École Pour l'Informatique et les Techniques Avancées – EPITA

Masters program

Course: Data Privacy by Design

Created & Coordinated by: M. Salman Nadeem

Data Privacy by Design (DPbD)

Course schedule (tentative)

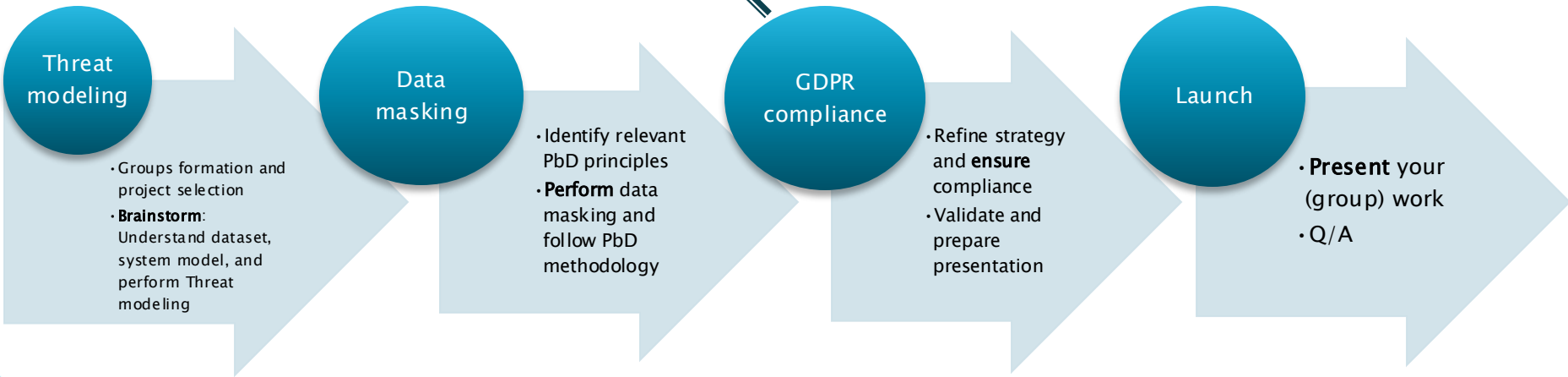
Date	No.	Topics	Duration (in hours)
(check Zeus platform for the date & time)	1	Introduction, DPbD fundamentals with case studies	3
(check Zeus platform for the date & time)	2	Data privacy risks, Crypto. Package, Data masking (Anonymization vs Pseudonymisation)	3
(check Zeus platform for the date & time)	**3**	**Privacy Enhancing Technologies (PETs), DPbD and General Data Protection Regulation (GDPR)**	**3**
(check **Zeus** platform for the date & time)	4	Recap, Conclusion	3
(check **Zeus** platform for the date & time)	5	Final evaluation	3
		Total	*15 hours*

GRADING criteria:
Class participation comprising attendance & reactivity: 10%
Class activities (quizzes): 30%
Final project (individual report & group presentation): 60%

Lecture 3 Outline

1. Note on Data masking
2. GDPR: Introduction & key definitions
3. GDPR: Scope and other aspects

QUIZ



Hands-on → Final project

PET (soft): Differential privacy

- Noise addition using a single value: epsilon (ϵ), which is a measure of how private a data release (output) is
  - Higher values of ϵ gives accurate, less private answers
  - Low-ϵ systems give highly random answers
- The outcome of any analysis on output dataset is essentially equally likely, independent of whether any individual joins, or refrains from joining, the input dataset; Used by: Apple, Microsoft, Google, Uber …

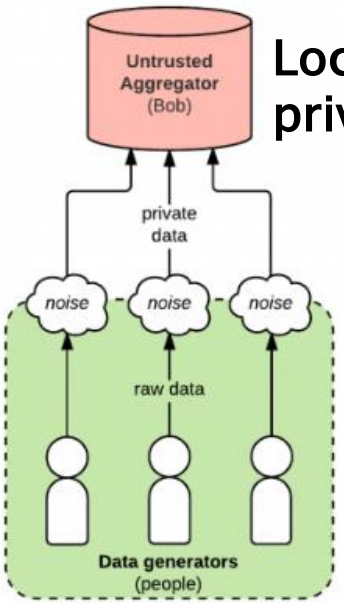
Formal expression

$$\Pr[\, A(D1) \in S \,] \leq \exp(\epsilon) \cdot \Pr[\, A(D2) \in S]$$

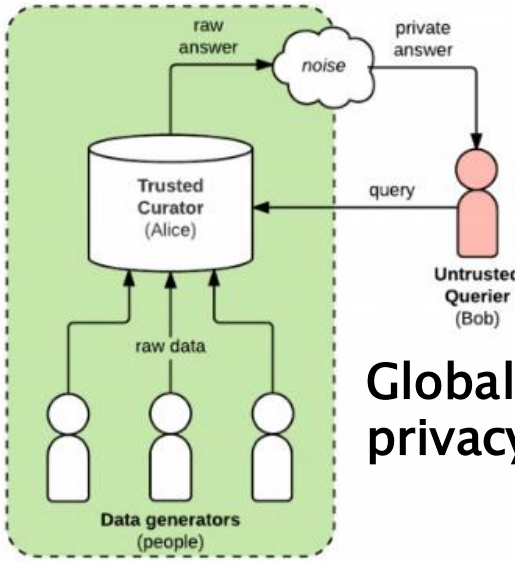
Two data sets: $\boldsymbol{D_1}$, $\boldsymbol{D_2}$
Mechanism/Algorithm: $\boldsymbol{A}$
All events/subsets: $\boldsymbol{S}$

The **algorithm** $\boldsymbol{A}$ is said to provide e-differential privacy, for all datasets (D_1, D_2), that differ on a single element (i.e., the data of one subject)…

$\boldsymbol{A}$ introduces randomness, such that we get epsilon (**ϵ**) differential privacy



Local privacy



Global privacy

PET (hard): Tor/Panoramix

LOW LATENCY

Cannot resist Global Adversary (assumes adversary cannot see both edges)

Web browsing, Instant Messaging, streaming

HIGH LATENCY

MIXMASTER / MIXMINION

Global Adversary resistance at the cost of latency (and long term patterns revealed)

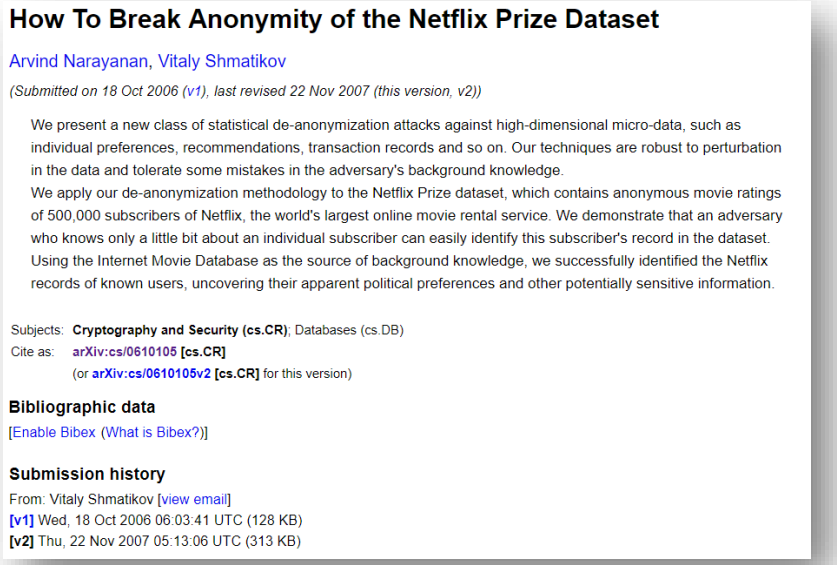
Email, Voting

Other examples: I2P, freenet

Panoramix
THE BEST OF BOTH WORLDS

Not all techniques work for all cases! (1/3)

- Netflix [Competition 'Prize' (2006)]
  - Competing teams had to create an algorithm to predict user ratings for films
  - Provided dataset included ~100M ratings, ~480k users for ~17k movies
    - Anonymization:
      - Replaced name of users with random chars
      - Replaced random ratings with fake one's



How To Break Anonymity of the Netflix Prize Dataset

Arvind Narayanan, Vitaly Shmatikov

(Submitted on 18 Oct 2006 (v1), last revised 22 Nov 2007 (this version, v2))

We present a new class of statistical de-anonymization attacks against high-dimensional micro-data, such as individual preferences, recommendations, transaction records and so on. Our techniques are robust to perturbation in the data and tolerate some mistakes in the adversary's background knowledge.
We apply our de-anonymization methodology to the Netflix Prize dataset, which contains anonymous movie ratings of 500,000 subscribers of Netflix, the world's largest online movie rental service. We demonstrate that an adversary who knows only a little bit about an individual subscriber can easily identify this subscriber's record in the dataset. Using the Internet Movie Database as the source of background knowledge, we successfully identified the Netflix records of known users, uncovering their apparent political preferences and other potentially sensitive information.

Subjects: **Cryptography and Security (cs.CR)**; Databases (cs.DB)
Cite as: **arXiv:cs/0610105 [cs.CR]**
(or **arXiv:cs/0610105v2 [cs.CR]** for this version)

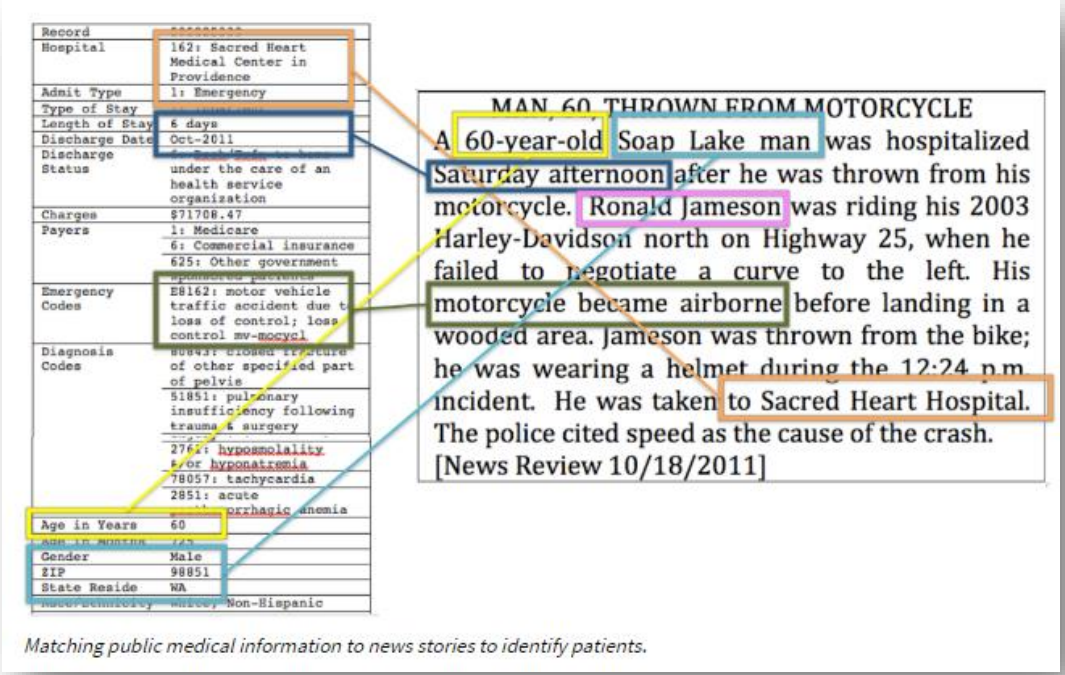
Bibliographic data
[Enable Bibex (What is Bibex?)]

Submission history
From: Vitaly Shmatikov [view email]
[v1] Wed, 18 Oct 2006 06:03:41 UTC (128 KB)
[v2] Thu, 22 Nov 2007 05:13:06 UTC (313 KB)

2007 -> Researchers successfully denonymized the Netflix dataset by combining it with the data of IMDB (Linkage attack)

Not all techniques work for all cases! (2/3)

- Another example of re-identification from the Journal of Technology Science that
  - An "anonymous" medical record is cross-referenced with a newspaper brief about a motorcycle crash
  - Patient in question is identified



Matching public medical information to news stories to identify patients.

Ref. https://techscience.org/a/2015092903/

Not all techniques work for all cases! (3/3)

- Many possible attacks exist!
  - Background information attack
  - Unsorted matching attack
  - Complementary release attack
  - Temporal attack
  - …

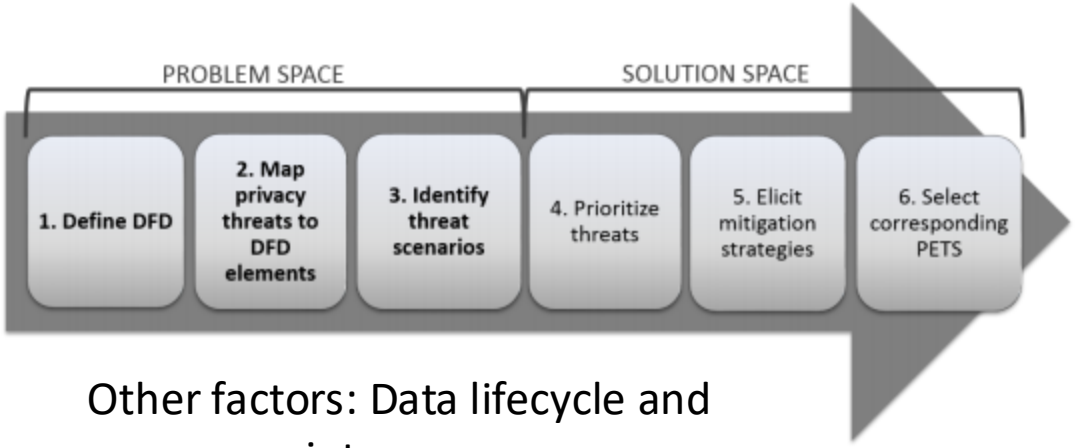
Carry out independent audits / reviews to ensure that the anonymized data-set is not vulnerable to de-anonymization attacks!

Note about Systematic approaches
→ For DPbD Activities 3 and 4

Suggestion: Use the LINDDUN Methodology when time and resources permit.

STRIPED	Spoofing	Tampering	Repudiation	Information disclosure	Privacy	Elevation of privilege	Denial of service
LINDDUN	Linkability	Identifiability	Non-repudiation	Detectability	Disclosure of information	Unawareness	Non-compliance

Scientific renown
Industry acceptance:
(ISO 27550, EDPS PbD opinion, ENISA PbD)



Other factors: Data lifecycle and maintenance, …

Note: Without sufficient security controls, all data privacy protections/guarantees will be ineffective!

General Data Protection Regulation (GDPR)

- A legal framework applicable directly in EU countries
  - Came into effect on **25th May 2018**
  - Repealed the previous European Directive (95/46/EC) from 1995 on Data Privacy
- Complemented by:
  - Do-not-call-me list
  - National Register Number
  - E-commerce laws
  - Cookie policy
  - ePrivacy
  - …

Why GDPR?

- Motivations behind the regulation:
  - Your data belongs to you!
  - It's a legitimate expectation that companies handle data with care
  - Companies must adapt to work with only the personal data they need for relevant purpose(s)
- Better control & enforcement

"Only a minority (15%) feel they have complete control over the information they provide online"

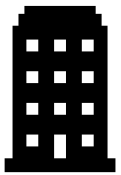
Special Eurobarometer 431

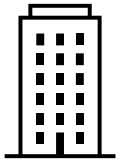
Source: https://data.europa.eu/data/datasets/s2075_83_1_431_eng?locale=en

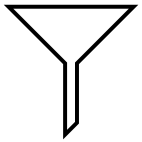
Key definitions & Roles

- **Personal data OR Personally Identifiable Information (PII):** Any information relating to an identifiable natural person (that can be directly or indirectly identified in particular by reference to an identifier)
- **Special categories of personal data** ("sensitive" data):
  - Racial or ethnic origin
  - Political opinions
  - Religious or philosophical beliefs
  - Trade-union membership
  - Data concerning health, sexual orientation, …
  - Genetic or biometric data

- **Data Subject**: an individual who is the subject of personal data = any individual consumer
- **Data Controller**: an entity that determines the purposes and means of processing personal data
- **Data Processor**: an entity responsible for processing personal data on behalf of a controller
- What is **Data Processing**?
  - Collecting, recording, holding, transferring or deleting personal data are examples
  - Carrying out any operation or set of operations on personal data

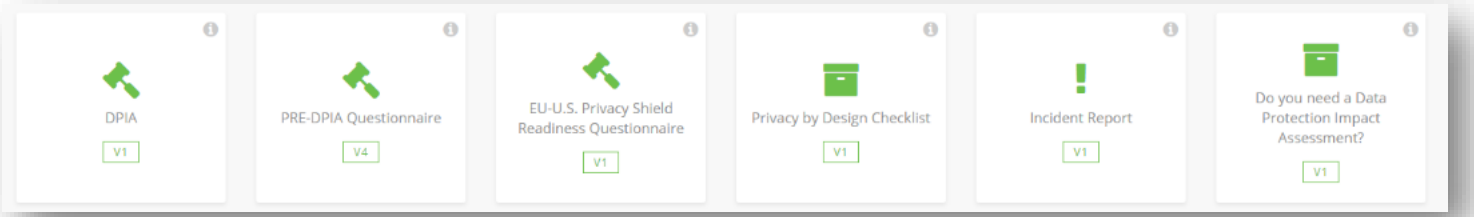






DPIA – Data protection impact assessment

- What?
  - Report assessing the risks and evaluating the technical and organizational measures
- When?
  - GDPR mandates a DPIA be conducted where data processing "is likely to result in a high risk to the rights and freedoms of natural persons"
  - Example tool (to conduct DPIA):



OneTrust – GPDR management tool

DPO, DPA and EDPB

- Data protection officer (DPO)
  - Official role as part of the accountability framework of the GDPR
  - Mandatory under certain circumstances e.g., Hospital: processing large sets of sensitive data
- Reports to national Data Protection Authority (DPA)
  - DPAs are co-operated by European Data Protection Board (EDPB): an independent European body whose purpose is to ensure consistent application of GDPR in EU

CNIL.
COMMISSION NATIONALE
INFORMATIQUE & LIBERTÉS

edpb
European Data Protection Board

Scope

- Apply to company or organization controlling or processing personal Data of EU residents
  - Where no EU presence exists, the GDPR will still apply whenever:
    1. An EU resident's personal data is processed in connection with goods/services offered to him/her
    2. The behavior of individuals within the EU is *"monitored"*

1 – Lawful basis for processing

- There are six available lawful bases for processing
  - No single basis is 'better' or more important than the others (and it all depend on your purpose and relationship with the data subject)

1. Consent
2. Contract
3. Legal obligation
4. Vital interests
5. Public task
6. Legitimate interests

You must determine your lawful basis before you begin processing, and you should document it

Your privacy notice should include your lawful basis for processing as well as the purpose(s) of the processing

2 – Data Privacy by design & by default

"The privacy of the data subject is taken into account from the start of the conception of products and services"

- Identify & Implement technical and organizational measures that protect personal data and apply the GDPR principles from start

"The standard options are privacy-friendly"

- Set to service/program settings/controls to the most privacy-friendly state by default e.g.,
  - A minimal amount of personal data is requested and processed
  - Permission is requested before processing personal data
  - …
- Covered thoroughly in earlier classes:
  - Data PbD: Objective & Strategies
  - Roadmap: Three phases:
    1. Define functionality & Elicit requirements, Draw a high-level system diagram
    2. Perform (4) activities
    3. Full-lifecycle protection: Repeat (when necessary), agile manner

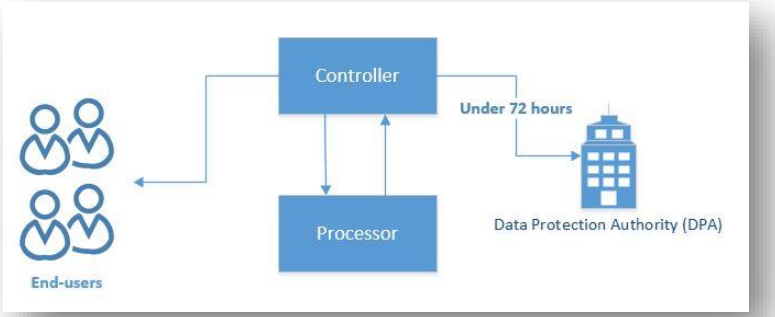
3 – Individual (subject) rights

- The procedures in place should ensure that they cover all the individual rights (in view of the set Legal basis):

	Right to access	Right to rectification	Right to erasure	Right to restriction	Right to portability	Right to Object
Consent	✓	✓	✓	✓	✓	~ (can withdraw consent)
Contract	✓	✓	✓	✓	✓	✗
Legal Obligation	✓	✓	✗	✓	✗	✗
Vital interests	✓	✓	✓	✓	✗	✗
Public task	✓	✓	✗	✓	✗	✓
Legitimate interests	✓	✓	✓	✓	✗	✓

- Response time should be no longer than one month (exceptions will have to be justified)
  - This necessitates technical implementations or readiness to handle data subject requests in due time

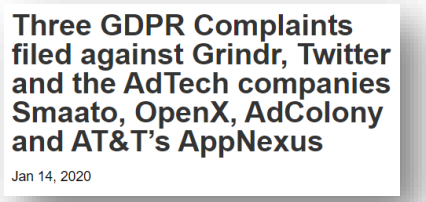
4) **Accountability and governance** has been elevated to a significantly greater scale, few necessary documentation examples are:
   - Documentation (Processing registry, …)
   - Data protection officers (DPO) (point of contact)
   - Contracts (Data processing agreement, …)
   - Data protection impact assessments (DPIA)
   - …
5) **Cross-border data transfers**: Any data processor/controller outside EU, will have to comply by GDPR
6) **Security**: GDPR requires personal data to be processed in a manner that ensures its security
7) **Data Breach**: GDPR introduces a duty on all organizations to report certain types of personal data breach to the relevant DPA under 72 hours (and also data subjects if the breach affects their personal data with necessary information and steps they should take)
   -> Record of any personal data breaches, regardless of whether you are required to notify or not, must be kept



8 – Sanctions & Fines

- Hard sanctions and fines can be imposed e.g., Periodic data protection audits
  - A fine up to €10 million or up to 2% of the annual worldwide turnover of the preceding financial year in case of an enterprise, whichever is greater e.g. where there has been an infringement of a given article(s) clauses of the GDPR
  - 2nd Level: In case of not reporting a data breach with-in the given time-frame
    - E.g., a fine up to €20 million or up to 4% of the annual worldwide turnover, whichever is greater
  - A warning in writing in cases of first and non-intentional non-compliance can given by the DPA too
  - …

Source: https://edition.cnn.com/2019/07/08/tech/british-airways-gdpr-fine/index.html



Source: https://noyb.eu/en/three-gdpr-complaints-filed-against-grindr-twitter-and-adtech-companies-smaato-openx-adcolony-and

Source: https://noyb.eu/en/irish-data-protection-authority-gives-eu-397-billion-present-meta

9 – Other aspects

- Children between the ages of 13 and 15 (inclusive), can provide their own consent only via their legal guardians (whoever holds parental responsibility for the child)
- Unclear areas:

Source

GDPR continues to evolve…

GDPR Mantra!

Data subjects rights have been widened!	• Adopt both your technical and administrative work-flows to orient them with data privacy at the core!
Document everything!	• Data handling procedures, incident response procedures, privacy by design procedures, data subject requests handling procedures…
Stay prepared!	• Design both your technical and administrative procedures to use them efficiently and effectively

Project: Data inventory and GDPR compliance

Individual report:

- **Section 6: Propose applicable solutions** for the threats outlined in section 3, leveraging your knowledge in Cryptography, Privacy Enhancing Technologies (PETs) and other references (e.g., OWASP Top 10)
- **Section 7**: In view of the goals and objectives of your project (company):
  1. Create a **spreadsheet of data inventory** (covering PII in all repositories e.g., Database, Logfiles, Backups) with following columns:

Personally Identifiable Information (PII)		Purpose (Legal basis for processing e.g., consent, contract necessity, legal obligation)
Direct identifiers	Quasi identifiers	

  2. **Determine data processing activities** and **fill** data processing register:
     → **Sheet number 2:**

Processing details				Purpose of the data processing	Special categories of personal data?
Name of the processing operation	N° / REF	Date of creation of the record form	Last update of the record form		Yes/No
(EXAMPLE) Payroll Management	1-Example	May 25, 2018	May 13, 2018	Payroll management, Calculation of remuneration, Calculation of the amount of payments sent to social security institutions.	No

  3. Write a **concise 1-page Privacy Policy** and **1-page Terms of Use**
  4. Propose a **privacy-focused strategy for obtaining and managing consent** for data processing e.g., a short textual description or a wireframe diagram (example)

NOTE: Collaborate on brainstorming as a group, but avoid copying and pasting your group members' work.

Lecture 3 ends here

- Course Slides: Go to MS Teams:
  'Data Privacy by Design'
  → "Files section"

- Send your questions by email OR via direct message using MS Teams

- Thank You!