Data Management

Dr Emma Murphy

Week 1, January 2021

OLLSCOIL TEICNEOLAÍOCHTA
BHAILE ÁTHA CLIATH
TU DUBLIN
TECHNOLOGICAL
UNIVERSITY DUBLIN

What is this module about?

The aim of this module is to **analyse** and **evaluate** the role of data management in an organisation or sector and the various **roles, processes, tools, techniques** and **requirements** that are involved in the data management function.

About me

- Teaching Areas: Databases, Data Management, UX Design
- Research Interests: Digital Health, Behaviour Change, Inclusive Design
- Previously worked as a Research Fellow at TCD
- Postdoctoral researcher in DCU and McGill University

Data Management

- Data is becoming a valuable asset for companies
- But the management of data is subject to huge challenges including meeting data **regulatory**, ethical, **legal** and **governance** requirements, maintaining **data privacy** and **ensuring the quality** of the data throughout the data lifecycle and over time.

Data Management

- No enterprise can be effective without high quality data. Today's organisations rely on their data assets to make more informed and more effective decisions.
- Leading organisations are using their data assets to create competitive advantages through greater knowledge of their customers, innovative uses of information and operational efficiencies.

Learning Objectives

- Demonstrate cognisance of the roles and responsibilities of data management stakeholders and be able to critically evaluate their significance in relation to relevant EU and International legislation.

- Analyse the critical issues for data protection and data governance in public and private sector organisations across different domains such as health, education, transport and energy.

Learning Objectives

- Identify the role data quality plays, discuss deficits and limitations and explain appropriate remedies

- Contribute to key ethical debates in data science and machine learning (i.e. data quality, bias in data, informed consent and privacy issues)

Learning Objectives

- Design and implement data balancing and fairness strategies to overcome bias in data
- Critically analyse different GDPR functions and responsibilities for organisations of various sizes
- Critically analyse the impact of data privacy needs on organisations and individuals

Learning Outcomes

1. Demonstrate cognisance of the roles and responsibilities of data management stakeholders and be able to critically evaluate their significance in relation to relevant EU and International legislation.
2. Analyse the critical issues for data protection and data governance in public and private sector organisations across different domains such as health, education, transport and energy.
3. Identify the role data quality plays, discuss deficits and limitations and explain appropriate remedies

Learning Outcomes

4. Contribute to key ethical debates in data science and machine learning (i.e. data quality, bias in data, informed consent and privacy issues)
5. Design and implement data balancing and fairness strategies to overcome bias in data

Learning Outcomes

6. Critically analyse different GDPR functions and responsibilities for organisations of various sizes
7. Critically analyse the impact of data privacy needs on organisations and individuals
8. Design and implement anonymisation strategies and examine related issues

Learning Outcomes

9. Design a data management plan for a complex multistakeholder prediction tool in the field of digital health.
10. Conduct a Data Protection Impact Assessment (DPIA)

Where is the module material?

- All class material will be available on Brightspace

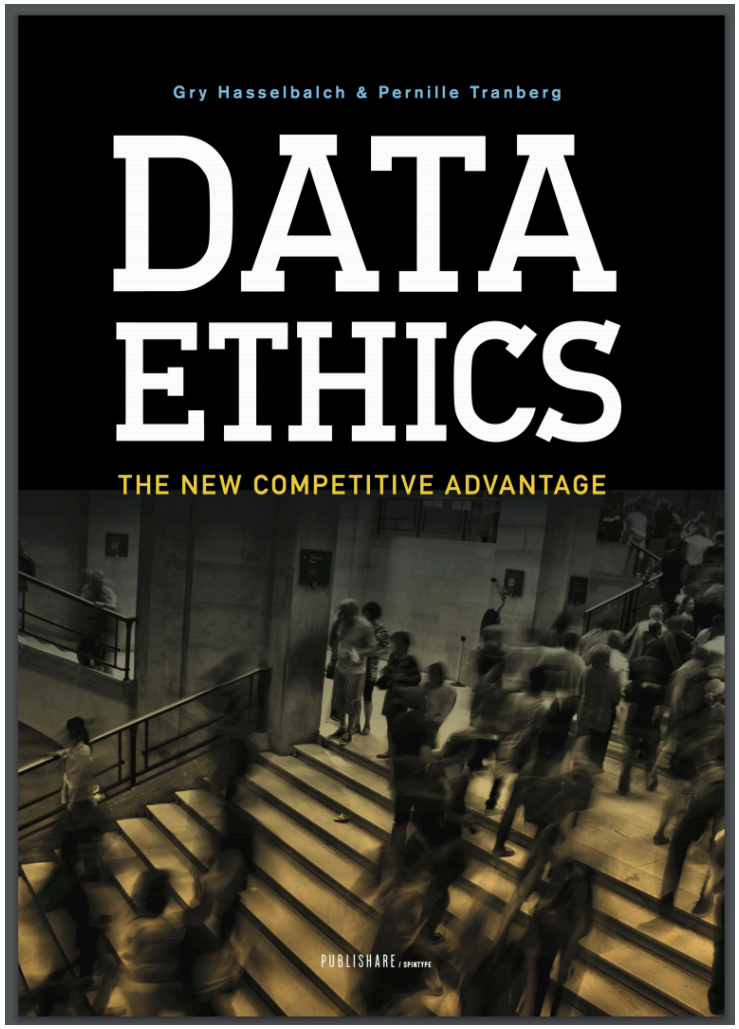
Module Name

- Data Management DATA9911: 2019-20

Module Code

- DATA991133611TU059-1920

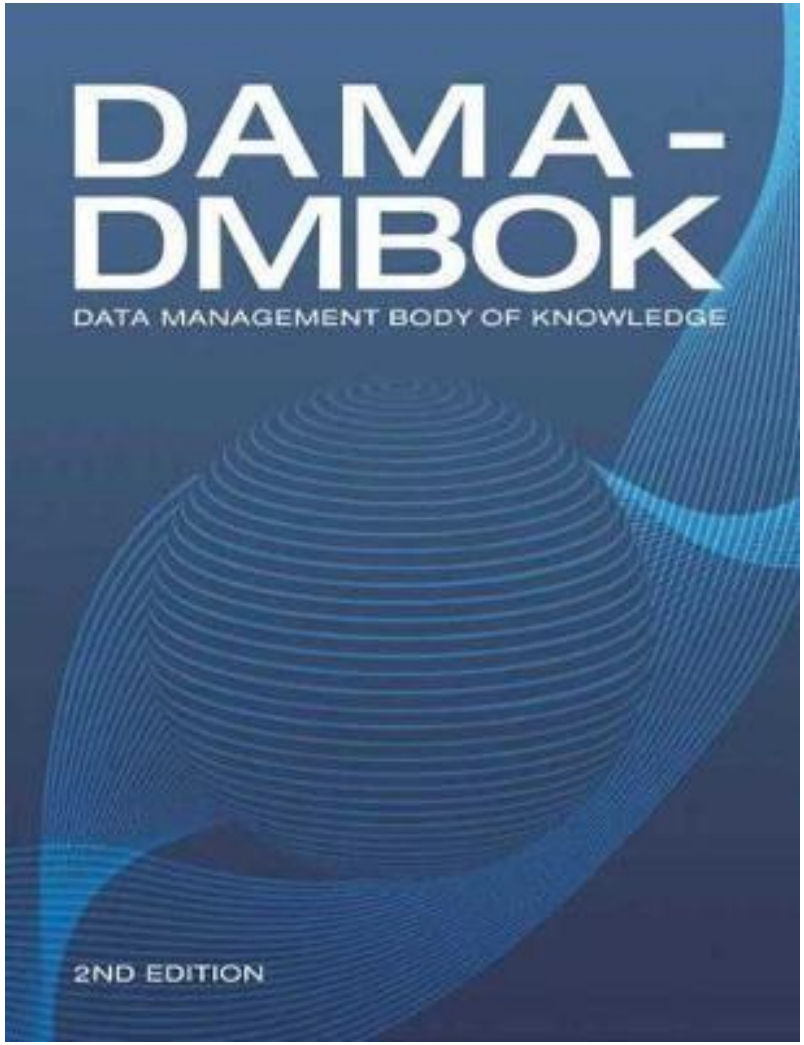
Key Texts



Hasselbach, G and Tranberg, P (2017) Data Ethics: The new competitive advantage, PubliShare, Copenhagen.

Available: https://dataethics.eu/wp-content/uploads/DataEthics-UK-original.pdf

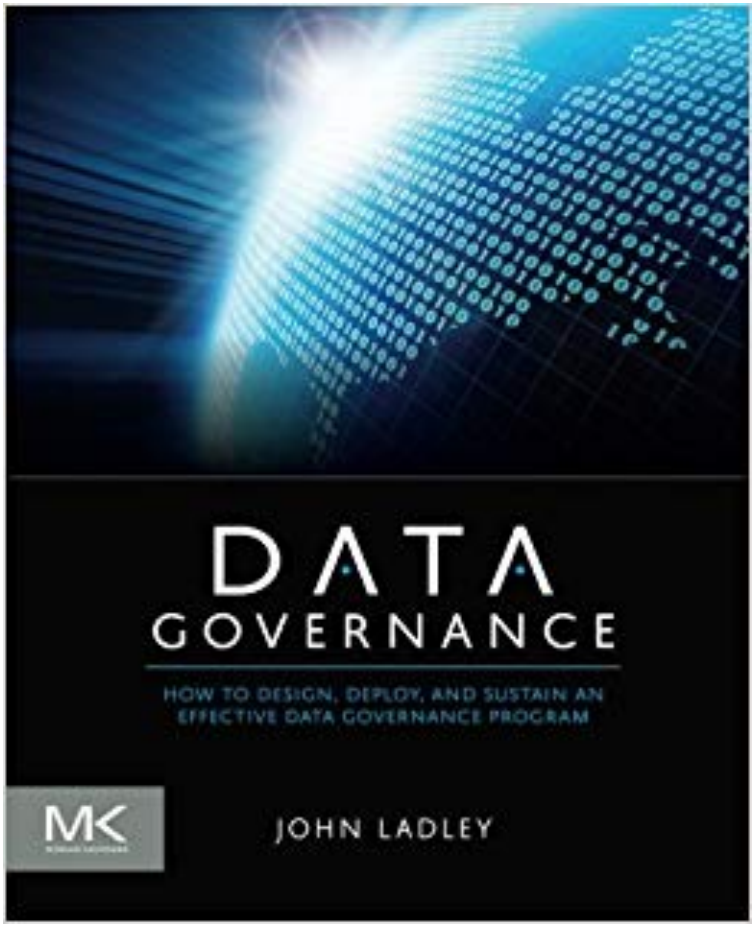
Key Texts



DAMA International (2017) DAMA DMBOK, DAMA DMBOK – Data Management Body of Knowledge, Technics Publications, New Jersey, pp 381–85

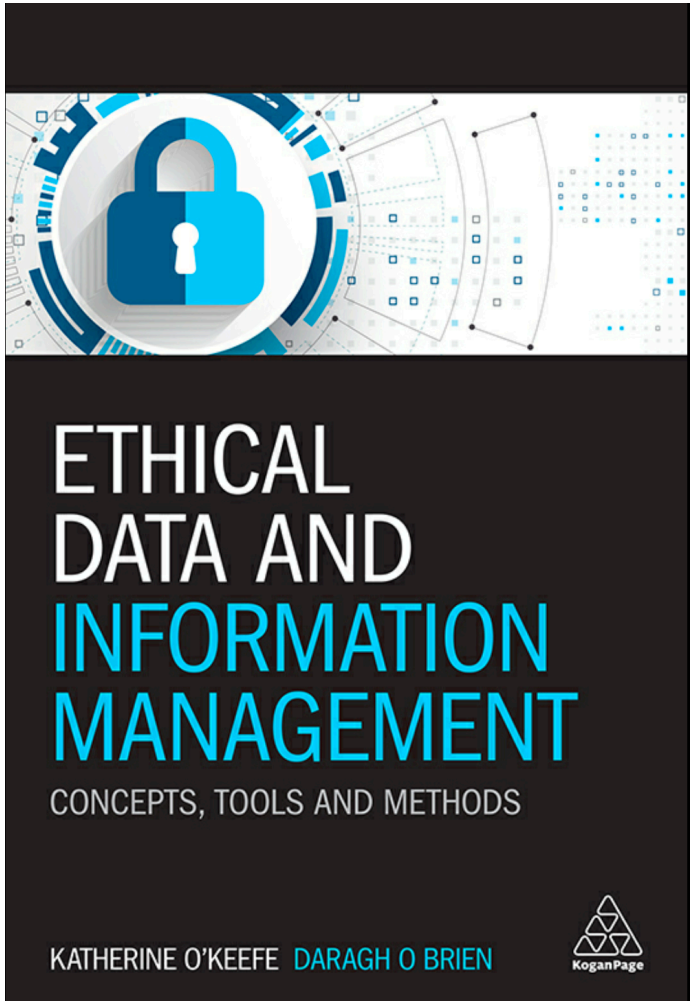
- Will shortly be available through the library in print

Key Texts



Ladley, J. (2018) *Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program*, Academic Press. 2nd Edition.

Key Texts



O'Keefe, K. and O'Brien, D. (2018) *Ethical Data and Information Management: Concepts, Tools and Methods* (1st. ed.). Kogan Page Ltd., GBR.

- Will shortly be available through the library in as an e-book

Contact Me

- Best way to contact me
  - Email: emma.x.murphy@tudublin.ie
  - Or talk to me during class

Assessment

- Assignment 1 (50%) - Week 5/7
- Research and Presentation = formative (not marked)– Week 5/6
- Individual Report = 50% of module mark – Week 7

- Assignment 2 (50%) – Week 12
- 2-part report based on case study

Course Overview

Week	Date (week starting)	Lecture
1	28/01/2021	Introduction and Overview
2	04/02/2021	Ethical Tools and Frameworks
3	11/02/2021	Data Management and the data lifecycle
4	18/02/2021	Data Governance
5	25/02/2021	Data Privacy
6	04/03/2021	Data Quality
7	11/03/2021	Ethics and Bias in Data
8	18/03/2021	Data Law and Regulation
9	25/03/2021	Data Security
		Easter Break
10	15/04/2021	Data Protection Impact Assessment
11	22/04/2021	Future Opportunities
12	29/04/2021	Revision and CA Support

Data is an asset

- “In today's digital infrastructure, data has become a company asset. It has a status similar to that of oil, steel and railways during the Industrial Revolution”. (Hasselbach and Tranberg, 2017)

Data Management Definitions

- ”The professional discipline of data management addresses the challenges of managing data and information as an enterprise asset, to better deliver value to an organization and its stakeholders. As with other asset management disciplines, this requires considering managing the data asset throughout the life cycle and considering its proper handling, from planning for acquiring or creation of the data through its maintenance and use and into its disposition once its purpose is concluded.”

- (O'Keefe and O’Brien, 2018)

Data Management - Definitions

- The Data Management Body of Knowledge (DMBOK) refers to Data Management as:

- “The development, execution, and supervision of plans, policies, programs, and practices that deliver, control, protect, and enhance the value of data and information assets throughout their lifecycles.” (DAMA, 2017)

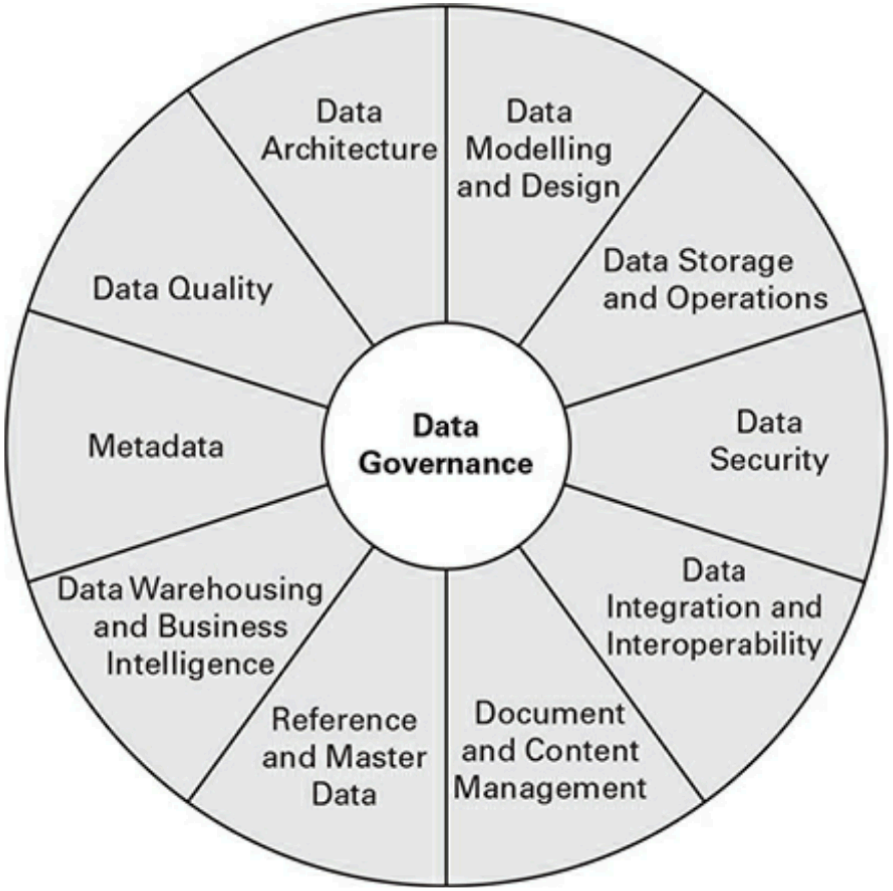
What Data Management is not

- DM is not Database Administration
- DM is not Database Design
- DM is not IT support
- DM is not Application Development
- DM is not Data Mining

But....

- If you are working in any of these roles (or with data) you need to understand that data management includes:
- The disciplines of development, execution, and supervision of plans, policies, programs, projects, processes, practices and procedures that control, protect, deliver, and enhance the value of data and information assets (DAMA, 2017)

Data Management Functions



Activity

- Take two minutes each to (re-)introduce yourselves to the class, and discuss any background or experience you have in the Data Management domain.

Data Law and Regulation

- “Rules of law of a society reflect it’s values” (Lasprogata, 2019)

- For example GDPR values privacy

Data Protection Day!

- The 28th January is the 40th Anniversary of Convention 108, "Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data" which was opened for signature on the same day in 1981.

- https://www.coe.int/en/web/data-protection/data-protection-day

GDPR – A reminder upfront!

- General Data Protection Regulation (GDPR) came into force across the EU on 25 May 2018 is a European Union-wide framework
- An accompanying Directive establishes data protection standards in the area of criminal offences and penalties. This is known as the law enforcement Directive.

https://www.citizensinformation.ie/en/government_in_ireland/data_protection/overview_of_general_data_protection_regulation.html

GDPR

- The GDPR and the law enforcement Directive provide for significant reforms to current data protection rules.
- They provide for higher standards of data protection for individuals and impose increased obligations on organisations that process personal data. They also increase the range of possible sanctions for infringements of these rules.

https://www.citizensinformation.ie/en/government_in_ireland/data_protection/overview_of_general_data_protection_regulation.html

GDPR – Personal data

- Under the GDPR, *personal data* is data that relates to or can identify a living person, either by itself or together with other available information. Examples of personal data include a person's name, phone number, bank details and medical history.
- A *data subject* is the individual to whom the personal data relates.
- Organisations that collect or use personal data are known as *data controllers* and *data processors*.

https://www.citizensinformation.ie/en/government_in_ireland/data_protection/overview_of_general_data_protection_regulation.html

GDPR – Sensitive Personal Data

- *Special category personal data* (known as *sensitive personal data* under previous Irish legislation) means personal data relating to any of the following:
  - The data subject's racial or ethnic origin, their political opinions or their religious or philosophical beliefs
  - Whether the data subject is a member of a trade union
  - The data subject's physical or mental health or condition or sexual life
  - Whether the data subject has committed or allegedly committed any offence
  - Any proceedings for an offence committed or alleged to have been committed by the data subject, the disposal of such proceedings or the sentence of any court in such proceedings

https://www.citizensinformation.ie/en/government_in_ireland/data_protection/overview_of_general_data_protection_regulation.html

GDPR

- The processing of special category data is prohibited unless the data subject has given their explicit consent before processing begins or the processing is authorised by law, for example, to protect the interests of a data subject, to comply with employment legislation or for reasons of public interest.
- Personal data relating to criminal convictions and offences may only be processed under the control of an official authority.

https://www.citizensinformation.ie/en/government_in_ireland/data_protection/overview_of_general_data_protection_regulation.html

GDPR Penalties

Serious infringements

- For the most serious infringements (for example, not having sufficient customer consent to process data or violating the core of privacy by design concepts) organisations can be fined up to 4% of their annual global turnover or €20 million, whichever is greater.

https://www.citizensinformation.ie/en/government_in_ireland/data_protection/overview_of_general_data_protection_regulation.html

GDPR Penalties

Lesser breaches

- Under the GDPR, organisations in breach of the Regulation can be fined up to 2% of their annual global turnover or €10 million, whichever is greater, for lesser breaches. Some examples of lesser breaches include: not having records in order, not notifying the supervisory authority and data subject about a breach or not an conducting impact assessment.

https://www.citizensinformation.ie/en/government_in_ireland/data_protection/overview_of_general_data_protection_regulation.html

Law and Technology

- “Law likes what is physical and present within boundaries of states, regions, and nations. It likes less what is intangible, such as that which is digital” (Lasprogata, 2019)

Law and Technology

- “Intensifying this misalignment is the fact that technology evolves at a rapid pace. Law, on the other hand, moves as slow as a snail. It is always catching up to technology” (Lasprogata, 2019)

Human Rights vs. Organisational Rights

- Legal tension that exists in this space with respect to human rights and organizational rights.

- There are many organizational players to consider that engage human data in different contexts, creating legal questions almost at every turn.
- (Lasprogata, 2019)

Examples of legal tensions

- If a person believes he was discriminated against as a result of algorithmic failure, will the law demand transparency of that algorithm to prevent future bias, or will the organization's intellectual property rights in the algorithm win?
- (Lasprogata, 2019)

Ethics

- At its simplest, ethics is a system of moral principles. They affect how people make decisions and lead their lives.
- Ethics is concerned with what is good for individuals and society and is also described as moral philosophy.
- The term is derived from the Greek word *ethos* which can mean custom, habit, character or disposition.

http://www.bbc.co.uk/ethics/introduction/intro_1.shtml

Ethics and Computer Science

- What does ethics have to do with the discipline of computer science?

Quite a lot…

- Bias and AI
- Facial recognition
- Censorship
- Rapid dissemination of misinformation AKA Fake news!

Ethics4EU

Search Twitter

#ETHICS4EU

Follow

Ethics4EU

@Ethics4EU

Ethics4EU is an Erasmus+ project which aims to develop new curricula, best practices and learning resources for digital ethics for computer science students.

sites.google.com/dit.ie/ethics4... Joined August 2019

503 Following **218** Followers

Ethics – key values

- What does it meant to be an ethical person?
- Values based on the well-being of others
  - (not causing) Suffering
  - Autonomy
  - Equality
- Values based on my own well-being
  - Virtue/Character Excellence
  - Trust

(Coloner, 2019)

Data Ethics

- Ethics = the study of human well-being
- Data ethics = the study of how data affects human well-being (positive and negative)
- (Coloner, 2019)

https://www.linkedin.com/learning/ethics-and-law-in-data-analytics/ethical-foundations-part-3?u=56744273

Data Ethics

- “Data is an asset, but it's also a risk.”

(Hasselbach and Tranberg, 2017)

Data ethics

- The tools and technologies we now have at our disposal have the potential to bring benefit or cause harm to people.
- All information that is processed, with very few exceptions, impacts on people in some way.
- the outcomes that result from the processing of data impact on people through an impact on privacy, a potential for bias in decision making, and an impact on governmental policies and climate change investments.

(O'Keefe and O'Brien, 2018)

Data impacts people

- Information that allows us to identify a person and make a determination about their eligibility for a loan,
- Or information that trains an artificial intelligence system that provides sentencing recommendations to judges,
- Or whether it is information about the performance of a car's engine in environmental impact tests,
- (O'Keefe and O'Brien, 2018)

Cambridge Analytica Scandal

- In 2018 the Observer and The Guardian broke the story that became the Cambridge Analytica scandal.
- It was the result of a year-long investigation in which reporter Carole Cadwalladr worked with ex-employee turned whistleblower Christopher Wylie to reveal how the data analytics firm that was behind Trump's 2016 campaign and played a role in Brexit, had used the data harvested from 87 million Facebook users without their consent.

https://www.theguardian.com/

Cambridge Analytica Scandal

- Cadwalladr's reporting led to the downfall of Cambridge Analytica and a public apology from Facebook's Mark Zuckerberg who was forced to testify before congress. Facebook has since lost $120 billion from its share price.

https://www.theguardian.com/

Cambridge Analytica Scandal

- https://www.theguardian.com/news/video/2018/mar/19/everything-you-need-to-know-about-the-cambridge-analytica-expose-video-explainer

- https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

Going beyond the legislation

- “Ethical companies in today's big data era are doing more than just complying with data protection legislation. They also follow the spirit and vision of the legislation by listening closely to their customers.” (Hasselbach and Tranberg, 2017)

Data ethics awareness

- “A company's degree of 'data ethics awareness' is not only crucial for survival in a market where consumers progressively set the bar, it's also necessary for society as a whole. It plays a similar role as a company's environmental conscience – essential for company survival, but also for the planet's welfare.
- …Yet there isn't a one-size-fits-all solution, perfect for every ethical dilemma. We're in an age of experimentation where laws, technology and, perhaps most importantly, our limits as individuals are tested and negotiated on a daily basis.”

(Hasselbach and Tranberg, 2017)

Activity

- What kinds of data-processing activity are you or your organizations * engaging in that could potentially raise ethical concerns or questions?

*If you cannot think of a work place example think of your experience as a student in the university, or in your interactions with commercial organisations as a user or consumer

Activity

- If something in your organisation's approach to managing data raised an ethical concern for you, how would you express that and who would you express it to?

Who is responsible

- Data management must be a shared responsibility across the organization
- Data Management Professionals in IT - Technical Aspects

- Business Data Stewards - Quality control and usage aspects of DM

Data Governance

- The DAMA DMBOK defines data governance as ‘the exercise of and authority and control (planning, monitoring and enforcement) over the management of data assets’ (DAMA International, 2017).

Next Week

- Ethical Frameworks

- Reading: Introduction and Chapter 1 of:

Hasselbach, G and Tranberg, P (2017) Data Ethics: The new competitive advantage, PubliShare, Copenhagen. Available: https://dataethics.eu/wp-content/uploads/DataEthics-UK-original.pdf

Also to think about….

- Over the next week think about work or social tasks that you undertake that require you to collect, process or provide data.
- Any ethical concerns?
- Any risks?

References

DAMA International (2017) DAMA DMBOK, DAMA DMBOK – Data Management Body of Knowledge, Technics Publications, New Jersey, pp 381–85

O'Keefe, K. and O'Brien, D. (2018) Ethical Data and Information Management: Concepts, Tools and Methods (1st. ed.). Kogan Page Ltd., GBR.

Hasselbach, G and Tranberg, P (2017) Data Ethics: The new competitive advantage, PubliShare, Copenhagen.

Lasprogata, E. (2019) *Ethics-and-law-in-data-analytics*, video recording, Linkedin Learning, https://www.linkedin.com/learning/ethics-and-law-in-data-analytics/welcome?u=56744273

Coloner, B. (2019) *Ethics-and-law-in-data-analytics*, video recording, Linkedin Learning, https://www.linkedin.com/learning/ethics-and-law-in-data-analytics/welcome?u=56744273