DATA IN MOTION

Model-driven Privacy and Data Transparency

20.07.2023

Model-driven Privacy and Data Transparency

Part Contents

Section Contents

The “Why”

- Organizations at all scale need to be complaint with privacy as well as data transparency rules;
- These rules apply to a variety of fields within a single organization, from contracts stipulation to actual production;
- Often organizations need to rely on multiple systems to handle privacy and data transparency in different areas;
- We would like to be able to combine those systems into a single one, capable of handling privacy and data transparency on a more general level.

The “How”

- Given a process (that we can always think in terms of a model), provide warnings on fields and/or combination of fields that might constitute a privacy risk, or that might have to be complaint with some data transparency standards.

But first...

- We need to understand the privacy as well as the data transparency standards that an organization has to be complaint with;
- We need to understand what work has been already done so far in this field.

Part Contents

Section Contents

GDPR: What is it?

- The General Data Protection Regulation (GDPR) is Europe's data privacy and security law;
- It has been put into effect on May 25, 2018;
- If you process the personal data of EU citizens or residents, or you offer goods or services to such people, then the GDPR applies to you even if you're not in the EU;
- The fines for violating the GDPR are very high.

GDPR: Terminology

Personal Data

Personal data is any information that relates to an individual who can be directly or indirectly identified. Names and email addresses are obviously personal data. Location information, ethnicity, gender, biometric data, religious beliefs, web cookies, and political opinions can also be personal data. Pseudonymous data can also fall under the definition if it's relatively easy to ID someone from it.

GDPR: Terminology

Data Processing

Any action performed on data, whether automated or manual. The examples cited in the text include collecting, recording, organizing, structuring, storing, using, erasing... so basically anything.

GDPR: Terminology

Data Subject

The person whose data is processed.

GDPR: Terminology

Data Controller

The person who decides why and how personal data will be processed. If you're an owner or employee in your organization who handles data, this is you.

GDPR: Terminology

Data Processor

A third party that processes personal data on behalf of a data controller. The GDPR has special rules for these individuals and organizations. They could include cloud servers like Tresorit or email service providers like Proton Mail.

GDRP: Data protection principles

If you process data, you have to do so according to seven protection and accountability principles outlined in Article 5.1-2:

1. **Lawfulness, fairness and transparency**
2. **Purpose limitation**
3. **Data minimization**
4. **Accuracy**
5. **Storage limitation**
6. **Integrity and confidentiality**
7. **Accountability**

GDPR: Accountability

The GDPR says data controllers have to be able to demonstrate they are GDPR compliant. Among the ways you can do this:

- Designate data protection responsibilities to your team.
- Maintain detailed documentation of the data you're collecting, how it's used, where it's stored, which employee is responsible for it, etc.
- Train your staff and implement technical and organizational security measures.
- Have Data Processing Agreement contracts in place with third parties you contract to process data for you.
- Appoint a Data Protection Officer (though not all organizations need one — more on that in this article).

GDPR: Data security

- You're required to handle data securely by implementing "appropriate technical and organizational measures."
- Technical measures mean anything from requiring your employees to use **two-factor authentication** on accounts where personal data are stored to contracting with cloud providers that use **end-to-end encryption**.
- Organizational measures are things like **staff trainings**, adding a **data privacy policy** to your employee handbook, or **limiting access to personal data** to only those employees in your organization who need it.
- If you have a data breach, you have 72 hours to tell the data subjects or face penalties. (This notification requirement may be waived if you use technological safeguards, such as encryption, to render data useless to an attacker.)

GDPR: Data protection by design and by default

- Everything you do in your organization must, “by design and by default," consider data protection.
- Practically speaking, this means you must consider the data protection principles in the design of any new product or activity.
- The GDPR covers this principle in Article 25.

GDPR: When you are allowed to process data

Article 6 lists the instances in which it's legal to process person data:

- The data subject gave you specific, **unambiguous consent** to process the data;
- Processing is necessary to execute or to prepare **to enter into a contract** to which the data subject is a party;
- You need to process it **to comply with a legal obligation** of yours;
- You need to process the data **to save somebody's life**;
- Processing is necessary **to perform a task in the public interest** or to carry out some official function;
- You have a **legitimate interest** to process someone's personal data.

GDPR: Consent

There are strict new rules about what constitutes consent from a data subject to process their information.

- Consent must be “freely given, specific, informed and unambiguous.”
- Requests for consent must be “clearly distinguishable from the other matters” and presented in “clear and plain language.”
- Data subjects can withdraw previously given consent whenever they want, and you have to honor their decision. You can’t simply change the legal basis of the processing to one of the other justifications.
- Children under 13 can only give consent with permission from their parent.
- You need to keep documentary evidence of consent.

GDPR: Data Protection Officers

Not every data controller or processor needs to appoint a Data Protection Officer (DPO). There are three conditions under which you are required to appoint a DPO:

1. You are a public authority other than a court acting in a judicial capacity.
2. Your core activities require you to monitor people systematically and regularly on a large scale. (e.g. You're Google.)
3. Your core activities are large-scale processing of special categories of data listed under Article 9 of the GDPR or data relating to criminal convictions and offenses mentioned in Article 10. (e.g. You're a medical office.)

GDPR: People's privacy rights

The GDPR recognizes a litany of new privacy rights for data subjects, which aim to give individuals more control over the data they loan to organizations. They are:

1. The right to be informed
2. The right of access
3. The right to rectification
4. The right to erasure
5. The right to restrict processing
6. The right to data portability
7. The right to object
8. Rights in relation to automated decision making and profiling.

Part Contents

Section Contents

Some investigation

- There seems to be no current automated solution with broad industrial applicability to carry out the task of checking GDPR compliance;
- There are, however, several efforts of trying to put the GDPR regulations into a more "schematic" way which can then be used by automated systems;
- The more promising I could find in terms of GDPR is the solution developed by Torre et al. in 1,2;
- Another interesting article I found is 3, which might be interested in the context of smart cities.

Model-Driven Engineering for Data Protection and Privacy

- In 1 the authors present an UML model of the GDPR regulations, developed together with a team of legal experts;
- They add a set of OCL constraints that represent the conditions an organization has to fulfill to be GDPR compliant;
- The OCL constraints can be then extended to match the specific regulations of the different member states, in such a way to pass from an abstract GDPR model to a concrete one;
- The document provides full description of the UML classes as well as the OCL code.

AI-assisted Approach for Completeness of Privacy Policies

- In 2 the authors present a working system, capable of checking completeness of privacy policies against GDPR;
- They develop a set of meta-data based on the recurrent and key terms of GDPR articles in terms of privacy policies;
- They build a NLP-ML algorithm to check the presence of such meta-data in privacy policy documents;
- Based on some constraints they give feedback on whether the document is complete according to the GDPR or not.
- It does not check for the actual meaning of the policy, it just checks whether a certain required topic is treated or not.

A Privacy-Aware Smart City

- In 3 the authors describe some potential privacy issues related to the development of smart cities;
- In particular they describe the citizens' privacy as a 5D matrix:
  - **Identity Privacy**
  - **Query Privacy**
  - **Location Privacy**
  - **Footprint Privacy**
  - **Owner Privacy**
- They provide examples in terms of smart cities service providers, and possible techniques to address these privacy issues.

Part Contents

Section Contents

What I learned

- It seems the GDPR is quite complex and such kind of tasks require the expertise of legal experts at some point, in order to create a meaningful model;
- GDPR covers a lot of different topics, so we should first clarify which aspects of it we want to focus on.

Section Contents

Next Steps

- If our scope, at least for now, is just determining some potentially “privacy-related” fields in a model, we could try a sort of simplified version of what they did in [2]:
  - From the definition of *Personal Data* we can try to derive some meta-data, which are just some keywords to identify potentially “privacy-related” fields;
  - From a given model, we would compute some sort of text similarity between its attributes and our meta-data and provide a feedback based on the result.
- For other aspects of GDPR, we should think which approach to test.

The Challenges I see

- Models attributes are not usually written using GDPR terminology, as it is instead the case for a privacy policy document;
- That means we would have to guess our meta-data and assume people name their models attributes with some common sense;
- When we talk about personal data, those are probably always the same, but if we want to provide any feedback on some other aspects of GDPR, these can be quite domain-specific (privacy policy, medical records, etc.), so we would need to think about a different set of meta-data for different cases;
- As far as I could see (but I spent only a day on it) common packages for NLP do not have any java-like lexicon (which would make hard applying these techniques for the search of word similarities in model attributes for instance).

Conclusion

Useful Links

OSGi Working Group

Working Group: www.osgi.org
WG Blog: www.osgi.org/blog
Twitter: @osgiwg
Bndtools: bndtools.org

Data In Motion

Web: www.datainmotion.com
Blog: datainmotion.com/blog
Twitter: @motion_data

Jürgen Albert

Email: j.albert@data-in-motion.biz

Mark Hoffmann

Email:
m.hoffmann@data-in-motion.biz