasks. Typical process involves decision points, that is, points in time when a decision is made that affects the way the process is executed. A process also involves several actors like executors (human actors, organizations), software systems and its modules, physical objects (equipment, materials, products, paper documents) and immaterial objects (electronic documents and electronic records) as resources and outcomes. Business process management (BPM) is the art and science of overseeing work in an organization to ensure consistent outcomes and take advantage of opportunities for improvement. In this context, the term "improvement" can have different meanings depending on the goals of the organization. Typical examples of improvement goals include reducing costs, reducing execution times, and reducing error rates. Improvement initiatives usually continuous. It is important that BPM is not about improving the execution of individual activities. Rather, it is about managing a whole chain of events, activities, and decisions that create value to the organization and its customers. These "chains of events, activities, and decisions" are called processes.

2. Types of Business Processes

Every organization has to manage a number of processes. From a management perspective, typical business processes can be categorized into three main types:

Primary Processes: These core activities directly contribute to creating and delivering a product or service to the customer. They are the revenue generators for the business. Examples of primary processes include:
Order fulfilment: Receiving customer orders, processing them, picking and packing items, and shipping them to the customer.
Product development: Designing, developing, testing, and launching new products or services.
Marketing and sales: Generating leads, qualifying leads, converting leads to customers, and promoting products or services.

Support Processes: These processes indirectly support the primary processes by providing essential services and resources. They ensure the smooth operation of the primary processes. Examples of support processes include:
Human resources: Recruiting, hiring, training, and managing employees.
Accounting and finance: Managing the company's finances, including accounts payable, accounts receivable, payroll, and budgeting.
Information technology (IT): Providing and maintaining computer systems, networks, and software applications.

Management Processes: These processes involve planning, organizing, directing, and controlling the organization's resources to achieve its goals. They ensure the effectiveness and efficiency of both primary and support processes. Examples of management processes include:

- **Strategic planning**: Setting the overall direction of the organization and developing long-term plans.
- **Risk management**: Identifying, assessing, and mitigating potential risks to the organization.
- **Performance management**: Setting goals, monitoring progress, and evaluating the performance of individuals, teams, and departments.

It's important to note that these categories are not mutually exclusive. Many business processes can span across multiple categories. For instance, the process of managing customer complaints might involve aspects of both primary processes (fulfilling customer orders) and support processes (customer service).

Typical examples of processes that can be found in most organizations include:

Order-to-cash: This is a type of process performed by a vendor, which starts when a customer submits an order to purchase a product or a service and ends when the product or service in question has been delivered to the customer and the customer has made the corresponding payment. An order-to-cash process encompasses activities related to purchase order verification, shipment (in the case of physical products), delivery, invoicing, payment receipt and acknowledgment.

Quote-to-order: This type of process typically precedes an order-to-cash process. It starts from the point when a supplier receives a "Request for Quote"

(RFQ) from a customer and ends when the customer in question places a purchase order based on the received quote. The order-to-cash process takes the relay from that point on. The combination of a quote-to-order and the corresponding order-tocash process is called a quote-to-cash process.

Procure-to-pay: This type of process starts when someone in an organization determines that a given product or service needs to be purchased. It ends when the product or service has been delivered and paid for. A procure-to-pay process includes activities such as obtaining quotes, approving the purchase, selecting a supplier, issuing a purchase order, receiving the goods (or consuming the service), and checking and paying the invoice. A procure-to-pay process can be seen as the dual of the quote-to-cash process in business-to-business interactions. Every procure-to-pay process has a corresponding quote-to-cash process on the supplier's side.

Issue-to-resolution: This type of process starts when a customer raises a problem or issue, such as a complaint related to a defect in a product or an issue encountered when consuming a service. The process continues until the customer, the supplier, or preferably both of them agree that the issue has been resolved. A variant of this process can be found in insurance companies dealing with "insurance claims". This variant is often called claim-to-resolution.

Application-to-approval: This type of process starts when someone applies for a benefit or privilege and ends when the benefit or privilege in question is either granted or denied. This type of process is common in government agencies, for example, when a citizen applies for a building permit or when a businessman applies for a permit to open a business (e.g. a restaurant). Another process that falls into this category is the admissions process in a university, which starts when a student applies for admission into a degree. Yet another example is the process for approval of a company's vacation or special leave requests.

3. Business Process Improvement

In general, the first question that a team embarking on a BPM initiative needs to clarify is "what business processes are we intending to improve"? Right at the outset and before the possibility of applying BPM is put on the table, there will already be an idea of what operational problems the team must address and what business processes are posing those operational problems. In other words, the team will not start from scratch. One must answer questions such as: Does the process start right from the moment when rental suppliers are selected? Does it end when the rented equipment is delivered to the construction site or does it end when the equipment is returned back to the supplier, or does it continue until the fee for equipment rental has been paid to the supplier? These questions might be easy or hard to answer depending on how much process thinking has taken place in the organization beforehand. If the organization has engaged in BPM initiatives before, it is likely that an inventory of business processes is available and that the scope of these processes has been defined, at least to some extent. In organizations that have not engaged in BPM before, the BPM team must start by at least identifying the processes that are relevant to the problem on the table, delimiting the scope of these processes, and identifying relations between these processes, such as for example part-of relations (i.e. one process being part of another process). This initial phase of a BPM initiative is termed process identification. This phase leads to a so-called process architecture, which typically takes the form of a collection of processes and links between these processes representing different types of relation.

In general, the purpose of engaging in a BPM initiative is to ensure that the business processes covered by the BPM initiative led to consistently positive outcomes and deliver maximum value to the organization in servicing its clients. Measuring the value delivered by a process is a crucial step in BPM. So before starting to analyze any process in detail, it is important to clearly define the process performance measures that will be used to determine whether a process is in "good shape" or in "bad shape".

Once a BPM team has identified which processes, they are dealing with and which performance measures should be used, the next phase for the team is to understand the business process in detail. We call this phase process discovery. Typically, one of the outcomes of this phase is one or several as-is process models. These as-is process models should reflect the understanding that people in the organization have about how work is done. Process models are meant to facilitate communication between stakeholders involved in a BPM initiative.

Once issues in a process have been analysed and quantified, the next phase is to identify and analyze potential remedies for these issues. At this point, the analyst will consider multiple possible options for addressing a problem. In doing so, the analyst needs to keep in mind that a change in a process to address one issue may potentially cause other issues down the road. Changing a process is not as easy as it sounds. People are used to work in a certain way and might resist changes. Furthermore, if the change implies modifying the information system(s) underpinning the process, the change may be costly or may require changes not only in the organization that coordinates the process, but also in other organizations.

Equipped with an understanding of one or several issues in a process and a candidate set of potential remedies, analysts can propose a redesigned version of the process, in other words a to-be process which would address the issues identified in the as-is process. This to-be process is the main output of the process redesign phase. Here, it is important to keep in mind that analysis and redesign are intricately related. There may be multiple redesign options and each of these options needs to be analysed, so that an informed choice can be made as to which option should be chosen.

Once redesigned, the necessary changes in the ways of working and the IT systems of the organization should be implemented so that the to-be process can eventually be put into execution. This phase is called process implementation. In the case of the equipment rental process, the process implementation phase would mean putting in place an information system to record and to track equipment rental requests, POs associated to approved requests and invoices associated to these POs. Deploying such an information system means not only developing the IT components of this system. It would also relate to training the process participants so that they perform their work in the spirit of the redesigned process and make the best use of the IT components of the system. More generally, process implementation may involve two complementary facets: organizational change management and process automation. Organizational change management refers to the set of activities required to change the way of working of all participants involved in the process.

4. Business Process Modeling

Business process models are important at various stages of the BPM lifecycle. Before starting to model a process, it is crucial to understand why we are modeling it. The models we produce will look quite different depending on the reason for modeling them in the first place. There are many reasons for modeling a process. The first one is simply to understand the process and to share our understanding of the process with the people who are involved with the process on a daily basis. Indeed, process participants typically perform quite specialized activities in a process such that they are hardly confronted with the complexity of the entire process. Therefore, process modeling helps to better understand the process and to identify and prevent issues. This step towards a thorough understanding is the prerequisite to conducting process analysis, redesign, or automation.

There are many languages for modeling business processes diagrammatically. One of the oldest ones is flowcharts. In their most basic form, flowcharts consist of rectangles, representing activities and diamonds, representing points in the process where a decision is made. More generally, we can say that regardless of the specific notation used, a diagrammatic process model typically consists of two types of nodes: activity nodes and control nodes. Activity nodes describe units of work that may be performed by humans or, software applications, or a combination thereof. Control nodes capture the flow of execution between activities. Although not all process modeling languages support it, a third important type of element in process models are event nodes. An event node tells us that something may or must happen, within the process or in the environment of the process, that requires a reaction, like for example, the arrival of a message from a customer asking to cancel their purchase order. Other types of nodes may appear in a process model, but we can say that activity, event, and control nodes are the most basic. Several extensions of flowcharts exist, like cross-organizational flowcharts, where the flowchart is divided into so-called swimlanes that denote different organizational units (e.g. different departments in a company).

UML Activity Diagrams are cross-organizational flowcharts. However, UML Activity Diagrams go beyond cross-organizational flowcharts by providing symbols to capture data objects, signals, and parallelism among other aspects. Other languages used for process modeling include data-flow diagrams and IDEF3, just to name two. It would be mind-boggling to try to learn all these languages at once. Fortunately, nowadays there are two widely used standard for process modeling, namely the Business Process Model and Notation (BPMN) and the Extended Event-driven Process Chains (eEPCs). eEPCs have some similarities with flowcharts, but they differ in that they treat events as first-class citizens.

5. Business Process Modeling Notation (BPMN)

The latest version of BPMN is BPMN 2.0. It was released as a standard by the Object Management Group (OMG) in 2011. In BPMN, activities are represented as rounded rectangles. Control nodes (called gateways) are represented using diamond shapes. Activities and control nodes are connected by means of arcs (called flows) that determine the order in which the process is executed. The model also shows the process participants involved in specifically the process where these participants is shown as a separate lane containing the activities performed by the participant in question. With over 100 symbols, BPMN is a fairly complex language. The core set of symbols provided by BPMN are events and activities. Events represent things that happen instantaneously (e.g. an invoice has been received) whereas activities represent work units with a duration (e.g. an activity to pay an invoice). The most elementary form of relation is that of sequence, which implies that one event or activity A is followed by another event or activity B. Events are represented by circles, activities by rounded rectangles, and arcs (called sequence flows in BPMN) are represented by arrows with a full arrowhead.

Recommended naming conventions:

- For activities and tasks, the label should begin with a verb in the imperative form followed by a noun, typically referring to a business object, e.g. “Approve order”. The noun may be preceded by an adjective, e.g. “Issue driver license”, and the verb may be followed by a complement to explain how the action is being done, e.g. “Renew driver license via offline agencies”.
- For events, the label should begin with a noun (this would typically be a business object) and end with a verb in past participle form, e.g. “Invoice emitted”. The verb is a past participle to indicate something that has just happened. Like activity labels, the noun may be prefixed by an adjective, e.g. “Urgent order sent”.
- To name a process model we should use a noun, potentially preceded by an adjective, e.g. “order fulfilment” or “claim handling” process. This label can be obtained by nominalizing the verb describing the main action of a business process, e.g. “fulfil order” (the main action) becomes “order fulfilment” (the process label). Nouns in hyphenated form, like “order-to-cash” and “procure-to-pay” indicating the sequence of main actions in the process, are also possible.

General verbs like “to make”, “to do”, “to perform” or “to conduct” should be replaced with meaningful verbs that capture the specifics of the activity being performed or the event occurring. Words like “process” or “order” are also ambiguous in terms of their part of speech. Both can be used as a verb (“to process”, “to order”) and as a noun (“a process”, “an order”). We recommend using such words consistently, only in one part of speech, e.g. “order” always as a noun. By following such naming conventions, we will keep our models more consistent, make them easier to understand for communication purposes and increase their reusability.

The process begins when a customer feels hungry, then he will pick up the phone and make a pizza order, a clerk answering the telephone in the shop will check the ingredients for the customer. The customer will order his pizza according to the list of ingredients provided. When the clerk receives the order from the customer, he will inform the baker to bake to pizza according to the customer's demand. The delivery boy takes the pizza when ready and delivers it to the customer. The customer will pay for the pizza on receiving it, and the delivery boy will provide the receipt. After consuming the pizza, the process ends.

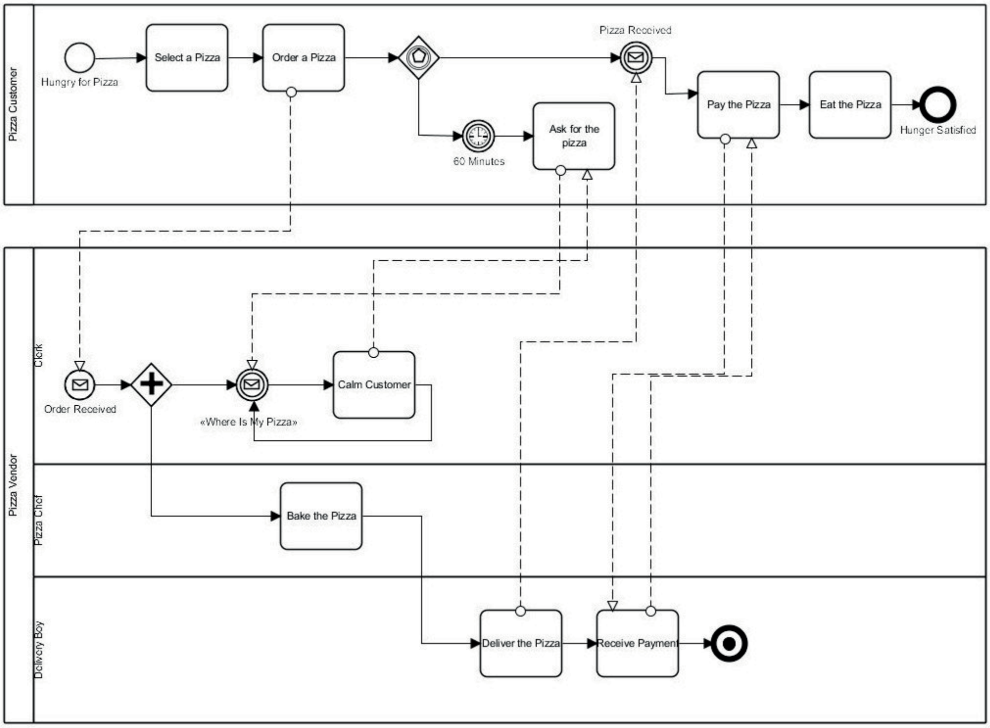


Figure 5.1. The Pizza Purchasing Ordering Process Model.

Source: https://www.businessprocessincubator.com/content/the-pizza-collaboration/

6. ARIS – eEPC (extended Event-driven Process Chain)

ARIS eEPC, a prominent methodology for modelling business processes, stands for Architektur integrierter Informationssysteme (Architecture of Integrated Information Systems) event-driven Process Chain. It provides a visual language with symbols and flowcharts to represent the steps, decisions, and participants involved in a business activity. The ARIS framework defines five different views: organization view, functional view, data view, control view and product/service view. ARIS House is shown in Figure 5.2. shows these five views and the relations between them. As depicted in the figure, the process models are the controlling dispatcher of resources and actions, they act as the central role in the ARIS house, associating with all the other views and integrating them in a uniformed way. A process model respecting these associations between different views is understandable among involved parties despite their different work perspectives and is therefore capable of providing better understandability, maintainability and scalability.

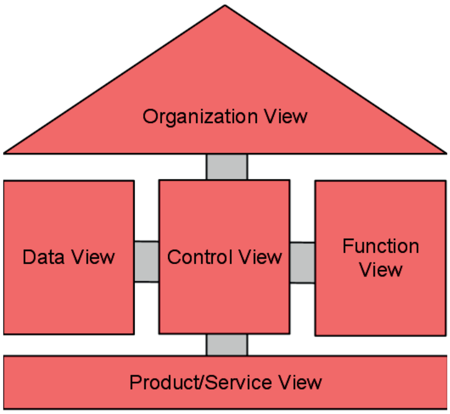


Figure 5.2. The ARIS House

Source: https://viewpointsonitarchitecture.wordpress.com/tag/aris/

ARIS eEPC is a valuable tool for the first two stages of BPM, allowing businesses to represent and analyze their workflows visually. The ARIS eEPC notation uses a set of symbols to depict different elements within a business process:

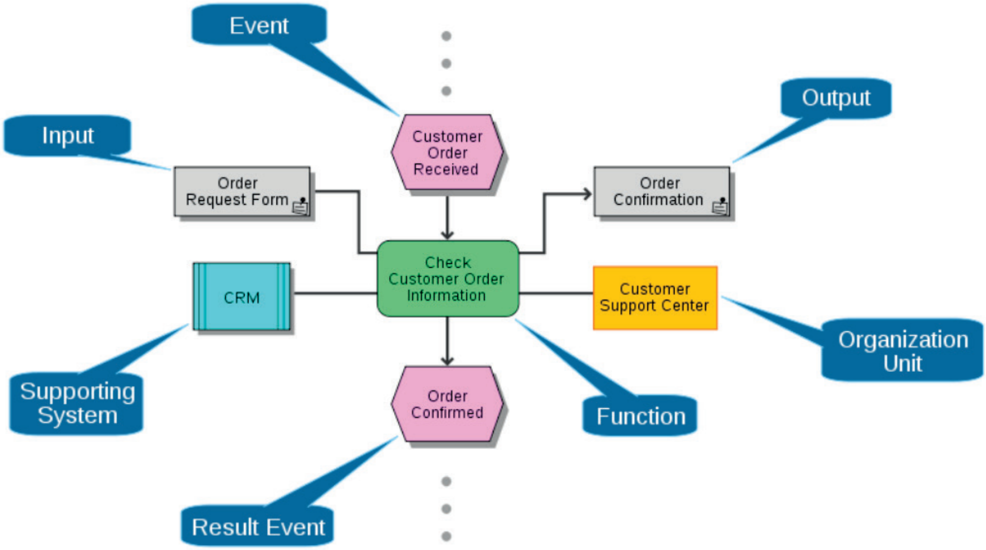


Figure 5.3. Notation example

Events (rounded rectangles): Represent a specific happening that triggers or ends a process or a particular function within it. Events can be starting points (“Order received”), intermediate milestones (“Payment confirmed”), or concluding points (“Product delivered”).

Functions (rectangles): Represent actions or tasks performed within the process. Functions can be manual activities completed by employees or automated steps executed by systems.

Connectors (arrows): Show the sequence and flow of the process, connecting events and functions. Solid lines indicate the normal flow, while dashed lines depict alternative paths based on decisions.

Organizational units (cylinders): Represent the departments, teams, or individuals responsible for carrying out specific functions.

Logical operators (diamonds): Model decision points or branching paths within the process. These include “AND” (all conditions must be met), “OR” (at least one condition must be met), and “XOR” (exclusive OR, only one condition can be true).

Data objects (documents): Represent information used or produced during the process.

Fig. x illustrates a detailed excerpt of a business process example that focuses the function “manufacture item”.

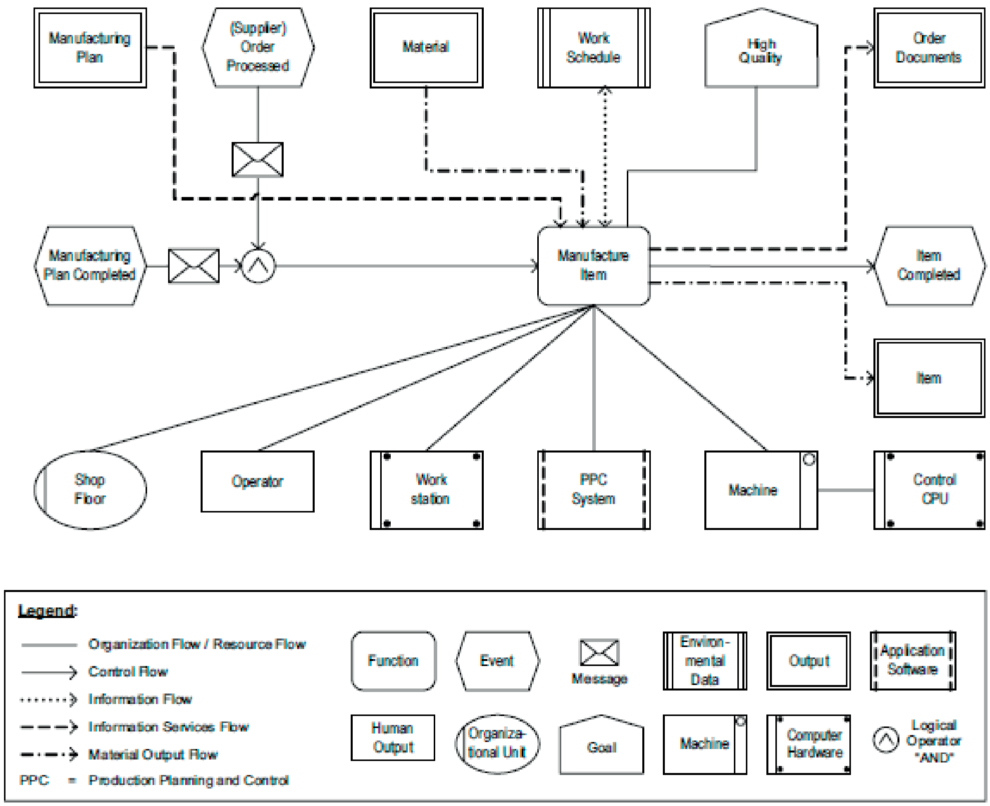


Figure 5.4. Business Process Example
Source: Scheer and Schneider, 2006

The displayed function is enhanced by event and message controls. This enables to describe the process sequence. Events describe condition changes and, e.g., characterize a function's beginning and the result. In addition to these simple events, there are also compounded events. In order to run the function "manufacture item" e.g. the planning needs be finished as well as the necessary parts need to be available. This necessity is expressed by the logical "AND" operator between the events. The method used to describe the process in Fig. 2 is called "Event-Driven Process Chain" (EPC) which was developed in 1992 at the Institute for Information Systems (IWi) at the University of Saarland together with SAP employees in an R&D project financed by SAP AG. The EPC method has become a standard in the field of Business Process Management (BPM), so it is the key component of SAP R/3's modeling concepts for business engineering and customizing.

Benefits of Using ARIS eEPC

Standardized Notation: Provides a common language for documenting and communicating business processes across the organization.

Visual Representation: Simplifies complex processes into easy-to-understand flowcharts, facilitating communication with both technical and non-technical audiences.

Improved Analysis: Enables identification of bottlenecks, redundancies, and opportunities for improvement.

Process Optimization: Supports the design and implementation of more efficient and effective workflows.

Collaboration: Facilitates collaboration between different departments involved in a process.

Limitations of ARIS eEPC

Focus on Events and Functions: May not capture all aspects of a process, such as data flows, roles, or performance metrics.

Complexity for Large Processes: Complex diagrams with numerous elements can become cluttered and difficult to understand.

Limited Automation Support: ARIS eEPC primarily focuses on process visualization, with limited support for directly translating models into executable workflows.

ARIS eEPC is just one component of the broader ARIS platform, a comprehensive suite of tools for enterprise architecture management. The platform offers additional functionalities like:

Data Modeling: Defines data structures and relationships used within the organization.

Organization Modeling: Represents the organizational structure and roles involved in business processes.

Performance Management: Provides tools for measuring and monitoring process performance metrics.

By combining ARIS eEPC with other ARIS components, organizations can gain a holistic view of their business operations and implement a more integrated approach to BPM.

ARIS eEPC remains a valuable tool for business process modeling due to its simplicity, clarity, and widespread adoption. While other methodologies like BPMN (Business Process Model and Notation) have emerged, ARIS eEPC offers a solid foundation for visualizing and analysing workflows, leading to process improvements and overall business efficiency.

7. Workflow applications

Asana *"https://asana.com/"*: A popular work management platform that helps teams organize tasks, projects, and communication in one place. It offers task assignments, due dates, file sharing, and progress tracking features. Great for collaborative projects and keeping track of individual workloads.

Monday.com *"https://support.monday.com/hc/en-us/articles/11065311570066-Get-started-monday-workflows"*: A highly customizable workflow application known for its visual boards and drag-and-drop functionality. Teams can create customized workflows for different processes, automate repetitive tasks, and collaborate in real-time. Ideal for teams that need a flexible platform to adapt to their specific needs.

Zapier *"https://zapier.com/"*: An automation tool that connects different web apps and services. You can create "Zaps" that trigger actions in one app based on events in another. This helps automate repetitive tasks and streamline workflows across different platforms. Useful for integrating various tools teams already use.

Kissflow *"https://kissflow.com/"*: A cloud-based workflow management solution that focuses on automating business processes. It allows users to design custom workflows with approval chains, data collection forms, and integrations with external systems. Well-suited for automating complex workflows within an organization.

Trello *"https://trello.com/"*: A simple and intuitive project management tool that uses Kanban boards to visualize workflows. Tasks are represented as cards that move across different stages (e.g., To Do, In Progress, Done). Trello offers basic functionality like checklists, attachments, and comments, ideal for straightforward workflow visualization and team collaboration.

Airtable *"https://www.airtable.com/"*: A database application that offers a spreadsheet-like interface but with greater flexibility. It allows users to create custom bases with different fields and record types. Airtable can be used to manage various workflows by setting up custom views and automations. Powerful for managing complex data sets within workflows.

Microsoft Power Automate *"https://www.microsoft.com/en-us/power-platform/products/power-automate"*: A cloud-based workflow automation tool included in Microsoft Microsoft 365. It allows users to automate tasks across various Microsoft services (e.g., Outlook, OneDrive) and connect them with external applications. Best suited for businesses heavily invested in the Microsoft ecosystem who want to automate tasks within their existing suite of tools.

REFERENCES

Buikis, A., Ciegis, R., & Fitt, A. D. (2003). Business modelling. Languages and tools. Progress in Industrial Mathematics at ECMI 2002, 5, 41–52. https://doi.org/10.1007/978-3-662-09510-2_4

Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2013). Fundamentals of business process management. IndraStra Global. https://doi.org/10.1007/978-3-642-33143-5

Gao, F., Derguech, W., & Zaremba, M. (2011). Extending BPMN 2.0 to enable links between process models and ARIS views modeled with linked data. BIS (Workshops), 41–52. https://doi.org/10.1007/978-3-642-25370-6_5

Scheer, A., & Schneider, K. (2006). ARIS — Architecture of integrated information systems. Architecture of Integrated Information Systems, 605–623. https://doi.org/10.1007/3-540-26661-5_25

CHAPTER 6: DATA SECURITY, DATA PROTECTION

(László Várallyai)

1. Basic safety concepts

We often use the term ‘security’ in many areas of life, but what do we mean by it, and what does it mean, for example, to be safe? The basics of our daily lives are in place today and are unlikely to change significantly. We also use the term ‘public safety’ in a similar sense. If crime is rare in our environment, we feel good about public safety, but if someone is robbed every day in our street, sooner or later, we start to fear that it could happen to us, and we urgently want to improve public safety. In both cases, it is a matter of security being a favourable state for the subject, which he does not expect to change but cannot exclude. In an idealised, idealised case, this state of affairs may persist forever and ever. Unfortunately, however, our world is not ideal, so, at every moment in time, there are many security threats. We feel safe to the extent that the world around us can prevent and detect threats and correct the adverse effects of events. If we accept that security exists when the threat is minimal, then security means the absence of vulnerabilities or protection from threats.

Safety is, therefore, the third requirement to be considered, alongside quality and reliability, to maintain long-term operation. In everyday terms, safety means a state free from hazards and without disruption. For IT systems, the most important thing is to achieve data security. There are three data **security requirements**:

- **Confidentiality can only be known by those authorised to know it, i.e., limited to thos**e entitled to know it.

- **Integrity**: something that is in its original state and complete.
- **Availability**: the availability of the necessary infrastructure and data to the user where and when it is needed.
- **Data security**: the establishment of confidentiality, integrity and availability of data stored in computer systems (does not address the security of applications and ancillary equipment such as uninterruptible power supplies)
- **IT security** refers to the security of data stored in information systems and the hardware and software resources used to process it. If the concept of 'data' is extended to 'information', this definition is equivalent to but narrower than, information security.
- **Security:** the confidentiality, integrity and availability of human or machine-formalised representations of facts, instructions, and ideas for transmission, processing or storage. If the concept of 'data' includes human formalisation (speech, lecture, conversation), it is equivalent to, but more extensive than, IT security.
- **Data protection**: the set of principles, rules, procedures, instruments and methods of processing which ensure the lawful processing of personal data and the protection of the individuals concerned.

1.1. Cyberspace

Why is there a need for information security at a societal level today? Today's societal systems, including those in the economy, government, local government, and the home, are dependent on information technology, and this dependence has become global with the interconnectedness of systems and the emergence of **cyberspace**.

Cyberspace is "a set of globally interconnected, decentralised, and growing electronic information systems and the social and economic processes that take the form of data and information expressed through these systems".

The cyberspace of Hungary is the part of the electronic information systems of global cyberspace that are located in Hungary and part of the social and economic processes in the form of data and information that are taking place in or directed towards Hungary or in which Hungary is involved, through the electronic systems of global cyberspace.

1.2. Security concept

The Common Criteria6, developed to assess the security of software systems, provide a comprehensive concept of security, considering the properties of systems. The concept includes the attacker, the attacks, the owner implementing the protection, the protection measures, and the elements to be protected.

Figure 6.1. Security concept

Source: www.compasscyber.com

The definitions of the main terms in the Figure 6.1. are given below:

- **Protective action**: action taken by organisational or technical means to reduce the likelihood of a threat occurring or reduce the damage that would be caused if it did occur. For example, firewalls, endpoint protection, implementation of security policies, user training, access control, etc.
- **Vulnerability**: The possibility of damage to the protected elements in the event of a successful attack by the threat source. In other words, a characteristic of the system to be protected that allows an attacker to exploit an inherent flaw or weakness to attack data, systems, services or other resources successfully.
- **Attack**: An attack is a process that implements a threat of some kind, usually based on one or more vulnerabilities, against the confidentiality, integrity and/or availability of resources.
- **Threat**: A threat is the possibility of an attack or breach of security directed at the resource under attack.
- **Risk**: Risk is the probability of the possibility that a threat will cause damage through an attack. A consequence may be financial, legal, reputational, human resource, etc. In determining the risk value, the probability of a given threat occurring and the magnitude of the damage consequence it causes are the main factors considered. Different methodologies use different calculation methods.

- **Protected elements**: the set of resources that achieve the mission/business objective or social purpose defined by the organisation's leadership (managers), including people, tools (IT and non-IT), documents, physical locations, processes and, last but not least, data, dedicated to the execution of IT tasks.

Based on the above definitions, the following conceptual statements can be read from the figure:

- The attacker wants to perform malicious activities on the elements to be protected.
- The owner wants to protect his defensible elements.
- The owner is aware of the vulnerabilities and applies protection measures.
- The protection measures reduce the risks.
- The vulnerabilities create risks.
- Protection measures act on vulnerabilities, mitigating their impact on the elements to be protected, either by reducing the damage consequences of attacks or the likelihood of their occurrence.
- The attacker tries to make the threat occur, which increases the risk.
- Threats exploit vulnerabilities.

Therefore, the attack mechanism is that the attacker looks for vulnerabilities in the IT system to be protected and tries to carry out attacks through them. The owner mitigates the risks by implementing protection measures to cover the vulnerabilities reported by the vulnerabilities. Security can then be measured in terms of the number of successful attacks, the quantification of the consequences of the damage, and the resources spent on protection. However, it is essential to note that there is no 100% security. If a risk is to be eliminated, this is typically achieved by eliminating the resource, system, or service to be protected.

The process itself, which implements the concept discussed above, is risk management. Risk management involves assessing threats directed at the organisation's assets, which can be systems, data, premises and even users. They also estimate how likely these threats will occur and, if they do, what the impact will be. Damage consequences can be measured along multiple impact dimensions to give the assessor a more accurate picture of the impact that a particular threat could have on the organisation. The resulting information indicates the risks that the organisation can decide to mitigate along four different concepts:

- **It is implementing risk mitigation measures**, i.e., controls that reduce the likelihood of an event occurring or its effects.
- It can also choose to **avoid risk** by not pursuing an activity that is considered risky so that the risk is also eliminated.
- You can also choose to **transfer risk**, but in this case, you are transferring the impact of the risk, not the risk. Typically, this includes insurance.
- **Risk can be accepted** if the organisation decides that the impact of the risk is at a level that is acceptable to it and that no further action is necessary.

Let's look at this through a practical example to make it more transparent. We plan to go skiing with the family. We intend to go by car, but we know we can't ski. The risk we are assessing is a possible broken leg. What can we do to reduce the risks involved? We have a few options in front of us:

- We go to training and prepare for the challenge. **This is a risk mitigation measure**.
- **We take out insurance**. If, despite the education, we still get up close and personal with the snow, we pass the cost of the rescue on to the insurer. (Not the damage, because we will be the injured.)
- **We accept the risk** and trust our luck.
- **We avoid the risk** and go to the beach in the summer instead.

2. Threats and attacks

Information is an asset whose presence or absence fundamentally impacts the feasibility and effectiveness of all our processes. It can increase efficiency if reasonable and cause complete uselessness or failure if it is terrible. It has become apparent during the IT dependency that the availability of quality information is essential for everyday life. It is clear that we can achieve more results with more relevant information, while with incorrect information, no process can produce the correct and maximally usable outcome. From an IT security point of view, information is generally considered the output of data processing and is represented on a computer storage medium. However, this is not the only way information can occur; we need only think of speech, telephone communications and information stored on paper, which may also need to be protected. Information has become such an essential and valuable element that it has been incorporated into information technology resources alongside hardware and software in all frameworks and standards. Therefore, information that is considered necessary must be protected alongside hardware and software. There is a growing view that the most critical resource is data. It is often the result of processes and calculations that cannot be reproduced or are challenging to reproduce. Let's take our family photo album as a simple example. Or that a video file of your child's first waking up and first steps, if you don't have more than one copy and it gets corrupted, is accidentally deleted or encrypted by a ransomware virus, and then chances are you'll never see it again because the source event cannot be replicated. By contrast, hardware or software can be reproduced, although of considerable value in monetary terms. I can buy a new laptop, and I can repurchase the software. The same cannot be said of individual data, documents, photos, files, or sound recordings.

These values have also been recognised by the attackers, who have grouped their attacks around two factors:

Destruction: damage to the attacked party by causing damage to resources (including information) necessary for its operations

Profiteering: to obtain for oneself, through the theft of resources, the benefit that can be obtained through the illegal use of another's resources (information theft, zombie network, etc.). A qualified case of this is identity theft, where the benefit is the attacker's and the punishment the attacked – unless he can prove his innocence. And in recent years, extortion has taken centre stage, carried out by attackers spreading blackmail viruses.

These two objectives are typically achieved through malware and other varied forms of attack.

2.1. Malicious software

Malicious software is any program that tries to do anything to our computers or data uploaded to the network without the owner's permission. The English version of the term (malware) comes from the abbreviation of “malicious software”. Malicious code is software that allows or implements unauthorised infiltration of computer systems or unauthorised activities that cause harm to users. They are created and sent for malicious purposes. System-level malicious code (rootkits) are commonly used to hide malware. Each malware can be classified as follows:

2.2. Viruses

Programs that spread by themselves when linked to other files or sent via email and can cause damage to computers. A particular subtype is ransomware, which encrypts files on the infected device (workstation, server, smartphone, etc.) and offers a ransom to obtain the key needed to unlock them. For more information, see ransomware.

2.3. Worms

A self-replicating computer program similar to a virus. However, while viruses attach themselves to or become part of other executable programs or documents, worms operate independently.

2.4. Trojan horses

Trojan horses named after the war machine used in the ancient siege of Troy, whereby a legitimate-looking download hides a program that eventually activates and causes incidents (e.g. downloads backdoors or launches malicious programs).

2.5. Backdoors

Backdoors are add-ons built into software that give selected individuals access to specific programs, computers, or data managed on them. Software developers deliberately build in some of these backdoors for service purposes. In contrast, others are built in by programming errors that allow attackers to bypass access rules and gain unauthorised access. In addition, attack programs are explicitly designed to open backdoors, usually distributed as part of viruses or spyware without the user's knowledge. Using these for attack purposes is dangerous because they can lead to the installation of malicious code in each case. The backdoor bypasses system security, so the protection that would otherwise be in place will not be adequate.

2.6. System-hiding programs

Malicious software designed to gain unrestricted, unauthorised and hidden access to computer resources. It is important to note that these programs bypass the established access protection system so that access gained here cannot be controlled at the system level.

2.7. Denial of Service

(DOS or Distributed Denial of Services – DDoS) attackers: a program running on one or more computers that sends many requests per second to a specified address without looking at or processing the responses sent. In this way, other users using the system do not receive replies to their actual requests due to the congestion of the computer under attack. In this way, if an Internet shop is attacked, for example, it is impossible to shop there, ergo there is a natural loss of revenue.

2.8. Spyware

Hidden programs that transmit data to an attacker without the user's knowledge or permission. They can be hidden as part of any application package that contains executable programs. In addition to computer programs, data-stealing programs written for smartphones have also appeared. Spyware can even log our keystrokes to steal our passwords, for example, or steal other information about us through our camera or microphone.

2.9. Ransomware

An attacker injects a program into a user's computer, phone, or any computer-based system that locks infected devices or encrypts valuable files, rendering them unusable. The program may also claim that it will only unlock for a fee. The victim is not guaranteed to get their data back after payment.

2.10. Cryptocurrency mining programs

An attacker installs a program on a user's computer, phone or other device – increasingly on powerful servers on corporate networks – that mines cryptocurrency for its host (who is not the user) without the user's knowledge. Cryptocurrency mining is a legal activity if the 'miner' does it on infrastructure owned by the user. From the user's perspective, cryptocurrency mining does not cause direct harm. Still, it can cause indirect harm by significantly slowing down the device running the program, which can degrade the user experience and consume significantly more power than it would during regular operation. In an enterprise environment, mining can cause more severe problems, as applications supporting business processes running on servers can slow down or even stop running, which in turn can cause monetary damage.

2.11. Junk mail

The name “spam” is borrowed from the name of a canned meat product (Spiced Pork and Ham) from an American company (Hormel Foods) that has been around since 1937. It has become the standard term for mass emails on the Internet, following a Monthy Python piece. A common feature of junk mail is that it promotes a product or service by unauthorised – and, in Hungary, among others, illegal – use of other people’s IT resources.

2.12. Adware

These are programs that can be downloaded and used free of charge and display advertisements on the user’s computer. They are also known as PUPs (Potential Unwanted Programs), often used to deliver malicious programs to the user’s computer.

2.13. Zombie network software

The English term (botnet) is derived from the words “robot” and “network”. IT jargon refers to a program that works remotely or automatically on an infected machine. It is possible that a user’s computer is part ofa botnet network and is working remotely without the user’s knowledge. This usually requires an online presence. The zombie network software can infect and control a computer without the owner’s permission. The zombie network software can be used to steal data, send spam or attack other computers, as it can be installed on the user’s computer undetected and carry out any activity.

The malware is most often delivered to the attacked computer via the Internet, which only requires the computer to be connected to the Internet. It is much less common for attackers to use physical means of attack, as this requires some personal presence, significantly increasing the risk of being caught. However, the following tools can be successfully used in an attack:

2.14. Keyboard keystroke loggers

Keyboard keystroke loggers are small hardware devices that an attacker plugs between the keyboard and the computer, with storage capacity to which the device records all keystrokes. An attacker can access sensitive information – typically system administrator passwords or other login credentials – by evaluating the device’s contents.

2.15. Hidden cameras

A small data recording device that can capture high-quality images and sound. Cameras may be continuous or motion/voice-activated, wired or radio controlled, or powered from their internal or electrical mains supply. An attacker can steal passwords or sensitive information while under surveillance. The disadvantage is the need for personal presence or physical location. In some cases, attackers can turn on a computer's built-in or connected webcams without the user's knowledge, use them as hidden cameras and steal data from the user's environment.

3. Typical forms of attack and methods

3.1. Pharming

An attacker redirects a user's traffic to a website to their website (or to a hacked website they own) by modifying some data on the user's computer so that the user can enter personal information – such as login details – without being suspected, thinking they are on the accurate site.

3.2. Downloading malware embedded in attachments

A widespread form of attack is to trick the user into downloading and opening an attractive attachment when the malware embedded in the attachment is activated – possibly while maintaining the pretence (e.g. document/image display, program running, etc.). This is how most ransomware and cryptocurrency mining programs get onto the victim's machine.

3.3. Phishing

The use by attackers of an image of an actual website (fake website) that looks no different from the original. Attackers use it to trick unsuspecting users into providing login or personal information while believing they are providing it on the original website. Advanced spoofed websites may also have a fake SSL certificate. Fake websites are accessed via links embedded in fake messages (e.g. a request for a data change from a system administrator via email, a password change request from a bank following a security incident, a billing statement from a service provider, etc.). This differs from hijacking phishing, where the attacker does not modify anything on the victim's machine.

3.4.Shoulder surfing

A direct surveillance technique where the attacker looks over the user's shoulder to obtain information. Shoulder surfing is effective in crowded places, when the user types in his PIN at an ATM (ATMs have small mirrors to detect this), or in public places such as an internet café or library, where he types in his customer security code, password, etc.

3.5. Social engineering

The attacker deceives the user about their identity to share information that the user would not otherwise be entitled to. For example, the attacker may impersonate a law enforcement officer or a system administrator. Still, it is not uncommon to see the newbie syndrome abused by the willingness to help newbies.

3.6. Data Leakage

Today, individuals and organisations generate a lot of electronic information and data daily. Communication channels and data carriers provide opportunities for users to handle and move this data and information. Data leakage is defined as events where data classified as confidential/secret (but certainly not public) is released from the protected control of the organisation due to the user's negligence or intentional action, and unauthorised access to this confidential data and information may occur.

3.7.Advanced Persistent Threat (APT)

This type of attack is characterised by the use of multiple, often interrelated, attack methods, preferably as stealthily as possible, even hidden for long periods, exploiting typically unknown vulnerabilities, attacking the target to carry out activities, such as data theft, the corruption of IT systems or other illegal activities.

3.8. Cybercrime

Cybercrime refers to crimes committed through cyberspace while using a computer. Examples include online phishing and theft of credit card details (name, number, expiry date, CVC).

4. Defending against malicious attacks

4.1. Access protection, password protection, authentication

Most attacks are committed over the network because the entire internet community sees an internet-connected computer, smart device and smartphone. In contrast, with a computer, the number of people physically entering the room containing the computer or, in the case of home appliances, the number of people physically entering our home is usually minimal. Whereas in the past, logical protection of networks (firewalls, content filtering, data leakage protection) was much more important than physical protection, today, in the age of portable devices, the physical protection of devices also needs serious attention. It is easy to lose a phone, which can be quickly snatched from your hand in a busy place. The vast majority of portable devices and laptops are stolen from cars.

For this reason, portable devices should not be left unattended in a car for even the shortest period, even in a locked place such as the boot. In busy urban areas (shopping malls, shopping centres, parks, schools), thieves will be looking for people who put what looks like a laptop in the boot and either steal it there or the next time they park. When the owner returns, the device will be in cold storage. Often, the owner doesn't even know where the device was stolen from.

In such cases, the best advice is to encrypt the backup storage, password-protect the device, and have a screen lock on the phone and a PIN code on the SIM card. Last but not least, don't store unsaved data on your portable device, as you can buy another phone, but you can never re-record a video of your child's first steps, for example.

In general, every network has someone who assigns and revokes file and device access. Suppose we need to give more than one person access to our own closed network. In that case, we have implemented a network administrator role, who is authorised to manage authentication, authorisation and accountability on the network and is responsible for maintaining the necessary access to data on the network. This typically means access to our home wireless network in a home environment. It is advisable first to change the default administrator (admin) password and then set an access password (wifi password), as radio signals do not stop at the wall. It may not be suitable if your neighbour is surfing the Internet over our wireless network.

There are many threats to the security of the network. You need to be aware that using an unsecured wireless network can allow your data to be intercepted by traffic interceptors or leaked by unauthorised parties, even without your knowledge. When designing password protection, the password must be sufficiently secure. It is recommended that good password management rules are followed, such as not sharing passwords with others, changing them periodically, using the correct password length, using a combination of appropriate password characters – letters, numbers and special characters, and not writing passwords in notebooks, on different files, on sticky notes or using the same password in more than one place.

When using passwords, we distinguish between three types of passwords:

- reusable password: you enter a password once on a given system and use it until the next time you change it.
- Time password (OTP): This password is generated by the user. It can only be used once after generation – typically, a token hardware device is required, or it is generated by the system you want to log in to and sent to you via some channel. The best examples are the SMS passwords banks use to access the Internet banking site and the SMS passwords used for transaction authentication (two-step authentication).
- biometric password: a physiological characteristic of a person (e.g. fingerprint, voice, retina, palm print, etc.)

Bad passwords do not provide security, as they cannot stop a potential attacker but can delay the attack by a moment or two because bad passwords can be cracked or guessed in seconds. Several analyses have been conducted at home and abroad to illustrate bad password usage habits.

From a user perspective, the security of operating systems typically refers to the security of files. File security can be discussed from several aspects through their associated operations. These are read, write, delete, and modify. The critical question is, who has these file permissions?

Read access is prevented by file encryption by encryption software – where you can open the file but cannot interpret it – or by password protection in a text editor, where you cannot open (unencrypt) the file without knowing the password. Password protection can be set for documents created by office software packages (text, spreadsheets, presentations, etc.) or compressed files.

The file can only be written in backing storage if you have the right to do so. Otherwise, the file you want to create will not be moved from memory or deleted when you close the program (if the program's memory management is set correctly). It is possible to write (modify) a file if the file is assigned to the user for modification-writing; otherwise, the user will not be able to save the modifications. It is also important to note whether there are any file elements that could be used in a malicious attack to write to the file or the system without the owner's knowledge – for example, macros.

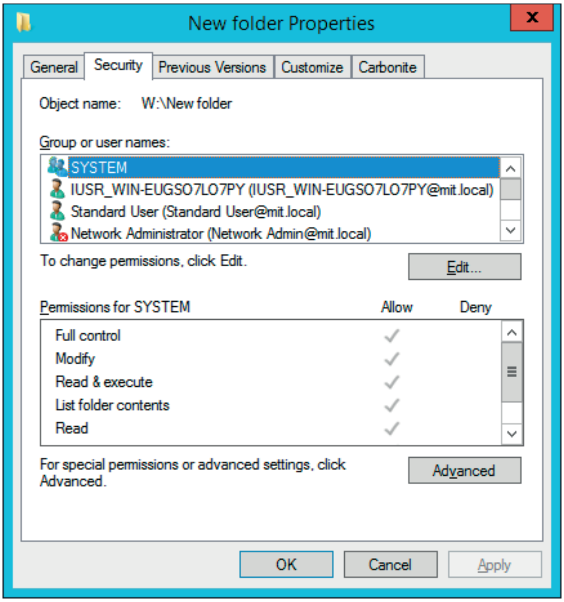


Figure 6.2. Setting access rights in the Windows operating system

Source: Own editing based on the Windows system

If you use a computer at home, it is recommended that each family member has a user account with a general level of access. The administrator account should only be used when necessary. This allows you to restrict access to each other's data in your operating system.

4.2. Firewalls

Firewalls are hardware or software devices that act at the boundary of a network to allow or deny incoming and outgoing data elements based on a predefined set of rules. In other words, firewalls enforce and adhere to the access rules we define for communication. There are many different levels and capabilities of firewalls. On the user side, the most important is the personal firewall.

A personal firewall is a software on your computer that allows or denies the running of individual applications and their network communications, in many cases in a self-learning system. A personal firewall is software that continuously runs on your computer. A personal firewall can either be part of the operating system or installed on your computer – for example, as part of a security suite.

A firewall's job is to protect the network from intrusion; in other words, it prevents unauthorised access to the network from an external location by enforcing predefined access protection. Restriction is done by rules that tell network traffic whether it is allowed or denied, which is why a firewall is a rule-based system. Additional rules can be created to manage incoming/outgoing network traffic if necessary. For example, when installing a new game program, ports that have been closed must be opened, i.e. allowed to use the game.

The goodness or badness of firewalls is determined by their ability to filter out unwanted traffic and allow expected traffic through all network levels. To do this, you need to be able to set rules for them, and there are many resources, forums, and guides on the Internet. After some experimentation, you can also learn how to create a secure environment independently.

4.3. Digital signature

A digital signature is an encrypted code that associates a person's identity with the signed file; in other words, it authenticates it. Authentication is the confirmation of the claimed identity, so authenticity means that the origin and the sender have not changed. More specifically, a digital signature is a mathematical sequence of numbers based on an asymmetric cryptographic algorithm generated by a digital signature scheme and used to ensure the message's authenticity (origin and integrity). The signature-creation data (secret key) used to create the digital signature will be paired with the signature-verification key (public key), which is included in the digital certificate by the CA after the signatory has been identified and authenticated. The digital certificate, therefore, certifies that the message's sender is who they claim to be. Digital certificates can also contain public keys and other authentication data such as name, city, address, personal identification, job title, department, etc.

Digital certificates can serve different purposes. Signing, encryption, authentication, personal, organisational, code signing and SSL certificates exist.

All certificates are structured in the same way; the differences lie in the content of the data and the purposes for which they are used.

Signing certificates are for digital signature purposes. They contain the signature verification data, which is used to create the digital signature with the signature creation data.

Steps to create and verify the digital signature on the receiving side:

- The data to be signed is extracted into a fixed (usually 160–512 bits) length extract,
- the digest is encrypted by the application using the signing algorithm and the secret key and becomes the digital signature,
- the initial verification of the signature is done automatically,
- the digital signature is attached to the data and sent to the recipient.

Digital signatures differ from public key encryption in that the secret key is used to sign the message, and the public key is used to verify the signature – the reverse is true for encryption.

4.4. Confidentiality tools on social networking sites

With the proliferation of social networking sites, such information and personal data can be leaked to the public network due to inappropriate or automatic default settings. It is essential to understand why confidential information should not be published on social networking sites and how to implement and monitor the protection settings for this information.

A controlled and thoughtful presence on social networking sites is also essential to avoid potential dangers such as cyberbullying, grooming, misleading/dangerous information, false identities, and using or accepting fraudulent links or messages. It is elementary for attackers to gain our trust if we provide intimate details on social networks, such as our nickname. Still, we can also get physically close to them if we disclose our holiday dates and address. The risk can be calculated accordingly because the potential number of virtual attackers is much larger than the number of attackers expected in real space.

There is less risk in providing information on musical interests and favourite television programmes, as this data is accessible from more places and is more widely known than a nickname. For a sex offender, every little piece of information we provide on social networking sites can make it easier to prepare for sexual exploitation, a known and dangerous threat here.

The result of making personal information publicly available on a social networking site is that the personal information can be viewed by anyone, indexed by search engines and even displayed to strangers as search results. When using social media, our data can be read not only by those who are friendly towards us but also by those who may not like some of our expressions and express this through cyberbullying. There are three ways to avoid this – or reduce the risks:

- we are careful when choosing our friends or deleting unpleasant friends, and
- we set our privacy settings as tightly as we can to ensure that no one other than our friends can read our posts or view our uploaded photos and
- be careful about who can contact us – if we don't need to, we should not allow anyone we don't know to contact us directly, only someone who has already been verified – in some way – as a friend by someone in our circle of friends.

Social networking sites offer users several recommended options. The settings are grouped around the following topics – for example – on one of the most popular social networking sites, Facebook: Who can see my things?

- Who can see my things?
- Who can see my contacts list?
- Who can contact me?
- Who can find me?

4.5. Online banking

Online banking is becoming more and more common these days. Banks are doing their utmost to ensure that their e-banking services are adequately protected and that their customer's data is protected. On the one hand, this is in their business interest, and on the other hand, the Financial Supervisory Authority is closely monitoring the efforts made by banks.

Internet banking, from the user's perspective, has two critical points. One is the login, and the other is the execution of the various financial transactions.

For Internet banking logins, domestic banks have offered a verified, so-called two-factor login for several years. This can be a one-time password sent by SMS, a hardware random number generator (token) provided by the bank, or a smart card login. Banks must accept various transactions only after some form of multi-factor authentication to prevent customers from being harmed or having their money stolen by an attacker. The simplest example is a transfer, where the bank sends a confirmation transaction code after submitting the order. The SMS includes the target account, the amount to be transferred and the transaction code at the end.

Since it will always be sent to the user's phone number, even if an attacker has managed to access the user's online bank, if the customer does not provide the transaction code for the transfer – because he realises that he did not want to transfer the wrong amount or the wrong amount, the transaction will not be completed, and there will be no financial loss. If this happens, we should inform our bank immediately and wait for them to investigate our report. Financial institutions are legally obliged to settle transactions proven to be incorrect or unauthorised.

To prevent the theft of users' Internet banking details and to prepare customers against this type of attack, all banks publish security warnings on their websites, which are worth heeding and applying.

4.6. Online shopping

Online shopping is becoming more and more common in the world. Electronic payment procedures are typically based on an electronic bank account, which may be linked to one or more bank cards. It is advisable to use trusted online shops where the shop does not handle your card details but redirects you to the bank's payment page, where the card payment is made, the merchant receives notification, and we receive the goods or services.

Several banks offer so-called virtual cards specifically for online payments. This solution is good because the card either does not physically exist or, if it does, does not have a magnetic stripe or chip on it, so it is unsuitable for cash withdrawals or POS terminal payments; only the card details are of interest. A virtual card is usually linked to your main account or has a sub-account. When we want to use the card, we have to increase the card limits in advance and reset to the fundamental limit immediately after the purchase or after 24 or 48 hours in case of a time limit. This way, even if our card details are stolen from a merchant's computer system, they cannot access the amount held in our account. Where the card is linked to a sub-account, we can spend as much as we have previously transferred to or hold in the sub-account.

One of the most common online payment methods is PayPal, which is a virtual account that also requires a credit card. If you want to use a very secure transaction, you can enter a virtual card to register and transfer all the funds to your PayPal account, which you can use to make purchases online. Meanwhile, our physical card and the money in our bank account are not at risk.

REFERENCES

Compass Cyber Security. (2024, January 30). Risk managers approach cyber security. Retrieved from https://www.compasscyber.com/blog/risk-managers-approach-cyber-security/

Common Criteria. (2017, August). Common criteria for information technology security evaluation, part 1: Introduction and general model (Version 3.1 CCMB-2017-04-001). Retrieved from https://www.commoncriteriaportal.org/files/ccfiles/CCPART1V3.1R5.pdf [Accessed: 2024, March 10]

Erdősi, P. M., & Solymos, Á. (2018). IT biztonság közérthetően. Neumann János Számítógép-tudományi Társaság. ISBN: 978-615-5036-12-5

Kormányhatározat. (2013, March 21). 51139/2013. (III. 21.) Korm. határozat Magyarország Nemzeti Kiberbiztonsági Stratégiájáról.

MTMT. (2024, January 10). Retrieved from http://real.mtak.hu/11147/

NAIH. (2024, January 12). Adatvédelmi szótár. Retrieved from https://www.naih.hu/adatvedelmi-szotar.html

White House. (2003, February). The National Strategy to Secure Cyberspace. USA.

CHAPTER 7: BLOCKCHAIN TECHNOLOGY AND CRYPTOCURRENCIES

(István Füzesi)

1. Overview of blockchain technology

Cryptocurrencies like Bitcoin use blockchain technology to record and authenticate transactions. In the blockchain system, network participants identify, record, and verify all transactions. The digital "double spend" problem (when the same digital unit of currency is tried to be sent or used more than once for the same purpose in digital payment systems). With traditional physical currencies, it is impossible to spend the same physical currency in two places simultaneously, as a single physical copy can only be spent in one place at a time. However, copying and creating digital copies can be quickly done with digital currencies.

Consequently, some people may try to send the same amount of money several times, taking advantage of the possibility of creating digital copies. The double spend problem is solved by blockchain technology through consensus-based network verification and transaction validation, enabling transparent and reliable transactions. In this way, network participants validate, verify, and confirm new transactions before they are added to the blockchain to prevent double-spending problems in blockchain cryptocurrencies.

A "trusted third party" is an entity or organisation that acts as an intermediary between two parties in financial transactions and helps them to carry out or verify those transactions. In financial transactions, parties often use the services of such a trusted third party to ensure the transaction's security, integrity and authenticity. A trusted third party may be, for example, a bank, payment service provider, financial institution or other institution that provides authentication, verification or guarantee for transactions. This entity may act as an intermediary in the transaction, verify the parties' identity or the transactions' authenticity, handle payments or data, and ensure the security and reliability of the transactions. However, with the advent of cryptocurrencies and blockchain technology, financial systems operate without a trusted third party. These systems are decentralized and based on blockchain technology, allowing direct transactions between parties without needing an intermediary or trusted third party. Blockchain technology captures and verifies transactions transparently and reliably, thus avoiding needing a traditional trusted third party in some instances.

For financial transactions, the benefits of cutting out intermediaries can include:

- Reduced costs: since there is no need for an intermediary or central institution to charge outstanding fees and commissions for transactions, parties can transact directly and at a lower cost.
- Faster transactions: Higher security and transparency: transactions are recorded publicly and transparently on the blockchain, reducing the possibility of fraud or manipulation. In addition, decentralized systems can be more resistant to cyber-attacks and manipulation.
- User control: users can gain greater control and ownership over their data or money. This means that users have complete control and management of their own money or information.

However, in some cases, the role and institution of a trusted third party may still be significant due to specific legal, financial or regulatory requirements.

To understand blockchain technology, it is first necessary to clarify the network types. A centralized network is an IT or data communication network with a central point of management or control. In this type of network, data, resources and decision-making processes originate from or are controlled by a central entity. In such a system, nodes (or participants) communicate directly or indirectly with the central entity, and the controlling entity primarily operates the network.

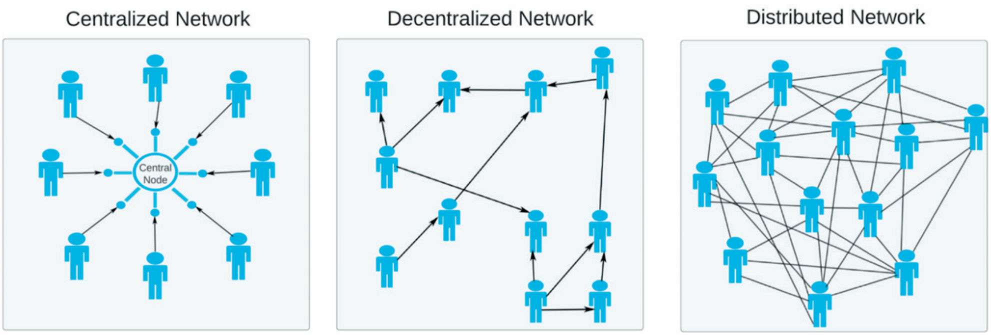


Figure 7.1. Network types

Source: Khoshavi et al., 2021

A decentralized network is one where control, decision making and resources are not concentrated in a single central unit but distributed across multiple system nodes. In this type of network, the participants communicate directly or indirectly with each other, and the participants operate the network.

A distributed network is a system in which control, resources and decision-making are spread across multiple nodes or computers. Distributed networks interact with each other directly or indirectly to accomplish tasks. Distributed networks aim to achieve scalability, redundancy, reliability and more efficient use of resources through coordination and cooperation between participants. A significant advantage of a distributed network is that the failure of one node does not cause any disruption to the system, and more remote, operational nodes immediately take over its tasks.

A blockchain is a distributed database created and maintained by nodes in a network. Nodes are key participants in the blockchain system and perform different functions on the network. A blockchain has two main types of nodes: full nodes and miners. The full nodes store and maintain a full copy of the blockchain. These nodes maintain complete control of the blockchain, monitor transactions, and contribute to verifying and accepting new blocks. Full nodes participate in the network control processes and help maintain the integrity of the blockchain. The miners are computers or systems that confirm transactions and add new blocks to the blockchain. Miners solve specific cryptographic tasks (usually in the context of Proof of Work (PoW) or Proof of Stake (PoS) consensus mechanisms) and receive rewards for successful solutions, such as cryptocurrency. Miners are responsible for creating and adding new blocks to the blockchain.

The consensus mechanism in a blockchain system is the process that helps participants agree on the state of the network, valid transactions and the addition of new blocks. The consensus mechanism is crucial because agreeing on changes or transactions between participants is difficult without a central authority overseeing everything in distributed systems. The most commonly used consensus mechanisms (other solutions exist nowadays):

- Proof of Work (PoW): PoW is the best-known consensus mechanism, and it is used, for example, by Bitcoin. In this system, miners' computers solve complex mathematical problems, and the first one to solve successfully is entitled to add a new block to the blockchain. The PoW system requires high computational power, and the miner who has done the most work is entitled to the prize.
- Proof of Stake (PoS): PoS is an alternative to PoW and measures the amount of cryptocurrency participants hold in the system. The more they hold, the more likely they are to be eligible to create new blocks and confirm transactions. The PoS consensus mechanism uses less energy than PoW.

Through communication and collaboration between nodes in a blockchain network, each node can verify and confirm transactions, maintain data integrity and reach a consensus on the system's state. The decentralized nature and consensus mechanisms make the blockchain system more resilient to failures or attacks from individual nodes while providing a single, trusted database for participants.

1.1. An example of how blockchain technology works in financial transactions

We can imagine that we want to buy a product with bitcoins. In this case, we see the Bitcoin address of the seller, and in our wallet application, we can specify that we want to send one Bitcoin to this address, for example. The transaction is recorded on the blockchain, which contains the sender and receiver addresses, the time of the transaction and the amount. The miners are responsible for checking that the amount we have designated is available in our wallet, that the amount is only spent once and that all other parameters are correct. Once one miner approves the transaction, the data is passed on to the other miners. If the other machines find it correct, the block is added to the blockchain and thus remains in a time-invariant state. Each node in different parts of the world contains a copy of the blockchain, so they constantly check each other. The owners of the volunteer nodes do not know each other, but the system constantly checks each other to ensure near-flawless operation. The only possible abuse would be to control more than 51% of the nodes. However, this is virtually impossible for a blockchain with tens or even hundreds of thousands of nodes worldwide.

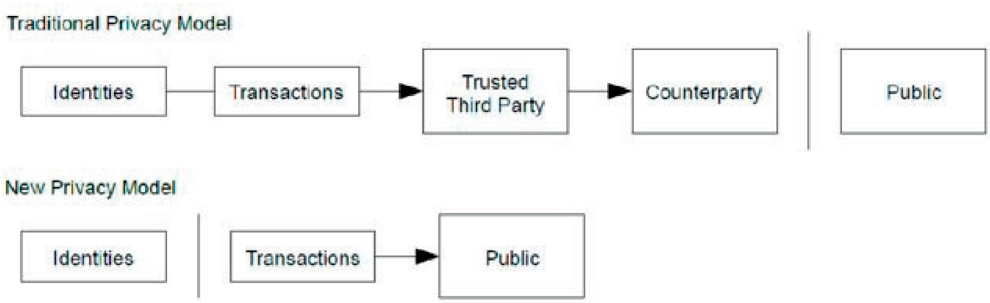


Figure 7.2. Models

Source: Presthus and O'Malley, 2017

Since transactions are broadcast to the whole network, they are entirely public. In contrast to traditional financial institutions, which protect the privacy of their customers by withholding information about their transactions, the blockchain system ensures this by not disclosing information about the owner of the addresses. In network and distributed database transactions, the payer and the payee appear only with their address. This address is nothing more than the public half of a cryptographic key pair, which allows the owner of the address (who alone holds the private key) to sign transactions issued against the balance on the address. The fundamental problem is to prevent someone from spending the money they have more than once. To prevent this, a block is added to the database every 10 minutes, where the new block is self-assembled by a randomly chosen node so there is no overspending. Transactions in the block, thus added to the database, are accepted by the other nodes as having occurred. When a node publishes a new block, it has to prove that it has solved a problematic cryptographic task that requires significant computational power, and the required rate is continuously regulated so that one node in the network can do this every 10 minutes or so. As a result, the majority of decisions on the network are determined by the underlying computational capacity.

1.2. Hash functions

Hash functions are mathematical algorithms that take, as input, a piece of data of arbitrary length (e.g., text or file) and generate a fixed-length code or hash value from it. These codes are usually shorter, of fixed length and unique to the input data. Hash functions are used for various purposes, such as data storage, password protection, digital signatures and other cryptographic applications. Hash functions must satisfy the following conditions:

- Deterministic: The same input always leads to the same hash value.
- Fast computation: hash functions must compute their output efficiently, even for large inputs.
- One-way: Recalculating the original input from the hash value should not be easy.
- Collision-resistant: It should not be easy to find two different inputs that lead to the same hash value (collision).

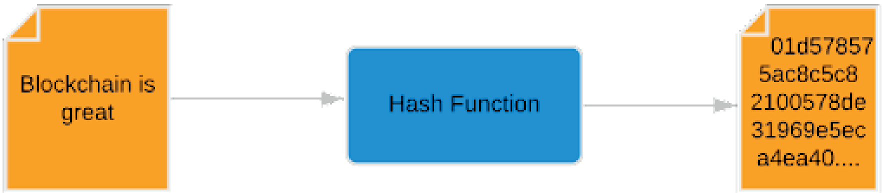


Figure 7.3. Simplified hashing flowchart

Source: https://chainkraft.com/what-is-hashing/

For example, generate a SHA-256 hash from a text. For example, take the text “Hello, world!”.First, convert the text “Hello, world!” to binary form. Then, we will use the binary data before applying the SHA-256 hash function. Using ASCII encoding, you can create binary data from the text “Hello, world!”. ASCII (American Standard Code for Information Interchange) is a character encoding system representing characters by binary numbers. Thus, each character is assigned a unique binary value.

“Hello, world!” -> 01001000 01100101 01101100 01101100 01101111 00101100 00100000 01110111 01101111 01110010 01101100 01100100 00100001

Then, we apply the SHA-256 hash function to this binary data. The SHA-256 hash function converts a binary sequence received as input into a hash value of 256-bit length. After the SHA-256 hash function is applied, the result is a hash value of 256-bit length, usually in hexadecimal format. An example of a SHA-256 hash value for the text “Hello, world!”:

2cf24dba5fb0a30e26e83b2ac5b9e29e1b161e5c-1fa7425e73043362938b9824

This hash value uniquely identifies the text “Hello, world!”; you will get the same result a zillion times if you repeat the same process. However, if you change just one bit in the input, the hash value generated will change radically. This feature is very useful in data integrity checking and other applications.

1.3. Mining

A “nonce” (number used once) in blockchain technology is data used by miners to create a new block in the blockchain and thus validate transactions. A nonce is a 32-bit integer used as an identifier during mining. During the mining process, miners must find a nonce that allows them to apply a hash function (usually SHA-256) to the current block header and thus generate a hash value for the block, which must meet specific difficulty criteria. The nonce must be chosen so the resulting hash value is low enough for the rest of the network to accept the block. Since the nonce can be chosen randomly, the miners try different nonce values until they find the value that allows the desired result to be obtained. This process is called the “mining difficulty” or “proof-of-work” algorithm. Nonce is an essential part of blockchain security and integrity, as it ensures that block creation is a more costly process, preventing blockchain manipulation or tampering.

The Bitcoin network adjusts the difficulty of mining to ensure an average block creation time of around 10 minutes. This is to maintain the stable operation of the blockchain and the fast processing of transactions. The difficulty of mining affects the nonce value that miners try to find when creating a block. The difficulty is controlled so that the first few bits of the block’s hash value are zero. The more zero bits miners try to achieve, the harder it will be to find the correct nonce value. Setting the difficulty is an automatic process that occurs after every 2016 block in the Bitcoin network. After each block creation, the network checks the average block creation time based on the last 2016 block. Suppose this time is shorter than the target of 10 minutes. In that case, the network will automatically increase the mining difficulty (by reducing the number of zeros in the block hash value) so that the next 2016 block can be created again in 10 minutes on average. If the average block creation time is longer, the difficulty is lowered. This mechanism allows the Bitcoin network to adapt to changes in the number of miners and changes in miner performance, maintaining the average block creation time.

The “51% attack” is a threat affecting blockchains’ security, such as the Bitcoin network. This attack occurs when an entity or a group can take control of more than 50% of a blockchain network with its total computing capacity. If an attacker controls more than 50% of the blockchain, it effectively gains complete control of the network. It would therefore be able to perform actions that could be detrimental to the stability and security of the network.

1.4. Smart contracts

Smart contracts are computer programs or pieces of code that operate and execute on blockchains in a planned way. These contracts automatically execute predefined terms and conditions without the need for intermediary or third-party intervention. Smart contracts are essentially composed of code and run on blockchain networks such as Ethereum. The main reasons for using smart contracts:

- Automation: smart contracts automatically execute specified conditions without the need for human intervention. This allows for the automation of transactions and processes.
- Simplicity and efficiency: smart contracts allow business processes to be simplified and made more efficient by eliminating intermediaries and administrative costs.
- Programming patterns and conditions: developers can program smart contracts in detail to behave exactly as they are intended. Conditions and patterns can be precisely defined in the code.
- Security: smart contracts also take advantage of the security benefits of blockchain technology, such as a distributed and unalterable database.
- Voluntariness: smart contracts are based on voluntariness, and the trust between the parties is anchored in the blockchain network.

Smart contracts can be used in a wide range of applications, including finance, real estate, legal agreements, insurance contracts and many other areas where automation of contracts and transactions is required.

Smart contract-based supply chain management process effectively automates the ordering, delivery, receipt and payment processes. Through the blockchain network, all transactions and status changes can be tracked transparently and reliably, minimizing the possibility of human error and fraud. It is important to note that the encryption and publication of smart contracts require a strong knowledge of blockchain and cryptography. If one does not have expertise in this area, it is worth involving an expert or developer in creating and testing the smart contract (Mohanta et al., 2018).

2. Cryptocurrencies

The concept of digital currencies has been on the minds of cryptographers and software developers for decades, but it was only in the 21st century that technology and society reached the point where these initiatives became real financial instruments. Although the concept of digital currencies was first conceived in the 1980s, cryptocurrency as we know it today began in 2009 with the creation of Bitcoin. In 2008, an unknown person or group, publishing under the pseudonym Satoshi Nakamoto, published a Bitcoin "white paper" detailing the workings of the digital currency and the underlying blockchain technology. The Bitcoin network was launched in 2009, and the first block, known as the "Genesis block", was mined. Bitcoin allows people to send and receive value directly without any central authority. Since then, there have been many different types of cryptocurrencies, each with unique characteristics and objectives.

In the world of cryptocurrencies, buying, storing and using are key aspects that allow people to interact with digital currencies. The most common way to buy cryptocurrencies is through various online exchanges such as Coinbase, Binance, or Kraken. These platforms allow users to exchange fiat currencies (e.g. USD, EUR) for cryptocurrencies. Another popular way to buy cryptocurrencies is to use peer-to-peer (P2P) platforms, where buyers and sellers can interact directly with each other. The storage of cryptocurrencies is a crucial consideration for anyone involved in the digital world. How cryptocurrencies are stored can significantly impact their security and availability. The devices that store cryptocurrencies are commonly called "wallets". They can be software-based, hardware-based or even paper-based:

- The "hot Wallets" are online wallets that are permanently connected to the internet, providing greater convenience and quick access to cryptocurrencies. They include mobile, desktop and web wallets. Although convenient, they can pose a security risk if not adequately protected.
- The "cold wallets" are offline wallets that are not permanently connected to the internet, thus providing more secure storage for cryptocurrencies. These include hardware wallets, which store cryptocurrencies on physical devices, and paper wallets, which contain private keys in printed form.
- The "hardware wallets" are physical devices, such as USB drives, that securely store and manage a user's private keys. These devices are highly secure as transactions are generated offline and inside the device.

Following the success of Bitcoin, several other cryptocurrencies, collectively known as "altcoins", have appeared on the market. These altcoins attempt to build on the foundations laid by Bitcoin, often offering new and innovative technologies or different use cases and features. Among altcoins are Ethereum, which focuses on running smart contracts and decentralized applications (dApps); Litecoin, which offers faster transaction times and lower transaction costs; and many others with different objectives and technological approaches. Bitcoin and altcoins together form a dynamic and rapidly evolving ecosystem that is constantly expanding with new digital currencies and blockchain-based projects. Bitcoin and altcoins play a critical role in the innovation and evolution of the cryptocurrency space, challenging traditional financial systems and exploring new possibilities for electronic money. Bitcoin remains the best-known and most market-capitalized cryptocurrency, serving as a store of value for many investors as "digital gold". On the other hand, Altcoins offer a wide range of innovations and specializations, allowing investors and users to participate in developing different aspects and applications of blockchain technology.

An ICO, or Initial Coin Offering, raises capital for blockchain and cryptocurrency projects. The essence of ICOs is that companies issue and sell new digital tokens to the public, often in exchange for Bitcoin, Ethereum or other cryptocurrencies. This process is similar to a traditional company's initial public offering (IPO) but significantly different. ICOs offer a decentralized, unregulated way to raise capital, allowing startups to bypass traditional banking and capital market funding routes. In ICOs, investors buy "tokens" that can grant them various rights in the project, such as a stake, the right to future profits, or access to certain services and products that the project offers. These tokens are usually issued on the project's blockchain and can function as a speculative instrument, as their value can vary depending on market supply and demand dynamics.

ICOs, at the height of their popularity, revolutionized the raising of capital in blockchain technology and cryptocurrencies, enabling small and medium-sized enterprises to raise significant amounts of money without relying on traditional capital markets. However, this form of unregulated capital raising also comes with many risks and challenges, such as high rates of fraud and project failure, which have attracted the attention of regulatory bodies worldwide (Feng et al., 2019).

2.1. Risk factors for cryptocurrencies

The volatility of cryptocurrencies is one of the most significant features and challenges in the digital currency market. This volatility attracts speculators and investors looking for quick profits while also posing a significant risk to those who do not manage their investments properly. The cryptocurrency market is relatively new and immature compared to traditional financial markets, such as stock or bond markets. Lower liquidity, less developed market infrastructure and lower market capitalization contribute to higher price volatility. Much of the value of cryptocurrencies is based on speculation, which means that prices are often driven by investor sentiment and expectations rather than the underlying economic factors associated with traditional financial instruments. Cryptocurrency prices are highly sensitive to news in the media and the influence of social media. A significant announcement, regulatory change or even a tweet from an influential personality can trigger significant price movements. The second risk is technological uncertainty. Cryptocurrencies and blockchain technology are still evolving, which could expose new vulnerabilities and flaws. These vulnerabilities threaten the technological infrastructure and the assets of investors and users. Historically, several major cryptocurrency exchanges have been victims of hacker attacks, with significant amounts of digital assets stolen. Another important risk is the possibility of market manipulation. Due to the immaturity of cryptocurrency markets and the lack of regulation, market manipulation is much more likely than in traditional financial markets. This includes “pump and dump” schemes, where investors artificially inflate the price of a cryptocurrency only to sell it at a higher price. Last but not least, the risks of storing cryptocurrencies are also significant. Digital assets are stored in online or offline “wallets”, and both have security challenges. Online wallets are convenient and easily accessible but more vulnerable to hacking attacks. Cold storage (offline wallets) is more secure but also poses risks, such as loss or failure of the device.

2.2. Legal environment of cryptocurrencies

Cryptocurrency regulation varies widely around the world. Some countries, such as Japan and Switzerland, have adopted relatively pro-cryptocurrency regulations, while others, such as China, have introduced strict restrictions or even banned cryptocurrency trading and mining altogether. The European Union is actively working to develop a single regulatory framework for cryptocurrencies that seeks to protect market integrity and users. In the United

States, regulation is still under development at several federal and state levels, under the jurisdiction of various regulatory bodies such as the SEC (Securities and Exchange Commission) and the CFTC (Commodity Futures Trading Commission).

The taxation of cryptocurrencies also varies from country to country. Many countries, including the United States and the United Kingdom, treat cryptocurrencies as assets, meaning that transactions with cryptocurrencies, such as trading, selling, or even using them as a means of payment, are taxable. Gains from cryptocurrencies may be taxable as capital gains and, in many cases, must be declared with the tax authorities. Users should inform themselves about their country's tax rules and seek expert advice if necessary.

The legal challenges and perspectives on cryptocurrencies change dynamically as technology and markets evolve. The challenge for regulators is to find the delicate balance that allows technological innovation while preserving market integrity and protecting users.

REFERENCES

Abed, S., Jaffal, R., Mohd, B. J., et al. (2021). An analysis and evaluation of lightweight hash functions for blockchain-based IoT devices. Cluster Computing, 24(3), 3065-3084. https://doi.org/10.1007/s10586-021-03324-1

Feng, C., Li, N., Wong, M. H. F., & Zhang, M. (2019, March 25). Initial coin offerings, blockchain technology, and white paper disclosures. Available at SSRN: https://ssrn.com/abstract=3256289 or http://dx.doi.org/10.2139/ssrn.3256289

Khoshavi, N., Tristani, G., & Sargolzaei, A. (2021). Blockchain applications to improve operation and security of transportation systems: A survey. Electronics, 10(5), 629. https://doi.org/10.3390/electronics10050629

Klinkmüller, C., Ponomarev, A., Tran, A. B., Weber, I., & van der Aalst, W. (2019). Mining blockchain processes: Extracting process mining data from blockchain applications. In Di Ciccio, C., et al. Business process management: Blockchain and Central and Eastern Europe forum. BPM 2019. Lecture notes in business information processing (Vol. 361, pp. 71–86). Springer, Cham. https://doi.org/10.1007/978-3-030-30429-4_6

Liu, Y., Tsyvinski, A., & Wu, X. (2022). Common risk factors in cryptocurrency. The Journal of Finance, 77(2), 1133–1177. https://doi.org/10.1111/jofi.13119

Mohanta, B. K., Panda, S. S., & Jena, D. (2018). An overview of smart contract and use cases in blockchain technology. 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Bengaluru, India, 2018, pp. 1–4. https://doi.org/10.1109/ICCCNT.2018.8494045

Presthus, W., & O'Malley, N. O. (2017). Motivations and barriers for end-user adoption of Bitcoin as digital currency. Procedia Computer Science, 121, 89-97. https://doi.org/10.1016/j.procs.2017.11.013

Stolbov, M., & Shchepeleva, M. (2020). What predicts the legal status of cryptocurrencies? Economic Analysis and Policy, 67, 273–291. https://doi.org/10.1016/j.eap.2020.07.011

Suratkar, S., Shirole, M., & Bhirud, S. (2020). Cryptocurrency wallet: A review. 2020 4th International Conference on Computer, Communication and Signal Processing (ICCCSP), Chennai, India, 2020, pp. 1–7. https://doi.org/10.1109/ICCCSP49186.2020.9315193

Wüst, K., & Gervais, A. (2018). Do you need a blockchain? 2018 Crypto Valley Conference on Blockchain Technology (CVCBT), Zug, Switzerland, 2018, pp. 45–54. https://doi.org/10.1109/CVCBT.2018.00011

Zhou, S., Li, K., Xiao, L., Cai, J., Liang, W., & Castiglione, A. (2023). A systematic review of consensus mechanisms in blockchain. Mathematics, 11(10), 2248. https://doi.org/10.3390/math11102248