SoK: Three Facets of Privacy Policies

Victor Morel, Raúl Pardo

▶ To cite this version:

Victor Morel, Raúl Pardo. SoK: Three Facets of Privacy Policies. Workshop on Privacy in the Electronic Society, Nov 2020, Virtual, France. hal-02267641v4

HAL Id: hal-02267641

https://inria.hal.science/hal-02267641v4

Submitted on 11 Sep 2020

SoK: Three Facets of Privacy Policies

Victor Morel, Raúl Pardo

RESEARCH REPORT
N° 9287
September 2020
Project-Teams Privatics

ISSN 0249-6399 ISRN INRIA/RR--9287--FR+ENG

SoK: Three Facets of Privacy Policies

Victor Morel, Raúl Pardo

Project-Teams Privatics

Research Report n° 9287 — September 2020 — 34 pages

Abstract: *Privacy policies* are the main way to obtain information related to personal data collection and processing. Originally, privacy policies were presented as textual documents. However, the unsuitability of this format for the needs of today's society gave birth to other means of expression. In this paper, we systematically study the different means of expression of privacy policies. In doing so, we have explored the three main categories, which we call *facets*, *i.e.*, natural language, graphical and machine-readable privacy policies. Each of these facets focuses on the particular needs of the communities they come from, *i.e.*, law experts, organizations and privacy advocates, and academics, respectively. We then analyze the benefits and limitations of each facet, and explain why solutions based on a single facet do not cover the needs of other communities. Finally, we set guidelines and discuss challenges of an approach to expressing privacy policies which brings together the benefits of each facet as an attempt to overcome their limitations.

Key-words: privacy policies, legal compliance, usability, enforcement

RESEARCH CENTRE
GRENOBLE – RHÔNE-ALPES

Inovallée
655 avenue de l'Europe Montbonnot
38334 Saint Ismier Cedex

SoK : Trois facettes des politiques de protection de vie privée

Résumé : Les *politiques de protection de vie privée* sont le principal moyen d'obtenir de l'information liée à la collecte et au traitement de données à caractère personnel. Ces politiques étaient originellement présentées comme des documents textuels. Cependant, la non-convenance de ce format aux besoins de la société actuelle a donné lieu à d'autres moyens d'expression. Dans ce rapport, nous étudions de manière systématique les différents moyens d'expression des politiques de protection de vie privée. Ce faisant, nous identifions trois catégories principales que nous nommons *dimensions*, *i.e.*, les politiques en langage naturel, la représentation graphique des politiques, et les politiques lisibles par les machines. Chacune de ses dimensions se concentre sur les besoins spécifiques de la communauté dont elle est issue, *i.e.*, respectivement les juristes, les organisations et les défenseurs de la vie privée, et les universitaires. Nous analysons ensuite les avantages et les limites de chaque dimension, et nous expliquons en quoi les solutions basées sur une seule dimension ne couvrent pas les besoins des autres communautés. Enfin, nous proposons une nouvelle approche pour exprimer les politiques de protection de vie privée qui réunit les avantages de chaque dimension, dans le but de surmonter leurs limites.

Mots-clés : politiques de vie privée, conformité légale, utilisabilité, mise en application

1 Introduction

As of today, the main way to obtain information related to personal data collection and processing is through *privacy policies.* Privacy policies are typically presented as textual documents describing details such as data collection, processing, disclosure and management. Organizations collecting personal data (in what follows *data controllers*, or DC) commonly use privacy policies to inform individuals (in what follows *data subjects*, or DS) about how personal data is handled. DS are often required to read these policies — even though it rarely occurs [57] — or are at least presumed to do so and to decide whether they accept the conditions. Alternatively, giving DS the possibility of describing their own privacy policies has recently gained in popularity. This approach gives DS the time to reflect on their choices, and the possibility to consult experts and pairs. Nonetheless, privacy policies are hard to understand [21], for DS [68] as for experts [89]. We use *DS policies* to refer to the privacy policies of individuals, and *DC policies* to denote the privacy policies of organizations collecting personal data.

Requirements and recommendations to express privacy policies come from different sources such as privacy regulations, authorities and organizations. For instance, in Europe, the General Data Protection Regulation (GDPR) [39] requires more transparency for data processing from DC, and guidelines have been issued by the WP29[1] to present their expectations [111]. These requirements are necessary for privacy policies to be compliant with the legislation. Recommendations for drafting policies have also been made by different organizations to improve their readability. For example, the National Telecommunications and Information Administration for mobile apps [72], and the WP29 for IoT devices [110]. Furthermore, Data Protection Authorities (DPAs) should be able to audit data processing systems, to ensure their compliance with the law and with the declared privacy policies.[2] All these requirements and recommendations can be summarized in three requirements:

- Privacy policies must be legally valid.
- Privacy policies must be understandable by all parties.
- Privacy policies must be enforceable and auditable in data processing systems.

Existing methods to express privacy policies address some of these requirements, but not all of them. Different methods have arisen from different needs, and they target different audiences — from expert to lay-users. For instance, legal privacy policies are often written as long and complex documents which are necessary in court, but that are not easy to understand for lay-users. As an attempt to simplify these legal documents, organizations work on summarized versions of privacy policies or use visual aids to help users understand the risks of having their data collected. Ensuring that data is processed according to the requirements in privacy policies is not an easy task either. Some work — coming mostly from academia — proposes an alternative format for privacy policies that can be read by computers. These solutions aim at bridging the gap between the textual legal requirements and their enforcement in the underlying system. Furthermore, some of these proposals are equipped with auditing tools which facilitate, for DPAs or DS, verifying that no violations of a privacy policy have occurred. Unfortunately, these solutions are not widely used.

In this work, we analyze the state-of-the-art methods on expressing privacy policies. We provide a comprehensive picture of existing proposals in order to identify gaps and challenges.

[1] WP29 stands for Working Party 29, an European advisory board.

[2] For instance, see the decision of the "Commission Nationale de l'Informatique et des Libertés" (CNIL), the French DPA, against Google LLC [25].

In doing so, we have studied the three main ways to express privacy policies: natural language, graphical and machine-readable; which we call the *facets* of privacy policies. Each of these facets have mostly arisen from different communities. Natural language comes from law experts, graphical from organizations and privacy advocates, and machine-readable from academics. Consequently, the content of this paper contextualizes knowledge often studied within different communities that have been working on similar issues, but with different objectives. Unsurprisingly, each facet mainly provides benefits to the specific community it was defined in. Therefore, we take the insights of our study to explore how the different facets of privacy policies can complement each other. This synergy includes the benefits of each facet and minimize their limitations. We discuss guidelines and challenges of combining all facets in a single policy, which we denoted as *multi-faceted privacy policies*, and study recent efforts in this direction.

More concretely, in bringing the above ideas to the forefront, our contributions are:

1. A categorization of existing work in each facet according to a privacy taxonomy (introduced in Section 2), and the specific features of each facet.

2. An in-depth study of the existing facets of privacy policies: i) privacy policies expressed in natural language (Section 3), ii) graphical privacy policies (Section 4), and iii) machine-readable privacy policies (Section 5). For each facet, we provide an overview of: its content; the available tools; its benefits; and its limitations.

3. Insights from the study on future research on designing privacy policies (Section 6).

Section 7 discusses related work and concludes the paper.

2 Systematization Methodology

For each facet, we study the content of the privacy policies, the available tools, benefits and limitations. The literature of privacy policies across the different facets is vast. An exhaustive analysis of all works is unfeasible and undesirable. Instead, we focus on highlighting representative work; hence we do not provide an all-encompassing reference to every related work.

We categorize the content in each facet according to a taxonomy of privacy policies. Several taxonomies have been proposed to categorize the content of privacy policies [96, 109, 77]. We use a slight variation of [109]. This taxonomy is appropriate for our purposes for two main reasons: 1) it was devised for a wide range of privacy policies and therefore reflects their content across facets; and 2) it encompasses most requirements of the current legislations and guidelines, such as the GDPR, the Fair Information Practice Principles (FIPPs) [40] in some cases, and the California Consumer Privacy Act of 2018 (CCPA) [97].

Taxonomy

Wilson *et al.* [109] proposed a taxonomy for privacy policies composed of the following items:[3] "*First Party collection*: How and why a service provider collects user information", "*Third Party collection*: How user information may be shared with or collected by third parties", "*Access, Edit, Delete*: If and how users may access, edit, or delete their information", "*Data Retention*: How long user information is stored", "*Data Security*: How user information is protected", "*Specific Audiences*: Practices that pertain only to a specific group of users (e.g., children, Europeans, or California residents)", "*Do-Not-Track*: If and how Do Not Track signals for online tracking and advertising are honored", "*Policy Change*: If and how users will be informed about changes

[3] We denote *item* a piece of information provided in a privacy policy.

to the privacy policy", "*Other*: Additional sub-labels for introductory or general text, contact information, and practices not covered by the other categories", and "*Choice Control*: Choices and control options available to users".

Our variation of Wilson *et al.*'s taxonomy accommodates it to the purposes of our study, but does not change the content. Concretely: i) we use *DS rights* to denote *Access, Edit, Delete* and *Choice Control* as they relate to DS rights in the sense of the GDPR (see Chapter III of the GDPR); ii) subsume *Specific Audiences* and *Do-Not-Track* under *Other* as they are occasional items; and iii) we observe that a legal requirement is missing in the taxonomy, even though it is often found in natural language privacy policies: the *legal basis* of processing, we will therefore add it to our taxonomy. These differences better accommodate recent regulations (for DS rights and Legal basis) and practices (for Other). Table 1 summarizes the taxonomy. Section 3.1 provides a detailed explanation of each taxonomy item together with illustrative examples.

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 1: 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.

3 Natural language privacy policies

Most legislations require notices expressed in natural language to inform DS about the collection and processing of their personal data: the use of natural language is necessary to ensure that the policy has legal value (*e.g.*, [39, Art. 13 & 14]). The ways these documents can be authored — *i.e.*, drafted automatically or written manually — and the manners to assist their authoring can vary greatly. We present the content expressed by natural language privacy policies in Section 3.1, the tools used to assist their authoring and to analyze existing natural language privacy policies in Section 3.2, the benefits in Section 3.3, and the limitations in Section 3.4.

3.1 Content

Natural language privacy policies are familiar to the public as they have been adopted by a large range of online services such as social networks, file hosting services, mobile applications, *etc.* Because these privacy policies are expressed in natural language, they are not restricted in terms of content; they cover all the items of our taxonomy (see Section 2). In the following we examine the taxonomy items in detail, and discuss what legal requirements appear explicitly in the GDPR, the FIPPs, or in the CCPA.

We focus on the requirements of the GDPR, the FIPPs, and the CCPA as they are the three main texts (legislations or guidelines) that determine the content required when informing DS of data collection and processing. The GDPR is the text regulating personal data collection and processing in the EU, and many countries consider it since it has an extraterritorial scope. The FIPPs are guidelines designed by the United States Federal Trade Commission's (FTC). They represent common principles regarding fair information practices. However, they have been criticised by scholars [22], and have not been updated in decades. The CCPA is more recent (2018), it is a privacy enhancing and consumer protection bill for Californians. It has been deemed "the strongest privacy controls of any state in the U.S" [2]. We also consider other widely-known regulations, such as health data for HIPAA [104, Notice and Other Individual Rights], or children for COPPA [41, Â§ 312.4]).

First Party Collection The most common item in natural language privacy policies is the first party collection, which describes *what* data is collected, *why* it is collected, and sometimes *how*. The type of data ranges from generic to more precise assertions, *e.g.*, respectively *we collect your data* and *your email address is collected.* Cookies [31] often have a distinct treatment, most likely because they are often collected by websites: it is common to find a dedicated paragraph for their management. Location information is often treated in a separate section as well because it can be collected from different sources — mobile applications, web browsers — or inferred from metadata — such as IP addresses. The purpose of processing often goes hand in hand with the type of data collected. It is also possible to find the collection mode in some natural language privacy policies: whether the data is collected automatically, by manual input of DS, or by any other means.

Informing about the type of data, the purpose of processing, the recipients and the means of collection is required by the GDPR and the FIPPs. The CCPA gives the right to request first party collection, but does not require it. HIPAA requires to inform of "the ways in which the covered entity may use and disclose protected health information". COPPA requires to inform of "what use, if any, the operator will make of the personal information collected".

Third Party Collection Third Party Collection is a common item in natural language privacy policies, and it is therefore usual to find the third-parties to whom data will be transferred: they can be advertisers, or other business partners. The notion of *sharing* can also refer to other DS and subsidiary companies. It is usually composed of the same content as First Party Collection.

Informing about third party collection is required by both the GDPR, the FIPPs, and COPPA. The CCPA gives the right to request categories of third parties, but does not require it except if data is sold to those third parties. HIPPA considers it implicitly (see [104, Who is Covered by the Privacy Rule]).

Legal basis Legal basis (or legal ground) is regularly found as a complement of the purpose of processing. [4] A common legal basis for processing is consent, which consists, for DC, in retrieving an authorization from DS to legally collect their data. Consent has to be informed and specific under the GDPR [39, Recital 32], and it still is often used as a legal basis. DC might consider the reading of their natural language privacy policies as a proper consent, without questioning the conditions to obtain consent [43]. Other legal basis can be found in natural

[4] Art. 13(1)(c) of the GDPR requires "The purposes of the processing for which the personal data are intended **as well as** the legal basis for the processing." (Highlights from authors)

language privacy policies, such as the necessity for the performance of a contract, compliance with legal obligations, protection of DS's vital interests or public interest, and the legitimate interests of a DC. These legal basis are listed in the GDPR [39, Art. 6], and major stakeholders generally consider cumulatively either all of them, or a large subset.

Informing about the legal basis is required by the GDPR, and not explicitly by the FIPPs which requires informing "whether the provision of the requested data is voluntary or required, and the consequences of a refusal to provide the requested information". Legal basis information is not required by CCPA, HIPPA, and COPPA.

DS Rights DS can exercise rights regarding their data, and natural language privacy policies now often mention the rights to access, rectify, port and erase data, likely due to the influence of the GDPR. DS rights can be seen more restrictively as possibilities to opt-in or opt-out. Thus natural language privacy policies present how to subscribe or unsubscribe to specific services.

Informing about DS rights is required by the GDPR, HIPAA and COPPA, but not by the FIPPs. In CCPA, only the right to opt-out is explicit (other rights are ensured but Californians do not have to be explicitly informed of them in a privacy policy).

Data Retention Natural language privacy policies often describe the period during which personal data will be stored. It can be a fixed value — *e.g.*, *30 days after data collection*, or variable — *e.g. as long as your account is active.* It often comes with the type of data, the purpose, and the legal basis of processing.

Informing about the retention time is required by the GDPR. COPPA addresses it but do not make it mandatory to inform about it. FIPPs, CCPA and HIPPA do not require it.

Data Security DC regularly explain in their policies how data is stored, if its communication is secured or its storage encrypted.

Informing about the security of data is required by the FIPPs, but not by the GDPR although it mentions that "Personal data should be processed in a manner that ensures appropriate security and confidentiality of the personal data" [39, Recital 39]. Similarly, COPPA states that "The operator must establish and maintain reasonable procedures to protect the confidentiality, security, and integrity of personal information collected from children." HIPPA requires to inform of the "entity's duties to protect privacy". CCPA does not require to inform of this item.

Policy Changes The modalities of notification in the case of a change in the privacy policy can also be observed. Notification is usually by email or within the service's interface, in some cases notifications are sent by regular mail or by phone. This item is not required by the GDPR, the FIPPs, CCPA, HIPPA, or COPPA.

Other We subsume the identity and contact of DC, requirements towards specific audiences such as children, and Do-Not-Track under this item. The DC usually provides its identity, as well as its contact details if the DS has to lodge a complaint. In many legislations [39, Recitals 38 & 58] [41] children have specific considerations. As a result, many natural language privacy policies include a dedicated section, even if it is only to mention that personal data of children under thirteen is not collected without parental consent.

The identity of DC is required by the GDPR, the FIPPs, and HIPAA, but not Do-Not-Track. As for requirements for children, they are specifically addressed by these legislations. CCPA requires more specifically to "Make available to consumers two or more designated methods for submitting requests for information", and does not address data collected from children.

COPPA requires to inform of the identity and contact of DC, and targets data collection related to children.

3.2 Tools

Several solutions exist to assist in the authoring of natural language privacy policies, offering different levels of automation. We denote these tools *authoring tools*. We distinguish *templates*, *generators*, and *retrievers*. Authoring tools are often tailored to websites and mobile app owners, and are constrained in terms of content. Furthermore, there exist *analysis tools*: software able to parse and analyze natural language privacy policies, in order to produce a machine-readable or a graphical version of a policy.

Templates and generators Tools such as Docracy [35], Termsfeed [102], SEQ Legal [100], and 3DCart [3] provide a *fill-in-the-gap* form, where the author writes appropriate terms in the fields. We refer to them as *templates*. These tools guide the authoring process, but do not provide additional functional such as validation of the input data. It is not possible, for instance, to check if the email of a service owner is valid. *Generators* comprise more complex tools where authors input data only once regarding relevant fields of a privacy policy, and the tool automatically generate the privacy policy. Most generators do not allow incorrect data: email addresses without `@` are highlighted, and the author cannot go further in the process and generate the policy. Generators also give the option of expressing the same policy according to different legal contexts. For instance, `privacypolicies.com` offers clauses specific to the GDPR or COPPA. A set of what we denote *light* generators have a restricted set of parameters, while another composed of more *detailed* generators offer a choice between a policy tailored to websites or mobile applications, with a more exhaustive list of items [62]. Privacy Policy Generator [82], Privacy Policy Online [32], or GetTerms [46] are examples of the former set, and `PrivacyPolicies.com` [84] and `FreePrivacyPolicy.com` [44] of the latter.

Retrievers Retrievers, such as those offered by Miao [70], Apolinarski *et al.* [10] and Yu *et al.* [113], automatically extract relevant information from code of mobile application using static code analysis or user behavior analysis. For instance, in [10] the authors analyze sharing behavior when using online collaboration tools. These prototypes work on Android applications, in which personal data management is structured around the concept of permissions [47]. These permissions define the type of data accessible by an application, such as contacts, content of text messages, or Wi-Fi management. A *retriever* analyses those permissions, and interprets them according to well-defined rules to author a natural language privacy policy. In that case, an author does not necessarily have to input any information in addition to the code: the *retriever* can parse the name of the DC, the permission requested by a service or the third-party libraries, and can convert this information into natural language. However, *retrievers* are often tailored to a specific solution — mobile applications in most cases — and could thus be difficult to implement in other ecosystems. Furthermore, they cannot automatically retrieve certain information, such as the purpose of collection or the retention time.

Analysis tools Analysis tools have been developed for over a decade. They focused on using Natural Language Processing or Information Extraction [26] to parse natural language privacy policies. An early work has been conducted by Brodie *et al.* [19]. The Usable Privacy Project lead by Sadeh [93] further investigated the automated classification of privacy policies. Within this project, Ammar *et al.* [8] conducted a pilot study for automatic text categorization. The Usable Privacy Project also developed a website privacy policy corpus [109], which will later be notably used by Polisis [50]. Zimmeck *et al.* [116] devised a hybrid solution combining machine learning classifiers (association rules) with crowdsourcing. Analysis tools achieve an accuracy averaging around 80%, which has not significantly improved since the first attempts.

The content produced by most authoring tools does not have legal value. Those tools do not provide legal advice, but rather general guidelines for policy authoring. These guidelines may be sufficient, but their legal validity is not guaranteed and should be verified by a lawyer. As an example, Iubenda advertises for its "Attorney-level compliance" [54], but advocates for a professional legal consultancy [55], and do not guarantee conformity with the law, which they claim "only a lawyer can do". In other words: DC are responsible for the compliance with the law. They have to ensure that their privacy policies address all legal requirements, and to enforce the claims made in their policies.

3.3 Benefits

Legal value Natural language privacy policies are the only type of privacy policies with legal value as it is the standard format for legal texts. [5] Most legislations require DC to provide a lawful statement detailing the processing of personal data, and natural language privacy policies often aim to fulfill this obligation. Lawyers rely on natural language to evaluate whether privacy policies are correctly drafted: these policies contain all details to determine whether there has been a violation. Also, lawyers use these policies to check compliance with data protection regulations. However, a document holding legal value is not necessarily compliant with the law. For instance, lawyers or DPAs may check that all items required by the legislation are provided to DS, and auditors can check that data processing is performed according to the policy. Legal compliance is twofold: with respect to the information requirements, and with respect to the actual processing.

3.4 Limitations

Ambiguity Natural language privacy policies can be ambiguous [88], as they may be interpreted in different ways. Reidenberg *et al.* [88, 87] presented privacy policies to privacy experts, law and policy researchers, who were ultimately unable to agree on some aspects of the policies. They proposed a crowd-sourcing annotation to tackle this issue, but admit that it would only provide a partial solution. This ambiguity is mainly due to the fact that a statement in natural language can be interpreted in different ways. Ambiguity has a direct impact on the understanding, the enforcement, and the auditability of privacy policies. Ambiguity is also an explanation of the inaccuracy of analysis tools described in Section 3.2.

Understanding McDonald and Cranor [68] showed that it would take 200 hours a year for an average US citizen to read all the natural language privacy policies of the online services she used. This is clearly impractical, and thinking that DS read privacy policies before using a service is a *fictio juris*. All the more, nowadays, it seems highly inconvenient to spend a significant amount of time before using an online service.

Enforcement & auditability Because they are currently ambiguous, natural language privacy policies are difficult to enforce: natural language lacks precise semantics, making it difficult to decide how data must be processed by the underlying system. Likewise, natural language

[5] It does not however mean than other facets do not have legal value, rather that natural language privacy policies are mandatory.

privacy policies can also be hurdles to auditing: it can be difficult for an independent authority to compare stated and existing processing.

Summary

Natural language privacy policies are the most used medium to express privacy policies, and the tools used to assist their production can be categorized into *templates*, *generators*, and *retrievers*. These tools are often tailored to specific solutions, such as website or mobile applications, therefore restricting their scope. Analysis tools are not accurate enough to be trusted blindly. Natural language privacy policies are necessary for legal compliance, but suffer in practice from ambiguity and understandability.

4 Graphical privacy policies

In the previous section, we analyzed natural language privacy policies. They are necessary for legal compliance, but they can mislead DS when they attempt to read them, as they are often difficult to understand. As a consequence, other representations focused on DS understanding have been devised: privacy policies can also be expressed with graphical (in the broad sense) representations, that we denote *graphical privacy policies*. Graphical privacy policies cover icons sets and standardized notices as well as solutions providing additional information, such as warnings or judgments, sometimes combined with simple text [91]. Note that graphical privacy policies are rarely specifications, rather additional representations of natural language privacy policies. For instance, Android Permissions [47] represent privacy policies combining icons and text according to XML specifications. Graphical privacy policies often come from privacy advocates, but this is not only the case, notably since the WP29 explicitly mentioned icons as appropriate to convey privacy notices in their guidelines for transparency [111]. Privacy notices are means to inform DS, and are often understood as what we denote graphical privacy policies. Therefore, we include them in our study. We categorize each work based on: i) the elements in the taxonomy presented in Section 3 that it captures; ii) its features, whether it is made of icons, complementary text, or variants of those; and iii) the intended audience of the language, *e.g.*, DS or DC. Table 2 in Section 4.4 summarizes our study.

4.1 Content

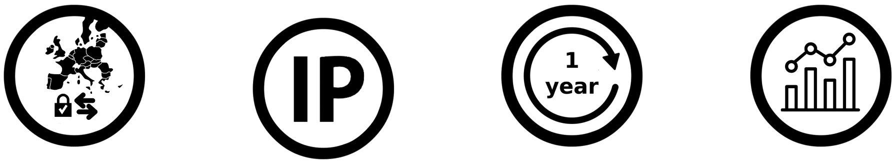
Based on their content, graphical privacy policies can be divided into three main types: *icons*, *standardized notices*, and *rating solutions*. This distinction emerged from an empirical analysis of the solutions studied, combined with impactful insights such as [28]. Graphical privacy policies based on icons intend to express the content of privacy policies, for DC and DS policies. Some of these icons try to cover all the items of the taxonomy introduced in Section 3. Other graphical privacy policies aim to express the same content as natural language privacy policies, but in a standardized and often comparable manner. Some graphical privacy policies provide rating information concerning certain aspects of privacy policies such as the transparency level or potential risks. These solutions were not devised to meet the same requirements as natural language privacy policies, the content of these graphical privacy policies in that respect is often restricted. In the following, we describe the content of graphical privacy policies according to the elements of the taxonomy and their type.

Sets of icons

The content of graphical privacy policies reviewed in this section lies in their icons, and sometimes in the simple explanations that comes with them. As an example, the symbol @ can represent collection of an email address, and a stylized calendar 📅 can represent retention time. But certain items are harder to express graphically. For instance, describing the legal basis of processing with the help of icons can easily be mistaken, and can mislead the intended audience instead of simplifying the understanding.

Usually, sets of icons (see [1, 37]) do not sufficiently express the items presented in the taxonomy. Instead, they express specific items. A solution such as Privicons [63] focuses on informing mail correspondents of how the data should be handled instead. Other solutions such as [92] include icons for selling, and second-use of data, but the type of data cannot be specified. Recently, Rossi and Palmirani [90] proposed a Data Protection icon set, named DaPIS. The interesting features of their approach is that they based the icon set on an ontology named PrOnto, and they tested their set in order to refine it. Moreover, it stands out by emphasizing items recently introduced in the legislations, such as DS rights or legal bases for processing. However, it does not consider the type of data.

Many privacy advocates also contributed to this area and provided numerous sets of icons (see [4, 86]). For instance, Mehldau [69] developed a set of 30 privacy icons describing the type of data, third-parties, the purpose of processing and the retention time. Recently, a set of privacy icons was designed by Privacy Tech [83]. This set considers many types of data, as well as advanced representations of sharing, such as adequacy transfer (see Figure 1).



(a) UE transfer adequacy

(b) Connection data

(c) One year conservation

(d) Audience measurement

Figure 1: Excerpt of the Privacy Tech icons

Another notable example of privacy icons are the android permissions [47], created by Google. They present icons combined with simple natural language. For each application installed on a mobile phone running Android, the permission manager presents a short graphical policy. Only little information is presented (the type of data collected, and processing in recent versions, but not the purpose for instance), and DS have to look into the natural language privacy policy in order to find more information.

Standardized notices

Another line of work considers the content described by the taxonomy as standardized notices. These standardized notices are often represented in tables [60], but the key concept is the common vocabulary among notices. Kelley *et al.* [60] represent policies in a table such as nutrition labels observed on food packaging. They present the fine-grained information in a table such as nutrition labels observed on food packaging. Polisis by Harkous *et al.* [50] can represent the natural language privacy policies as a combination of icons, highlights of the corresponding paragraphs in the natural language privacy policies, and a flow diagram (see Figure 2). Because Polisis relies on supervised machine-learning, *i.e.*, on a labeled corpus, it classifies natural language privacy policies in a standardized way. Such summarization has been accomplished by PrivacyCheck as well [114].

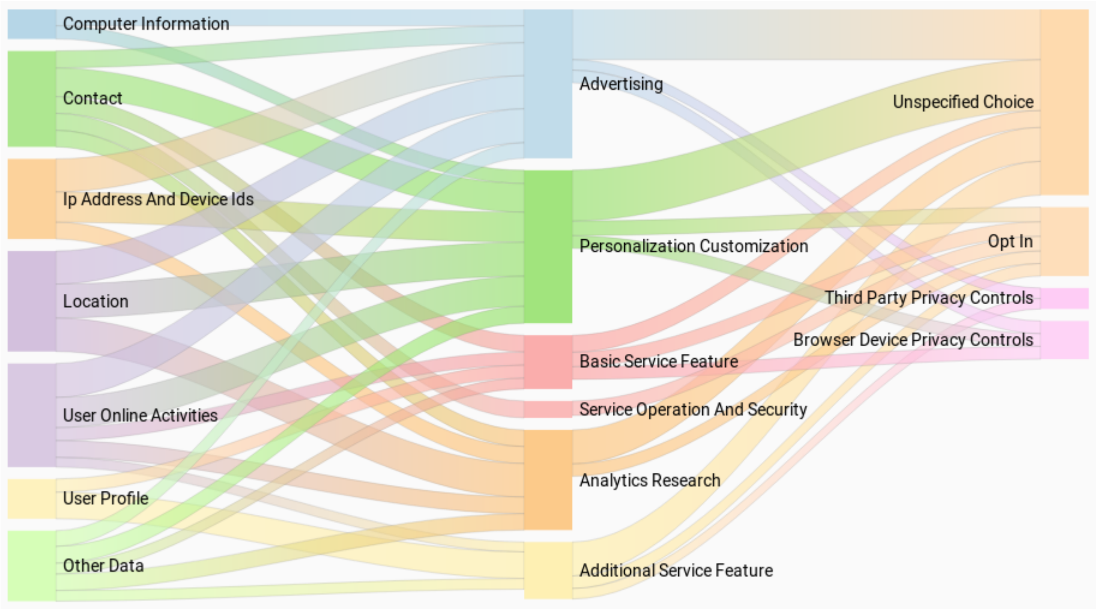

Computer Information
Contact
Ip Address And Device Ids
Location
User Online Activities
User Profile
Other Data
Advertising
Personalization Customization
Basic Service Feature
Service Operation And Security
Analytics Research
Additional Service Feature
Unspecified Choice
Opt In
Third Party Privacy Controls
Browser Device Privacy Controls

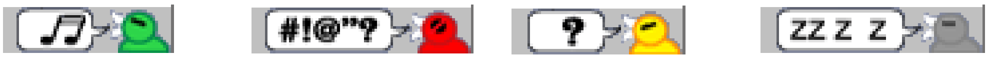
Figure 2: Example of a flow diagram in Polisis

Emami-Naeini *et al.* [38] conduct a survey in order to rank the factors of IoT devices purchase. They determine that security and privacy were among the most important factors of purchase, and consequently developed an IoT privacy label to improve information visualization. Cranor analyzes the impact of the development of standardized mechanisms of notice and choice in [28], and more specifically the efforts conducted around P3P. Cranor reconsiders the advances made in standardization, [6] as well as the limitations, lack of adoption and enforcement.

Rating solutions

Certain graphical privacy policies do not focus on expressing the items in our taxonomy, but present extra information related to privacy policies [108, 99, 49], often a judgment of the risk level associated to a DC policy, or a comparison between DS and DS policies. We denote them *rating solutions.*

Privacy Bird [29] is one of the first rating solution. It consists of a colored bird, where the color indicates the matching (green for a match between the DS policy and the website's DC policy, red for conflict, yellow for uncertain, gray when disabled) (see Figure 3).



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

Figure 3: Privacy Bird

[6] Note that icons are considered as part of standardization efforts in [28].

It is represented as an add-on for Internet Explorer, restricted to Microsoft Windows. A dedicated website [29] provides an explanatory tour as well as a feature named *privacy finder*: Privacy Bird is then used as an indicator when browsing the web [20, 28]. Privacy finder displays the search results of a search engine, combined with the analysis of Privacy Bird. The bird was placed alongside search results, and was influential in the choice of shopping websites [103], notably when the items being purchased were likely to have privacy concerns [36]. DS could rank the results according to the matching between their DS policy and the websites' DC policies. In [75], Pardo & Le Métayer present a web interface to inform DS about the potential risks of their privacy policies. The interface is composed of a user-friendly form for DS to input their privacy policies and a set of risk analysis questions, *e.g.*, "Can company X collect my data?". Additionally, DS may introduce risk assumptions in order to specify possible misbehaviors that the collecting parties can perform. In Section 5, we describe in detail the underlying privacy language and the automatic risk analysis. The ToS;DR initiative helps DS understanding the risks associated to a DC policy [101]. It started in 2011 during the Chaos Communication Camp. ToS;DR comprises not only icons, but also results from crowdsource analyses in simple language. The idea of the project is to assess the data practices of web services by giving them badges, awarded by the project's community. Once a service has enough badges to assess the level of protection of their terms for users, a class is automatically assigned by pondering the average scores.

4.2 Tools

Tools for representing graphical privacy policies are tailored to the web, and are often found as add-ons for web browsers.

Privacy Bird is represented as an add-on for Internet Explorer, restricted to Microsoft Windows. ToS;DR is also an add-on, for both Firefox and Chrome, as it ranks policies based on crowdsourced analyses by a community directly within the web browser. A website has been built to present Polisis [80], and add-ons for Chrome [78] and Firefox [79] are available (the add-ons redirect to the corresponding part of the website). The add-on "Disconnect Privacy Icons“ [34], in collaboration with TRUSTe, which evolved from Raskin's set of icons for Mozilla [86], provided an interactive and comprehensive view of privacy policies within the browser. The add-on would display icons according to a website privacy policy if the website complies with the solution.

All of these tools focus on the web, whereas the IoT is left unchallenged in that respect.

4.3 Benefits

Graphical privacy policies cannot be seen as legal commitments because they lack precise meaning, but they have other benefits: they can foster understanding.

Designed for lay-user understandability Many solutions coming from privacy advocates (*e.g.*, [69, 86, 4, 85]) aim to provide intelligible information to lay-users. These solutions were built with the will to popularize natural language privacy policies, and were designed to be understood quickly and take simplicity account. It is also the case for academic solutions such as the privacy labels [60], which "allows participants to find information more quickly and accurately". Based on the principle that existing natural language privacy policies do not convey intelligible information about data collection and processing, Kelley *et al.* strove to provide a universal solution. Graphical privacy policies can also convey intelligible notices for scientists and physicians using sensitive datasets, such as the DataTags [99, 15]. Their solution includes a simplified interface for access requirement to medical data, as this type of data is mostly restricted to medical practitioners and researchers.

Attempts were made to analyze what icons were recognizable and to measure their reliability. Egelman *et al.* [37] crowdsourced privacy indicators for the IoT. In their study, they found out that some icons are well-recognized (*e.g.*, the camera symbol was recognized by more than 95% of participants as representing *video recording*), while others are not (only 3.6% recognized the *voice command & control* icon). Kelley *et al.* [61] conducted a user study, to refine their privacy label. They compared the accuracy of information retrieval between their proposition and natural language privacy policies in natural language. As a result, they purposely combined simple natural language to prevent confusion, notably for the terms *opt-in* and *opt-out.* [7] A promising attempt to measure understandability has been conducted in [90]. They performed three evaluations of their icon set in order to improve the recognition of icons. However, they regret the lack of diversity in the participants' panel, notably for the educational level.

4.4 Limitations

Ambiguity Though accessible to and often designed for lay-users, graphical privacy policies may be interpreted in different ways, thus leading to ambiguities (see Section 4.1). The same icon can be interpreted in different ways according to the differences in culture, education level, or context *etc.* For instance, a euro symbol € can represent the commercial use of collected data, or that DS will be paid for having her data collected. Little has been done to produce a reasonably recognized set of icons for privacy — *e.g.* validated by a user study — despite the attempts of [37] and [1, Chapter 15] to see what were the most recognizable icons, and of [61] to provide a graphical policy where results could be found accurately. The three stages evaluation of Rossi and Palmirani [90] however leads the path to less ambiguous graphical privacy policies.

Incompleteness Graphical privacy policies are limited by their restricted scope. As seen in Section 4.1, existing graphical privacy policies are not as expressive as natural language privacy policies, due to the limited number of icons available. Some aspects are rarely mentioned, others only in complementary text and not in the graphical part of the policy. One aspect in particular is never mentioned in graphical privacy policies (policy change), and another is considered in only one work (legal basis).

Claim over legal compliance Some graphical privacy policies, such as cookie consent notices, have been used to claim legal compliance to retrieve consent. Degeling *et al.* [31] observed that a significant part (16%) of websites added cookie consent notices after the GDPR, but these notices do not always comply with transparency requirements according to Utz *et al.* [106] as they tend to use Dark Patterns [48] to lure DS into giving their consent.

Summary

Graphical privacy policies are promising for conveying summarized versions of natural language privacy policies, and they can rely on user-friendly tools to be adopted. However, they should come with explanations to ensure human understanding and mitigate their restricted content. See Table 2 for a visual and global overview of our study on graphical policies.

[7] Note that they also test the speed of retrieval, as well as comparisons between DC policies in addition to information retrieval.

	Icons	Simple text	Rating	Standardized notice	DC	DS	1^{st} party	3^{rd} party	DS rights	Data retention	Data security	Legal basis
Privacy Bird [29]	✓		✓		✓	✓	—	—	—	—	—	—
Rundle [92]	✓	✓				✓	◐	○	●	○	◐	○
Mehldau [69]	✓				✓		●	◐	○	●	●	○
Privacy Commons [4]	✓				✓		◐	◐	○	○	○	○
Privacy Nutrition Label [60]		✓		✓	✓		●	◐	○	○	○	○
Primelife [1, Chapter 15]	✓				✓		—	—	—	—	—	○
Raskin [86]	✓				✓		◐	◐	○	●	○	○
Privicons [63]	✓					✓	○	◐	○	○	○	○
Privacy wheel [108]		✓	✓		✓		○	◐	○	○	◐	○
Android permissions [47]	✓	✓			✓		◐	○	○	○	○	○
"Is this thing on?" [37]	✓				✓		◐	○	○	○	○	○
Datatags [15]		✓	✓		✓		○	○	○	○	●	○
Hagan [49]		✓	✓		✓		◐	◐	◐	◐	○	○
Polisis [50]	✓	✓		✓	✓		●$_a$	●$_a$	●$_a$	●$_a$	●$_a$	○
Privacy Tech [83]	✓				✓		●	◐	○	●	◐	○
DaPIS [90]	✓				✓		◐	◐	●	○	○	●
IoT label [38]		✓		✓	✓		●	●	○	●	●	○
ToS DR [101]		✓	✓		✓		●$_a$	●$_a$	●$_a$	◐$_a$	◐$_a$	○
	Features				Type		Content					

Table 2: Categorization of graphical privacy policies. We use the subscript $_a$ to denote that a solution extensively uses natural language besides graphical representations. The column indicates: *Features,* whether the solution is made of icons or provides ratings about policies, and whether it provides explanation in simple natural language; *Type (Type of policy),* whether the solution expresses a DC or a DS policy; *Content,* whether the solution can express the different items enumerated in Section 3. We use ● to denote that the solution can express most or all values; ◐ to denote that the solution expresses few values of the items, and is mostly insufficient; ○ to denote that the solution cannot express the item; and "—" to denote that the material does not permit judging whether the solution can express this item or not. Note that some items of the taxonomy are omitted since no solution includes them.

5 Machine-readable privacy policies

Many efforts have been devoted to the expression of *machine-readable privacy policies* — *i.e.*, privacy policies that can be automatically processed by computers. Most of these efforts were made by academics, and result in what has been called *privacy languages.* According to Kasem *et al.* [59], a privacy language is "a set of syntax and semantics that is used to express policies". Many privacy languages have been proposed in the past twenty years [59, 107]. We review here the different ways in which privacy languages are used to express machine-readable privacy policies. In particular, we categorize each work based on: i) the elements in the taxonomy presented in Section 3 that it captures; ii) the type of enforcement mechanism it uses and whether it has been implemented; iii) additional tools for policy analysis or comparison; iv) the intended audience of the language, DS or DC; and v) whether it is intended to be used by lay-users. Table 3 summarizes our study.

5.1 Content

Here we describe the content that machine-readable privacy policies include. This content is determined by the syntax of the privacy language. Many languages are defined using machine-readable formats such as XML so that they can be automatically processed by machines. Other languages, however, are based on mathematical definitions (*e.g.*, logical languages), thus enabling the possibility of reasoning about them — these languages can easily be expressed in machine-readable formats due to the lack of ambiguity. Another important factor is the target audience of a language, *i.e.*, DC, DS or both. In what follows, we describe the content of machine-readable languages (according to the items defined in Section 3), the format used to express the policies and their target audience.

Access control languages such as XACML [9] and RBAC [94] have been among the first languages used for the specification of machine-readable privacy policies. Typically, these policies include the datatype to which they apply, and the set of entities with access privileges. Some extensions such as GeoXACML [66] include conditions depending on geolocation information. Usage control (UCON) [76, 81] extends access control with the notion of *obligations*, *i.e.*, actions to be executed after data has been received — *e.g.*, "do not transfer data item *i* to Service X" or "remove data on 25/05/2018". These obligations make it possible to express items such as retention time, purpose and allowed data transfers. The Obligation Specification Language (OSL) [51] is a fully-fledged UCON language. OSL leverages Digital Right Management systems (DRMs) [42] to enforce the obligations in UCON policies. Both access control and UCON are used by DC to define their policies, and do not offer mechanisms for DS to express their policies.

Several languages focused on expressing privacy policies have been developed. One of the first is the "Platform for Privacy Preferences" (P3P) [27]. P3P appeared as a policy language for the web. P3P policies are specified in XML format, and include: purpose, retention time, and *conditions*. Conditions may be opt-in and/or opt-out choices for DS, or preferences based on enterprise data — *e.g.*, DS's credit or service usage. Many extensions to P3P have been proposed [64, 11, 7] — for instance, E-P3P [11] extends P3P with obligations *à la* UCON. After P3P, new languages with similar syntax have been proposed: "Enterprise Policy Authorization Language" (EPAL) [12], "An Accountability Policy Language" (A-PPL) [13], "Customer Profile Exchange" (CPExchange) [18], "Privacy Rights Markup Language" (PRML) [115], "Purpose-to-Use" (P2U) [56] and "Layered Privacy Language" (LPL) [45]. These languages are similar than P3P in terms of content, but bring numerous enhancements in terms of usability and enforcement (see Section 5.2).

Formal privacy languages (*formal languages* in the following) comprise a different approach to express privacy policies. CI [73], PrivacyAPIs [67], SIMPL [65], PrivacyLFP [33], S4P [17], QPDL [107], and PILOT [75] are languages that have their syntax and semantics defined mathematically. More precisely, they use formal languages such as *Linear Temporal Logic* [53], *First-Order Logic* [53] or *Authorization Logic* [5]. However, not all of these formal languages have the same focus. S4P, SIMPL and PILOT are focused on expressing DS and DC policies. Similarly to the languages above, it is possible to express types of data, conditions, purpose, retention time and allowed data transfers. Conditions are often more sophisticated than that of the previous languages as they are based on logical languages. On the other hand, CI, PrivacyAPIs and PrivacyLFP focus on encoding privacy regulations such as HIPAA [104], COPPA [41] or GLBA [105]. As a consequence, their expressive power is greater than languages focusing on DS and DC policies. They include temporal operators that make it possible to express policies about past and future events. For example, Barth *et al.* [16] express the following statement from COPPA "[...] an infant can only send identifiable information to a website, if her parent have previously sent their consent for data collection". QPDL is a meta-language to reason about privacy languages. While privacy policies can be expressed in QPDL, it is not its intended use — it was designed to formally reason about policy languages. Jeeves [112] is a programming language with built-in support for a limited form of privacy policies. It allows programmers to specify confidentiality conditions based on the execution context. For instance, in a double-blind conference management system, only organizers can see paper authors until the review process is completed.

5.2 Tools

In this section, we describe the mechanisms used to enforce machine-readable privacy policies, and existing tools to compare and perform analyses on policies for instance.

Formal Semantics Formal languages give meaning to their privacy policies by means of *formal semantics*. Typically, these semantics define what events may be executed depending on the privacy policies selected by the actors interacting in the system. SIMPL, PILOT, S4P and CI use trace semantics, *i.e.*, they define the sequences of events (traces) that respect the privacy policies of DS and DC. Jeeves has its semantics formalized using lambda calculus [23]. Rei semantics is defined in Prolog [24]. Though precise and unambiguous, there is a gap between the definition of formal semantics and the real implementation. Nevertheless, this gap may be very small, *e.g.*, Jeeves lambda calculus semantics were implemented as a Scala library, Rei's semantics are encoded in Prolog, and PILOT semantics are implemented as a Promela model [52].

Informal Semantics Access control, UCON and privacy dedicated languages have their enforcement mechanisms specified as W3C specifications, specification languages such as UML, or they are simply implemented using a general purpose programming language. All these languages use *request evaluation engines* to enforce privacy policies. Request evaluation engines take a *data request* and evaluate whether the requester may access the data based on the privacy policies. The content of data requests depends on the language. For instance, in RBAC, data requests contain type of data and the role of the requester. If the role of the requester matches one of the roles allowed by the policy associated with the data, then data can be accessed. Most languages do not have mechanisms to enforce that data will be used according to the policies after being collected — *e.g.*, checking that data is deleted before the retention time — but there are some exceptions. LPL erases automatically data from the central repository after the retention time has elapsed. UCON-based languages use DRM to enforce obligations.

Policy comparison For some languages, algorithms have been devised to automatically compare policies. The goal is to determine, given two policies, which one is more restrictive. For example, a policy that allows data processing for research purposes during 7 days is more restrictive than a policy that allows data processing for advertisement and research during 90 days. Policy comparison is necessary to mechanize consent management. If a DC policy is more restrictive than a DS policy, then DS privacy preferences are satisfied. EPAL, P3P and PILOT include algorithms and tools to compare policies. CI, SIMPL and S4P follow a different approach. They define how restrictive a policy is, based on its semantics. [8] Languages that do not distinguish

[8] Using trace semantics it is possible to compare policies based on the set of traces satisfying the policy. The less traces a policy satisfies, the more restrictive it is.

DS and DC policies — such as RBAC, A-PPL, or OSL — do not have algorithms to compare policies. This is not surprising, their goal is to enforce a policy typically defined by DC or system administrators.

Analysis tools Formal languages often come with tools to perform different types of automatic analyses. PILOT uses model-checking [14] to perform risk analysis. Given a DS policy and a set of risk assumptions, such as "Company X may transfer data to Company Y", it is possible to automatically answer questions such as "Can Company Z use my data for advertisement?". Rei comes with a Prolog interface where queries such as the above can be asked. PrivacyAPIs also uses model-checking to automatically verify properties about the privacy regulation HIPAA. It can, for instance, be used to determine who can access patients medical files depending on their content or role.

5.3 Benefits

Machine-readable privacy policies have four main benefits: 1) they can be automatically enforced; 2) they can be audited; 3) it is possible to reason about their correctness; and 4) they make it possible to automate certain procedures. In what follows we explain each of these benefits in detail.

Enforcement As opposed to natural language or graphical policies, machine-readable policies can be automatically enforced. As described in Section 5.2, all policy languages have the means to guarantee that data is accessed according to the policies. Languages based on request evaluation are easier to implement (cf Section 5.2), and are consequently more widespread. Typically, every party holding personal data must implement a part of the request evaluation engine. In general, the implementation of formal languages is more complicated. They require tracking actions applied to the data, or inferring what are the purposes for which data is used — as opposed to simply control access to data.

Auditability Machine-readable privacy policies make it possible to audit whether data is being handled according to the privacy policies. This functionality is of great value for DPAs. Auditing mechanisms are typically implemented as logs recording the operations performed on sensitive data. For instance, EPAL requires to create an audit trail of access to keep track of whom has accessed personal data. In A-PPL, on the other hand, it is possible to specify *auditable operations* such as read or delete, and the enforcement records in a log every time that such operations occur.

Correctness The lack of ambiguity in policy languages makes it possible to precisely reason about their correctness, *i.e.*, that data is handled as stated in the privacy policies. This is specially true for formal languages. Their semantics can be used to formally prove certain correctness properties. For example, S4P, SIMPL and PILOT have been used to prove global properties such as "data is never used after its retention time", or, "data is always used according to DS policies". Moreover, languages focused on modeling privacy regulation — CI, PrivacyAPIs and PrivacyLFP — can be used to find inconsistencies in the regulation (if any). For example, it was possible using PrivacyAPIs to find unexpected ambiguities in HIPPAA.[9] It is important to remark that there exists a gap between the formal semantics and its implementation — technical details not modeled in the semantics may lead to unforeseen violation of the properties.

[9] These ambiguities were also found by commenters four years after it was enacted [67].

Therefore, formal languages should include auditing mechanisms, as the languages mentioned in the previous paragraph.

Automation Machine-readable privacy policies allow for automating procedures such as policy communication and consent management. Automatic policy communication facilitates transparency by boosting DS awareness about how their data is being handled — notably in ubiquitous systems where passive data collection is the norm [30]. Automatic consent management can empower DS if managed in a protective way — *e.g.*, by mitigating the burden of choice [98] — and facilitate the retrieving of an informed consent for DC. Cunche *et al.* [71] devise a generic framework to manage informed consent in the IoT, using DS and DC policies based on PILOT [75]. Automatic communication of privacy policies also makes possible a negotiation of privacy choices: DC and DS can interact more quickly by means of machines.

5.4 Limitations

The main limitations of machine-readable privacy policies are their lack of adoption and usability. As adoption relies among other things on human-understandability, understandable and usable policies seems to be a condition *sine qua non* for their adoption.

Human understandability One of the most recurring criticism of machine-readable privacy policies is their lack of human understandability. Only a handful of languages such as SIMPL, LPL or PILOT take it into account: they include a natural language version of each policy. It is however questionable whether they can actually be understood. To put things into perspective, the OECD [74] conducted a study showing that two third of adults from developed countries cannot conduct a medium-difficulty task related to ICT environments. Although privacy management was not mentioned in the study, it is a medium-difficulty task, and solutions tackling privacy management must consider information-illiteracy. Machine-readable privacy policies should be expressed in languages close to natural language in order to be understood, or be complemented by friendly interfaces.

Lack of adoption Another pitfall for machine-readable privacy policies is their lack of adoption. It is arguably a consequence of poor human understandability. Most of the work done on privacy languages had few or no impact, apart from P3P. With the other solutions stemming from it (APPEL, E-P3P, . . .) and the extension Privacy Bird for Internet Explorer, P3P obtained recognition out of the academic scope. It has been an official set of specifications of the W3C supported by the web browser Internet Explorer. Note that other languages were published as specifications by companies [18, 9] and can therefore be considered as having had some recognition. On the other hand, most formal languages lack a practical and scalable implementation, making it difficult to use them in practice. Usability, implementation and widespread recognition are a rare combination in privacy languages.

Summary

Machine-readable privacy policies can provide means to express unambiguous privacy policies, and can be enforced as well as audited by authorities. However, they are often unintelligible for lay-users, which results in a lack of adoption. We provide a visual and global overview of machine-readable policies in Table 3.

	Usability	Syntax	Enforcement	Implemented	Tools	DS	DC	Time	Space	1^{st} party	3^{rd} party	DS rights	Data security	Data Retention
P3P [27]		XML	Informal	✓	Comparison	✓	✓	◐	○	●	○	○	○	◐
CPExchange [18]		XML	Informal				✓	○	○	●	○	○	◐	●
PRML [115]	✓	XML	Informal				✓	○	○	●	◐	○	◐	◐
APPEL [64]	✓	XML	Informal	✓		✓	✓	◐	○	●	○	○	○	◐
E-P3P [11]		XML	Formal			✓	✓	◐	○	◐	◐	○	◐	◐
Rei [58]		Formal	Formal		Analysis		✓	◐	◐	◐	◐	○	○	◐
Xpref [7]	✓	XML	Informal	✓		✓		◐	○	●	○	○	○	◐
XACML [9]		XML	Informal	✓			✓	○	○	◐	○	○	○	○
EPAL [12]		XML	Informal	✓	Comparison		✓	○	○	●	◐	○	○	◐
CI [16]		Formal	Formal			✓	✓	◐	◐	◐	●	◐	○	○
SIMPL [65]	✓	Formal	Formal			✓	✓	◐	○	●	●	●	◐	●
S4P [17]	✓	Formal	Formal			✓	✓	◐	○	●	●	○	○	◐
Jeeves [112]		Formal	Formal	✓			✓	◐	○	◐	◐	○	◐	○
P2U [56]	✓	XML	Informal			✓		○	○	●	●	○	○	●
QPDL [107]		Formal	Formal			✓	✓	◐	●	●	●	◐	●	○
RBAC [94]		XML	Informal	✓			✓	○	○	◐	○	○	○	○
OSL [51]		Formal	Formal	✓			✓	◐	○	●	●	○	●	●
GeoXACML [66]		XML	Informal	✓			✓	○	●	◐	◐	○	○	○
A-PPL [13]		XML	Informal				✓	○	◐	●	●	◐	◐	◐
LPL [45]	✓	XML	Informal	✓		✓	✓	○	○	●	●	○	◐	●
PrivacyAPIs [67]		Formal	Formal		Analysis	✓	✓	◐	○	●	●	●	●	●
PrivacyLFP [33]		Formal	Formal			✓	✓	◐	○	●	●	●	●	●
PILOT [75]	✓	Formal	Formal		Analysis	✓	✓	●	●	●	●	◐	○	●
			Features			Audience		Conditions				Content		

Table 3: Categorization of privacy languages. The columns indicate: *Usability,* whether the language is intended to be understood by DS; *Syntax,* whether the syntax of the language defined in XML or a formal language; *Enforcement,* whether the language has a formally or informally defined enforcement; *Implemented,* whether the language has been implemented; *Tools,* the type of available tools for the language; *Audience,* whether the language can describe a DS or a DC policy; *Conditions,* whether the language supports conditional rules describing when and/or where data may be collected; *Content,* whether the language can express the items described in Section 3.1. We use ● to denote that the item is explicitly included in the language; ◐ to denote that the item is partially supported; and ○ the item is not present in the language and cannot be encoded. Some items of the taxonomy are omitted since no solution includes them.

6 Insights

In this section, we provide several insights that we identified as a result of our study.[10] We show that each facet is tailored to a specific audience, and that this is both 1) what makes it beneficial, but also 2) an obstacle to the compliance with all the requirements stated in Section 1 (*i.e.*, legal validity, understandability by all parties, and enforceability through auditable mechanisms). In Section 6.1, we discuss why a single facet cannot comply with every requirement. In Section 6.2, we put in perspective the works which attempt to overcome the limitations of mono-faceted solutions, and provide guidelines for designing privacy policies that aim to cover the three facets. Finally, in Section 6.3, we discuss the coverage of the items in our taxonomy by privacy policies

[10] We refer the reader to Table 4 in Appendix A for an overview of the juxtaposition of the results combining all facets.

in each facet.

6.1 Limitations of mono-faceted solutions

A single facet cannot cover all the requirements of privacy policies. This is due to the tension between: i) the suitability for lawyers, ii) the suitability for lay-users, and iii) the automatic enforcement by machines. Concretely, there are details that only have meaning in one facet and are irrelevant in others. Details specific to natural language privacy policies are used by lawyers to check that the policy complies with privacy protection regulations — such as the GDPR. In general, lay-users may not have the knowledge to fully understand these details, which makes it less accessible for them. Likewise, natural language privacy policies do not include low level details related to the enforcement of the policies by a machine — those details are often unnecessary for law enforcement and require lawyers to be familiar with those technicalities. Graphical privacy policies have the objective of being understood by a general audience, but this form of privacy policies have no use for lawyers or enforcement by machines. Machine-readable privacy policies aim at being enforced by machines. They are written in a machine-readable format, and they include all the necessary details for the underlying system to enforce them. These details make them difficult to understand by humans, and are, consequently, unsuitable for lawyers and lay-users.

Illustrative example Consider the icon (1 year) from Privacy Tech Icons (see Section 4) that denotes that data is deleted after 1 year, and this excerpt of Facebook's privacy policy: "[...] 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.*" Facebook's policy is more precise than the icon: it refers to concrete data which is produced after a certain user action. However, these details may not be of prime interest for some lay-users, at least in a first stage. For instance, they may not know what a log entry or a query are. Hence, this level of detail may be counter-productive for lay-users. Yet, this information is required to determine whether Facebook is processing data according to the policy. Thus, it cannot be omitted for legal purposes or for users who may be interested in more detailed information. Consider also an excerpt of the policy in APPEL-P3P: `<retention-time days=182 xmlns=".../P3P/retention-time/"/>`. This policy includes details that are not present in the natural language policy above: The format of the policy, XML; the parameter `xmlns`, required so that the computer can retrieve the set of possible values for the element in the policy (the XML namespace); and the fact that retention time must be specified in days. These details are often irrelevant for law enforcement, and make the policy difficult to understand for lawyers and lay-users.

6.2 Multi-faceted privacy policies

In many cases, limitations in one facet can be addressed by other facets. For instance, natural language privacy policies may use graphical privacy policies to enhance readability, and machine-readable privacy policies to be automatically enforced. In this section, we study *multi-faceted privacy policies*, *i.e.*, policies that combine any of the three facets studied in this paper: natural language, graphical, and machine-readable. We present existing works on multi-faceted privacy policies, discuss challenges in defining them, and provide guidelines on addressing those challenges.

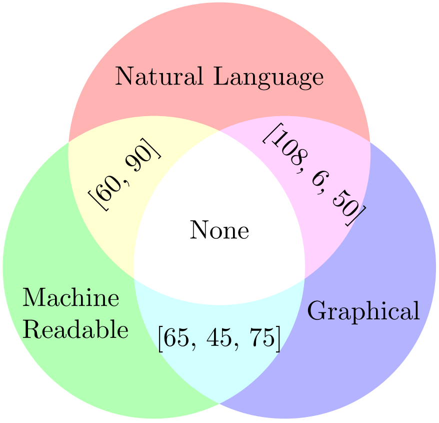


Figure 4: Works on multi-faceted privacy policies grouped by combination of facets.

Existing works on multi-faceted privacy policies.

Several initiatives are already proposing multi-faceted solutions. Harkous *et al.* [50] combine natural language and graphical privacy policies, bringing together the accessibility of icons and simple text with the legal value of a natural language privacy policy. Similarly, [108, 6] follow a layered approach to combining graphical and natural language policies. Users are first presented with a graphical or simplified version of the policy, which can be refined in several steps all the way to the legal notice. SIMPL [65], PILOT [75] and LPL [45] combine natural language and machine-readable privacy policies. They provide an enforceable policy that helps DS to better understand their choices and provide informed consent — as required by the GDPR. Similarly, [60, 90] add graphical representations for P3P policies, resulting in intelligible and enforceable privacy policies.

Figure 4 summarizes the landscape of work on multi-faceted policies. Note that existing solutions combine at most two types of privacy policies, and their number is small. Notably, no existing solution encompasses the requirements for legal compliance, understandability, and enforceability.

Guidelines and challenges

Here we set guidelines and discuss challenges in designing multi-faceted privacy policies, ideally covering all three facets. We describe two approaches: *unified* and *compound*.

Unified In a unified approach, a core facet is defined and the remaining facets are generated from the core using a *policy generator*. Natural language is a suitable candidate as core facet — it is a legal requirement and cannot be omitted. Machine-readable privacy policies could also be considered as core facet as natural language could automatically be generated from them. Graphical representations are not suitable for this purpose; they lack the precision required to generate natural language or machine-readable privacy policies. The main challenge of this approach is ensuring the soundness of the generated policies. Existing solutions compromise the expressiveness of a facet to generate policies in other facets. For example, SIMPL uses constrained natural language so that privacy policies can be enforced by a machine. As mentioned above, each facet has details that are not addressed by other facets (see Section 6.1). Thus, another requirement for a policy generator is to include the details specific to the facets generated,

	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 5: Coverage of taxonomy items by different types of privacy policies, and the privacy legislations in Table 1.

although absent in the core facet. To sum up, the challenge for a unified approach is to ensure that the facets generated from the core keep what makes them beneficial and preserve the meaning of the core facet.

Compound The compound approach consists in taking mono-faceted policies and using them together. For instance, we could use legal text, icons from DaPIS, and P3P. Most existing solutions work this way. Unlike the unified approach, the facets in the compound approach include the specific details of each facet — insofar that it combines existing mono-faceted solutions. The main challenge here is ensuring consistency between the different facets, *e.g.*, does the machine-readable facet accurately represents the legal text? Currently, consistency is manually checked by designers as it is a very difficult task to automate. Tool support and strict guidelines may systematize this process and reduce errors. The machine-readable facet is suitable for full automation, but we argue that the natural language and graphical facets require expert supervision (lawyers and designers, respectively). In a nutshell, the main challenge is to devise methods ensuring consistency between the different facets.

6.3 Missing taxonomy items

As we saw in the previous sections, not all items of our taxonomy are covered by existing solutions. Here we quantify the coverage of each item by the solutions we have studied. The aim of this section is to shed light on "forgotten" items, and hopefully guide future research on these directions.

Figure 5 summarizes our results. 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. For instance, the item 1st party is completely covered by 69% of policy languages (forth column, first row in Figure 5).

The 1st and 3rd party taxonomy items have the best coverage in our study. Probably due to the fact that they express the most relevant information regarding data collection and processing for DS. These items are followed (in terms of coverage) by Data Retention and Data Security, which are absent from almost 50% of the languages studied — except for Data Retention in machine-readable privacy policies. DS Rights is absent in around 70% of the graphical and machine-readable policy languages. Most likely, because they refer to information difficult to express graphically, and they are outside the realm of what machine-readable language are designed for. Finally, legal basis and policy change are absent from all the studied work. Possibly due to their absence in most legislations; 60% and 80% respectively.

7 Final discussion

Related work This work is not the first to propose an overview of the different manners to express privacy policies. Schaub *et al.* [95] present the requirements and best practices for presenting privacy *notices*. Their work focuses on providing users impactful notices. In other words, they study how well they understand the messages conveyed by the privacy notices. Cranor [28] describes the notice and choice mechanisms, what P3P attempted to do to palliate issues raised by these principles, and why it failed in doing so. These solutions focus on the design of privacy policies to enhance usability for lay-users. In our work, we focus on connections of the graphical and machine-readable privacy policies with legal requirements (*i.e.*, natural language privacy policies). Although both of these articles consider the machine-readibility of privacy notices, we highlight benefits that they did not consider, such as the possibility it offers for enforcement, auditing or automatic consent management.

To the best of our knowledge, this work is the first to systematically study privacy policies based on the different means of expression or “facets”, and how these facets must be combined to provide legally valid, usable and enforceable privacy policies.

Conclusion In this paper, we have studied the different ways to express privacy policies: in natural language, with graphical representations, and using machine-readable means. We have categorized the existing work in each facet according to a taxonomy of privacy policies, as well as their specific features. Additionally, we have studied the benefits and limitations of each facet, and we have shown that the limitations of one facet can be addressed by the benefits of the other facets. We have studied the combination of different facets, which overcomes limitations by bringing together the benefits of each facet, and provided guidelines to design multi-faceted privacy policies. We have made explicit the degree of coverage of the items in our taxonomy by the surveyed work, thus shedding light on future research directions on the design of mono- and multi-faceted privacy policies. We envision this work as an effort to facilitate and boost collaborative work between the legal domain, design, and computer science, and to provide a big picture of how transparency can be ensured through privacy policies.

Acknowledgments

This work has been partially funded by the CHIST-ERA project UPRISE-IoT (User-centric Privacy and Security in the IoT) and the ANR project CISC (Certification of IoT Secure Compilation). We address special thanks to Daniel Le Métayer for his insights and comments on the early stages of this article.

References

[1] Privacy and Identity Management for Life, 2011. URL http://link.springer.com/10.1007/978-3-642-20317-6.

[2] California's New Privacy Law: It's Almost GDPR in the US, 2018-07-02. URL https://www.bankinfosecurity.com/californias-new-privacy-law-its-almost-gdpr-in-us-a-11149.

[3] 3DCart. Create an online store with 3dcart store builder., 2019-03-26. URL https://www.3dcart.com/personalized-policy.html.

[4] Helton Aaron. Privacy Commons Icon Set .:aaron.helton:., 2009. URL https://web-beta.archive.org/web/20090601215200/http://aaronhelton.wordpress.com/2009/02/20/privacy-commons-icon-set.

[5] Martín Abadi. Logic in access control. In *Proceedings of 18th IEEE Symposium on Logic in Computer Science (LICS 2003), 22-25 June 2003, Ottawa, Canada*, page 228, 2003. doi: 10.1109/LICS.2003.1210062.

[6] Marty Abrams and Malcolm Crompton. Multi-layered privacy notices: A better way. 2 (1):1–4, 2005.

[7] Rakesh Agrawal, Jerry Kiernan, Ramakrishnan Srikant, and Yirong Xu. An XPath-based preference language for P3P. In *Proceedings of the 12th International Conference on World Wide Web*, pages 629–639. ACM, 2003. URL http://dl.acm.org/citation.cfm?id=775241.

[8] Waleed Ammar, Shomir Wilson, Norman Sadeh, and Noah A. Smith. Automatic categorization of privacy policies: A pilot study. 2012. URL http://repository.cmu.edu/lti/199/.

[9] Anne Anderson, Anthony Nadalin, B. Parducci, D. Engovatov, H. Lockhart, M. Kudo, P. Humenn, S. Godik, S. Anderson, S. Crocker, et al. Extensible access control markup language (xacml) version 1.0. 2003. URL http://courses.cs.vt.edu/cs5204/fall05-kafura/Papers/Security/XACML-Specification.pdf.

[10] Wolfgang Apolinarski, Marcus Handte, and Pedro Jose Marron. Automating the Generation of Privacy Policies for Context-Sharing Applications. pages 73–80. IEEE, 2015-07. ISBN 978-1-4673-6654-0. doi: 10.1109/IE.2015.18. URL http://ieeexplore.ieee.org/document/7194273/.

[11] Paul Ashley, Satoshi Hada, GÃ¼nter Karjoth, and Matthias Schunter. E-P3P privacy policies and privacy authorization. In *Proceedings of the 2002 ACM Workshop on Privacy in the Electronic Society*, pages 103–109. ACM, 2002. URL http://dl.acm.org/citation.cfm?id=644538.

[12] Paul Ashley, Satoshi Hada, GÃ¼nter Karjoth, Calvin Powers, and Matthias Schunter. Enterprise privacy authorization language (EPAL). 2003.

[13] Monir Azraoui, Kaoutar Elkhiyaoui, Melek Önen, Karin Bernsmed, Anderson Santana de Oliveira, and Jakub Sendor. A-PPL: An Accountability Policy Language. In *Data Privacy Management, Autonomous Spontaneous Security, and Security Assurance - 9th*

International Workshop, DPM 2014, 7th International Workshop, SETOP 2014, and 3rd International Workshop, QASA 2014, Wroclaw, Poland, September 10-11, 2014. Revised Selected Papers, volume 8872 of *Lecture Notes in Computer Science*, pages 319–326, 2014.

[14] Christel Baier and Joost-Pieter Katoen. *Principles of Model Checking.* MIT Press, 2008. ISBN 978-0-262-02649-9.

[15] Michael Bar-Sinai, Latanya Sweeney, and Merce Crosas. DataTags, data handling policy spaces and the tags language. In *Security and Privacy Workshops (SPW), 2016 IEEE*, pages 1–8. IEEE, 2016. URL `http://ieeexplore.ieee.org/abstract/document/7527746/`.

[16] Adam Barth, Anupam Datta, John C. Mitchell, and Helen Nissenbaum. Privacy and contextual integrity: Framework and applications. In *2006 IEEE Symposium on Security and Privacy (S&P'06)*, pages 15–pp. IEEE, 2006. URL `http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1624011`.

[17] Moritz Y. Becker, Alexander Malkis, and Laurent Bussard. S4P: A generic language for specifying privacy preferences and policies. 2010. URL `http://www.msr-waypoint.com/pubs/122108/main.pdf`.

[18] Kathy Bohrer and Bobby Holland. *Customer Profile Exchange (Cpexchange) Specification.* 2000. URL `http://mail.ctiforum.com/standard/standard/www.cpexchange.org/cpexchangev1_0F.pdf`.

[19] Carolyn A. Brodie, Clare-Marie Karat, and John Karat. An empirical study of natural language parsing of privacy policy rules using the SPARCLE policy workbench. In *Proceedings of the Second Symposium on Usable Privacy and Security - SOUPS '06*, page 8. ACM Press, 2006. ISBN 978-1-59593-448-2. doi: 10/b3tswp. URL `http://portal.acm.org/citation.cfm?doid=1143120.1143123`.

[20] Simon Byers, Lorrie Faith Cranor, Dave Kormann, and Patrick McDaniel. Searching for Privacy: Design and Implementation of a P3P-Enabled Search Engine. In David Martin and Andrei Serjantov, editors, *Privacy Enhancing Technologies*, volume 3424, pages 314–328. Springer Berlin Heidelberg, 2005. ISBN 978-3-540-26203-9 978-3-540-31960-3. doi: 10.1007/11423409_20. URL `http://link.springer.com/10.1007/11423409_20`.

[21] F. H. Cate. The Limits of Notice and Choice. 8(2):59–62, 2010-03. ISSN 1540-7993. doi: 10/cgjkcd.

[22] Fred H Cate. The Failure of Fair Information Practice Principles. page 38, 2008.

[23] Alonzo Church. An unsolvable problem of elementary number theory. *American Journal of Mathematics*, 58(2):345–363, 1936. ISSN 00029327, 10806377.

[24] William F. Clocksin and Christopher S. Mellish. *Programming in Prolog (4. ed.).* Springer, 1994. ISBN 978-3-540-58350-9.

[25] CNIL. The CNIL's restricted committee imposes a financial penalty of 50 Million euros against GOOGLE LLC, 2019-01-21. URL `https://www.cnil.fr/en/cnils-restricted-committee-imposes-financial-penalty-50-million-euros-against-google-llc`.

[26] Elisa Costante, Yuanhao Sun, Milan Petković, and Jerry den Hartog. A machine learning solution to assess privacy policy completeness:(short paper). In *Proceedings of the 2012 ACM Workshop on Privacy in the Electronic Society*, pages 91–96. ACM, 2012. URL http://dl.acm.org/citation.cfm?id=2381979.

[27] Lorrie Cranor, Marc Langheinrich, Massimo Marchiori, Martin Presler-Marshall, and Joseph Reagle. The platform for privacy preferences 1.0 (P3P1. 0) specification. 16, 2002. URL https://elearn.inf.tu-dresden.de/hades/teleseminare/wise0405/Act.%208%20Models%20Languages%20Pierangela/Materials/P3P.pdf.

[28] Lorrie Faith Cranor. Necessary but not sufficient: Standardized mechanisms for privacy notice and choice. 10:273, 2012. URL http://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/jtelhtel10&section=22.

[29] CyLab Usable Privacy and Security Laboratory. Privacy Bird, 2019-03-26. URL http://www.privacybird.org/.

[30] Anupam Das, Martin Degeling, Daniel Smullen, and Norman Sadeh. Personalized Privacy Assistants for the Internet of Things. 2018, 2018. doi: 10.1109/MPRV.2018.03367733.

[31] Martin Degeling, Christine Utz, Christopher Lentzsch, Henry Hosseini, Florian Schaub, and Thorsten Holz. We Value Your Privacy ... Now Take Some Cookies: Measuring the GDPR's Impact on Web Privacy. 2019. doi: 10/gfxgxm. URL http://arxiv.org/abs/1808.05096.

[32] Daniel DelPercio. Privacy Policy Online (2011), 2019-03-26. URL http://www.PrivacyPolicyOnline.com.

[33] Henry DeYoung, Deepak Garg, Limin Jia, Dilsun Kirli Kaynar, and Anupam Datta. Experiences in the Logical Specification of the HIPAA and GLBA Privacy Laws. In *Proceedings of the 2010 ACM Workshop on Privacy in the Electronic Society, WPES 2010, Chicago, Illinois, USA, October 4, 2010*, pages 73–82, 2010.

[34] Disconnect. Privacy Icons, 2016-03-04. URL https://web.archive.org/web/20160304013156/https://disconnect.me/icons.

[35] Docracy. An open source privacy policy for mobile apps, 2012-07-24. URL https://web.archive.org/web/20171124185357/https://blog.docracy.com/post/27931026976/an-open-source-privacy-policy-for-mobile-apps.

[36] Serge Egelman, Janice Tsai, Lorrie Faith Cranor, and Alessandro Acquisti. Timing is everything?: The effects of timing and placement of online privacy indicators. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 319–328. ACM, 2009. URL http://dl.acm.org/citation.cfm?id=1518752.

[37] Serge Egelman, Raghudeep Kannavara, and Richard Chow. Is This Thing On?: Crowdsourcing Privacy Indicators for Ubiquitous Sensing Platforms. pages 1669–1678. ACM Press, 2015. ISBN 978-1-4503-3145-6. doi: 10.1145/2702123.2702251. URL http://dl.acm.org/citation.cfm?doid=2702123.2702251.

[38] Pardis Emami-Naeini, Henry Dixon, Yuvraj Agarwal, and Lorrie Faith Cranor. Exploring How Privacy and Security Factor into IoT Device Purchase Behavior. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, pages 1–12. ACM Press, 2019. ISBN 978-1-4503-5970-2. doi: 10/gf5d6v. URL http://dl.acm.org/citation.cfm?doid=3290605.3300764.

[39] European Parliament. General Data Protection Regulation, 2016-04-26.

[40] Federal Trade Commission. FAIR INFORMATION PRACTICES IN THE ELECTRONIC MARKETPLACE. page 208, 2000-06.

[41] Federal Trade Commission. Children's Online Privacy Protection Rule; Final Rule, 2013-01-17. URL `https://www.ftc.gov/system/files/documents/federal_register_notices/2013/01/2012-31341.pdf`.

[42] Joan Feigenbaum, Michael J. Freedman, Tomas Sander, and Adam Shostack. Privacy engineering for digital rights management systems. In *Security and Privacy in Digital Rights Management, ACM CCS-8 Workshop DRM 2001, Philadelphia, PA, USA, November 5, 2001, Revised Papers*, volume 2320 of *Lecture Notes in Computer Science*, pages 76–105. Springer, 2001. doi: 10.1007/3-540-47870-1_6. URL `https://doi.org/10.1007/3-540-47870-1_6`.

[43] ForbrukerrÃ¥det. Deceived by Design, 2018-06-27. URL `https://fil.forbrukerradet.no/wp-content/uploads/2018/06/2018-06-27-deceived-by-design-final.pdf`.

[44] FreePrivacyPolicies.com. Free Privacy Policy Generator & Template with GDPR - Free Privacy Policy, 2019-03-26. URL `https://www.freeprivacypolicy.com/`.

[45] Armin Gerl, Nadia Bennani, Harald Kosch, and Lionel Brunie. LPL, Towards a GDPR-Compliant Privacy Language: Formal Definition and Usage. *Trans. Large-Scale Data- and Knowledge-Centered Systems*, 37:41–80, 2018.

[46] GetTerms. Getterms.io, 2019-03-26. URL `http://getterms.io/`.

[47] Google. Android Permissions overview, 2019-03-26. URL `https://developer.android.com/guide/topics/permissions/overview`.

[48] Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. The Dark (Patterns) Side of UX Design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, pages 1–14. ACM Press, 2018. ISBN 978-1-4503-5620-6. doi: 10/gfxvpz. URL `http://dl.acm.org/citation.cfm?doid=3173574.3174108`.

[49] Margaret D. Hagan. User-Centered Privacy Communication Design. 2016. URL `https://www.usenix.org/system/files/conference/soups2016/wfpn16-paper-hagan.pdf`.

[50] Hamza Harkous, Kassem Fawaz, RÃ©mi Lebret, Florian Schaub, Kang G. Shin, and Karl Aberer. Polisis: Automated Analysis and Presentation of Privacy Policies Using Deep Learning. 2018-02-07. URL `http://arxiv.org/abs/1802.02561`.

[51] Manuel Hilty, Alexander Pretschner, David A. Basin, Christian Schaefer, and Thomas Walter. A Policy Language for Distributed Usage Control. In *Proceedings of the 12th European Symposium On Research in Computer Security, ESORICS'07*, volume 4734 of *Lecture Notes in Computer Science*, pages 531–546. Springer, 2007. ISBN 978-3-540-74834-2.

[52] Gerard J. Holzmann. *The SPIN Model Checker - Primer and Reference Manual.* Addison-Wesley, 2004. ISBN 978-0-321-22862-8.

[53] Michael Huth and Mark Dermot Ryan. *Logic in computer science - modelling and reasoning about systems (2. ed.).* Cambridge University Press, 2004.

[54] Iubenda. Features — Compliance Solutions, 2019-03-26. URL https://www.iubenda.com/en/features.

[55] Iubenda. Terms of service, 2019-03-26. URL https://www.iubenda.comhttps://www.iubenda.com/en/user/tos.

[56] Johnson Iyilade and Julita Vassileva. P2U: A Privacy Policy Specification Language for Secondary Data Sharing and Usage. pages 18–22. IEEE, 2014-05. ISBN 978-1-4799-5103-1. doi: 10.1109/SPW.2014.12. URL http://ieeexplore.ieee.org/document/6957279/.

[57] Carlos Jensen and Colin Potts. Privacy policies as decision-making tools: An evaluation of online privacy notices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 471–478. ACM, 2004. URL http://dl.acm.org/citation.cfm?id=985752.

[58] Lalana Kagal. Rei. 2002. URL http://ebiquity.umbc.edu/get/a/publication/57.pdf.

[59] Saffija Kasem-Madani and Michael Meier. Security and privacy policy languages: A survey, categorization and gap identification. 2015. URL https://arxiv.org/abs/1512.00201.

[60] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W. Reeder. A nutrition label for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, page 4. ACM, 2009. URL http://dl.acm.org/citation.cfm?id=1572538.

[61] Patrick Gage Kelley, Lucian Cesca, Joanna Bresee, and Lorrie Faith Cranor. Standardizing privacy notices: An online study of the nutrition label approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1573–1582. ACM, 2010. URL http://dl.acm.org/citation.cfm?id=1753561.

[62] Jan Kolter and GÃ¼nther Pernul. Generating User-Understandable Privacy Preferences. pages 299–306. IEEE, 2009. ISBN 978-1-4244-3572-2. doi: 10.1109/ARES.2009.89. URL http://ieeexplore.ieee.org/document/5066486/.

[63] Ulrich KÃ¶nig and Jan Schallaboeck. Privacy preferences for E-Mail messages. 2012. URL https://tools.ietf.org/html/koenig-privicons-03.txt.

[64] Marc Langheinrich, Lorrie Cranor, and Massimo Marchiori. Appel: A p3p preference exchange language. 2002. URL https://www.w3.org/TR/P3P-preferences/.

[65] Daniel Le MÃ©tayer. A formal privacy management framework. In *International Workshop on Formal Aspects in Security and Trust*, pages 162–176. Springer, 2008. URL http://link.springer.com/chapter/10.1007/978-3-642-01465-9_11.

[66] Andreas Matheus and J Herrmann. Geospatial Extensible Access Control Markup Language (GeoXACML). *Open Geospatial Consortium Inc. OGC*, 2008.

[67] Michael J. May, Carl A. Gunter, and Insup Lee. Privacy APIs: Access Control Techniques to Analyze and Verify Legal Privacy Policies. In *Proceedings of the 19th IEEE Computer Security Foundations Workshop, CSFW'06*, pages 85–97. IEEE Computer Society, 2006. ISBN 0-7695-2615-2.

[68] Aleecia M. McDonald and Lorrie Faith Cranor. The cost of reading privacy policies. 4:543, 2008. URL http://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/isjlpsoc4&section=27.

[69] Matthias Mehldau. Icons of privacy (original), 2007. URL `https://netzpolitik.org/wp-upload/data-privacy-icons-v01.pdf`.

[70] Daniela Yidan Miao. PrivacyInformer: An Automated Privacy Description Generator for the MIT App Inventor, 2014. URL `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1029.2434&rep=rep1&type=pdf`.

[71] V. Morel, M. Cunche, and D. Le MÃ©tayer. A generic information and consent framework for the iot. In *Proceedings of the 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications(TrustCom)*, pages 366–373, 2019. doi: 10.1109/TrustCom/BigDataSE.2019.00056.

[72] National Telecommunications and Information Administration. Short Form Notice Code of Conduct to Promote Transparency in Mobile Apps Practices, 2013. URL `https://www.ntia.doc.gov/files/ntia/publications/july_25_code_draft.pdf`.

[73] Helen Nissenbaum. Privacy as contextual integrity. 79:119, 2004. URL `http://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/washlr79&section=16`.

[74] Organisation for Economic Co-operation and Development. Skills matter: Further results from the survey of adult skills, 2016. OCLC: ocn953634518.

[75] Raúl Pardo and Daniel Le Métayer. Analysis of privacy policies to enhance informed consent. In *Proceedings of the 33rd Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy*, volume 11559 of *Lecture Notes in Computer Science*, pages 177–198, 2019. doi: 10.1007/978-3-030-22479-0_10.

[76] Jaehong Park and Ravi S. Sandhu. The UCON$_{\text{ABC}}$ Usage Control Model. *ACM Trans. Inf. Syst. Secur.*, 7(1):128–174, 2004.

[77] Niklas Paul, Welderufael B. Tesfay, Dennis-Kenji Kipker, Mattea Stelter, and Sebastian Pape. Assessing Privacy Policies of Internet of Things Services. In Lech Jan Janczewski and Mirosław Kutyłowski, editors, *ICT Systems Security and Privacy Protection*, volume 529, pages 156–169. Springer International Publishing, 2018. ISBN 978-3-319-99827-5 978-3-319-99828-2. doi: 10.1007/978-3-319-99828-2_12. URL `http://link.springer.com/10.1007/978-3-319-99828-2_12`.

[78] Polisis. Chrome Polisis, 2019-03-26. URL `https://chrome.google.com/webstore/detail/polisis/bkddolgokpghlbhhkflbbhhjghjdojck`.

[79] Polisis. Firefox Polisis, 2019-03-26. URL `https://addons.mozilla.org/en-US/firefox/addon/polisis/`.

[80] Polisis. Polisis, 2019-03-26. URL `https://www.pribot.org/polisis`.

[81] Alexander Pretschner, Manuel Hilty, and David A. Basin. Distributed Usage Control. *Commun. ACM*, 49(9):39–44, 2006.

[82] Privacy Policy Generator. Privacy Policy Generator, 2019-03-26. URL `https://privacypolicygenerator.info/`.

[83] Privacy Tech. Privacy icons, 2018. URL `https://www.privacytech.fr/privacy-icons/`.

[84] PrivacyPolicies.com. Privacy Policy Generator: Free, GDPR, CalOPPA - PrivacyPolicies.com, 2019-03-26. URL https://www.privacypolicies.com/.

[85] Aza Raskin. Making Privacy Policies not Suck, 2010. URL http://www.azarask.in/blog/post/making-privacy-policies-not-suck/.

[86] Aza Raskin. Privacy Icons - MozillaWiki, 2011. URL https://wiki.mozilla.org/Privacy_Icons.

[87] Joel R. Reidenberg, Travis Breaux, Lorrie Faith Cranor, Brian French, Amanda Grannis, James T. Graves, Fei Liu, Aleecia McDonald, Thomas B. Norton, and Rohan Ramanath. Disagreeable privacy policies: Mismatches between meaning and users' understanding. 30:39, 2015. URL http://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/berktech30&section=6.

[88] Joel R. Reidenberg, Jaspreet Bhatia, Travis Breaux, and Thomas B. Norton. Automated comparisons of ambiguity in privacy policies and the impact of regulation. 2016. URL http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2715164.

[89] Joel R. Reidenberg, Jaspreet Bhatia, Travis D. Breaux, and Thomas B. Norton. Ambiguity in Privacy Policies and the Impact of Regulation. 45(S2):S163–S190, 2016-06. ISSN 0047-2530, 1537-5366. doi: 10/gdcdzm. URL https://www.journals.uchicago.edu/doi/10.1086/688669.

[90] Arianna Rossi and Monica Palmirani. DaPIS: An Ontology-Based Data Protection Icon Set. pages 181–195, 2019. ISSN 0922-6389. doi: 10/gf7fbn. URL http://www.medra.org/servlet/aliasResolver?alias=iospressISBN&isbn=978-1-61499-984-3&spage=181&doi=10.3233/FAIA190020.

[91] Arianna Rossi, Rossana Ducato, Helena Haapio, and Stefania Passera. When Design Met Law: Design Patterns for Information Transparency. page 43, 2019.

[92] Mary Rundle. International Data Protection and Digital Identity Management Tools, presentation at IGF 2006. 2006.

[93] Norman Sadeh, Alessandro Acquisti, Travis D. Breaux, Lorrie Faith Cranor, Aleecia M. McDonald, Joel R. Reidenberg, Noah A. Smith, Fei Liu, N. Cameron Russell, Florian Schaub, et al. The usable privacy policy project, 2013. URL http://ra.adm.cs.cmu.edu/anon/usr0/ftp/home/anon/isr2013/CMU-ISR-13-119.pdf.

[94] Ravi S. Sandhu, Edward J. Coyne, Hal L. Feinstein, and Charles E. Youman. Role-Based Access Control Models. *IEEE Computer*, 29(2):38–47, 1996. ISSN 0018-9162.

[95] Florian Schaub, Rebecca Balebako, Adam L. Durity, and Lorrie Faith Cranor. A design space for effective privacy notices. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*, pages 1–17, 2015. URL https://www.usenix.org/conference/soups2015/proceedings/presentation/schaub.

[96] Daniel J. Solove. A taxonomy of privacy. 154:477, 2005. URL http://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/pnlr154&section=20.

[97] State of California. Assembly Bill No. 375 California Consumer Privacy Act, 2018-06-28.

[98] Cass R. Sunstein. Choosing Not to Choose. 2014. ISSN 1556-5068. doi: 10/gftmr3. URL http://www.ssrn.com/abstract=2377364.

[99] Latanya Sweeney, MercÃ¨ Crosas, and Michael Bar-Sinai. Sharing sensitive data with confidence: The datatags system. 2015. URL http://techscience.org/a/2015101601/.

[100] Alasdair Taylor. Privacy policy, 2012-01-23T08:05:39+00:00. URL https://seqlegal.com/free-legal-documents/privacy-policy.

[101] Terms of Service; Didn't Read. Terms of Service Classification, 2019-03-26. URL https://tosdr.org/classification.html.

[102] Termsfeed. Generic Privacy Policy template, 2019-03-26. URL https://www.termsfeed.com/assets/pdf/privacy-policy-template.pdf.

[103] Janice Y. Tsai, Serge Egelman, Lorrie Cranor, and Alessandro Acquisti. The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study. 22(2):254–268, 2011-06. ISSN 1047-7047, 1526-5536. doi: 10/cxhgzz. URL http://pubsonline.informs.org/doi/abs/10.1287/isre.1090.0260.

[104] United States Congress. Health Insurance Portability and Accountability Act, 1996. URL https://www.hhs.gov/sites/default/files/privacysummary.pdf.

[105] United States Congress. Gramm—Leach—Bliley Act, 1999.

[106] Christine Utz, Martin Degeling, Sascha Fahl, Florian Schaub, and Thorsten Holz. (Un)informed Consent: Studying GDPR Consent Notices in the Field. page 18, 2019.

[107] Jasper van de Ven and Frank Dylla. Qualitative Privacy Description Language. In *Annual Privacy Forum*, pages 171–189. Springer, 2016. URL http://link.springer.com/chapter/10.1007/978-3-319-44760-5_11.

[108] Bibi Van den Berg and Simone Van der Hof. What happens to my data? A novel approach to informing users of data processing practices. 2012. doi: 10.5210/fm.v17i7.4010. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2100417.

[109] Shomir Wilson, Florian Schaub, Aswarth Abhilash Dara, Frederick Liu, Sushain Cherivirala, Pedro Giovanni Leon, Mads Schaarup Andersen, Sebastian Zimmeck, Kanthashree Mysore Sathyendra, N. Cameron Russell, et al. The creation and analysis of a website privacy policy corpus. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2016. URL http://www.aclweb.org/anthology/P/P16/P16-1126.pdf.

[110] WP29. Opinion 8/2014 on the Recent Developments on the Internet of Things. 2014.

[111] WP29. Guidelines on transparency under Regulation 2016/679, 2017-12-15.

[112] Jean Yang, Kuat Yessenov, and Armando Solar-Lezama. A language for automatically enforcing privacy policies. page 85. ACM Press, 2012. ISBN 978-1-4503-1083-3. doi: 10.1145/2103656.2103669. URL http://dl.acm.org/citation.cfm?doid=2103656.2103669.

[113] Le Yu, Tao Zhang, Xiapu Luo, and Lei Xue. AutoPPG: Towards Automatic Generation of Privacy Policy for Android Applications. pages 39–50. ACM Press, 2015. ISBN 978-1-4503-3819-6. doi: 10.1145/2808117.2808125. URL http://dl.acm.org/citation.cfm?doid=2808117.2808125.

[114] Razieh Nokhbeh Zaeem, Rachel L. German, and K. Suzanne Barber. PrivacyCheck: Automatic Summarization of Privacy Policies Using Data Mining. 18(4):1–18, 2016. ISSN 1533-5399, 1557-6051. doi: 10.1145/3127519. URL `https://dl.acm.org/doi/10.1145/3127519`.

[115] Zero-knowledge. Privacy Rights Markup Language Specification. 2001.

[116] Sebastian Zimmeck and Steven M. Bellovin. Privee: An Architecture for Automatically Analyzing Web Privacy Policies. USENIX Association, 2014. ISBN 978-1-931971-15-7. URL `https://www.usenix.org/system/files/conference/usenixsecurity14/sec14-paper-zimmeck.pdf`. OCLC: 254320948.

A Appendix

Table 4 provides a condensed overview of the content, benefits, limitations and tools that we have studied for each facet. The content columns are a coarsed grained presentation of the results in Figure 5. The three last columns list the benefits, limitations and tools of each facet.

	1st Party	3rd Party	Legal basis	DS rights	Data retention	Data security	Policy change	Others			
Natural Language	✓	✓	✓	✓	✓	✓	✓	✓	- Legal value	- Ambiguity - Understandability - Enforceability	- Templates - Generators - Retrievers - Analysis tools
Graphical	✓	✓	✓	✓	✓				- Understandability	- Ambiguity - Incompleteness	- DS Notification - Visual comparison
Machine Readable	✓	✓		✓	✓				- Enforcement - Auditability - Correctness - Automation	- Understandability - Lack of adoption	- Enforcement engines - Formal Semantics - Policy comparison - Analysis tools
	Content								Benefits	Limitations	Tools

Table 4: Summary of each facet's content, benefits, limitations and tools. The ✓symbol indicates that at least one work on the corresponding facet captures the taxonomy item.

RESEARCH CENTRE
GRENOBLE – RHÔNE-ALPES

Inovallée
655 avenue de l’Europe Montbonnot
38334 Saint Ismier Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399