Research Data Management & Ethics workshop

EUVIP Methods Skill-Building Workshops

(08/05/24)

Nicolas Bocquet
nicolas.bocquet@uclouvain.be
Office: A241 Leclercq (2nd floor)

Plan

1. Context and historicity of the data protection regulation
2. Definitions
3. GDPR and scientific research
4. Best practices
5. Q&A

Who am I?

- PhD researcher
  - @ Institut de sciences politiques Louvain-Europe (ISPOLE), UCLouvain & University of Geneva
- PhD thesis
  - Privacy in the digital age: comparative analysis of data protection policy dynamics in Belgium, Canada and Switzerland
- Research interests
  - Digital technology and democracy, governance and regulation of technology, privacy, personal data, surveillance, social movements

⚠️ Not a lawyer/legal expert point of view on these issues…

⚠️ Work mainly with qualitative tools and methods

1. Context and historicity of the data protection regulation

1. Context and historicity of the data protection regulation

- Over the last three decades, Internet and ICTs developments have increased the **number of ways in which (personal) data can be collected, stored, processed and shared**



- In this context, **various regulations** – including the famous **2016 European GDPR** – have emerged to regulate personal data collection and processing



- The GDPR is considered to be **the most significant regulatory change worldwide** in terms of privacy/data protection to date



- In Belgium, before the GDPR, the collection and processing of personal data was regulated by the **1992 Privacy Law**



1. Context and historicity of the data protection regulation

- **Scientific researchers are one of the target groups** of these regulatory policies governing the use of personal data
  - “Scientific research” mentioned 2 times in the 1992 Privacy Law
  - “Scientific research” mentioned 33 times in the 2016 GDPR



- The number of **legal requirements for personal data controllers** is increasing, and scientific research is also becoming increasingly regulated



- Initially, the will to ensure that researchers comply with these (new) standards was relatively weak, but **stricter rules on ethics and data protection are now beginning to be imposed trough the main – Belgian and European – funding bodies**



1. Context and historicity of the data protection regulation

- Data protection regulation or the **success of the 'privacy paradigm'**
  - 60s/70s **Cold War context** → liberal democracies start to be afraid of population surveillance in the Soviet Union
  - *"How to prevent an authoritarian shift in liberal democracies (similar as what is happening on the other side of the wall)?"*
  - Meetings of **international experts** (OECD) coming from the liberal tradition → atomistic conception of society (Locke, Mill)
  - A paradigm designed **by and for men**: *private life* vs. *public life*
  - Consensus:
    - ✔ **Privacy as a precondition for democracy** (counter-powers, secret ballot, source of other fundamental freedoms, …)
    - ✔ **The state** (and its armed force, **the police**) = main threat to privacy









1. Context and historicity of the data protection regulation

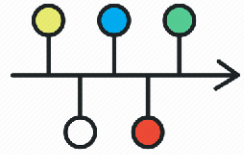




1. Context and historicity of the data protection regulation







- **<u>Problem framing</u> : bureaucratic tendency** to collect more and more personal data (rationalisation process)
- **<u>Objective of all policy since then</u>** : **giving individuals back control** over the personal data that public (and then private) organisations collect and process about them
  - 1973: 1st national data protection law in Sweden
  - 90s web and ICTs developments → need to also regulate private organisations → 1992 first Belgian Privacy Law
  - After 9/11 attacks: ‘security’ vs ‘privacy’
  - 2000s: surveillance capitalism
  - 2013 Snowden revelations → new impetus for regulation
  - 2016: EU GDPR

1. Context and historicity of the data protection regulation

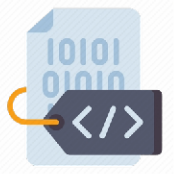


2. Definitions

2. Definitions

- **Personal data**: “any information related to an identifiable person”
  - Name, phone number, address, date of birth, e-mail address, voice recording, national identification number, …
    - ⚠ Cross-referencing data often reveals identity: e.g. age + gender + address → identification of the person
    - ⚠ Trivial digital data can reveal identity: e.g. IP address or mac address
- **Metadata**: “any data that provides information about other data"
  - It can be technical (time, location, camera of an image in the file) or manual (additional information written in relation to an interview)
    - ⚠ Metadata can also easily reveal a person's identity





2. Definitions

- **Encryption**: “cryptographic method that makes it impossible for anyone who does not have the encryption key to understand a document”



⚠ The aim of cryptography is not to hide information, but to protect it from unwanted access



⚠ Encryption algorithms are not unbreakable... but they do protect your data for many years to come → always ask yourself "*what happens if my research data leaks?"*

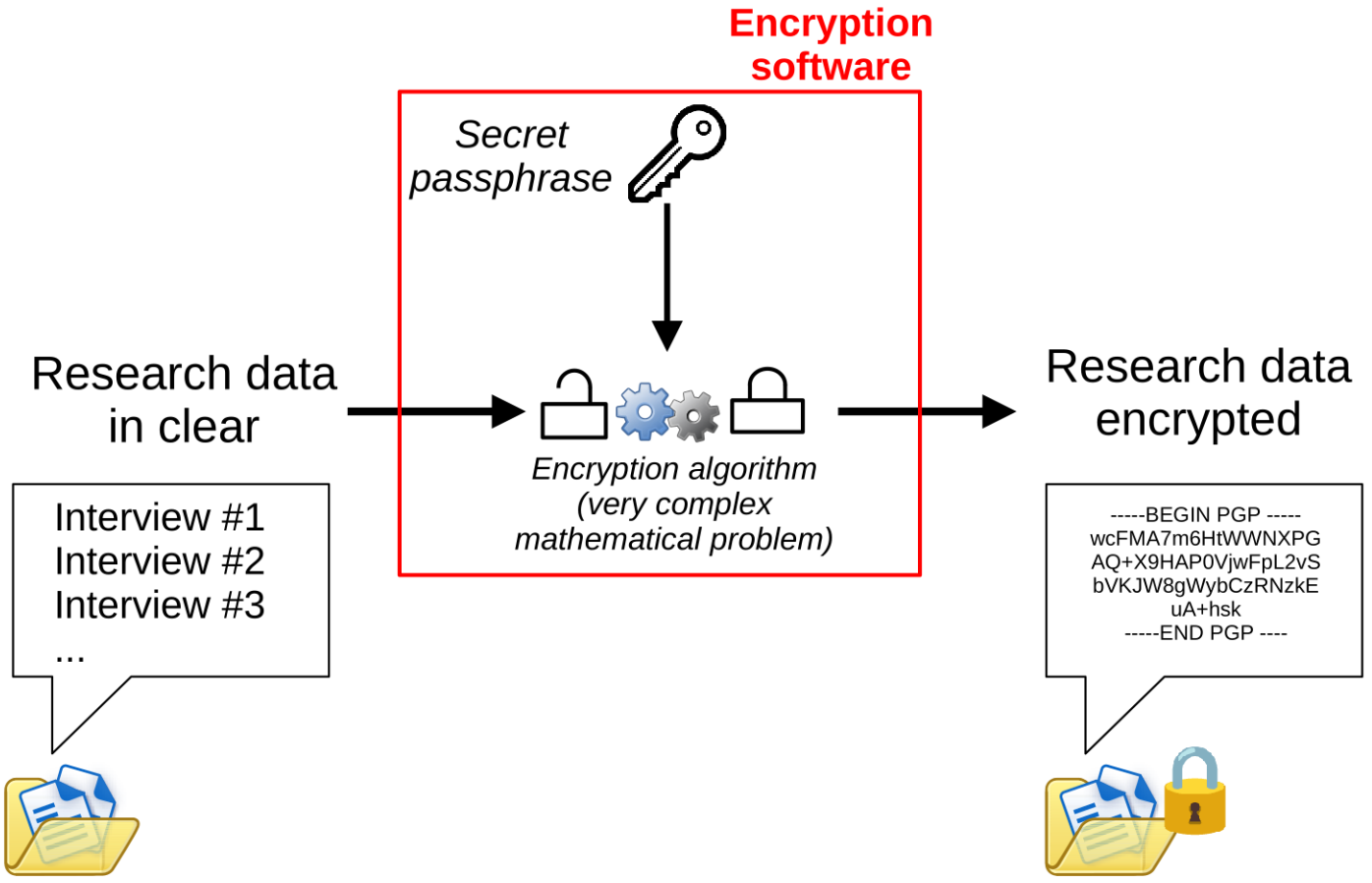


⚠ Essential to use *free* and *open-source* encryption software (e.g. *VeraCrypt*), otherwise you can't be sure that there isn't a backdoor in the software that would allow to access the data without the passphrase





Encryption


Encryption software
Secret passphrase
Research data in clear
Interview #1
Interview #2
Interview #3
...
Encryption algorithm (very complex mathematical problem)
Research data encrypted
-----BEGIN PGP -----
wcFMA7m6HtWWNXPG
AQ+X9HAP0VjwFpL2vS
bVKJW8gWybCzRNzkE
uA+hsk
-----END PGP ----

2. Definitions

- **Pseudonymisation**: “de-identification procedure by which the personally identifiable information are replaced by one or more artificial identifiers or pseudonyms”



⚠️ Data pseudonymisation is not always enough to protect a person's identity



⚠️ Cross-referencing pseudonymised data with other data often reveals identity → need to encrypt data and store it on different devices



- **Anonymisation**: “process of <u>permanently</u> removing personally identifiable information from data sets, so that the people whom the data describe remain anonymous”



⚠️ Cross-referencing anonymous data with other data often reveals identity → need to encrypt data and store it on different devices



⚠️ Very difficult to permanently delete part of the data so that it can be considered anonymous forever



3. GDPR and scientific research

3. GDPR and scientific research

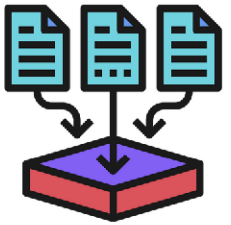


General Data Protection Regulation

- Very long European regulation adopted in 2016 and implemented since 2018
- Considered as **the most heavily lobbied European text** in the history of the EU, in particular from the GAFAM
- The GDPR policy process already started before the 2013 Snowden revelations → willingness to **regulate further personal data use after Snowden**
- While the text obviously targets GAFAM, it puts **everyone on an equal footing, including researchers** (EU was afraid of a possible economic backlash from the US)

3. GDPR and scientific research

Purpose of processing (art. 5), purpose limitation (art. 89) and consent form



"*Personal data shall be collected* ***for specified, explicit and legitimate purposes*** *and* ***not further processed in a manner that is incompatible with those purposes*** *('purpose limitation')* […]"

- Need for a **consent form** in which:
  - The **purpose of processing** is specified (not allowed to lie unless there is a methodological exception)
  - The **duration of storage** is indicated (e.g. 4 years for my PhD research) → you can't keep data forever
  - The **possible re-use of personal data** for similar scientific purposes is specified (provide as much information as possible about other possible re-use) → without this, you cannot re-use the personal data collected for other research...



3. GDPR and scientific research

Sensitive data (art. 4 + 9)



- The following personal data is considered ‘sensitive’ and is **subject to specific processing conditions**:
  - Personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs;
  - Trade-union membership;
  - Genetic data, biometric data processed solely to identify a human being;
  - Health-related data;
  - Data concerning a person’s sex life or sexual orientation
  - (To a certain extent, data collected on children)



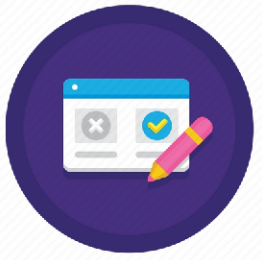
- Sensitive data can be collected by researchers **only if there is explicit consent for this particular processing** in the consent form from the data subjects



- **Need to be very careful with sensitive data** because 1) data subjects may be exposed to repression and 2) researchers face also the greater risk of being prosecuted for mistakes involving this type of data...

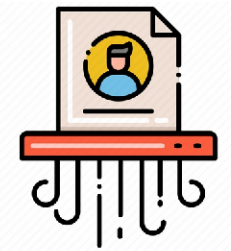
3. GDPR and scientific research

Right to rectification (art. 16) and right of erasure (art. 17)



"The data subject shall have the right to obtain from the controller ***without undue delay*** *the* ***rectification of inaccurate personal data*** […] *the right to have incomplete personal data completed, including by means of* ***providing a supplementary statement****"*

- Data subjects have the **right to request rectification** of their personal data or to **provide additional information** (e.g. after an interview)



"The data subject shall have the right to obtain from the controller the ***erasure of personal data*** *concerning him or her* ***without undue delay*** […]"

- Data subjects may **withdraw consent** → obligation to **erase** personal data **UNLESS erasure prevents research** (e.g. longitudinal studies with 4 people to see changes in voting behaviour, etc.) → trade-off between individual interests and research
- **Children** must have a **facilitated procedure** for withdrawing their consent when they become adults

3. GDPR and scientific research

Security of processing (art. 32)



"[…] *the controller and the processor shall implement* ***appropriate technical and organisational measures to ensure a level of security appropriate to the risk****, including inter alia as appropriate:*

- *The* ***pseudonymisation*** *and* ***encryption*** *of personal data;*
- […]"



- **Data minimisation principle** (collect only the data you need)
- **Anonymity protection** (unless it is necessary to know the person's identity for research purposes) → non-anonymous storage is acceptable for the purpose announced (with encryption security), but erasure or anonymisation is then mandatory



⚠️ Security of personal data in the **cloud** → more difficult to control who has access to online data → encrypt data before putting it online



3. GDPR and scientific research



<u>Data sharing (e.g. international research team)</u>

- Obligation to **inform data subjects of sharing in the consent form**

- Need to guarantee **data security** (encryption to ensure that only you and those authorised have access to the data)



- Not allowed to share personal data from Europeans to **countries that do not have the 'same level of protection'** (EEA + GDPR adequacy)



⚠️ US does not have an adequacy decision for the moment…



⚠️ Possibility of sharing data with countries that do not have the same level of protection (or adequacy decision) only **if data subjects give their explicit consent**



4. Best practices



- Collect the **minimum amount of personal data** necessary for the research



- **Anonymise all research data you can** that does not need to be identified with an individual



- **Encrypt all research data** to limit risks



- Put only the minimum amount of personal data online in the cloud, always **prioritise local external hard drives**



- Don't travel every day with **data you don't need every day**



- Implement these data protection best practices **as early as possible** to avoid wasting time during the research process later and to limit unpleasant surprises…



- **<u>Safest practice</u>**: use **encrypted external hard drives**, **locked** at home or in the office (in two different locations if possible) and **used only for research data** (no problems with online interception or hacking and, in the event of physical access, e.g. in case of robbery, personal data is protected by encryption)

Thank you for listening!

