GDPR strategies — A guide for survival

by Rigo Wenning (ERCIM/W3C legal counsel)
19 May 2020
Three possible strategies:

• Do nothing (flying under the radar of the DPA)
• Entering the Anonymisation – De-Anonymisation game
• Contextual consent gathering and intelligent data handling that preserves knowledge about legal grounds for processing
Do nothing

Risk depends on use case
Benefits

• Simplicity: Processing data without limits and filters and constraints
• New insights: Derive new knowledge from your data
• Cost: No time/money/investment in additional safeguards
• administrative fines up to €20 Mio, or in the case of an undertaking, up to 4% of the total worldwide annual turnover of the preceding financial year, whichever is higher
• Depending on type of service: Loss of trust & clients:
  – If Facebook-like service, trust is not really important
  – If more like banking & shopping it is a major risk
• Penal fines e.g. in France (Art. 226-16 Code pénal ss)
  – Up to five years in prison

Caveats
Anonymization buzzwords

- K - Anonymity (Samarati 1999) (at least k people in a sample)
- L - Diversity
- T - Closeness
- Differential Privacy (Removing outliers by adding noise)
The anonymity Problem

- The amount of records is increasing every day
- Data is de-identified by removing an explicit identifier (e.g. eMail, SSN)
- Other data with identifiers can be correlated with the de-identified data for re-identification
- Identities can be linked with de-identified information => re-identification
Issues

• The nicer the reconstruction attacks, the stronger the need for € to be high
• The more noise or means to obfuscate, the less quality and entropy in the data
• Aggregation and mining risk pure whitewashing if the result is used to target/discriminate against individuals
The anonymization trap

• The better the reconstruction attack, the more data must be stripped
• Data has less and less information
• Data value is destroyed
• Data becomes meaningless
Getting to consent

- Simplicity of consent is difficult
- Contextual interfacing is key for consent
- Asking simple questions needs mastering of one's own workflow
Advantages

• High quality data and user involvement
• Legal compliance made easy
• Usable data lakes without liability risk
• Creation of data value chains with others while respecting policy & GDPR
Screen restrictions and simple messages

Sport and dining recommendations at your fingertips

Sign up for free for intelligent lifestyle recommendations tailored to you.

At least one of the following data categories is needed to make recommendations.

Location

Your smartphone and our GSM network allow to track your location.

Yes, I agree that BeFit uses my location data based on my smartphone.

Yes, I agree that BeFit uses my heartbeat information.

Yes, I agree that BeFit uses my location data based on my smartphone.

Yes, I agree that BeFit uses my age.

Yes, I agree that BeFit uses the calories I consume.

BeFit
Gather Metadata

- Metadata from interaction with consent interface
- Metadata from protocol chatter
- Cookies and identifiers
- Policies applicable to this particular interaction
- Environmental data (data controller, third parties involved)
Sticky policies using Linked Data

DB Data ➔ URI ➔ DB Meta Data

API
The SPECIAL Consent, Transparency and Compliance framework
Need for Vocabularies to express those things
eye. Existing Vocabularies

- **FOAF**, **vCard** and **schema.org** offer vocabularies to model personal data
- **DICOM** can be used for health data with lots of attributes for fitness
- **NeoGeo, GeoSPARQL** or **WGS84 Geo Positioning** can be used to express location data.
- **P3P** WG had developed an RDF vocabulary with purposes
- **ODRL** t has a model that allows the expression of actions, prohibitions, and obligations to describe consent semantics, use of data and access.
- **OWL Time** can express time and duration for processing and retention.
- **PROV** is a good starting point to model source and quality of data and the source of the consent needed for processing.
The SPECIAL Policy Language Starting Points

Data collection

Purpose

Notifications

Place of storage

Data retention

Data sharing

Transparency & Compliance

Who
When
Why
How
What
Where
# Data Privacy Controls and Vocabularies

A W3C Workshop on Privacy and Linked Data

17–18 April 2018, WU Vienna, Vienna, Austria, Europe

<table>
<thead>
<tr>
<th>Report</th>
<th>Intro</th>
<th>How to participate</th>
<th>Logistics</th>
<th>Program committee</th>
<th>People</th>
<th>Workshop schedule</th>
</tr>
</thead>
</table>
What would you like to standardise (rank in order) V2?

1. Taxonomy of regulatory privacy terms (including all GDPR terms)
2. Taxonomy for personal data
3. Taxonomy of purposes
4. Taxonomy of disclosure
5. Glossary of terms
6. Metadata related to the details of the anonymisation
7. Requirements for evaluation of privacy statement
8. Log vocabulary
9. Taxonomy of linkage operations
10. Taxonomy of relevant contexts
11. Taxonomies of human behaviour
Data Privacy Vocabularies & Controls Community Group (DPVCG)

• A Group to do coordination und harmonisation
• A platform for those interested in Privacy enabling technologies
• Development of a vocabulary for data protection & Interoperability
• https://www.w3.org/community/dpvcg/
Semantics centered around processing
Centered around data subject
If you want to know more: Get involved

https://www.w3.org/community/dpvcg/
Attributions

• Sonar Principle by Georg Wiora (Dr. Schorsch) [CC BY-SA 3.0]
• Quadcopter by Simon Waldherr [CC BY-SA 3.0]
• Bird by Malikmak333 [CC BY-SA 4.0]
• Jail Bars by Antonu [CC BY-SA 3.0]