GDPR – Challenges for Reconciling Legal Rules with Technical Reality

Mirosław Kutyłowski,[1] Anna Lauks-Dutka,[1] Moti Yung[2]

[1] Department of Fundamentals of Computer Science, Wrocław University of Science and Technology, Wrocław, Poland

[2] Columbia University, New York, USA and Google LLC

ESORICS 2020

General Data Protection Regulation

M. Kutyłowski,
A. Lauks-Dutka,
M. Yung

What is the GDPR?

- key element of **information protection**
- **strict legal requirements**, aimed to guarantee **control over processing** personal data
- privacy **by-design**, **revocable** consent to process data

effectively in force since May 2018

Why so important?

- activities in EU
- offering services and good to users in EU (regardless of origin)

M. Kutyłowski,
A. Lauks-Dutka,
M. Yung

Data Implicit Metastructure

Data Subject

the person whose data is processed

Legal Basis

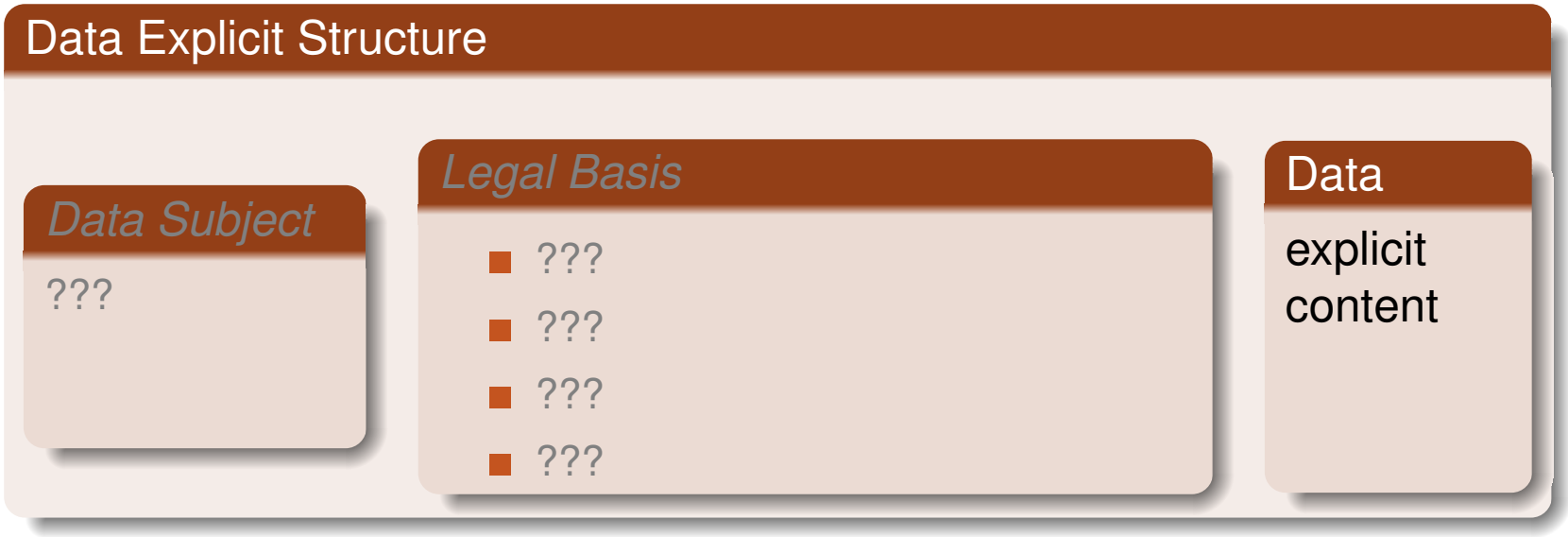
- consent of data subject, or
- legal obligation, or
- ...

Data

explicit content

- easy **testing legality** of data processing (who, what, how to process)

M. Kutyłowski,
A. Lauks-Dutka,
M. Yung

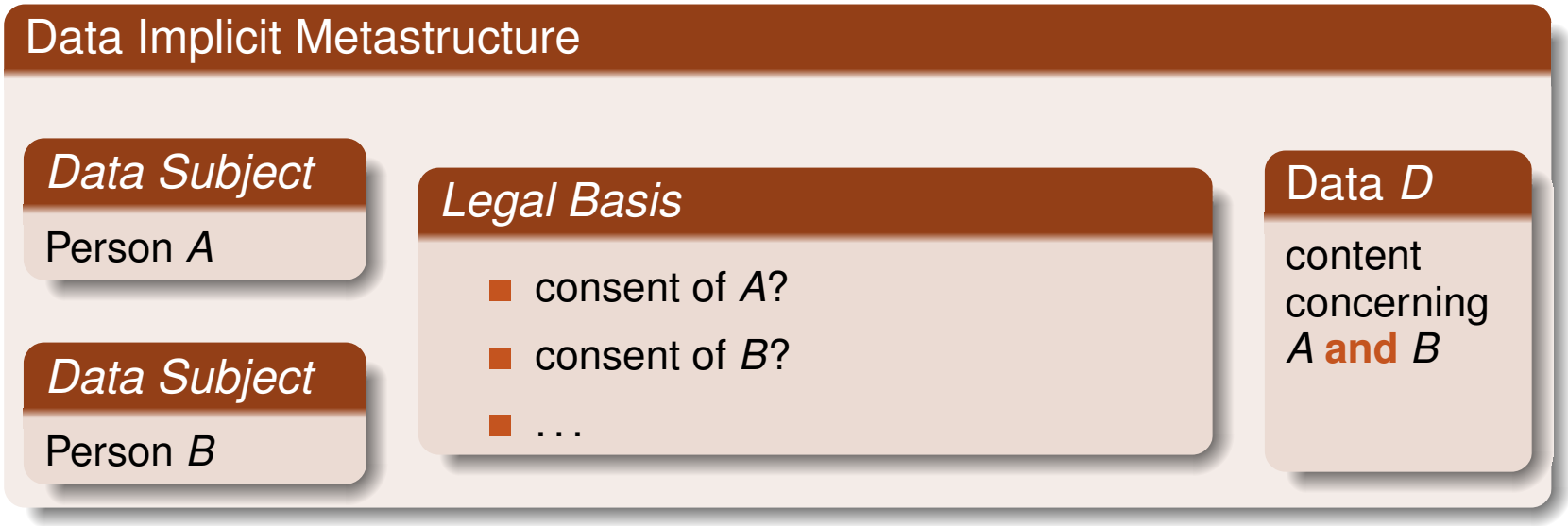


Belief of GDPR authors:

- **implicit components should be easy to derive** from the data or its context
- **only a reasonable effort** necessary

GDPR Challenges

Shared personal data



- data D concerns identifiable data subjects A and B
- one **cannot split** D into D_A (concerning A) and D_B (concerning B) without changing semantics[1]

[1] examples to come

M. Kutyłowski,
A. Lauks-Dutka,
M. Yung

Consent problem

1. a **consent** of which party (A or B or (A and B)) is **required to process** D according to GDPR?
2. what to do if there are consents' **discrepancies**?

Example:

A* requests to store *D, while ***B* asks to remove *D***
Deadlock:

option 1: D is erased following the request of $B \Rightarrow$ the right of A to protection of her data from erasure is **violated**

option 2: D is kept following the request of $A \Rightarrow$ the right-to-be-forgotten of B is **violated**

option 3: consent of all data subjects is required to process $\Rightarrow$ both storing and erasing are **illegal**

Solution attempt - reverting to the basic model

M. Kutyłowski,
A. Lauks-Dutka,
M. Yung

Splitting process

Split a dataset into data chunks so that

- **a single data subject** for each chunk,
- **the original semantics is preserved**

Problems

- the conversion process should be **automatic** or semi-automatic
- the data might be logically inseparable **without changing semantics**

Pseudonymization as a silver bullet?

- if there are k data subjects in data D, then create ***k* pseudonymized copies of *D***
- all data subjects **pseudonymized** in D_A **except for** data subject A

Semantically neutral pseudonymization

M. Kutyłowski,
A. Lauks-Dutka,
M. Yung

Ineffective pseudonymization problem

It might be infeasible...

Example

A medical record:

"Alice suffers the same symptoms as her brother Bob"

Pseudonymization attempts

- *"Bob"* $\rightarrow$ *pseudonym X*:
  "Alice suffers the same symptoms as her brother *X*"
  - if Alice has a single brother, *X* becomes an **identifiable** person, so pseudonymization is not effective
- *"her brother Bob"* $\rightarrow$ *pseudonym X*:
  "Alice suffers the same symptoms as *X*"
  - **semantic difference** - we lose the genetic context
- *"Y suffers the same symptoms as her brother X"*:
  - **useless** for the medical treatment of Alice

Implicit personal data processing

M. Kutyłowski,
A. Lauks-Dutka,
M. Yung

Example

1. Alice and Bob have published: *"Alice and Bob earn together x EUR"* in a public dataset D_1.
2. Later Alice gives her consent to publish M': *"Alice earns y EUR"* in a public cloud D_2 run by P_2.

Implicit personal data processing

M. Kutyłowski,
A. Lauks-Dutka,
M. Yung

Example

1. Alice and Bob have published: *"Alice and Bob earn together x EUR"* in a public dataset D_1.
2. Later Alice gives her consent to publish M': *"Alice earns y EUR"* in a public cloud D_2 run by P_2.

Problems

- can the provider P_2 of D_2 **publish** M' following the request of Alice? – **this would mean publishing information *"Bob earns* $x - y$ *EUR"***
- is P_2 obliged to perform a **semantic analysis** of the request having in mind **privacy violations of third parties**?
- what is the necessary **scope** of the semantic analysis?
- how far is P_2 **responsible** for misclassification?

Classification as "personal data"

M. Kutyłowski,
A. Lauks-Dutka,
M. Yung

Personal data

D falls into the category of **personal data** $\iff$ it concerns an **identifiable person**

Relatively easy case: positive decision

data containing **explicit identifiers** of data subjects $\Rightarrow$ *personal data*

Frequently hard case: negative decision

a proof of impossibility of identification **is not just an absence of explicit identifiers** in the data

impossibility proofs are generally harder as they concern **all possible ways of identification**

M. Kutyłowski,
A. Lauks-Dutka,
M. Yung

Decision context

what is the right **decision context** from the point of view of GDPR:

1. just the data by itself?
   - easy to implement but so trivial to violate GDPR goals without violating the rules
2. data in the context of all existing datasets?
   - simply unrealistic
3. information available to the data processor?
   - what is available? E.g. in case of an unlimited access to database run by a **third party**?

Personal Data and Encryption in the European General Data Protection Regulation, Gerald Spindler, Philipp Schmechel, *Journal of Intellectual Property, Information Technology and E-Commerce Law*, 2016 **no reaction!**

M. Kutyłowski,
A. Lauks-Dutka,
M. Yung

Temporal validity of a decision

how to insure that the classification *personal/non-personal* is **up to date**:

- continuous monitoring?
- periodic monitoring?
- event driven?

Consequences of processing "non-personal data"

M. Kutyłowski,
A. Lauks-Dutka,
M. Yung

Processing non-personal data

GDPR does not concern processing non-personal data
so their processing is not restricted by GDPR?

Consequences of processing "non-personal data"

M. Kutyłowski,
A. Lauks-Dutka,
M. Yung

Processing non-personal data

GDPR does not concern processing non-personal data
so their processing is not restricted by GDPR? **NO!**

Non-personal → personal data conversion and transfer problem

Party *A*

Non-personal dataset D

transfer of D →

Party *B* (Country with no GDPR)

1. $D' :=$ De-anonymization(D)
2. publish personal data D'

Can anybody be accused of GDPR violations?

- **neither** party *B*
  - – as long as *B* does not offer goods or services in Europe
- **nor** party *A*
  - *A* has not transferred any personal data

Data aggregation and anonymization

M. Kutyłowski,
A. Lauks-Dutka,
M. Yung

Example

- party *A* holds a dataset *D* containing personal data of its clients collected according to GDPR
- *A* **aggregates** data *D* by computing the average amount of money spent by the clients of *A*

Data aggregation and anonymization

M. Kutyłowski,
A. Lauks-Dutka,
M. Yung

Example

- party A holds a dataset D containing personal data of its clients collected according to GDPR
- A **aggregates** data D by computing the average amount of money spent by the clients of A

Challenges

1. does the **result** of an aggregation operation fall into the category of **personal data**?
2. more general: at which moment the aggregated data **looses** its attribute **personal data**?
3. is aggregation **processing** of **personal data** (assuming that its inputs are personal data)?

Data aggregation and anonymization

M. Kutyłowski,
A. Lauks-Dutka,
M. Yung

Example

- party A holds a dataset D containing personal data of its clients collected according to GDPR
- A **aggregates** data D by computing the average amount of money spent by the clients of A

Challenges

1. does the **result** of an aggregation operation fall into the category of **personal data**?
2. more general: at which moment the aggregated data **looses** its attribute **personal data**?
3. is aggregation **processing** of **personal data** (assuming that its inputs are personal data)?

Detailed problem - **the Right-to-Anonymize Data**

Is it legal from the point of view of GDPR to create a dataset $Anon(D)$ by **anonymization of all data records of D**?

Reality of GDPR

Reality of GDPR

Business

- many efforts aiming to achieve compliance with GDPR
- ... or at least **collect arguments about *due diligence*** for the case of a conflict with data protection supervision authorities

- some branches of information processing industry in a state of **paralysing legal risks** – e..g. **AI companies in Europe**

Reality of GDPR

Supervising authorities

- limited guidance on “how to implement GDPR” and interpretation of its rules,
  ... well, this is not an easy task as we have seen
- **threat of misusing** power for particular economical and political advantage
- **strict position of supervising authorities** – example: EDPS versus EASO and a decision banning processing (anonymized) data by EASO (fighting smugglers)
- **... and tolerance elsewhere** – e.g. paparazzi . . .

Reality of GDPR

Protection level achieved

- GDPR called a **Paper Tiger** – useless against severe violations by clever adversaries,
- ... **not much real impact and improvement** of situation of an **average person**,
- ... **but** annoying questions about cookies, problems to access information, ...

to some extent GDPR is busy with problems created by GDPR

Reality of GDPR

M. Kutyłowski,
A. Lauks-Dutka,
M. Yung

GDPR as a Holy Grail

with a few exceptions[a] the IT community is passive:

- R&D on how to comply with the GDPR
- almost no critics and feedback to the authorities

[a] e.g. Center for Data Innovation

GDPR as an evolving regulation

EU report from June 24:

> *Communication - two years of application of the General Data Protection Regulation*

`https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0264`

EU Report

M. Kutyłowski,
A. Lauks-Dutka,
M. Yung

Report focus

- **complimentary regulations** in EU – **incompatibility problems still unresolved** – the devil is in details
- **supervisory authorities** – it seems that **cooperation need to be improved** – different approaches, unharmonized guidelines, ...
  a company active in many EU countries – which guidelines to follow?
- **no representatives in EU** – of companies offering goods and services in EU (they are obliged to have!)
  no problem for big corporations to have representatives, for SME a serious cost
  To become compliant the simplest solution would be to block IP addresses from EU.
  Are we going to build a European Wall?

EU Report

M. Kutyłowski,
A. Lauks-Dutka,
M. Yung

Report's section "The application of the GDPR to new technologies"

The GDPR, having been conceived in a technology neutral way, is based on principles, and ***is therefore designed to cover new technologies as they develop.***

Our opinion: as we have shown, **the GDPR model is a severe limitation for development of new technologies.**

It is seen as an ***essential and flexible tool*** *to ensure that the development of new technologies is in compliance with fundamental rights.*

Our opinion: we are **not convinced about flexibility**, definitely a proper implementation of the GDPR principles requires deep rethinking many elements of IT systems – **the process might be very costly and time consuming**

EU Report

M. Kutyłowski,
A. Lauks-Dutka,
M. Yung

Report's section "The application of the GDPR to new technologies"

> *The data protection and privacy legislative framework* ***proved its importance and flexibility during the COVID-19 crisis****, notably in relation to the design of the tracing apps and other technological solutions to fight the pandemic.*

Our opinion: **GDPR contributed a lot** to defer creating tracking apps that would collect data on citizens not really necessary for fighting the epidemics

However,

- the first European initiatives have been not compliant with GDPR.
- COVID-19 can still be used as an excuse for collecting data in unrestricted way.

EU Report

M. Kutyłowski,
A. Lauks-Dutka,
M. Yung

Report's section "The application of the GDPR to new technologies"

> *Future challenges lie ahead in clarifying how to apply the* ***proven principles*** *to specific technologies such as artificial intelligence, blockchain, Internet of Things or facial recognition which require a monitoring on a continuous basis.*

Our opinion: **"proven principles"** sounds like **lack of interest for rethinking the basic principles and resolving the incompatibility** between the current law and emerging technologies

- no revision of GDPR planned, only some soft approach when SME are concerned,
- the next revision of GDPR in 2024!

Solution proposals

Different approaches

M. Kutyłowski,
A. Lauks-Dutka,
M. Yung

Steps – technology driven approach

1. identify needs
2. analyse what is **doable** from the **technical, economical and social point of view**
3. formalize legal requirements as a pragmatic compromise between different factors
4. adjust the systems

Different approaches

M. Kutyłowski,
A. Lauks-Dutka,
M. Yung

Steps – technology driven approach

1. identify needs
2. analyse what is **doable** from the **technical, economical and social point of view**
3. formalize legal requirements as a pragmatic compromise between different factors
4. adjust the systems

Steps – law driven approach

1. identify needs
2. formulate goals and corresponding legal rules
3. **let the technicians to find a solution**
4. create supervision authorities, collect fines,...

It does not work this way: **what if there is no reasonable technical solution??**

Shared personal data

M. Kutyłowski,
A. Lauks-Dutka,
M. Yung

Rule: Progressive/regressive data processing

Each data record should have a field or multiple fields **"data subject"**.

The operations on personal data should be classified as:

- **progressive** (creates a new information contents) – a **consent of all** data subjects is **necessary**,
- **regressive** (strictly limited to erasing information contents) – a request/withdrawal of the consent by **just one** data subject is enough to legitimize the operation.

→ clear situation for data processors

→ warning for users: a co-owner of shared personal data can erase it without asking anybody

Consent

M. Kutyłowski,
A. Lauks-Dutka,
M. Yung

Rule: Extended context of a consent rule

A **consent** should be understood as the right to process data **regardless** of the **context** that may emerge **outside**

→ eliminates infeasible analysis by data processor

→ only a data subject might be aware of all contexts of the consent

Rule: "Personal data" as an attribute of data & processing party

A data shall be considered a **personal data** by a party processing it

$\Longleftrightarrow$

this party can **identify** a physical person related to these data.

→ implementable, reduces the risk but still protects privacy

Processing non-personal data

M. Kutyłowski,
A. Lauks-Dutka,
M. Yung

Rule: Impact of processing non-personal data

A party X processing **non-personal data** is responsible for **all consequences** of that processing from the point of view of GDPR.

Particular case: Admissibility of data transfer rule

A may **send** data D to B $\iff$ A can **reasonably** assume that either:

- D is **not-personal data** for B or any potential partner of B, or
- B **complies** with the GDPR obligation and has the right to keep this data.

Data aggregation

M. Kutyłowski,
A. Lauks-Dutka,
M. Yung

Rule: Narrow definition of data processing

A data processing P where **personal data** are included in the **input** of P shall **not** be understood as **processing of personal data,** if the **output** of P (explicit and implicit) does **not** contain **personal data**.

→ **freedom for data analysis** as long as the output does not violate the rights and freedoms of people

Final remarks

M. Kutyłowski,
A. Lauks-Dutka,
M. Yung

many (detailed) issues left open

examples:

1. what to do with personal data processed by a party that becomes inactive (abandoned data)?
2. how does the GDPR regulation apply to a party holding a share of a personal data according to a secret sharing scheme?
3. P2P technology, quorum systems, . . . versus GDPR
4. right-to-be-forgotten and distributed ledger technology
5. ...

It's time:

- to focus on **rethinking** the general **paradigms** of GDPR!
- to seek for **improvements** and **better legal solutions** based on realistic privacy needs and computing goals!

the EU review on GDPR is a step forward, but most of the work to be done!

Thank you for your attention!

Disclaimer: while pointing to key problems regarding implementation of GDPR, our voice in discussion should be regarded as “amicus curiae brief”. In no way we attempt to undermine the necessity of personal data protection – one of key cybersecurity components.