Data Lifecycle, Quality & Ethics, Security & Legislation Presentation

SCDV45: CW1

Callum Richards

Data Lifecycle

Data Lifecycle: The stages that data goes through from its creation and collection to its eventual disposal.

Key stages of the data lifecycle include:

- Creation and Collection
- Storage
- Processing
- Analysis
- Disposal

Examples of data lifecycle models include:

- The DAMA-DMBOK lifecycle model
- The CRISP-DM lifecycle model

Fundamentals of data storage in databases

Database: An organised collection of structured information stored electronically.

Types of database:

- Relational (SQL)
- Non-Relational (NoSQL)

Database Management System (DBMS) examples:

- MySQL (Relational)
- MongoDB (Non-Relational)

Database structuring and normalisation.

Maintaining data quality challenges

Data Quality: How accurate, complete, reliable, and relevant the data is for its intended purpose.

Potential challenges when maintaining quality data can include:

- Inaccurate, incomplete or inconsistent data
- Duplicate data
- Poor data governance

Using poor data quality can have a variety of negative impacts on an organisation, making it highly important to maintain quality.

Ethical considerations

Handling data comes with ethical responsibilities that must be considered to build trust between users and organisations.

These considerations include:

- Privacy and Protection
- Transparency
- Security

The Cambridge Analytica Scandal is a real-world case where ethics were not considered, causing concerns about privacy of data.

Preventative and mitigation strategies for data quality and ethical issues

Possible methods to ensure data quality:

- Implement data validation methods
- Commit regular monitoring of data and audits

Possible methods to ensure ethical practices:

- Implement clear data ethics policies
- Transparency and user empowerment

Security policies and their impact on data

Implementing security policies will increase the security of data and prevent possible cyber-attacks compromising information.

Security policies that can be implemented include:

- Access control
- Data encryption
- Multi-factor authentication and passwords

Data management legislation

Data management legislation are laws that data holders must comply with when collecting and storing personal information.

Legislation examples:

- General Data Protection Regulations (GDPR)
- Data Protection Act 2018

Breaking the regulations can lead to investigations from the Information Commissioner's Office (ICO) which can then result in further legal sanctions and large fines.

Potential policies to sustain data quality, ethical standards and security

Employ a data governance framework: A set of rules that ensures that data is collected and processed responsibly. Also helps in maintaining data quality.

Employ staff training: Educating staff on best practises to keep data storage and management ethical, legal and secure.

Security and incident response: Implementing MFA systems, data encryption and incident response plans to prevent and mitigate issues arising from security breaches and system errors.

References

Ataccama (2024) *Ataccama ONE Platform*, *Ataccama*. Available at: https://www.ataccama.com/platform (Accessed: 7 February 2025).

Caldwell, R. (2023) *Liquid Web*, *Liquid Web*. Available at: https://www.liquidweb.com/blog/database-audit/ (Accessed: 7 February 2025).

Gov.uk (2018) *Data Protection Act*, *GOV.UK*. www.gov.uk. Available at: https://www.gov.uk/data-protection (Accessed: 8 February 2025).

Government Data Quality Hub (2021a) *Hidden costs of poor data quality*, *GOV.UK*. Available at: https://www.gov.uk/government/news/hidden-costs-of-poor-data-quality (Accessed: 6 February 2025).

Government Data Quality Hub (2021b) *What is data quality?*, *GOV.UK*. Available at: https://www.gov.uk/government/news/what-is-data-quality (Accessed: 6 February 2025).

OptimizeMRO (2024) *DAMA-DMBOK: A Comprehensive Framework for Data Management*, *OptimizeMRO*. Available at: https://www.optimizemro.com/blog/dama-dmbok-a-comprehensive-framework-for-data-management/ (Accessed: 4 February 2025).

Patel, A. (2016) *Introduction of DBMS (Database Management System) | Set 1 - GeeksforGeeks*, *GeeksforGeeks*. Available at: https://www.geeksforgeeks.org/introduction-of-dbms-database-management-system-set-1/ (Accessed: 5 February 2025).

Saltz, J.S. (2021) 'CRISP-DM for Data Science: Strengths, Weaknesses and Potential next Steps', *2021 IEEE International Conference on Big Data (Big Data)* [Preprint].

Schneble, C.O., Elger, B.S. and Shaw, D. (2018) 'The Cambridge Analytica Affair and Internet-mediated Research', *EMBO reports*, 19(8). Available at: https://doi.org/10.15252/embr.201846579.

Wolford, B. (2018) *What is GDPR, the EU's new data protection law?*, *GDPR.eu*. Available at: https://gdpr.eu/what-is-gdpr/ (Accessed: 8 February 2025).