

3: Data Management and Security

Topic Outcomes

- Explain methods of data management and data security
- Explain data privacy
- Explain the rights and responsibilities of data subjects and data owners

Data Management

Taking care of data is a difficult task, even when data quality rules are put in place. This is where data management comes in. It involves the care of data throughout its lifecycle, to ensure it can be treated as an asset, like any item of value.

Data management refers to activities involved in treating data as an asset through its lifecycle and **Data Governance** refers to the practices and processes which ensure the formal management of data as an asset. Data governance can be considered the glue that holds data management processes together.

Data Management is a group of activities relating to the planning, development, implementation and administration of systems for the acquisition, storage, security, retrieval, dissemination, archiving and disposal of data. Such systems are commonly digital, but the term equally applies to paper-based systems where the term records management is commonly used. The term embraces all forms of data, whether these datasets are simple paper forms, the contents of relational databases, multimedia datasets such as images, or scientific data such as seismic records of the UK land mass.

Data Management implies the definition of policies, roles, processes and responsibilities throughout the company on the definition and management of data. An effective data governance model requires a complete structure that facilitates collaboration between technology and business. The goal of Data Management is to increase the value of an organisation's data through data governance.

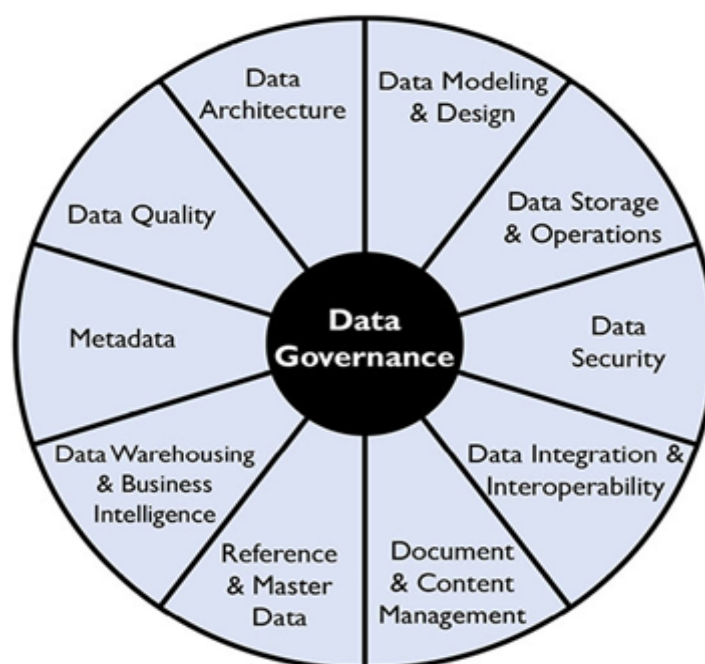
A Gartner survey identified a new figure in many organisations, the CDO or Chief Data Officer.

This CDO is an agent of change and aims to enhance the value of the data. Data exists to extract business value and improve decision making. These roles belong to the top management of the organisation and innovation. The CDO will be responsible for all data governance and will prepare the organisation for the digital transformation that will involve the integration of data, previously separated into silos.

The tasks of the CDO are:

- To define data governance
- To prepare and involve the entire organisation to expand data culture.
- The CDO will assist in decision making by ensuring that data is a source of quality information and implementing all decisions to make the information credible.

Corporate data management is about effectively managing the entire data lifecycle. The below framework is one way in which to think about data management. This is the DAMA framework, an international structure that oversees the structure of data management. It is a compilation of good practices for data management, identifying 11 functions for successful data management.



Data architecture: the standards for how data is collected, stored and used.

Data modelling and design: a map of how the data relates to the real world.

Data storage and operations: the life cycle of storage including backing up, archiving and deleting.

Data security: protecting the data from unauthorised access or loss

Data integration and interoperability: how data moves between different systems.

Documents and content: caring for data in an unstructured form.

Reference and master data: looking after some of the most critical datasets.

Data warehouse and business intelligence: managing the data used for reporting.

Metadata: understanding what data is available.

Data quality: defining rules to check for issues with quality and then fix them.

Data governance: the policies, processes and responsibilities that encompass treating data as an asset.

Technologies for Data Management

The selection of technologies used in a Data Management Initiative will vary depending on the specific needs and infrastructure of each organisation. However, some of the most common and used technologies relate to databases, modeling, data quality and data integration.

In terms of databases, the 1970s saw the introduction of the relational database management approach, which became a reliant and efficient, structured approach to managing data. As the volume of data stored in databases increased, this required different storage models and by the 1990s different types of relational data management tools and systems continued to dominate. However more recently, systems have been developed that are less reliant on the relational database structure. For example Hadoop is an open source framework for handling unstructured data using a parallel processing technique called MapReduce, originally developed by Google. Later developments include a series of NoSQL and NewSQL databases, which support scalability. Traditional relational database management systems are not capable of handling big data.

Regarding modeling, conceptual modeling in particular has been focused on the organisation of data and is important for big data. Data quality is one of the most important problems in data management. A database system typically aims to support the creation, maintenance, and use of large amounts of data, focusing on the quantity of data. In contrast to traditional data management tasks, data quality management enables the detection and correction of errors in the data, syntactic or semantic, in order to improve the quality of the data and hence, add value to business processes. When data is pulled from multiple sources, it can be used to analyse and generate meaningful insights, however this will require different levels of data integration. Data integration can be a major area of focus for organisations. In today's business world, it is typical that enterprises run different but coexisting information systems. Employing these systems, enterprises struggle to realise business opportunities in highly competitive markets. In this setting, the integration of existing information systems is becoming more and more indispensable in order to dynamically meet business and customer needs while leveraging long-term investments in existing IT infrastructure.

Big data management challenges:

There are many challenges and problems associated with big data project management, the most notable of which are:

- Data-driven culture: The data must be considered objectively, without relying on intuition. This is evident from the successful
- supply-chain management of companies such as Apple, Google, and Wal-mart.
- Business goals: Tie the project to the goals of the business. Ensure that the big data project matches well with the needs of the
- business application.
- Recruit proven analytical talent: Acquire the needed analytical talent. Innovations, as identified by such talent, can come from
- performing diagnostics, predictive, and prescriptive analytics.
- Understand big data solutions: Solutions can involve Hadoop ecosystems, Spark, NoSQL, cloud, in-memory computing, and
- data virtualisation.
- Security: As with regular data management projects, data security is always a major challenge.
- Leadership: Leadership is required throughout an entire data analytic life cycle.

Data Security & Privacy

Data security is relevant to protect intellectual property rights, commercial interests, or to keep sensitive information safe.

Arrangements need to be proportionate to the nature of the data and the risks involved. Data that contain personal information should be treated with higher levels of security than data which do not, as the safeguarding of personal data is dictated by national legislation, the Data Protection Act 2018, which states that personal data should only be accessible to authorised persons.

Some of key threats to data security generally, can be summarised quite simply as: corruption, loss, destruction, deletion, hacking and damage. This covers intentional and unintentional acts that threaten data security.

One of the challenges of the rapidly evolving field of data science and the short development life cycles involved is the difficulty of ensuring that new methods and products are secure or if they have problems and flaws in their security. Severe security loopholes can be common. Due to the increasing centralisation of storage for large quantities of data, these sites of cloud infrastructure and big data applications become appealing as targets for malicious attacks. This includes in particular an 'advanced targeted attack' (ATA) or an 'advanced persistent threat' (APT) which involve significant

effort and knowledge, as well as time to get access to a system or to data. These attacks can often utilise zero-day exploits, which are unknown to the security industry, making it unlikely that signature-based systems will detect them. Detection is further complicated where attackers aim to be as stealthy as possible and some data science tools such as machine learning and growth of publicly accessible data, are also available for use by cyber criminals, and used for improving their methods and strategies of attack. For example, the ability to profile people based on their activities on social media and determining what type and style of social engineering attacks makes them do something they do not want to do would be very useful to cyber criminals.

Data Privacy

Data protection focuses on the handling, storage and usage of personal data to maintain a person's right to privacy. Data security focuses on protecting data from unauthorised access. However, if security is compromised and a data breach has occurred, then it is likely that privacy has also been compromised.

Individual privacy: the right to keep one's personal matters and relationships secret.

Information privacy: the right to have some control over how your personal information is collected and used.

Data Protection: a legal framework to protect individual privacy by focusing on data privacy.

It is important to actively manage your privacy online otherwise more information may be being shared than is necessary. The kind of information that is often being stored, and possibly shared is more than just name and email addresses. It could be:

- Geographic location
- Web browsing habits
- Websites visited
- Products bought online
- Illnesses searched for online
- Devices used to connect to the internet
- Reading habits and history
- Food preferences
- Political views

This type of information allows companies to build up profiles of individuals and use them for targeting products and services. It is important to regularly review your privacy settings and the information that online companies are storing about you. Most have easy ways to exercise your rights and access this information.

Keeping data secure

Measures that can be taken to keep data secure include:

- making regular backups of files (backup copies should be stored in fireproof safes or in another building)
- protecting yourself against viruses by running anti-virus software
- using a system of passwords so that access to data is restricted
- safe storage of important files stored on removable disks, eg locked away in a fireproof and waterproof safe
- allowing only authorised staff into certain computer areas, eg by controlling entry to these areas by means of ID cards or magnetic swipe cards
- always logging off or turning terminals off and if possible locking them
- avoiding accidental deletion of files by write-protecting disks using data encryption techniques to code data so that it makes no apparent sense

Physical security

Controlling access to buildings, rooms, cabinets where data, computers, media or hardcopy materials are held

Logging the removal of, and access to, media or hardcopy material in store rooms

Transporting sensitive data only under exceptional circumstances, even for repair purposes; for example, giving a failed hard drive containing sensitive data to a computer manufacturer may cause a breach of security

Network security

- Not storing sensitive data such as those containing personal information on servers or computers connected to an external network, particularly servers that host internet services
- Firewall protection and security-related upgrades and patches to operating systems to avoid viruses and malicious code

Security of computer systems and files

- Locking computer systems with a password
- Ensuring computer software is up-to-date
- Protecting servers by power surge protection systems through line-interactive uninterruptible power supply (UPS) systems

- Implementing password protection and controlled access to data files, for example 'no access', 'read only', 'read and write' or 'administrator-only' permission
- Controlling access to files, folders or entire hard drives encryption
- Not sending personal or confidential data via email or other file transfer means without first encrypting them
- Destroying data in a consistent manner when needed: deleting files and reformatting a hard drive will not prevent the possible recovery of data; consult our guidance on data disposal
- Imposing non-disclosure agreements for managers or users of confidential data

Password management best practices:

- Use a strong password
- Store passwords in a password manager
- Use multi-factor authentication (MFA)
- Use biometrics where possible
- Do not reuse passwords between different accounts
- Change passwords regularly
- Do not use personal information that can be guessed, or words that can be found in a dictionary
- Do not use consecutive letters on a keyboard
- Do not share passwords with other people and if you do, change it immediately

Multi-factor authentication

Where available, this provides an extra level of security by requiring multiple factors:

- Knowledge: something you know e.g. password
- Possession: something you have e.g. phone
- Inherence: something you are e.g. biometrics
- Location: somewhere you are e.g. a building

Biometrics

A person's physical characteristics can be used to authenticate access. Although convenient, the use is highly debated, as they cannot be replaced if compromised, unlike passwords.

Examples include facial recognition, iris recognition, fingerprint scanners and hand geometry.

Data security and cloud storage

Cloud-based storage such as Google Drive, Dropbox, OneDrive or iCloud are easy to use, but not necessarily permanent or secure. Cloud-based storage is usually overseas and,

therefore, not subject to UK law, consequently its use could be in violation of the UK Data Protection Act 2018 (DPA) and/or the General Data Protection Regulation, which require that personal and sensitive data should not be transferred to other countries without adequate protection.

Cloud data storage should not be used for high-risk information such as files that contain personal or sensitive information or that have a very high intellectual property or commercial value. While file encryption safeguards data files to a certain degree, it does not negate the requirements of the DPA.

Alternatives are secure FTP (SFTP) servers, secure content management systems set up and controlled by an institution or secure workspaces.

Data Protection Legislation

Data protection and data security, although related, are not the same.

Data protection concerns: 1) personal data 2) data subjects

Personal data is information relating to an identifiable living individual. Whenever personal data is processed, collected, recorded, stored or disposed of, it must be done within the law. Personal data can also be 'sensitive', and falls under stricter protections.

Data subjects are the identified or identifiable living individual to whom the personal data relates.

Legislation

Legislation is one of the most important instruments of government in organising society and protecting citizens. It determines, amongst other, the rights and responsibilities of individuals and authorities to whom the legislation applies.

These laws help protect your data and information; they outline how organisations can carry out direct marketing; and they relate to how you can access information from public authorities.

THE EUROPEAN CONVENTION ON HUMAN RIGHTS (ARTICLE 8)

The European Convention on Human Rights 1953 (ECHR) is an international treaty that was designed to protect human rights and fundamental freedoms in Europe.

Article 8 states:

1 Everyone has the right to respect for his private and family life, his home and his correspondence.

There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for

the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.

It is important that our personal dignity and autonomy and how we interact with others, in both private and in public, is protected. Having respect for one's private life includes:

- Respect for an individual's sexuality
- The right to personal autonomy and physical and psychological integrity, ie the right not to be physically interfered with
- Respect for private and confidential information, particularly the storing and sharing of such information
- The right not to be subject to unlawful state surveillance
- Respect for privacy when one has a reasonable expectation of privacy; and the right to control the dissemination of information about one's private life, including photographs taken covertly

Article 8 also refers to family life and respect for the home. This is to ensure that unlawful surveillance is not carried out on individuals or their family members.

There are, as would be expected, restrictions to these rights. While they are afforded in most situations, they may be limited. These limitations, however, must be in accordance with law and for one or more of the following legitimate aims:

- The interests of national security
- The interests of public safety or the economic wellbeing of the country
- The prevention of disorder or crime
- The protection of health or morals
- The protection of the rights and freedoms of others

FREEDOM OF INFORMATION (SCOTLAND) ACT

The Freedom of Information (Scotland) Act 2002 (or 'FOISA') came into force on 1 January 2005. Under FOISA, a person who requests information from a Scottish public authority is entitled to have it. This right is subject to certain conditions and exemptions, which are set out in the act.




The Scottish Government considers this act as an essential part of a democratic and open government, and it applies to all local authorities, NHS, colleges and universities. The act also applies to companies owned by a public authority, and private companies that may be carrying out contract work for a public authority.

Exemptions from the Act

Not all information that is held by a public authority must be given out. There are some exemptions. The categories that would be exempt from a FOISA would be relating to matters such as national security, police investigations and the formation of government policy.

GENERAL DATA PROTECTION REGULATION 2018

The General Data Protection Regulation (GDPR) replaced the Data Protection Act (2008) in May 2018. It is designed to overhaul how businesses process and handle data. The new regulation applies in the UK and the EU. It is designed to give greater protection and rights to individuals and their data. The GDPR also applies to all companies worldwide that process personal data of European Union (EU) citizens. The Data Protection Act 2018 brought into force the requirements of the GDPR in the UK, and you can explore how this relates to data protection in other countries [here](#).

General Data Protection Regulation	Data Protection Act
	
<ul style="list-style-type: none">• Took effect: May 25, 2018• After Brexit applies to: EU Citizens in UK	<ul style="list-style-type: none">• Took effect: May 23, 2018• After Brexit applies to: UK Citizens
<p>All rights to personal data belong to citizens</p> <p>Age of consent: 16</p> <p>Criminals' personal data can only be handled by authorities</p> <p>Fixed fines for companies in violation</p> 	<p>Exceptions for data used for scientific or statistical purposes</p> <p>Age of consent: 13</p> <p>Criminals' personal data are not limited to authorities</p> <p>Unlimited fine for companies that identify individuals from anonymous data</p>

Personal and sensitive data is covered by GDPR. Personal data, as mentioned earlier, is any data or information that can identify a person. Sensitive data covers religious views, political views, genetic data, sexual orientation and more. These definitions can relate to any information that can be collected through automated processes, but the GDPR now also includes data that can be considered pseudonymised. This means data that is less

identifiable; for example, where a name is replaced with a unique number and, therefore, reduces concerns of data sharing and data retention.

Article 5 of the GDPR states the main principles of the act. Personal data shall be:

- Processed lawfully, fairly and transparently
- Only collected and used for lawful purposes
- Adequate, relevant and not excessive for that purpose
- Accurate and up to date
- Stored no longer than necessary
- Kept secure, and its integrity and confidentiality protected

There is also a new accountability principle, which means companies and organisations need to be able to demonstrate compliance with these principles.

Rights of the Individual

The GDPR states eight main rights that individuals have. These are:

- The right to be informed
- The right of access
- The right to rectification
- The right to erasure
- The right to restrict processing
- The right to data portability
- The right to object
- Rights relating to automated decision-making profiling

The threshold for consent to use individuals' data is much higher than in the previous Data Protection Act. Any request for consent must be very clear, easy to read and in a plain language that the data subject can understand. In practice, if consent is needed to collect and use your personal data, it is likely that an express opt-in will be the least that is required.

Consent:

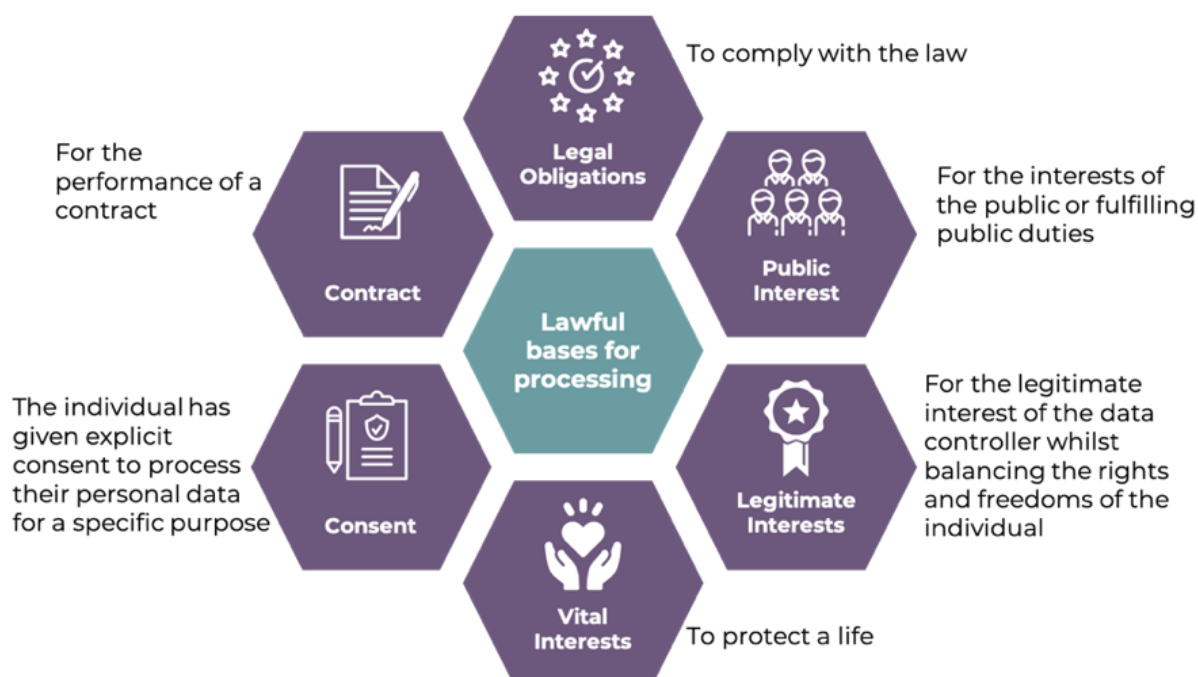
- Can be withdrawn at any time (and it must be easy to do)
- Is not always appropriate (eg, if data would be processed anyway)
- Won't be valid if there is too much of an imbalance of power (eg, if an employer asks an employee for consent, the employee may feel obliged to grant it even if they do not want to)
- Needs to remain valid (meaning it may need to be asked for again after a given period of time)

Controllers and Processors

The GDPR applies to companies that process and control personal data. It divides individuals and organisations into two categories — a controller, and a processor. The table below summarises the responsibilities of each under the GDPR.

Controller	Processor
Determines the purpose of the personal data being used.	Maintains records of personal data, and the processing activities carried out on that data.
Responsible for making sure the Processor is complying with the GDPR.	Needs to follow a specific set of legal obligations placed on them by the GDPR.
Has the legal liability and is responsible for any data breach.	In order to process data, there must be a legal reason to do so

The figure below shows the valid reasons for processing data and at least one must be in place to allow personal data to be processed. The lawful basis for processing the data should always be documented in the Privacy Notice.



Data stored outside Europe

Our data is stored across many countries. Cloud-based storage means that our personal data can now be stored on a server in a country outside of Europe. Countries outside of Europe may have very different laws to our own when it comes to keeping data secure. The GDPR states that data can only be transferred to a country, or international organisation, when an adequate level of protection is guaranteed.

As well as the legal obligations companies and organisations have regarding personal data, there are also the ethical considerations they must keep in mind when handling personal, and sometimes sensitive, data. This is covered in the next topic.

References

Council of Europe. European Convention for the Protection of Human Rights and Fundamental Freedoms, as amended by Protocols Nos. 11 and 14, 4 November 1950, ETS 5. Available at: <https://www.refworld.org/docid/3ae6b3b04.html>

Fan, W. and Geerts, F. 2012. Foundations of Data Quality Management. Synthesis Lectures on Data Management. Available at: <https://www.morganclaypool.com/doi/abs/10.2200/S00439ED1V01Y201207DTM030>

Kadadi, R. Agrawal, C. Nyamful and R. Atiq, "Challenges of data integration and interoperability in big data," *2014 IEEE International Conference on Big Data (Big Data)*, Washington, DC, USA, 2014, pp. 38-40, doi: 10.1109/BigData.2014.7004486.

ICO. 2021. What is the Freedom of Information Act 2000? Available at: <https://ico.org.uk/for-organisations/guide-to-freedom-of-information/what-is-the-foi-act/>

Roman, V. 2019. Data Management Strategy: Introduction. Towards Data Science. Available at: <https://towardsdatascience.com/data-management-strategy-d3ce6db599c1>

Scottish Government. Freedom of Information (Scotland) Act 2002. Available at: <https://www.legislation.gov.uk/asp/2002/13/contents>

Storey, V. and Song, Y. 2017. Big data technologies and Management: What conceptual modeling can do. *Data & Knowledge Engineering*. 108, 57-67.

Stranieri, S. Global Data Privacy Laws: US, EU, China And More. Progress. Available at: <https://blog.ipswitch.com/global-data-privacy-laws-us-eu-china-and-more>

Tellenbach et al. (2018) Security of Data Science and Data Science for Security, in Braschler, Stadelmann, Stockinger (Eds.). *Applied Data Science - Lessons Learned for the Data-Driven Business*. Amsterdam: Springer.

UK Government. 2018. Data Protection Act 2018. Available at: <https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted>

Ziegler, P. and Dittrich, K. 2007. Data Integration — Problems, Approaches, and Perspectives. In: Krogstie, J. et al. (eds). *Conceptual Modelling in Information Systems Engineering*. Berlin: Springer. (pp.39-58).