



THE UNIVERSITY OF  
MELBOURNE

**PREVIEW ONLY.  
NOT FINAL SLIDES.  
PLEASE COME TO THE LECTURE TO TAKE PART IN  
THE 'FLIPPED CLASSROOM' ACTIVITIES.**

*Week 9/S1/2023*

# Data Governance

**Marc Cheong**

School of Computing and Information Systems  
Centre for AI & Digital Ethics  
The University of Melbourne  
marc.cheong [at] unimelb.edu.au



*what utilitarianism is  
maximum X, minimize 2. context of use*



THE UNIVERSITY OF  
MELBOURNE

## Learning Outcomes

(4) Arguments people (users) can get more utility

1. Define the concept of data governance from an organisational perspective.
2. Understand how laws such as the famous GDPR (EU) help protect users.
3. Discover how your data is stored and managed by the organisations you entrust them with (and also organisations you never knew had access to), as well as your rights.
4. Critically evaluate research and storage practices on large amounts of data.

From utilitarianism → less friction → lower maintenance costs  
save money → remove the use and throw culture  
remove the use and throw culture  
more job opportunities → phone repair specialist  
cheaper than buy new phone

(5) Argument 3: Environmental

cheaper, lower maintenance costs  
Cheaper than my new one

This material has been reproduced and communicated to you by or on behalf of the University of Melbourne pursuant to Part VB of the Copyright Act 1968 (the Act).

The material in this communication may be subject to copyright under the Act.

Any further copying or communication of this material by you may be the subject of copyright protection under the Act.

Do not remove this notice



## Related Reading

This module on Data Governance has two readings, with very interesting (and relevant) contemporary studies which encompass data science, policy, and ethics.

Data and its (dis)contents: A survey of dataset development and use in machine learning research.

Amandalynne Paullada, Inioluwa Deborah Raji, Emily M. Bender, Emily Denton, Alex Hanna. *arXiv [cs.LG]*, 9 Dec. 2020.

<https://arxiv.org/abs/2012.05345>

What if Facebook goes down? Ethical and legal considerations for the demise of big tech.

Carl Öhman, Nikita Aggarwal. *Internet Policy Review* 9(3), 1-21, 2020.

<https://policyreview.info/articles/analysis/what-if-facebook-goes-down-ethical-and-legal-considerations-demise-big-tech>



# Outline

1. What is data governance? The ‘textbook definition’.
2. Data governance in practice. The ‘practitioner’s view’.
3. From organisations to users.
4. Laws that protect us (GDPR at a glance).
5. Organisations: ‘*A Tale of Beacons and Pixels*’.
6. Users: Your rights, freedoms, and countermeasures.
7.  *Case Study & Reflection:* A new problem - Generative AI and large language/image models
8.  *Case Study & Reflection:* Does your data outlive you/the tech giants?
9.  *Case Study & Reflection:* A ‘set’ of data versus a dataset.



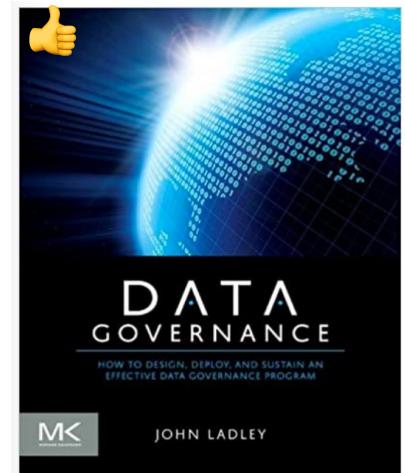
## The ‘textbook’ definition #1

To appreciate the various ethical issues related to data governance, we first need to find out what it means!

"Data governance represents the program used by [an organisation] to manage the organi[s]ational bodies, policies, principles, and quality that will ensure access to accurate and risk-free data and information..."

- John Ladley (2012). *Data Governance: How to Design, Deploy and Sustain an Effective Data Governance Program (The Morgan Kaufmann Series on Business Intelligence) 1st Edition.* Morgan Kaufmann.

Image source: Amazon/Morgan Kaufmann





# The ‘textbook’ (industry) definition #2

A screenshot of a CIO Australia website article. The header includes the CIO Australia logo, a navigation bar with links to Leadership, The CIO Show, CIO Summit, CIO50 Awards, and Insider, and a search icon. The main content area shows a breadcrumb trail "Home > Data Management", a "FEATURE" section, and a large title: "What is data governance? A best practices framework for managing data assets". Below the title is a brief description: "Data governance defines roles, responsibilities, and processes for ensuring accountability for and ownership of data assets across the enterprise." A circular profile picture of Thor Olavsrud is shown next to his name and title: "Senior Writer, CIO | 11 FEBRUARY 2020 22:00 AEDT".

Home > Data Management

FEATURE

## What is data governance? A best practices framework for managing data assets

Data governance defines roles, responsibilities, and processes for ensuring accountability for and ownership of data assets across the enterprise.

By **Thor Olavsrud**  
Senior Writer, CIO | 11 FEBRUARY 2020 22:00 AEDT

“The Data Governance Institute defines data governance as “a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods.” ...

“The Data Management Association (DAMA) International defines data governance as the “planning, oversight, and control over management of data and the use of data and data-related sources.” ...  
– Olavsrud (2020)

Image source: CIO Magazine

PREVIEW ONLY. NOT FINAL SLIDES.



# Data governance in practice. *The 'practitioner's view'.*

Image by DALLE-2



# What tasks are there in data governance?

Let's see what a Data Governance Officer does at UniMelb

<https://secure.dc2.pageuppeople.com/apply/TransferRichTextFile.ashx?sData=Fwg6i4Eli-DgfNqV-PbYvXeEOtAs0yXwPaCK7AQYFAk1mpHE7B7I4z2iK7fPjnJtkHDeTmWnOXw~>

## ***Key Responsibilities***

- ▶ Define and document a data governance framework (people, process and tools) and implementation plans.
- ▶ Contribute to the on-going implementation of the data governance framework across the University.
- ▶ Provide consultation on governance and data management practices across the stakeholder community within the University.
- ▶ Develop data management processes in accordance with the data governance framework.



# What tasks are there in data governance?

Let's see what a Data Governance Officer does at UniMelb

<https://secure.dc2.pageuppeople.com/apply/TransferRichTextFile.ashx?sData=Fwg6i4Eli-DgfNqV-PbYvXeEOtAs0yXwPaCK7AQYFAk1mpHE7B7I4z2iK7fPjnJtkHDeTmWnOXw~>

- ▶ Drive the implementation of data management processes across the University e.g. data stewardship.
- ▶ Monitor and review data assets and procedural controls in order to continually improve data management practices.
- ▶ Report on implementation of data governance framework and compliance across the University.



From  
organisations to  
users.

Image by DALLE-2



# USers

You might have noticed the pun there – we are users of many services run by many different organisations.

Social media?

- Twitter/Facebook/TikTok...

Search giants?

- Google...

Device/platform manufacturers?

- Apple/Google...



Also, we may be stakeholders even without consciously knowing!

Image source: Medium / @storjproject / lantechsystems

PREVIEW ONLY. NOT FINAL SLIDES.



# Our relationship with tech orgs. (us ❤️ tech?)

Point to ponder: is this relationship symmetrical?

- Consider, say, your relationship with a small business vendor selling coffee, and paying cash. The data you share with the vendor is just a coffee loyalty card on paper, your usual orders, etc. The vendor might use the data for e.g. how many total loyalty cards were used in a week to predict sales, your usual orders to personalise their service to you/thank you for being a loyal customer, etc.
- Your only records are:



Image source: Designwizard.com

PREVIEW ONLY. NOT FINAL SLIDES.



# Our relationship with tech orgs. (us 💔 tech!)

Some scholars call for a ‘duty’ (a la Immanuel Kant’s philosophy) for us to thwart microtargeting.

JOURNAL OF MEDIA ETHICS  
2018, VOL. 33, NO. 3, 133–148  
<https://doi.org/10.1080/23736992.2018.1477047>



## Social networks, the 2016 US presidential election, and Kantian ethics: applying the categorical imperative to Cambridge Analytica’s behavioral microtargeting

Ken Ward

Department of Communication, Lamar University, Beaumont, Texas, USA

**Reflection:** Are we truly helpless as we constantly depend on big tech to stay connected, conduct business transactions, learn and study, especially during the Covid pandemic?

PREVIEW ONLY. NOT FINAL SLIDES.



Image source: The Pokemon Company

## Audience activity III [5-10 mins]

Let's help you practice  
philosophical argument!

SIMON USES PHILOSOPHICAL ARGUMENT.





# Laws that protect us.

*(& GDPR at a glance).*

Image by DALLE-2





# Disclaimer: I am not a lawyer



The information provided in this mini-lecture is summarized from various sources to explain how laws such as the GDPR and local laws protect users (at a glance).

This lecture won't make you an expert in the GDPR 😊

Image source: CAPCOM / Cinemablend.



# Remember these?

These GDPR banners seem to have popped up everywhere after mid-2018...

A composite image showing two examples of GDPR cookie consent banners. The top banner is from Cookiebot, featuring a cookie icon, the text "THIS WEBSITE USES COOKIES", and a detailed explanation of how cookies are used for personalisation, advertising, and analytics. It includes buttons for "Necessary cookies", "Allow selection", and "Allow all cookies". The bottom banner is from JQueryScript, titled "Cookies &amp; Privacy", stating that the website uses cookies to ensure the best experience. It includes "More information" and "Accept" buttons.

Image source: Cookiebot / JQueryScript

PREVIEW ONLY. NOT FINAL SLIDES.

~~highly examinable~~



# The EU General Data Protection Regulation (GDPR)

Read more here:

<https://gdpr.eu/what-is-gdpr/>

The cookie popups?

“The data subject gave you specific, **unambiguous consent** to process the data.”

-- summary by Proton Technologies AG (2021)

Also: EU ePrivacy Directive  
(<https://www.cookiebot.com/en/cookie-law/>)

**GDPR.EU** Home Checklist FAQ GDPR News & Upd

## Data protection principles

If you process data, you have to do so according to seven protection and accountability principles outlined in [Article 5.1-2](#):

1. **Lawfulness, fairness and transparency** — Processing must be lawful, fair, and transparent to the data subject.
2. **Purpose limitation** — You must process data for the legitimate purposes specified explicitly to the data subject when you collected it.
3. **Data minimization** — You should collect and process only as much data as absolutely necessary for the purposes specified.
4. **Accuracy** — You must keep personal data accurate and up to date.
5. **Storage limitation** — You may only store personally identifying data for as long as necessary for the specified purpose.
6. **Integrity and confidentiality** — Processing must be done in such a way as to ensure appropriate security, integrity, and confidentiality (e.g. by using encryption).
7. **Accountability** — The data controller is responsible for being able to demonstrate GDPR compliance with all of these principles.



# The EU General Data Protection Regulation (GDPR)

Amongst other things, any “organization that processes the personal data of people in the EU must comply with the GDPR...” (<https://gdpr.eu/faq/>)

(all quotes below are from <https://gdpr.eu/checklist/> )

- “data protection is something you now have to consider whenever you do anything with other people's personal data”, not just an afterthought
- “People have the right to see what personal data you have about them and how you're using it”  
Though in practice, this is tricky...
- “You are also required to quickly communicate data breaches to your data subjects” and not just keep quiet about it!
- *NB: A ‘right’ to explainability – this is the subject of many scholarly works in legal studies and digital ethics. (Our very own Prof Tim Miller will be able to provide the technical details about what it means to have explainable AI)*



# In Australia...



<https://www.oaic.gov.au/privacy/australian-privacy-principles-guidelines/>

The screenshot shows the homepage of the Australian Government Office of the Australian Information Commissioner (OAIC). The top navigation bar includes links for 'About us', 'Privacy', 'Freedom of information', 'Information policy', and 'Consumer Data Right'. Below the navigation, a breadcrumb trail shows 'Home / Privacy'. The main title 'Australian Privacy Principles guidelines' is displayed, along with an 'RSS feed' icon. A call-to-action button says 'Download the complete print version (combined July 2019) PDF'. A 'Filter by all tags' section lists various terms like 'access', 'anonymity and pseudonymity', 'APP1', 'APP10', etc., each enclosed in a small yellow box. To the right, three document summaries are listed:

- Summary of version changes to APP guidelines** (22 July 2019)  
Chapters of the APP guidelines are updated individually. This page contains archived versions of each chapter, and notes on the changes between versions for each chapter.  
[Read more ↗](#)
- Preface** (24 February 2014)  
Preliminary pages and preface by John McMillan.  
[Read more ↗](#)
- Chapter A: Introductory matters** (22 July 2019)  
The purpose of the APP guidelines and which organisations and agencies the

PREVIEW ONLY. NOT FINAL SLIDES.



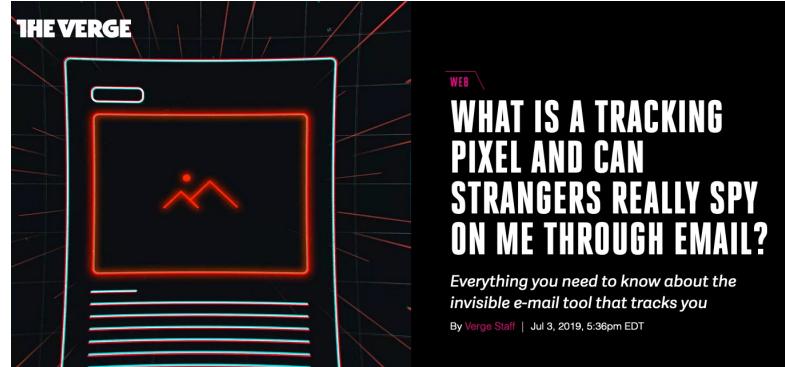
# Organisations, and Beacons and Pixels

Image by DALLE-2



# Implications

Your web-browsing life ‘off’ social media could potentially be picked up and used for ads, targeting, etc (and vice versa).



**Facebook is not the only one.**

**Consider a shopping website and a news website using a tracking pixel system XYZ.**

From Wikipedia ([https://en.wikipedia.org/wiki/Web\\_beacon](https://en.wikipedia.org/wiki/Web_beacon))

- “The image could be as small as a single pixel... (thus the name “tracking pixel”).
- “When a user opens the page or email where such an image was embedded, they might not see the image, but their web browser ... would...
- “...typically includes its IP address... existence of cookies previously sent by the host server ... [which can] associate it with a session identifier or tracking token that uniquely marks the interaction.”

Image source: The Verge

PREVIEW ONLY. NOT FINAL SLIDES.



# Users: *Your rights, freedoms, and countermeasures.*

Image by DALLE-2



# Our ethical duty?

We have a right to privacy!

- With events such as the Cambridge Analytica scandal (using Facebook data) reminding us...
- ... and initiatives by Firefox/Apple to encourage us.

From Ward (2018)'s claim that we have a duty to avoid microtargeting

- I might add, phenomena such as algorithmic personalisation (leading to Pariser's 2011 *filter bubble* phenomenon) has shown us that it's not just *us*, but those in our network are also at stake.

Some more philosophy...

JOURNAL OF MEDIA ETHICS  
2018, VOL. 33, NO. 3, 133–148  
<https://doi.org/10.1080/23736992.2018.1477047>

Routledge  
Taylor & Francis Group

Check for updates

Social networks, the 2016 US presidential election, and Kantian ethics: applying the categorical imperative to Cambridge Analytica's behavioral microtargeting

Ken Ward

Department of Communication, Lamar University, Beaumont, Texas, USA



# Ethics of care and *existentialism*

Some [more] philosophy:

- Simone de Beauvoir is a prominent feminist philosopher and also an **existentialist philosopher**.
  - [Important: we \*do not\* cover existentialism in the syllabus/assessment, but it's good for you to know]
- Some philosophers think of “Beauvoir as a **precursor of the ethic of care tradition** in feminist ethics...”
- ‘**strong parallels to the approach to ethics subsequently labeled the "ethic of care"** by feminists are nevertheless discernible. ...  
Arp suggests that Beauvoir's work furnishes an important contribution to discussions of these issues’  
(Fisher, 2000; in her introduction to Arp, 2000).



Image source: Wikipedia

PREVIEW ONLY. NOT FINAL SLIDES.



# Ethics of care and *existentialism*

We are *free* only if we recognise our responsibilities to others in promoting their *freedom to pursue their own projects*.

– Simone de Beauvoir (*Pyrrhus and Cinéas*, 1944).

This is one of the aspects covered in Beauvoir's existentialism...

...which can be applied to, say, our ethical analysis of the CA micro-targeting issue as covered before.

If what we do online (e.g. FB activity) can then be used to **microtarget those in our social network** (i.e. not 'just' us), then did we just deny others' freedom?

**Tough ethical question in the same vein as Kant's CI.**

**Reflection: this again links back to Simon's lectures on ethics**

**– there are many schools of thought, and not 'one' single answer.**

Image source: Wikipedia

PREVIEW ONLY. NOT FINAL SLIDES.



JOURNAL OF MEDIA ETHICS  
2018, VOL. 33, NO. 3, 133–148  
<https://doi.org/10.1080/23736992.2018.1477047>

Routledge  
Taylor & Francis Group



Social networks, the 2016 US presidential election, and Kantian ethics: applying the categorical imperative to Cambridge Analytica's behavioral microtargeting

Ken Ward

Department of Communication, Lamar University, Beaumont, Texas, USA



# 💡 Case Study: *A new problem - Generative AI and large language/ image models*

Image by DALLE-2



# Recall Week 6: “Lack of transparency in the data sets?”

🤖 “lack of transparency in the data sets used to train generative AI systems” (ChatGPT, 2023).

Plagiarism must also be considered, Cheong says. If the software scrapes existing content from the web, could copyright become an issue? What if the algorithm learns from [discriminatory material](#)?

“Imagine an up-and-coming artist with exceptional memory and attention to detail set off to learn the works of thousands of artists worldwide, learning to create new images in the style of any artist ‘by request’ – from van Gogh to an amateur painter – with little effort,” says Cheong. “Any artist who might have had their work viewed by this new maestro will be concerned that the same could happen to their art.”

The screenshot shows a news article from The Sydney Morning Herald. The header includes the university logo, the title "Recall Week 6: ‘Lack of transparency in the data sets?’", and a search bar. The main content features a heading "Dress code: Does AI technology belong in fashion?", a byline "Nell Geraets", and a date "November 12, 2022 – 5.00am". There are also "Save" and "Share" buttons, and font size controls.

Geraets, N. (2022, November 11). Dress code: Does AI technology belong in fashion? *The Sydney Morning Herald*. <https://www.smh.com.au/business/companies/dress-code-does-ai-technology-belong-in-fashion-20221108-p5bwh2.html>



# 💡 Case Study: *Does your data outlive you/the tech giants?*

Content warning: One of the legal case studies within discusses the topic of death.

Image by DALLE-2



# Reading: Öhman & Aggarwal (2020)



INTERNET POLICY REVIEW  
Journal on internet regulation

Volume 9 | Issue 3

---

## What if Facebook goes down? Ethical and legal considerations for the demise of big tech

**Carl Öhman**

*Oxford Internet Institute, University of Oxford, United Kingdom, carl.ohman@oii.ox.ac.uk*

**Nikita Aggarwal**

*Faculty of Law, University of Oxford, United Kingdom*

Published on 11 Aug 2020 | DOI: 10.14763/2020.3.1488



# Case Study: Dominicé & Haux (2020)

Santander Art and Culture Law Review 2/2020 (6): 251-260  
DOI: 10.4467/2450050XSNR.20.018.13021

## LEGAL COMMENTARIES

Antoinette Maget Dominicé\*

antonette.maget@lmu.de  
orcid.org/0000-0001-9056-4544  
Institute for Art History  
LMU Munich  
Zentnerstrasse 31  
DE 80798 Munich, Germany

Dario Henri Haux\*

dario.haux@unilu.ch  
orcid.org/0000-0003-1889-6231  
Faculty of Law  
University of Lucerne  
Frohburgstrasse 3  
P.O. Box 4466  
CH 6002 Lucerne, Switzerland

The Decision of the German Federal Court of Justice against Facebook:  
Opportunity to Define Digital Heritage?

### LEGAL COMMENTARIES

Antoinette Maget Dominicé and Dario Henri Haux

**Abstract:** The German Federal Court of Justice recently ruled (27.08.2020 – III ZB 30/20) that Facebook must grant parents direct access to the account of their deceased daughter. At the same time, the parents are prohibited from actively using the account. In this way, the judges established binding standards for the use of social network accounts of deceased users. Beyond inheritance and data protection law, the judgment provides an opportunity to prompt ongoing discussions about sustainable ways of safeguarding, as well as providing access to, digital content. Against the backdrop of a jurisprudence sensitized to the humanities, the two authors encourage a reflection on “spaces”, “containers”, and more generally on the significance of digital media for our everyday lives and future generations.

**Keywords:** digital inheritance, digital heritage, access, law & humanities, German Federal Court of Justice



# Case Study: Dominicé & Haux (2020)

Facts in brief, quoted from the paper:

Role of tech companies?

- Dominice & Haux (2020):  
**"In response to this [first judgment in the paper]... Facebook sent a USB stick containing a PDF file of more than 14,000 pages to the parents. These pages, which were static content, included unstructured data copied from the account..."**
- "Here the question arose whether the rights and obligations set out by the Berlin regional court in 2015 had been met. While Facebook claimed to have met the requirements, the family still felt that the "access" provided was insufficient. Hence, they brought the case back to the Berlin regional court..."
- "Whilst they should not be able to use the Facebook account actively, the parents should nevertheless be able to explore the content on the original platform. 15 In so ruling, the judges made clear that according to their legal evaluation, the obligations related to "access" under the judgment of the regional court of 17 December 2015 had not been fulfilled..."



# Reflection.

Role of tech companies?

Who owns your digital life?

As consumers, what can we do with our data?

Open question: What happens when technologies / companies shut down?

**Link to this week's tutorial: what if an important archiving tool, used by Internet historians, gets taken down from the tech provider hosting it (Github)?**



# 💡 Case Study: *A ‘set’ of data versus a dataset.*

Image by DALLE-2



# Reading: Paullada et al (2020)

---

## Data and its (dis)contents: A survey of dataset development and use in machine learning research

---

Amandalynne Paullada  
Department of Linguistics  
University of Washington

Inioluwa Deborah Raji  
Mozilla Foundation

Emily M. Bender  
Department of Linguistics  
University of Washington

Emily Denton  
Google Research

Alex Hanna  
Google Research

### Abstract

Datasets have played a foundational role in the advancement of machine learning research. They form the basis for the models we design and deploy, as well as our primary medium for benchmarking and evaluation. Furthermore, the ways in which we collect, construct and share these datasets inform the kinds of problems the field pursues and the methods explored in algorithm development. However, recent work from a breadth of perspectives has revealed the limitations of predominant practices in dataset collection and use. In this paper, we survey the many concerns raised about the way we collect and use data in machine learning and advocate that a more cautious and thorough understanding of data is necessary to address several of the practical and ethical issues of the field.

We have data of/about ourselves online.

To what extent should researchers use our ‘set of data’ as a dataset? And what are the implications?

- Harms? E.g. Representational?
- Autonomy?
- Consent?
- Use, reuse, and research ethics?  
Tying it back to Data Governance.

**This case study ties in with many other facets of digital ethics and responsible AI research.**

**We do not have all the answers – but hope this case study is a starting point for your own reflection.**



THE UNIVERSITY OF  
MELBOURNE

# Thank you

---