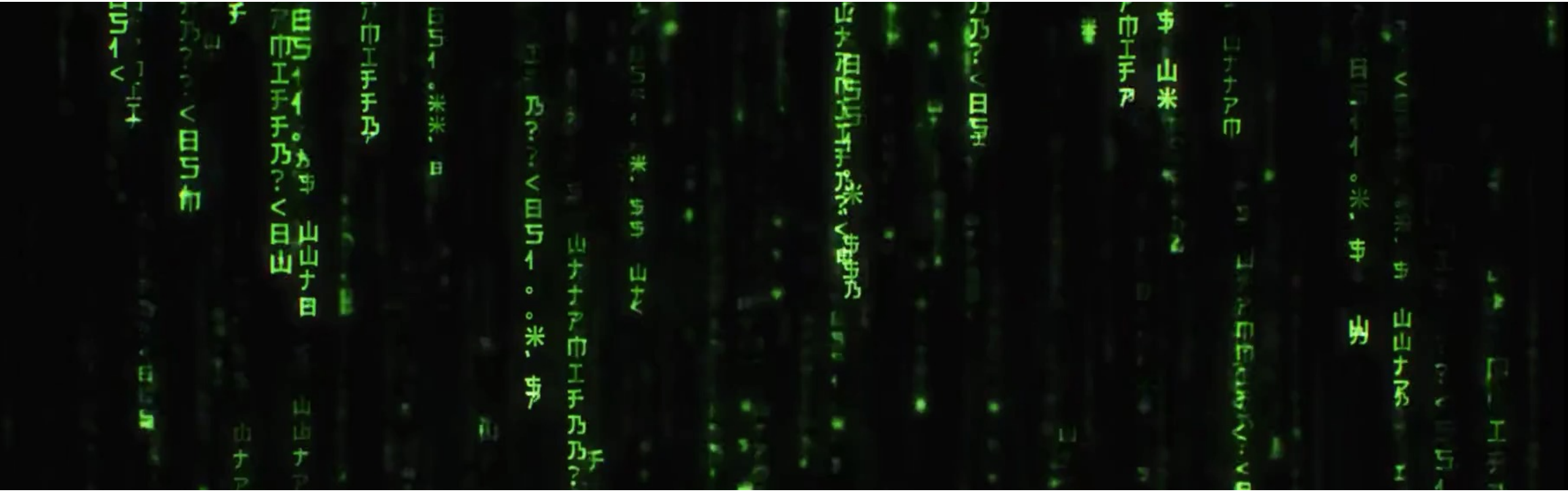
Data, Politics and Society

W6 – Data Protection Legislation



Where we at?

Part I: Data and its role in society

W1
W2
W3

Data: The Good, The Bad, The Ugly

W4
W5

Societal and environmental impacts of data and technology

Where we at?

Part II: Mitigating the risks of working with large-scale datasets

W6
W7

Regulations and governance

W8
W9

Crowdsourcing, VGI, and Geographic Citizen Science

W10

Critical Data Studies

Today

- Privacy and confidentiality
- General Data Protection Regulation (GDPR) and Data Protection Act
- GDPR in the information age

Data protection

Phillips and Knoppers 2019:

- Data protection refers to “a set of legal rules that aims to protect the rights, freedoms, and interests of individuals, whose personal data are collected, stored, processed, disseminated, destroyed etc.”
- The ultimate objective is to ensure “fairness in the processing of data and, to some extent, fairness in the outcomes of such processing.”

Why are we worried about data?

- From cookies to contract-tracing apps, our right as individuals to privacy are being challenged everyday by our use of technology and their pervasive behaviours of collecting data on us.
- We often give our personal data to services in exchange for goods – but we have had little to no control on how this data is used or shared.

Privacy as a right

1948 Universal Declaration of Human Rights Article 12 United Nation:

“No one shall be subjected to arbitrary interference with their privacy, family, home or correspondence, nor to attacks upon their honour and reputation. Everyone has the right to the protection of the law against such interference or attacks."

Why do we care as researchers?

The study of human subjects requires that the interests of individual privacy and data confidentiality are balanced against social benefits of research access and utility.

Why do we care as researchers?

The study of human subjects requires that the interests of individual privacy and data confidentiality are balanced against social benefits of research access and utility.

Terminology

- Privacy: “encompasses not only the famous ‘right to be left alone,’ or keeping one’s personal matters and relationships secret, but also the ability to share information selectively but not publicly.” (President’s Council of Advisors on Science and Technology, 2014).
- Confidentiality is “preserving authorized restrictions on information access and disclosure, including means for protecting personal privacy and proprietary information.” (McCallister, Grance, and Scarfone, 2010).
- Utility: Data utility is the value resulting from data use.

Privacy as Contextual Integrity

Nissenbaum 2019:

“Privacy, defined as contextual integrity (CI), is preserved when information flows generated by an action or practice conform to the legitimate contextual informational norms: it is violated when they are breached”

A ‘flexible’, context-dependent, framework to think about privacy and flows of information.

Privacy as Contextual Integrity

1) Privacy is the appropriate flow of personal information.
2) Appropriate flows conform with contextual informational norms (‘privacy norms’).
3) Five parameters define privacy norms: subject, sender, recipient, information type, transmission principle.
4) The ethical legitimacy of privacy norms is evaluated in terms of: a) interests of affected parties, b) ethical and political values, 3) contextual functions, purposes and values. Example:

   - patient-doctor confidentiality versus data analytics with societal benefits
   - healthcare context must be promoted to avoid patients getting fearful information flows to wrong parties

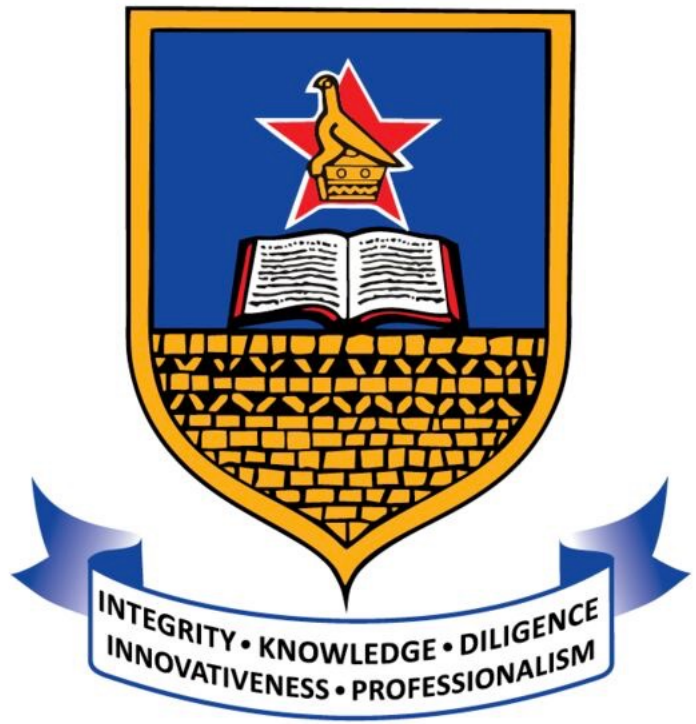
Privacy and confidentiality in the information age

Who have you shared personal data with in the last 12 months – and who do you trust with your data?


UCL


Home Office


UNIVERSITY OF
OXFORD



UNIVERSITY OF ZIMBABWE


TILBURG UNIVERSITY


Sainsbury's


F
Foxtons

Ethics of data

- The use of personal data is an essential tool for research in many settings, including academic, state, public and commercial bodies. But too much personal information can overstep our right to privacy as we become identifiable in the dataset.
- Data that can be linked with another dataset can also lead to reducing us to our individual persons and become identifiable in the dataset.
- A useful way of thinking about maintaining privacy is the notion of regulating the appropriate flow of information to prevent this re-identification from happening (i.e. privacy as Contextual Integrity).

Why regulation?

- *Risk of data leakages*: breaching our expectations of confidentiality by leaking (purposefully or not) our data to others.
- *Risk of identification*: combining and linking large amounts of our information which could lead to re-identification and a breach of our privacy rights.
- *Risk of harm*: the use of our data against us to prevent us from accessing certain products or services, or from others misusing our data for harmful purposes.
- *Risk of discrimination*: Related to the former, using our data to discriminate against specific groups. This is theoretically already illegal under Anti-Discrimination Law / equality and freedom human rights.

GDPR

- The GDPR (General Data Protection Regulation (2016) is the EU regulation covering people's personal data. Active since May 2018 in the UK through implementation in the Data Protection Act.
- Designed to protect individuals' personal information in an era of mass digital data use.
- The Data Protection Legislation imposes much tougher restrictions on how personal data is used and it applies to organisations who use personal data of individuals in the EU.

GDPR

- Concerned with personal data: any information relating to an *identified* or *an identifiable individual*.
- Enhances the rights of individuals over their own data in an era of ‘Big Data’.
- Focus on potential harm rather than ‘whether a piece of information is shared’.
- Restrictions on collection of data as well as on processing of data that fall under the GDPR regulations.
- Almost all organisations deal with some kind of personal data (e.g. employment records), so effects of its implementation have been widespread.
- Applies to any data collection or processing taking place in the EU.

GDPR

Concerned with personal data:

- address and contact details
- credit scoring history
- correspondence to and from an individual

GDPR

Greater protection for sensitive data:

- sex life or sexual orientation
- genetic data
- biometric data
- racial or ethnic origin data
- political opinions
- religious or philosophical beliefs
- trade-union membership
- data related to physical or mental health

GDPR

Personal data can be processed if at least one of the following conditions ("lawful basis") is met:

- consent has been given explicitly
- contract (e.g. specific steps required before signing)
- information is required by law (legal obligation)
- protection of individual's vital interest
- reasons of substantial public interest (public task)
- legitimate interest

Terminology

Terminology:

- *Data subject*: individual to which the personal data pertains.
- *Data controller*: means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data, i.e. the ‘why’ and ‘how’ of processing personal data
- *Data processor*: means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller.

Enhanced rights of data subjects

Under GDPR, data subjects have the right to:

- access their personal data
- rectification
- erasure
- restrict processing
- data portability
- object to processing
- object to automated decisions
- to be informed

Data protection by design and default

GDPR expects organisations to put data security and privacy foremost when designing new systems and ways of processing:

- Ensure systems are secure and respect privacy rights by design, and from the outset.
- Minimise the collection of personal data to only that which is necessary for the purpose of processing.
- Anonymise and encrypt wherever possible.
- Use privacy enhancing technologies.

Responsibilities and accountabilities

The GDPR creates new and enhanced requirements for organisations to:

- Document their data processing activities.
- Provide evidence of how personal data is protected.
- Be transparent, lawful and fair in their use of personal data.
- Have a legal basis for processing personal data, e.g. consent or a legal obligation.

Principle-based processing

When processing personal data, the following principles should be observed:

- Lawfulness, fairness, and transparency
- Purpose limitation
- Data minimisation
- Accuracy
- Storage limitation
- Integrity and confidentiality
- Accountability

Lawfulness, fairness, and transparency

For UCL this means that.

- To not process data in a way that is unduly detrimental, unexpected or misleading to the individuals concerned; and;
- Be clear, open and honest with people from the start about how UCL uses their personal data.

To meet these requirement: issue privacy notices

At UCL

Lawfulness of processing of personal data: public task

- Teaching and learning for undergraduate and postgraduate courses
- Ethical research
- Ancillary functions to support core purposes

Additional condition needs to be fulfilled for the processing of special category personal data, e.g. ‘legitimate interest’

Breach

- Keep a company record of the breach
- Notify all staff affected
- Notify the relevant department (UCL: Information Security Group) to assess the incident and report it within 72 hours
- Data Protection Officer will notify the relevant supervisory authority.

Compliance

Failure to comply with GDPR:

- A business could be fined up to €10,000,000 or 2% of global income if, for example, they fail to maintain records of processing or report breaches.
- A business could be fined up to €20,000,000 or 4% of global income if the violation relates to fundamental issues (e.g. individuals' rights, or conditions for consent).
- Additionally: rights to audit, warnings, order a controller to comply with a data subject's request; impose a temporary or definitive limitation including a ban on data processing.

Quiz

Go to www.menti.com and use the code 6404 3139

Is it enough?

Wachter 2019:

- A focus on data collection – but no real protection during or after analysis. Ignores the potential for unforeseen threats to privacy can arise after collection owing to inferential analytics.
- GDPR does not regulate how and according to which parameters the data is assessed and evaluated.

Is it enough?

Wachter 2019:

- GDPR does not apply to anonymised data; re-identification risks?
- Unclear whether inferences fall under the definition of ‘personal data’.
- ”Everything is potentially sensitive data, we just do not know it yet” (p.7)

Is it enough?

- Previously, national statistical agencies had the capacity and the mandate to make dissemination decisions.
- They assessed the risk, they understood the data user community and the associated utility from data releases.
- They had the means to address the legal, technical, and statistical issues associated with protecting confidentiality.

Is it enough?

- Researchers would minimise risk in data collection by: anonymising the data and asking human subjects for the consent to use their data.
- Due to ‘Big Data’ these approaches have become obsolete to a certain extent but have not been replaced by an alternative framework.

Is it enough?

- Larger amounts of data increase the risk of identification due to richer detail and larger groups of people having access.
- What is the legal framework when the ownership of data is unclear?
- EU Digital Services Act (October 2022)?

What about open science?

- Open methodology, open source, open data, open access, open peer review, open educational resources.
- But: his goes against the idea of data protection.
- Synthetic data through microsimulation (e.g. Lomax *et al.* 2022)?

Conclusion

- General Data Protection Regulation, Data Protection Act.
- Categories used in legislation (personal, non-personal, sensitive, non-sensitive) do only reflect the nature of the data at the time it is collected, but ignore its subsequent usage and potential transformations. But: ”safe research with sensitive data”?
- Also: the various types of biases embedded within human-generated data are not accounted for within the legal frameworks regulating the use of such data and ethical usage depends on the ethics and moral compass of those using and applying these data and algorithms.

Seminar preparation

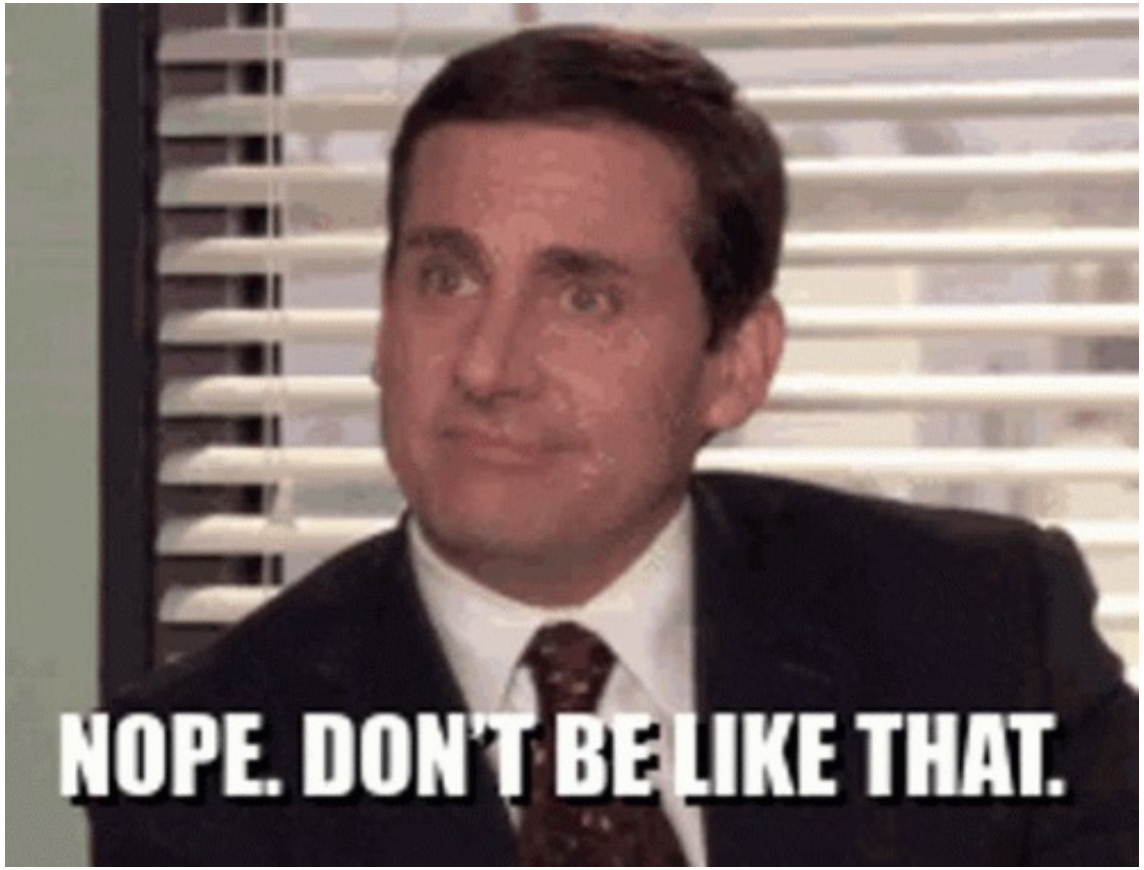
In this week's seminar, we will turn to the final coursework assignment by conducting a Guided Marking exercise.

On Moodle you will find four examples of commentary articles that students submitted in previous years. In preparation for the seminar, please skim-read these four articles and try to assign them a grade. Use the marking matrix to guide you.

Assignment submission

Please submit with your Exam Candidate ID only, e.g. *ABCXYZ.pdf*.

Assignment submission



Continuous Module Dialogue

Go to www.menti.com and use the code 4341 8068

Questions

Justin van Dijk

j.t.vandijk@ucl.ac.uk