SoK: Three Facets of Privacy Policies

WPES2020 online

Victor Morel & Raúl Pardo

Inria

victor.morel@inria.fr

9th November 2020

Introduction

Privacy policies must be:

- Legally valid
- Understandable
- Enforceable (auditable)

Three requirements $\rightarrow$ three *ways* to express privacy policies or **facets**

Methodology

Taxonomy item	Description	GDPR	FIPPs	CCPA	HIPAA$_a$	COPPA$_b$
First Party collection	Type of data collected, purpose and collection mode.	●	●	◐	●	●
Third Party collection	Type of data collected, purpose and collection mode for third parties.	●	●	◐	◐	●
Legal basis	Ground on which is determined the lawfulness of processing.	●	◐	○	○	○
DS rights	Rights of the DS, *e.g.*, right to access, to rectify, to port or erasure.	●	○	◐	●	●
Data Retention	Duration of data storage	●	○	○	○	◐
Data Security	Modalities of protection of data, *e.g.*, encrypted communication and storage.	◐	●	○	●	◐
Policy Change	Modalities of notification for policy changes.	◐	○	○	○	○
Other	Other items such as identity of DC, information related to Do-Not-Track, to children ...	●/ ◐	●/ ◐	●/ ○	◐	●

Table: Summary of our taxonomy with the legal requirements of items. We use ● to denote *Required explicitly*; ◐ to denote *Addressed but not required*; and ○ to denote *Absent*. The subscript $_a$ means that HIPAA only considers health data. The subscript $_b$ means that COPPA only considers personal information from children, and notice must be addressed to parents.

Outline

Content of natural language privacy policies

Presentation

Natural language privacy policies are textual documents used to inform about personal data collection and processing.

Typical example

"[...] when you search for something on Facebook, you can access and delete that query from within your search history at any time, but the log of that search is deleted after six months."

Content of natural language privacy policies II

- What kinds of information do we collect?
- How do we use this information?
- How is this information shared?
- How do the Facebook Companies work together?
- What is our legal basis for processing data?
- How can you exercise your rights provided under the GDPR?
- Data retention, account deactivation and deletion
- How do we respond to legal requests or prevent harm?
- How do we operate and transfer data as part of our global services?
- How will we notify you of changes to this policy?
- How to contact Facebook with questions

Figure: Menu of Facebook's privacy policy

Natural language privacy policies tools

Templates and generators

- Fill-in-the-gap forms for templates
- Input verification for generators

Retrievers

- Automatic extraction of information
- Tailored to mobile applications (permissions)

Analysis tools

- NLP to parse existing policies
- See Polisis[a] for instance

[a] Harkous et al., "Polisis".

Benefits & Limitations

Benefits

- Legal value

Limitations

- Ambiguity
- Understanding
- Enforcement & auditability

Outline

Content of graphical privacy policies

Sets of icons

Aim to cover the items of the taxonomy

(a) UE transfer adequacy

(b) Connection data

(c) One year conservation

(d) Audience measurement

Figure: Excerpt of the Privacy Tech icons

Content of graphical privacy policies

Standardized notices

Aim to express content in a standardized and often comparable manner

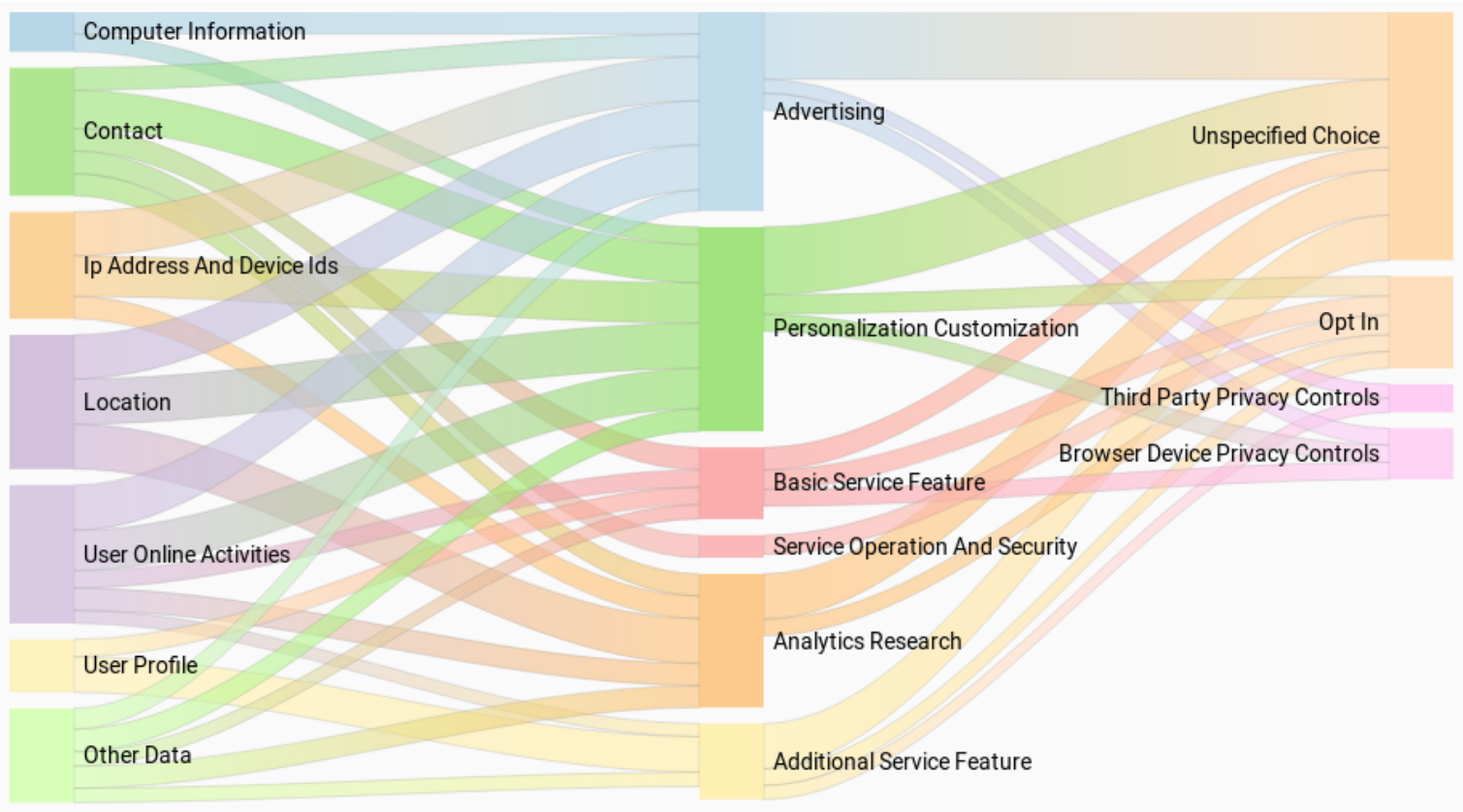
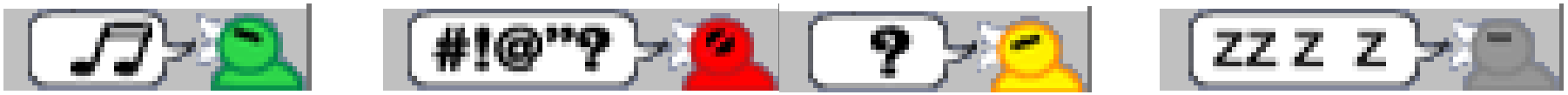


Figure: Example of a flow diagram in Polisis

Content of graphical privacy policies

Rating solutions

Provide rating information concerning certain aspects of privacy policies such as the transparency level or potential risks



(a) Matching policies (b) Conflicting policies (c) Uncertain decision (d) Add-on disabled

Figure: Privacy Bird[1]

[1] CyLab Usable Privacy and Security Laboratory, *Privacy Bird*.

Benefits & Limitations

Benefits

- Designed for lay-user understandability

Limitations

- Ambiguity
- Incompleteness
- Claim over legal compliance

Outline

Content of machine-readable privacy policies

- Machine-readable privacy policies are mostly *privacy languages*
- Content is based on the syntax of the language
  - Access-based control
  - P3P and its derivatives
  - Formal languages
  - Languages modelling privacy regulations ...

Typical example

```
<retention-time days=182 xmlns=".../P3P/retention-time/"/>
```

Content of machine-readable privacy policies II

Another example: Pilot

$$
\begin{aligned}
\textit{Pilot Privacy Policy} &::= (\textit{datatype}, \textit{dcr}, \textit{TR}) \\
\textit{Data Communication Rule}\,(\textit{dcr}) &::= \langle \textit{condition}, \textit{entity}, \textit{dur} \rangle \\
\textit{Data Usage Rule}\,(\textit{dur}) &::= \langle \textit{Purposes}, \textit{retention\_time} \rangle \\
\textit{Transfer Rules}\ (\textit{TR}) &::= \{\textit{dcr}_1, \textit{dcr}_2, \ldots\}
\end{aligned}
$$

Machine-readable privacy policies tools

Formal semantics

Semantics define what events may be executed depending on the privacy policies selected by the actors interacting in the system

Informal semantics

Specifications use request evaluation engines to enforce privacy policies

Policy comparison

1. Research purposes, 7 days
2. Research and advertisement for 90 days

Which one is more restrictive? Number 1!

Benefits & Limitations

Benefits

- Enforcement
- Auditability
- Correctness
- Automation

Limitations

- Human understandability
- Lack of adoption

Outline

Limitations of mono-faceted solutions

A single facet cannot cover all the requirements of privacy policies

- Natural language privacy policies have legal argon:
  "[...] when you search for something on Facebook, you can access and delete that query from within your search history at any time, but the log of that search is deleted after six months."
- Graphical privacy policies have no use for lawyers or enforcement by machines: 

- Machine-readable privacy policies include technical details that may confuse lay-users and lawyers:
  `<retention-time days=182 xmlns=".../P3P/retention-time/"/>`

Multi-faceted privacy policies

Limitations in one facet can be addressed by other facets

- Pilot[a] combines natural language and machine-readable privacy policies
- [b] add graphical representations to machine-readable policies

[a] Pardo and Le Métayer, "Analysis of Privacy Policies to Enhance Informed Consent".

[b] Kelley et al., "A Nutrition Label for Privacy".

Two approaches to design multi-faceted privacy policies:

Unified

A core facet is defined and the remaining facets are generated from the core using a policy generator

Compound

Consists in taking mono-faceted policies and using them together

Multi-faceted privacy policies II

$(Facebook, \ldots,$
$\{ (Internal\_Companies, \ldots),$
$(External\_Parties, \ldots)\})$

"We share information globally, both internally within the Facebook Companies and externally with our partners and with those you connect and share with around the world in accordance with this Policy. Information controlled by Facebook Ireland will be transferred or transmitted to, or stored and processed in, the United States or other countries outside where you live for the purposes as described in this Policy.[...]"

Figure: Multifaceted Privacy Policy: a compound example

Missing taxonomy items

	Graphical Policies			Machine-Readable Policies			Required by Legislations		
	●	◐	○	●	◐	○	●	◐	○
1st party	33%	33%	34%	69%	31%	0%	80%	20%	0%
3rd party	17%	50%	33%	47%	26%	27%	60%	40%	0%
Legal basis	5%	0%	95%	0%	0%	100%	20%	20%	60%
DS Rights	22%	5%	73%	13%	17%	70%	60%	20%	20%
Data Retention	28%	11%	61%	30%	43%	27%	20%	20%	60%
Data Security	22%	22%	56%	17%	30%	53%	40%	40%	20%
Policy Change	0%	0%	100%	0%	0%	100%	0%	20%	80%

Figure: Coverage of taxonomy items by different types of privacy policies, and the privacy legislations. Each cell of the heat map shows the percentage of the studied works (in a given facet) that cover completely (●), partially (◐) or neither (○) an item of the taxonomy.

Conclusion

To conclude

- We have studied the different ways to express privacy policies
  - In natural language
  - With graphical representations
  - Using machine-readable means
- Each work is categorized according to a taxonomy
- We have studied the combination of different facets
- Towards collaborative work between the legal domain, design, and computer science?