Common Ethical Challenges (i)

for Data Practitioners and Users

** based on Introduction to Data Ethics module (Part 2) developed by Shannon Vallor, Ph.D.*

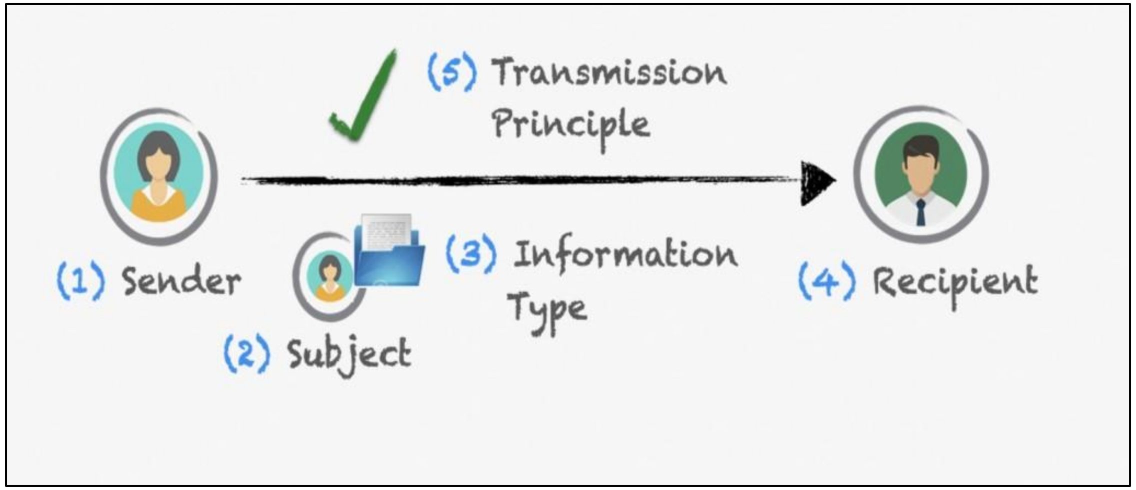
1. Ethical Challenges in Appropriate Data Collection and Use

How can we properly acknowledge and respect the purpose for, and context within which, certain data was shared with us or generated for us?

- In a medical context, a patient may share their medical records with their doctor.
- It would be odd for a doctor to ask the salary information of the patient in this context.
- However, if a person makes a loan application, it would be appropriate to share financial information.

Contextual Integrity Theory*

- Privacy is not absolute, depends heavily on **context**.



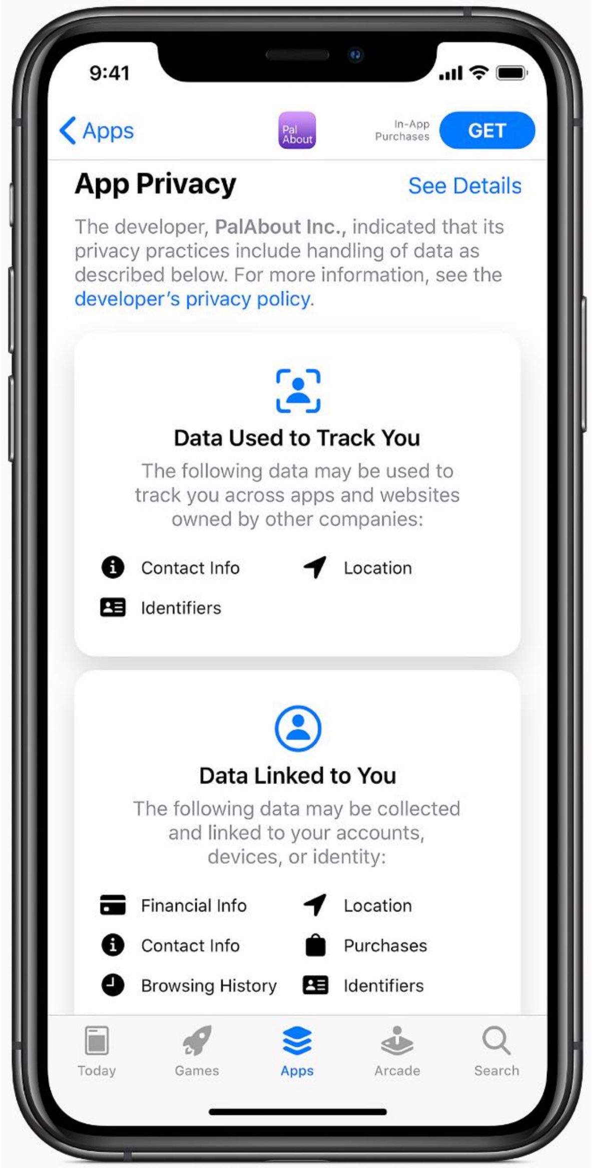
ACTIVE RESEARCH FIELD

* *Nissenbaum, Helen. "Privacy as contextual integrity." Wash. L. Rev. 79 (2004): 119.*

How can we avoid unwarranted or indiscriminate data collection?

- We should not collect data randomly.
- We should justify why certain pieces of information are collected:
  - GDPR: Data minimization
  - GDPR: Purpose should be specified

iOS: Privacy Nutrition Labels



App Privacy

See Details

The developer, **PalAbout Inc.**, indicated that its privacy practices include handling of data as described below. For more information, see the developer's privacy policy.

Data Used to Track You

The following data may be used to track you across apps and websites owned by other companies:

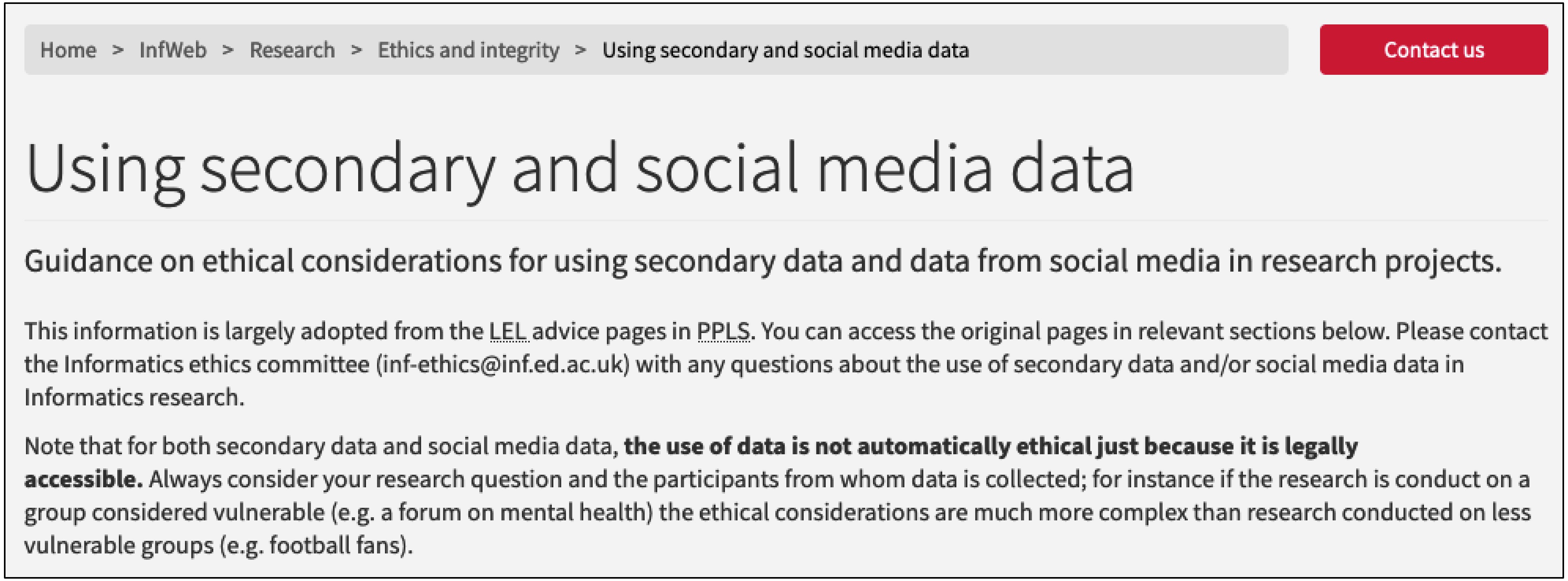
- Contact Info
- Location
- Identifiers

Data Linked to You

The following data may be collected and linked to your accounts, devices, or identity:

- Financial Info
- Location
- Contact Info
- Purchases
- Browsing History
- Identifiers

School of Informatics -- Ethics Procedure*



Home > InfWeb > Research > Ethics and integrity > Using secondary and social media data
Contact us

Using secondary and social media data

Guidance on ethical considerations for using secondary data and data from social media in research projects.

This information is largely adopted from the LEL advice pages in PPLS. You can access the original pages in relevant sections below. Please contact the Informatics ethics committee (inf-ethics@inf.ed.ac.uk) with any questions about the use of secondary data and/or social media data in Informatics research.

Note that for both secondary data and social media data, **the use of data is not automatically ethical just because it is legally accessible.** Always consider your research question and the participants from whom data is collected; for instance if the research is conduct on a group considered vulnerable (e.g. a forum on mental health) the ethical considerations are much more complex than research conducted on less vulnerable groups (e.g. football fans).

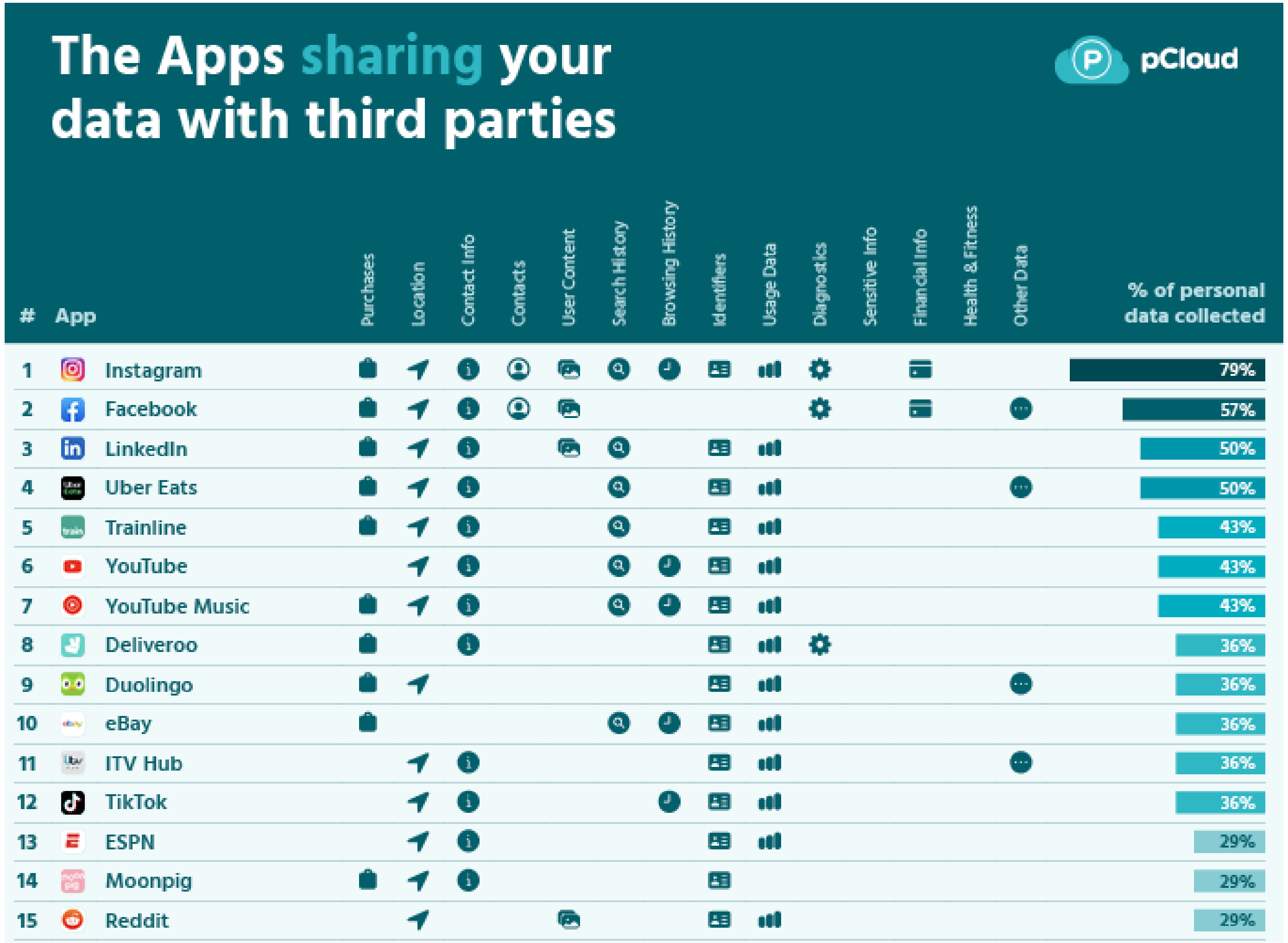
[* https://web.inf.ed.ac.uk/infweb/research/ethics-and-integrity/using-secondary-social-media-data](https://web.inf.ed.ac.uk/infweb/research/ethics-and-integrity/using-secondary-social-media-data)

Have we adequately considered the ethical implications of selling or sharing subjects’ data with third-parties?

- We need a **policy** to define how we control data.
  - GDPR: Data Protection Impact Assessment
- We need to have mechanisms to enforce such policies (e.g. Auditing Guidelines).
- We should be careful about the third-parties we work with. Do they disseminate subjects' data any further?

Apps: Third-Parties

- Instagram has 1 billion monthly active users.
- Note that not only your personal data is shared, information about your friend network is also collected.



The Apps sharing your data with third parties

pCloud

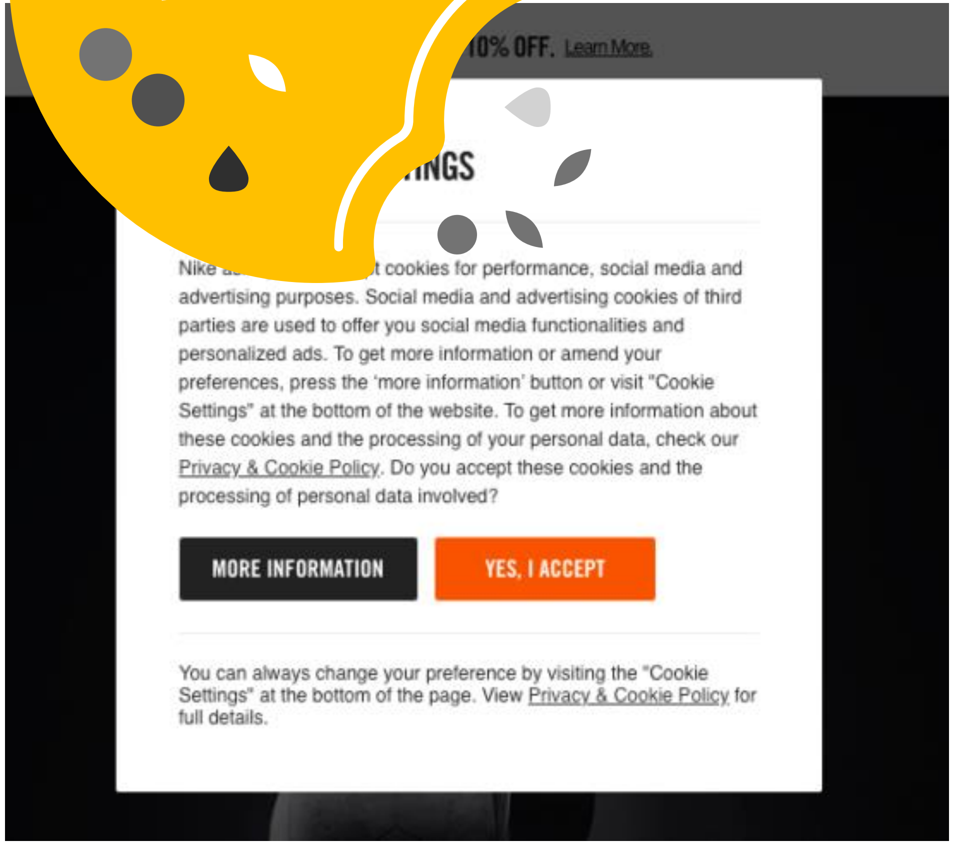
#	App	Purchases	Location	Contact Info	Contacts	User Content	Search History	Browsing History	Identifiers	Usage Data	Diagnostics	Sensitive Info	Financial Info	Health & Fitness	Other Data	% of personal data collected
1	Instagram	●	●	●	●	●	●	●	●	●	●		●			79%
2	Facebook	●	●	●	●	●					●		●		●	57%
3	LinkedIn	●	●	●		●	●		●	●						50%
4	Uber Eats	●	●	●			●		●	●					●	50%
5	Trainline	●	●	●			●		●	●						43%
6	YouTube		●	●			●	●	●	●						43%
7	YouTube Music	●	●	●			●	●	●	●						43%
8	Deliveroo	●		●					●	●	●					36%
9	Duolingo	●	●						●	●					●	36%
10	eBay	●					●	●	●	●						36%
11	ITV Hub		●	●					●	●					●	36%
12	TikTok		●	●				●	●	●						36%
13	ESPN		●	●					●	●						29%
14	Moonpig	●	●	●					●							29%
15	Reddit		●			●			●	●						29%

https://cybernews.com/privacy/instagram-is-sharing-79-of-your-personal-data-with-third-parties/ published on 06 April 2021.

Have we given data subjects appropriate forms of choice in data sharing?

- Opt-in vs Opt-out Privacy settings

Nike: Cookie Settings

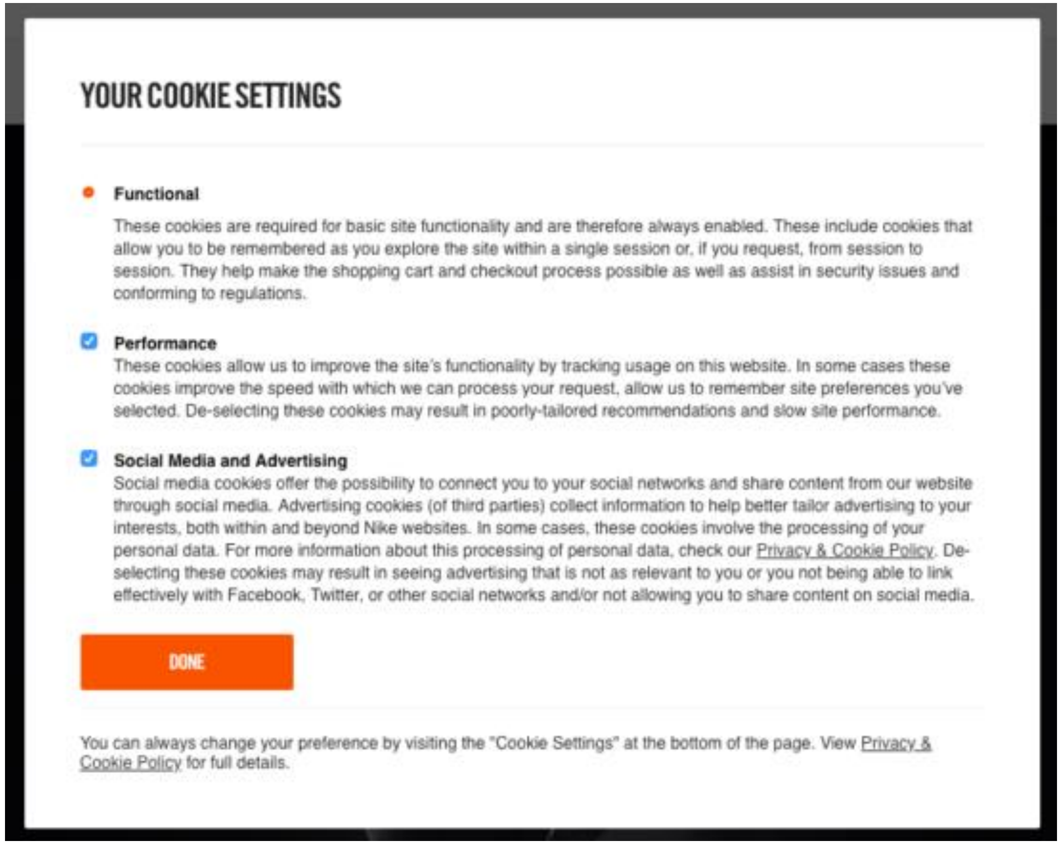


10% OFF. Learn More.

INGS

Nike a... t cookies for performance, social media and advertising purposes. Social media and advertising cookies of third parties are used to offer you social media functionalities and personalized ads. To get more information or amend your preferences, press the 'more information' button or visit "Cookie Settings" at the bottom of the website. To get more information about these cookies and the processing of your personal data, check our Privacy & Cookie Policy. Do you accept these cookies and the processing of personal data involved?

You can always change your preference by visiting the "Cookie Settings" at the bottom of the page. View Privacy & Cookie Policy for full details.



YOUR COOKIE SETTINGS

● **Functional**

These cookies are required for basic site functionality and are therefore always enabled. These include cookies that allow you to be remembered as you explore the site within a single session or, if you request, from session to session. They help make the shopping cart and checkout process possible as well as assist in security issues and conforming to regulations.

☑ **Performance**

These cookies allow us to improve the site's functionality by tracking usage on this website. In some cases these cookies improve the speed with which we can process your request, allow us to remember site preferences you've selected. De-selecting these cookies may result in poorly-tailored recommendations and slow site performance.

☑ **Social Media and Advertising**

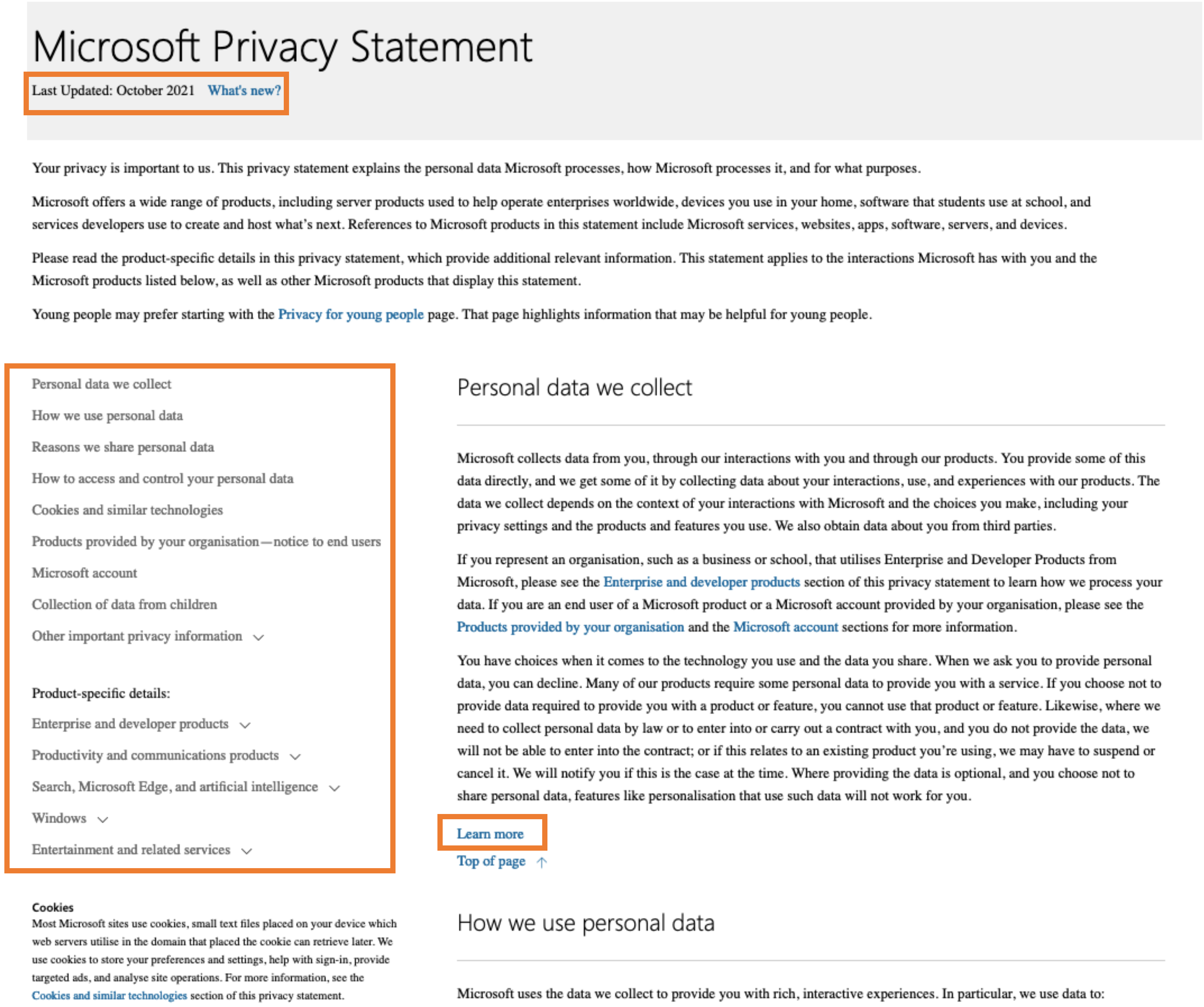
Social media cookies offer the possibility to connect you to your social networks and share content from our website through social media. Advertising cookies (of third parties) collect information to help better tailor advertising to your interests, both within and beyond Nike websites. In some cases, these cookies involve the processing of your personal data. For more information about this processing of personal data, check our Privacy & Cookie Policy. De-selecting these cookies may result in seeing advertising that is not as relevant to you or you not being able to link effectively with Facebook, Twitter, or other social networks and/or not allowing you to share content on social media.

You can always change your preference by visiting the "Cookie Settings" at the bottom of the page. View Privacy & Cookie Policy for full details.

Are the terms of our data policy laid out in a clear, direct, and understandable way, and made accessible to all data subjects?

- Most privacy policies takes a legal perspective since they are written by lawyers.
- Privacy policies are not updated frequently to match data practices.
- How to make privacy policies machine readable is an **open research question**.

Good Example



Microsoft Privacy Statement

Last Updated: October 2021 What's new?

Your privacy is important to us. This privacy statement explains the personal data Microsoft processes, how Microsoft processes it, and for what purposes.

Microsoft offers a wide range of products, including server products used to help operate enterprises worldwide, devices you use in your home, software that students use at school, and services developers use to create and host what's next. References to Microsoft products in this statement include Microsoft services, websites, apps, software, servers, and devices.

Please read the product-specific details in this privacy statement, which provide additional relevant information. This statement applies to the interactions Microsoft has with you and the Microsoft products listed below, as well as other Microsoft products that display this statement.

Young people may prefer starting with the Privacy for young people page. That page highlights information that may be helpful for young people.

Personal data we collect
How we use personal data
Reasons we share personal data
How to access and control your personal data
Cookies and similar technologies
Products provided by your organisation—notice to end users
Microsoft account
Collection of data from children
Other important privacy information ⌄

Product-specific details:
Enterprise and developer products ⌄
Productivity and communications products ⌄
Search, Microsoft Edge, and artificial intelligence ⌄
Windows ⌄
Entertainment and related services ⌄

Cookies
Most Microsoft sites use cookies, small text files placed on your device which web servers utilise in the domain that placed the cookie can retrieve later. We use cookies to store your preferences and settings, help with sign-in, provide targeted ads, and analyse site operations. For more information, see the Cookies and similar technologies section of this privacy statement.

Personal data we collect

Microsoft collects data from you, through our interactions with you and through our products. You provide some of this data directly, and we get some of it by collecting data about your interactions, use, and experiences with our products. The data we collect depends on the context of your interactions with Microsoft and the choices you make, including your privacy settings and the products and features you use. We also obtain data about you from third parties.

If you represent an organisation, such as a business or school, that utilises Enterprise and Developer Products from Microsoft, please see the Enterprise and developer products section of this privacy statement to learn how we process your data. If you are an end user of a Microsoft product or a Microsoft account provided by your organisation, please see the Products provided by your organisation and the Microsoft account sections for more information.

You have choices when it comes to the technology you use and the data you share. When we ask you to provide personal data, you can decline. Many of our products require some personal data to provide you with a service. If you choose not to provide data required to provide you with a product or feature, you cannot use that product or feature. Likewise, where we need to collect personal data by law or to enter into or carry out a contract with you, and you do not provide the data, we will not be able to enter into the contract; or if this relates to an existing product you're using, we may have to suspend or cancel it. We will notify you if this is the case at the time. Where providing the data is optional, and you choose not to share personal data, features like personalisation that use such data will not work for you.

Learn more

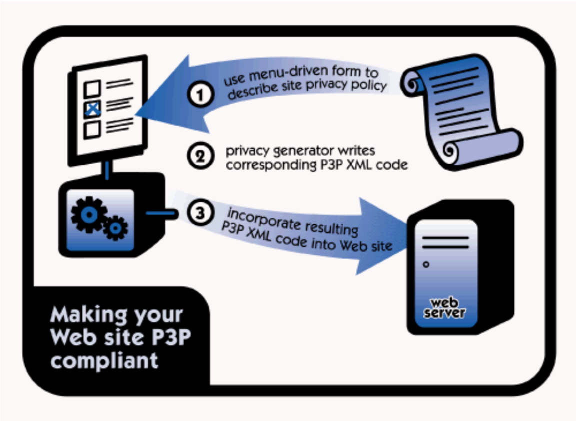
Top of page ↑

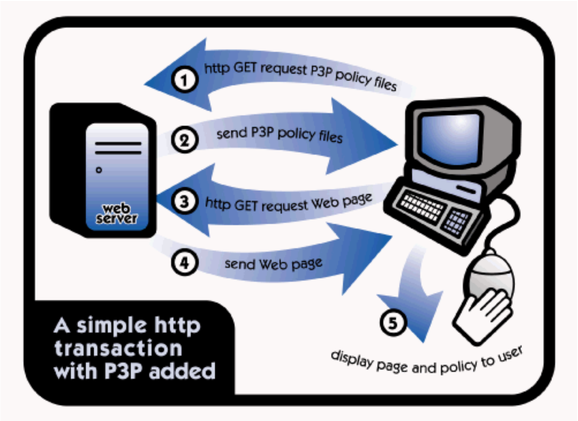
How we use personal data

Microsoft uses the data we collect to provide you with rich, interactive experiences. In particular, we use data to:

P3P: The Platform for Privacy Preferences Project

- P3P enables Websites to express their privacy practices in a standard format. P3P supports both machine- and human-readable formats.
- The idea is to automate decision-making process for the users.

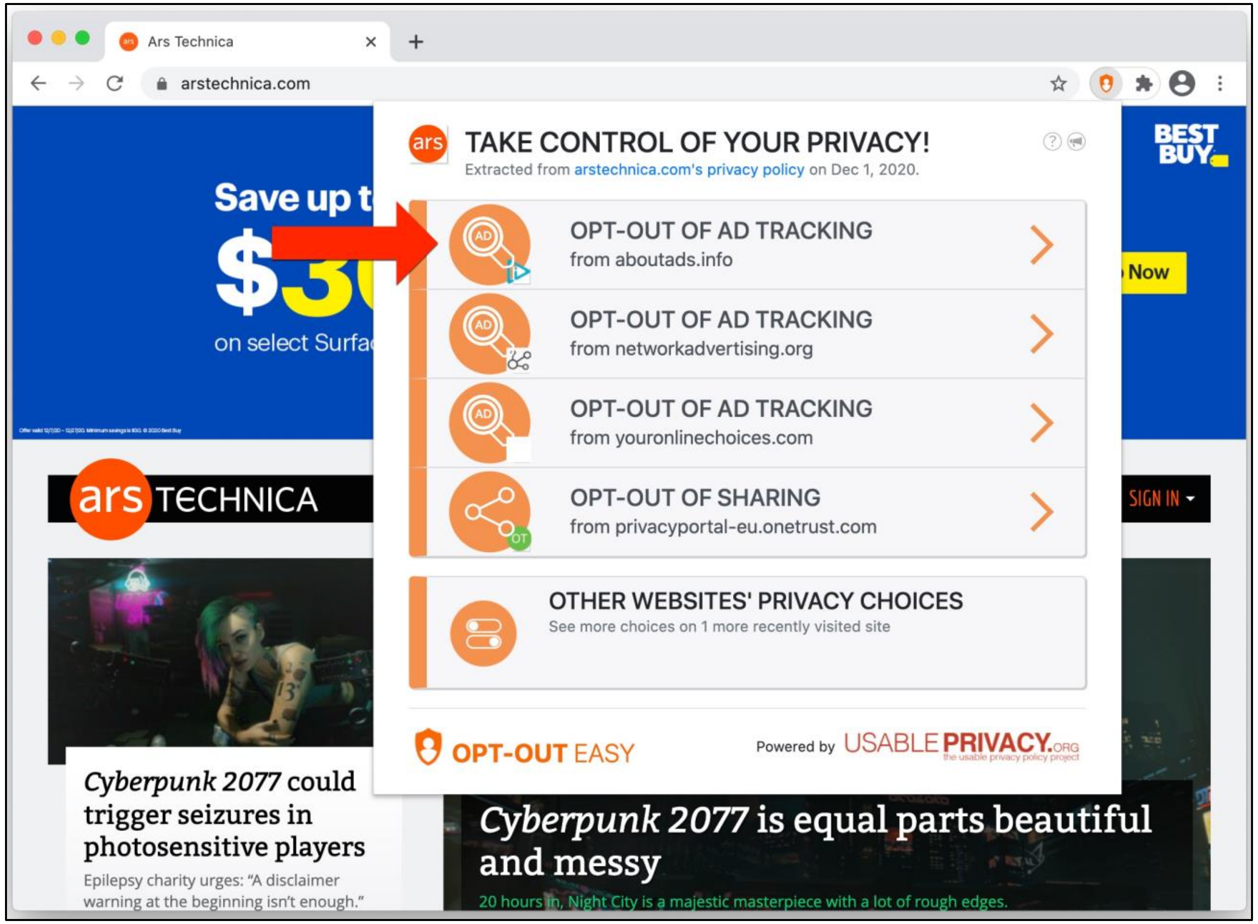

1 use menu-driven form to describe site privacy policy
2 privacy generator writes corresponding P3P XML code
3 incorporate resulting P3P XML code into Web site
web server
Making your Web site P3P compliant


1 http GET request P3P policy files
2 send P3P policy files
3 http GET request Web page
4 send Web page
5 display page and policy to user
web server
A simple http transaction with P3P added

Cool idea, but it did not work...

https://www.w3.org/P3P/

Opt-Out Easy --- A Browser Plugin*



TAKE CONTROL OF YOUR PRIVACY!

Extracted from arstechnica.com's privacy policy on Dec 1, 2020.

OPT-OUT OF AD TRACKING
from aboutads.info

OPT-OUT OF AD TRACKING
from networkadvertising.org

OPT-OUT OF AD TRACKING
from youronlinechoices.com

OPT-OUT OF SHARING
from privacyportal-eu.onetrust.com

OTHER WEBSITES' PRIVACY CHOICES
See more choices on 1 more recently visited site

OPT-OUT EASY

Powered by USABLE PRIVACY.ORG the usable privacy policy project

Cyberpunk 2077 could trigger seizures in photosensitive players

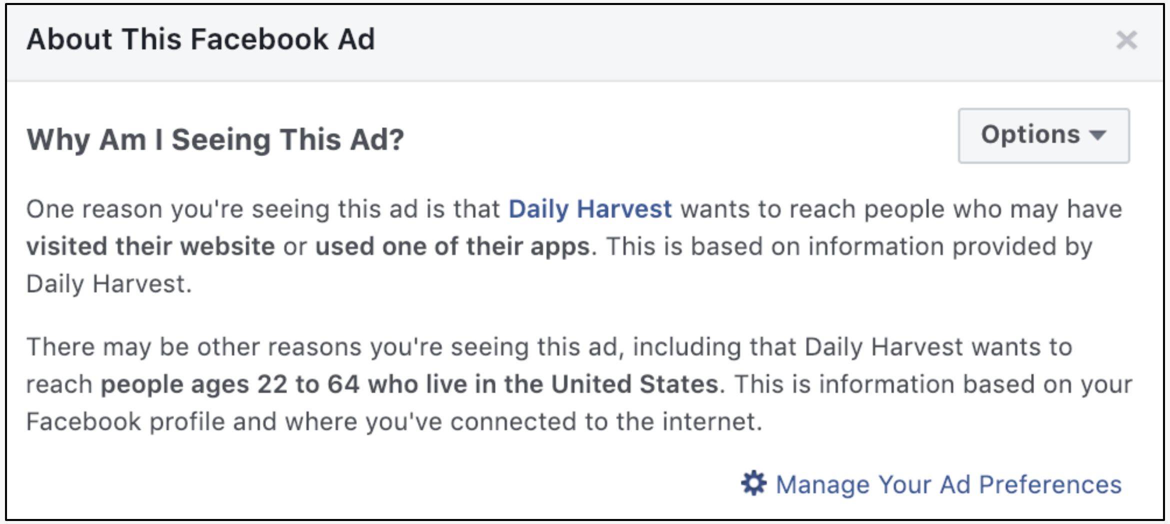
Epilepsy charity urges: "A disclaimer warning at the beginning isn't enough."

Cyberpunk 2077 is equal parts beautiful and messy

20 hours in, Night City is a majestic masterpiece with a lot of rough edges.

* *Vinayshekhar Bannihatti Kumar, Roger Iyengar, Namita Nisal, Yuanyuan Feng, Hana Habib, Peter Story, Sushain Cherivirala, Margaret Hagan, Lorrie Cranor, Shomir Wilson, Florian Schaub, and Norman Sadeh. 2020. Finding a Choice in a Haystack: Automatic Extraction of Opt-Out Statements from Privacy Policy Text. In Proceedings of The Web Conference 2020 (WWW '20). Association for Computing Machinery, New York, NY, USA, 1943–1954.*

Are data subjects given clear paths to obtaining more information or context for a data practice?



Why Am I Seeing This Ad?

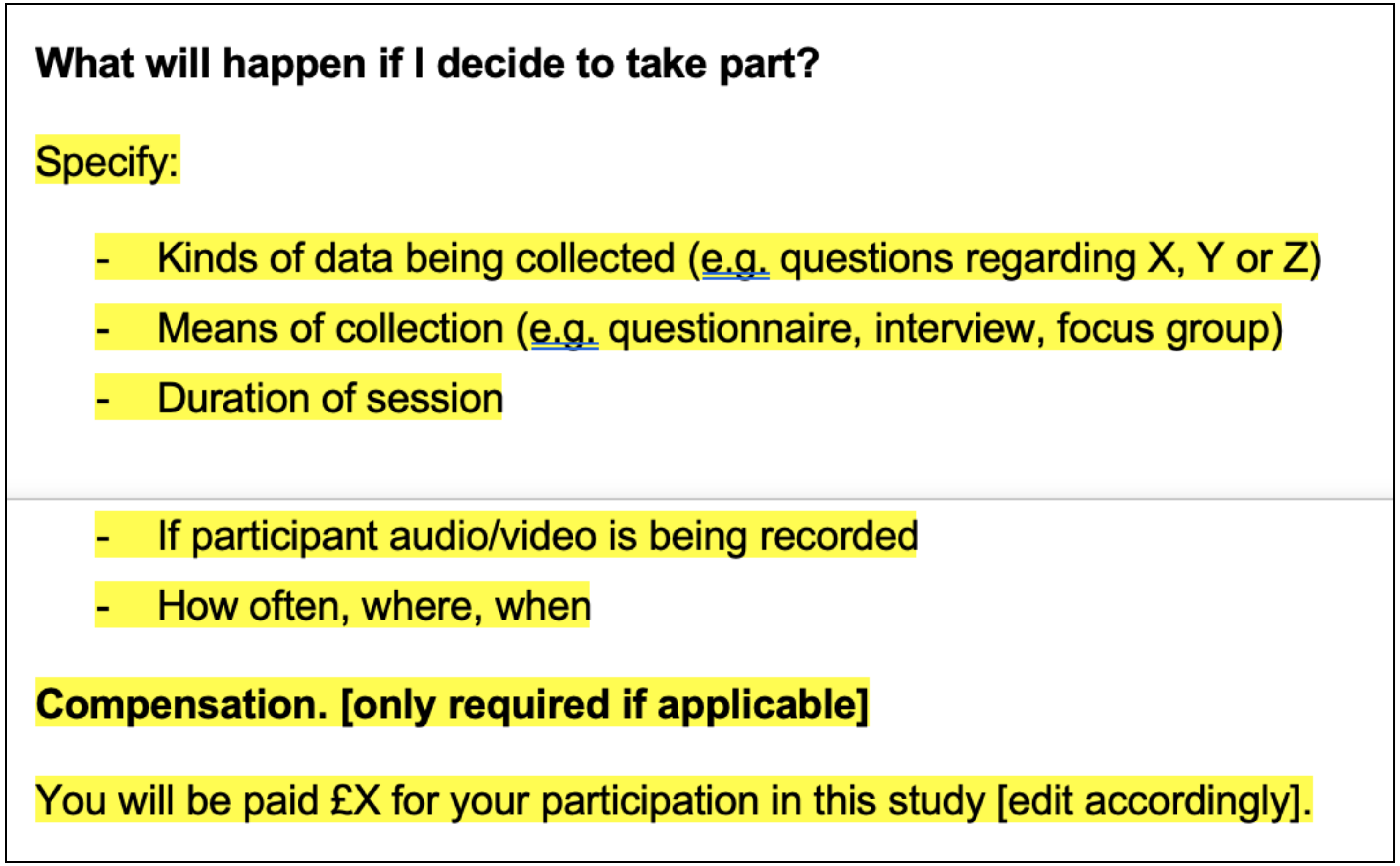
One reason you're seeing this ad is that Daily Harvest wants to reach people who may have **visited their website** or **used one of their apps**. This is based on information provided by Daily Harvest.

There may be other reasons you're seeing this ad, including that Daily Harvest wants to reach **people ages 22 to 64 who live in the United States**. This is information based on your Facebook profile and where you've connected to the internet.

Are data subjects being appropriately compensated for the benefits/value of their data?

- Compensation could be in the form of **money** (e.g., vouchers).
- Participation could also be **voluntary**.

Participation Information Sheet (PIS)



What will happen if I decide to take part?

Specify:

- Kinds of data being collected (e.g. questions regarding X, Y or Z)
- Means of collection (e.g. questionnaire, interview, focus group)
- Duration of session
- If participant audio/video is being recorded
- How often, where, when

Compensation. [only required if applicable]

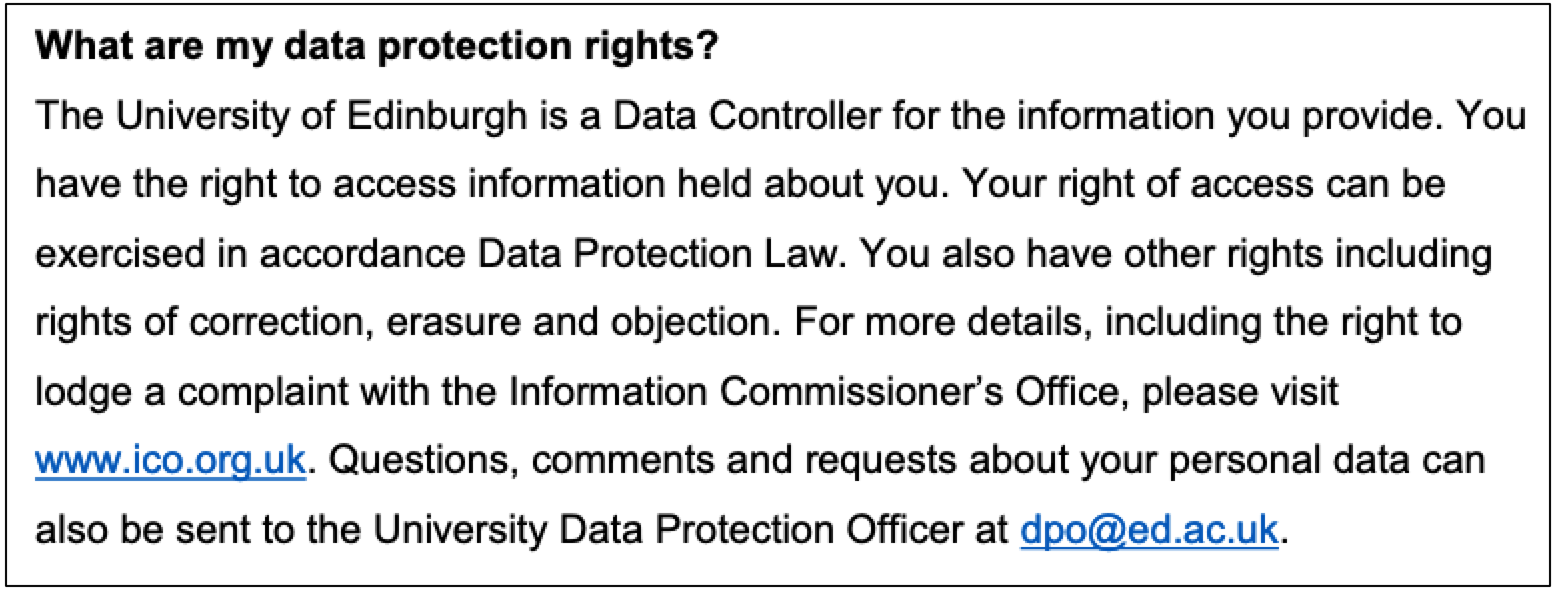
You will be paid £X for your participation in this study [edit accordingly].

https://web.inf.ed.ac.uk/infweb/research/ethics-and-integrity/ethics-resources

Have we considered what control or rights our data subjects should retain over their data?

- The users should be able to withdraw, correct or update their shared data.
- The data collector should make this clear in their privacy policies.

Participation Information Sheet (PIS)



What are my data protection rights?

The University of Edinburgh is a Data Controller for the information you provide. You have the right to access information held about you. Your right of access can be exercised in accordance Data Protection Law. You also have other rights including rights of correction, erasure and objection. For more details, including the right to lodge a complaint with the Information Commissioner's Office, please visit www.ico.org.uk. Questions, comments and requests about your personal data can also be sent to the University Data Protection Officer at dpo@ed.ac.uk.

https://web.inf.ed.ac.uk/infweb/research/ethics-and-integrity/ethics-resources

2. DATA STORAGE, SECURITY AND RESPONSIBLE DATA STEWARDSHIP

How can we responsibly and safely store personally identifying information?

- The data collectors should make it clear how they **store** their data.
  - Ethics Boards in organizations
  - Ethics committees at universities

Have we reflected on the ethical harms that may be done by a data breach, both in the short-term and long-term, and to whom?

- Stakeholders can be **direct** or **indirect**.
- It is very difficult to **estimate the correct audience** who could be affected by a data breach.
- A **risk estimation** should be done in any case.

What are our concrete action plans for the worst-case-scenarios?

- **Mitigation strategies** should be set up-front.
- The specified protocols should be applied in case of any incidents.

Have we made appropriate investments in our data security/storage infrastructure (relative to our context and the potential risks and harms)?

- A good infrastructure requires investment. Not all organizations can have it.
- The context in which data collection happens matters a lot. For example, medical data would be more sensitive compared to data collected via a temperature sensor.

Research Data Service

- For sensitive data, some external providers exist:
  - Edinburgh International Data Facility, Dataloch, Lothian Research Safe Haven
- Active Data Storage
  - DataStore
- Version Control
  - GitLab, Subversion
- Collaboration
  - DataSync, Wiki Service, SharePoint
- High-computing
  - Edinburgh Compute and Data Facility (ECDF)
- Other options...

https://www.ed.ac.uk/information-services/research-support/research-data-service/during

What privacy-preserving techniques do we rely upon, and what are their various advantages and limitations?

- Data pseudonymization
- Data anonymization
- Obfuscation
- Differential privacy

…

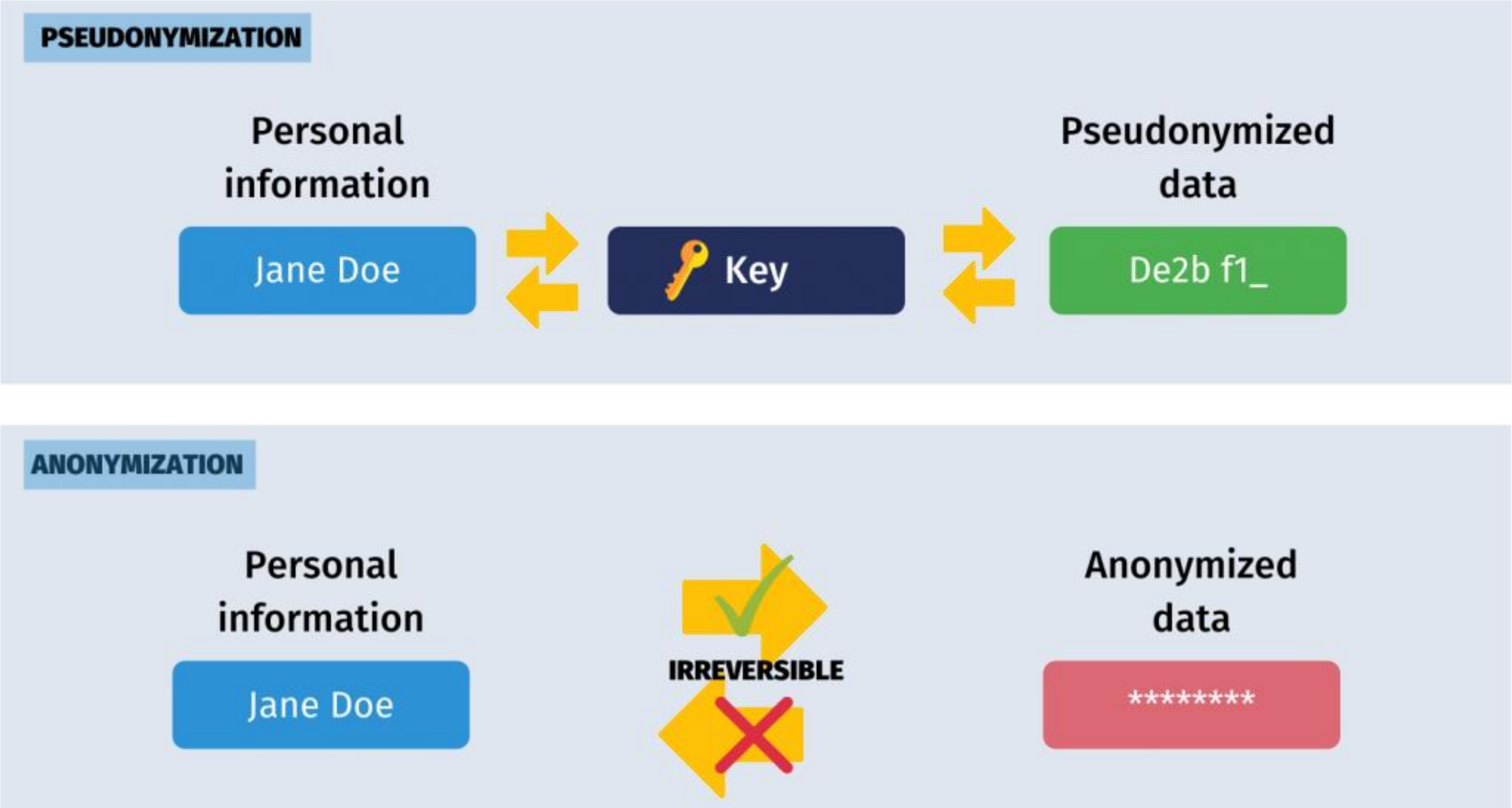
GDPR Data Pseudonymization

In Article 4(5) of the GDPR, the process of pseudonymization is defined as:

“the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.”

A pseudonym is personal data according to the GDPR since the process is reversible.

Pseudonymization vs Anonymization



https://dataprivacymanager.net/pseudonymization-according-to-the-gdpr/

Code Obfuscation

- Making the code less "hackable"
- It may destroy code readability, increase computation time

```
int i=1, sum=0, avg=0
while (i = 100)
{
    sum+=i;
    avg=sum/i;
    i++;
}
```

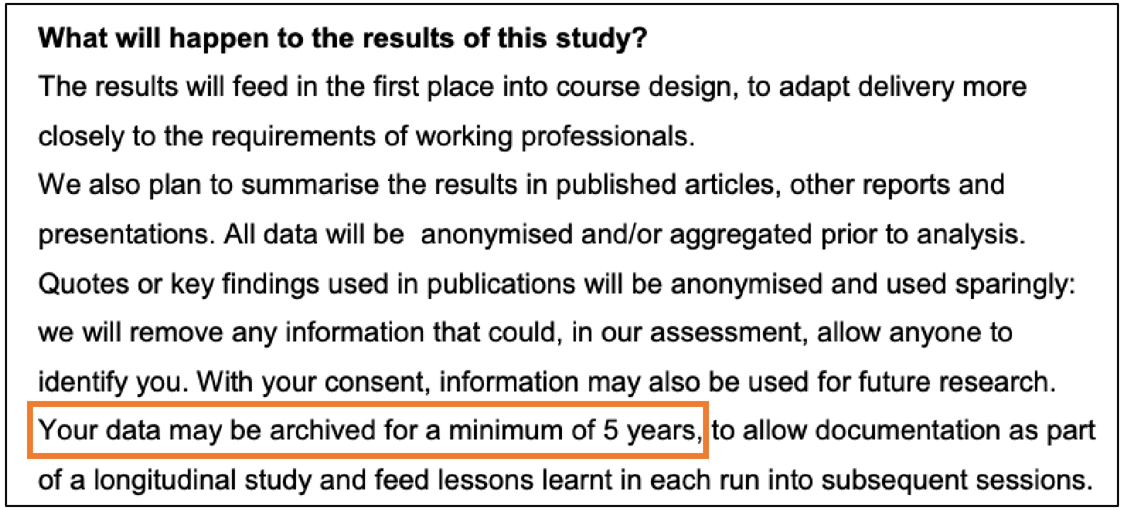
```
int random = 1;
while (random != 0)
{switch (random)
{case 1:
{
    i=0; sum=1; avg=1;
    random = 2;
    break;
}
case 2:
{
    if (i = 100)
        random = 3;
    else random = 0;
        break;
}
case 3:
{
    sum+=i;  avg=sum/i ; i++;
    random = 2;
    break;
}}}
```

https://www.freecodecamp.org/news/make-your-code-secure-with-obfuscation/

What are the ethical risks of long-term data storage? How long we are justified in keeping sensitive data, and when/how often should it be purged?

- Other limitations:
  - the availability of space,
  - the cost of storage,
  - protection of confidential information.
- Funding agencies have policies regarding the minimum length of data retention.

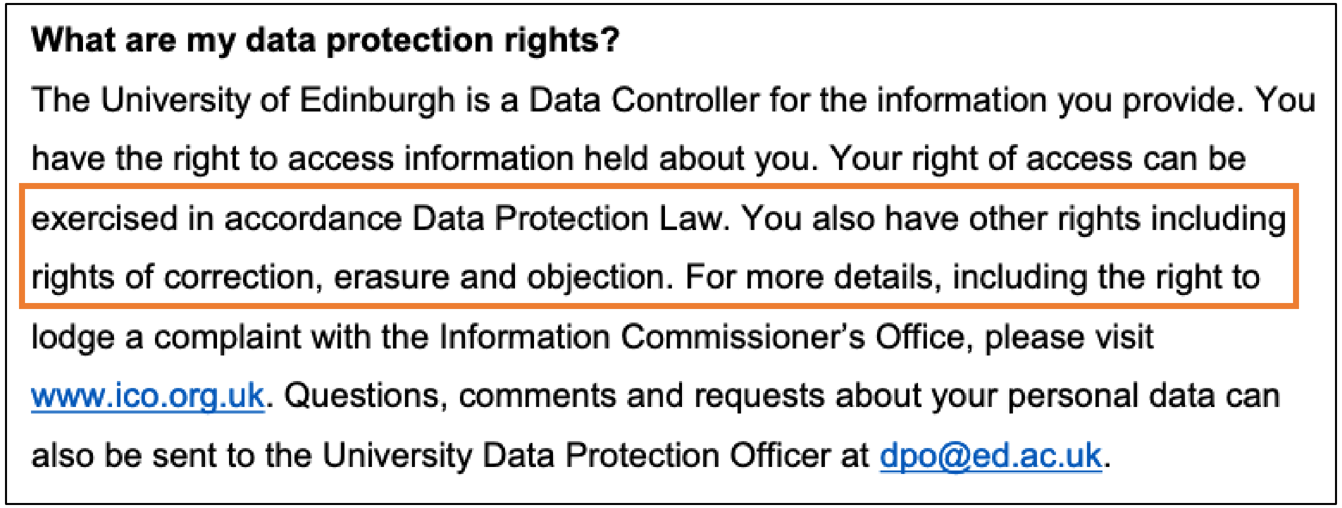
Participation Information Sheet (PIS)



What will happen to the results of this study?

The results will feed in the first place into course design, to adapt delivery more closely to the requirements of working professionals.

We also plan to summarise the results in published articles, other reports and presentations. All data will be anonymised and/or aggregated prior to analysis. Quotes or key findings used in publications will be anonymised and used sparingly: we will remove any information that could, in our assessment, allow anyone to identify you. With your consent, information may also be used for future research. Your data may be archived for a minimum of 5 years, to allow documentation as part of a longitudinal study and feed lessons learnt in each run into subsequent sessions.

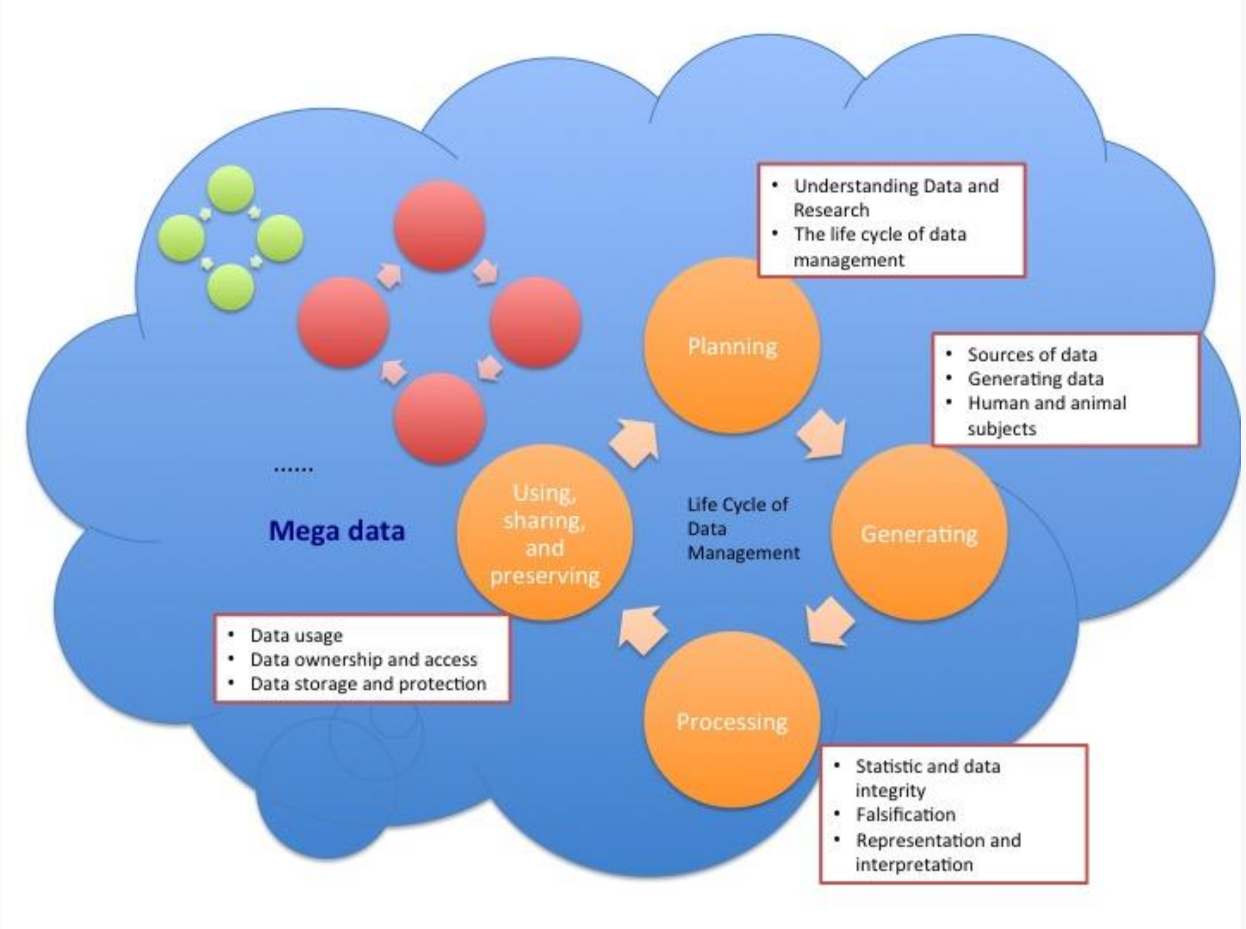


What are my data protection rights?

The University of Edinburgh is a Data Controller for the information you provide. You have the right to access information held about you. Your right of access can be exercised in accordance Data Protection Law. You also have other rights including rights of correction, erasure and objection. For more details, including the right to lodge a complaint with the Information Commissioner's Office, please visit www.ico.org.uk. Questions, comments and requests about your personal data can also be sent to the University Data Protection Officer at dpo@ed.ac.uk.

https://web.inf.ed.ac.uk/infweb/research/ethics-and-integrity/ethics-resources

Do we have an end-to-end plan for the lifecycle of the data we collect or use?



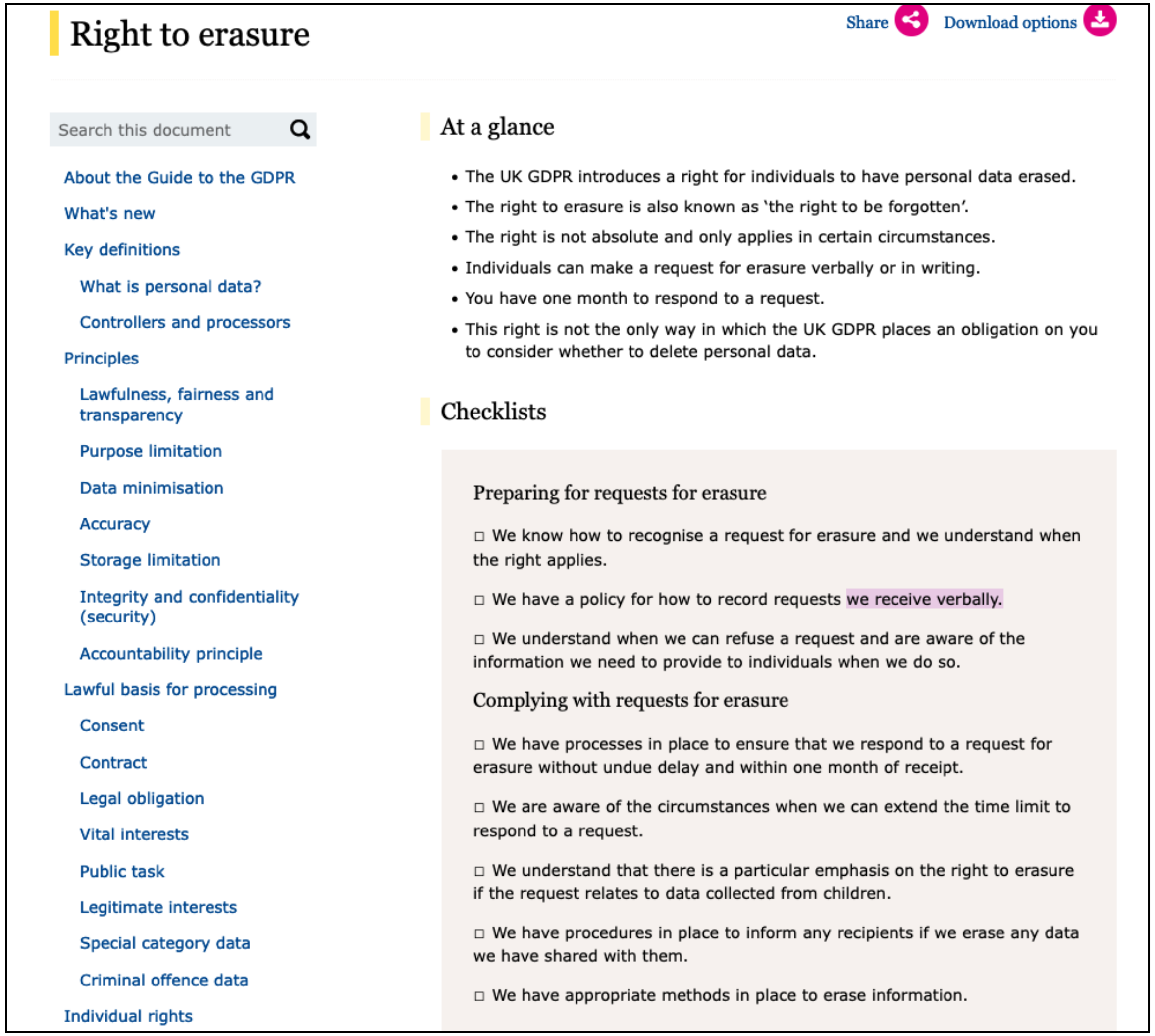
Mega data
......
Planning
Generating
Processing
Using, sharing, and preserving
Life Cycle of Data Management
• Understanding Data and Research
• The life cycle of data management
• Sources of data
• Generating data
• Human and animal subjects
• Statistic and data integrity
• Falsification
• Representation and interpretation
• Data usage
• Data ownership and access
• Data storage and protection

https://sites.psu.edu/ethicsofdatamanagement/

What measures should we have in place to allow data to be deleted, corrected, or updated by affected/interested parties?

- Protocols, protocols, protocols…

ICO – GDPR Guidance



Right to erasure

Share Download options

Search this document

At a glance

- The UK GDPR introduces a right for individuals to have personal data erased.
- The right to erasure is also known as 'the right to be forgotten'.
- The right is not absolute and only applies in certain circumstances.
- Individuals can make a request for erasure verbally or in writing.
- You have one month to respond to a request.
- This right is not the only way in which the UK GDPR places an obligation on you to consider whether to delete personal data.

Checklists

Preparing for requests for erasure

□ We know how to recognise a request for erasure and we understand when the right applies.

□ We have a policy for how to record requests we receive verbally.

□ We understand when we can refuse a request and are aware of the information we need to provide to individuals when we do so.

Complying with requests for erasure

□ We have processes in place to ensure that we respond to a request for erasure without undue delay and within one month of receipt.

□ We are aware of the circumstances when we can extend the time limit to respond to a request.

□ We understand that there is a particular emphasis on the right to erasure if the request relates to data collected from children.

□ We have procedures in place to inform any recipients if we erase any data we have shared with them.

□ We have appropriate methods in place to erase information.

Summary

Ethical Challenges in Appropriate Data Collection and Use

- Purpose of data collection, context, dissemination of data, choice in data sharing, compensation, control/rights...

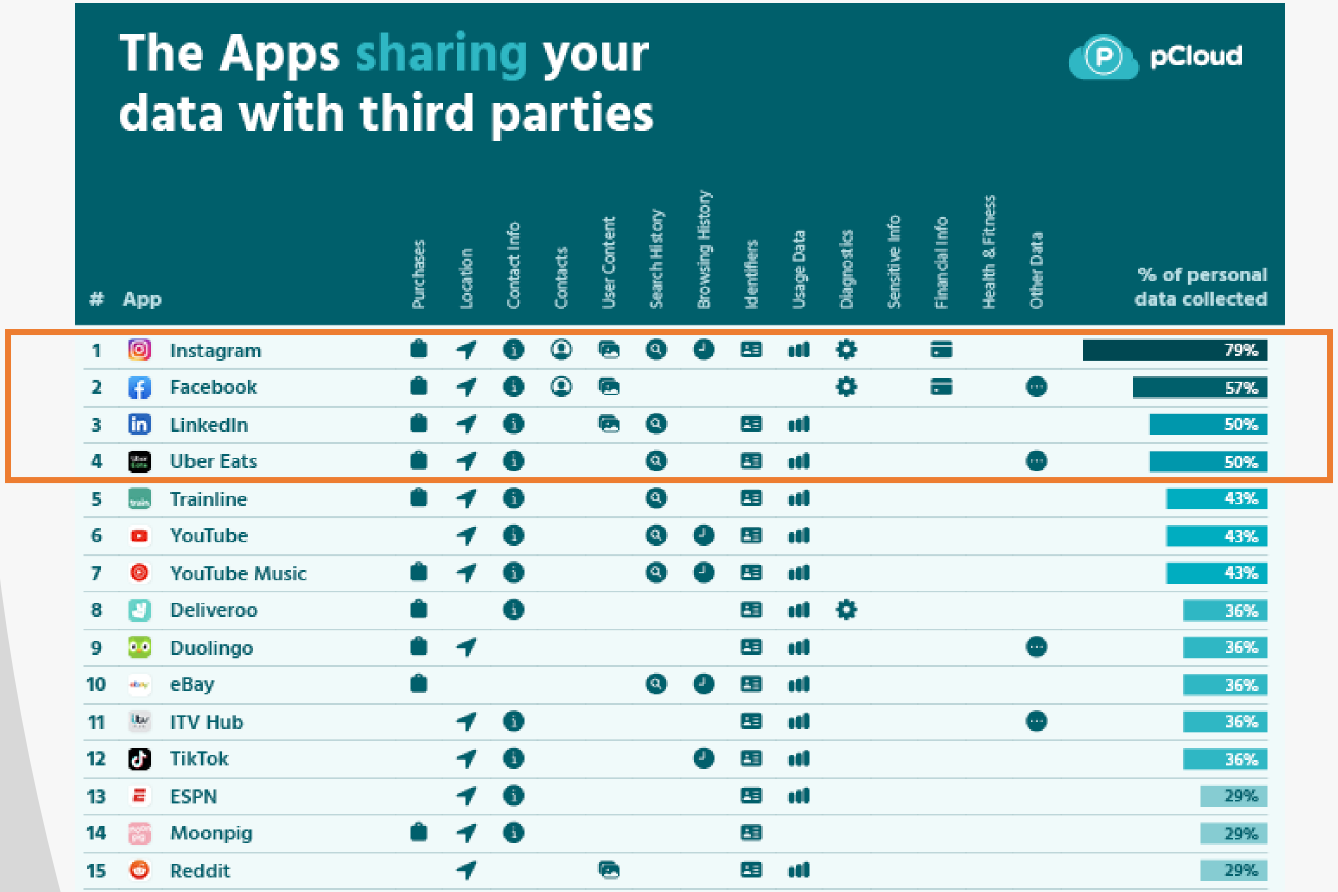
Data Storage, Security and Responsible Data Stewardship

- Storage of data, risk estimation, mitigation strategies, privacy-preserving techniques, ethical risks of keeping data longer...

Exercise

* Pick one of the top four apps on the list

* Discuss two ethical challenges covered in this lecture.



The Apps sharing your data with third parties

pCloud

#	App	Purchases	Location	Contact Info	Contacts	User Content	Search History	Browsing History	Identifiers	Usage Data	Diagnostics	Sensitive Info	Financial Info	Health & Fitness	Other Data	% of personal data collected
1	Instagram	●	●	●	●	●	●	●	●	●	●		●			79%
2	Facebook	●	●	●	●	●					●		●		●	57%
3	LinkedIn	●	●	●		●	●		●	●						50%
4	Uber Eats	●	●	●			●		●	●					●	50%
5	Trainline	●	●	●			●		●	●						43%
6	YouTube		●	●			●	●	●	●						43%
7	YouTube Music	●	●	●			●	●	●	●						43%
8	Deliveroo	●		●					●	●	●					36%
9	Duolingo	●	●						●	●					●	36%
10	eBay	●					●	●	●	●						36%
11	ITV Hub		●	●					●	●					●	36%
12	TikTok		●	●				●	●	●						36%
13	ESPN		●	●					●	●						29%
14	Moonpig	●	●	●					●							29%
15	Reddit		●			●			●	●						29%

https://cybernews.com/privacy/instagram-is-sharing-79-of-your-personal-data-with-third-parties/ published on 06 April 2021.