



Three Dimensions of Privacy Policies

Victor Morel, Raúl Pardo

► To cite this version:

Victor Morel, Raúl Pardo. Three Dimensions of Privacy Policies. [Research Report] RR-9287, Inria - Research Centre Grenoble – Rhône-Alpes; CITI - CITI Centre of Innovation in Telecommunications and Integration of services. 2019. hal-02267641v1

HAL Id: hal-02267641

<https://inria.hal.science/hal-02267641v1>

Submitted on 19 Aug 2019 (v1), last revised 11 Sep 2020 (v4)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Three Dimensions of Privacy Policies

Victor Morel, Raúl Pardo

**RESEARCH
REPORT**

N° 9287

August 2019

Project-Teams Privatics



Three Dimensions of Privacy Policies

Victor Morel, Raúl Pardo

Project-Teams Privatics

Research Report n° 9287 — August 2019 — 42 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 others means of expression. In this report, we systematically study the different means of expression of privacy policies. In doing so, we have identified three main categories, which we call *dimensions*, i.e., natural language, graphical and machine-readable privacy policies. Each of these dimensions focus 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 dimension, and explain why solutions based on a single dimension do not cover the needs of other communities. Finally, we propose a new approach to expressing privacy policies which brings together the benefits of each dimension 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

Les trois dimensions des politiques de la 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 data collection and processing is through *privacy policies*. Typically, privacy policies are 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, see Appendix A) commonly use privacy policies to inform individuals (in what follows *data subjects*, or DS, see Appendix A) about how personal data is handled. DS are required to read these policies — even though it rarely occurs [52] — and decide whether they accept the conditions. Alternatively, giving DS the possibility of describing their own privacy policies, that we denote *DS privacy policies*, has recently gained in popularity. This approach gives DS the time to reflect on their choices, and perhaps the possibility to consult experts and pairs. Nonetheless, privacy policies in their current format are hard to understand, for DS [64] as for experts [84]. In the sequel, we use *DS policies* to denote the privacy policies of individuals, and *DC policies* to refer to the privacy policies of organizations collecting personal data (see Appendix B for a precise definition).

Requirements and recommendations to express privacy policies come from different sources such as privacy regulations, authorities and organizations. For instance, the General Data Protection Regulation (GDPR) [33] — the legal framework governing personal data collection and processing in Europe since May 2018 — requires more transparency for data processing from DC, and guidelines have been issued by the WP29¹ [104] to present their expectations. 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 [67] for mobile apps privacy policies, and the WP29 [105] for IoT devices privacy policies. Furthermore, authorities such as Data Protection Authorities (DPA in the sequel) should be able to audit data processing systems, to ensure their compliance with the law and with the declared privacy policies — *e.g.*, see the decision of the “Commission Nationale de l’Informatique et des Libertés” (CNIL), the French DPA, against Google LLC [21]. All these requirements and recommendations can be summarised in three requirements:

- Privacy policies must be legally valid.
- Privacy policies must be understandable by all parties (including lay-users).
- Privacy policies must be effectively enforced through auditable mechanisms.

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 summarised versions of privacy policies (that do not have legal value) or use visual aids to help users understand the risks of having their data collected. Another way forward in this direction is the use of techniques for extracting relevant information from existing privacy policies through Natural Language Processing (NLP) [102]. However, it is not powerful enough to be applicable to existing privacy policies yet. Ensuring that data is processed according to the requirements in privacy policies is not an easy task either. Some works — coming mostly from academia — propose an alternative format for privacy policies that can be read by computers. These works aim at bridging the gap between the textual legal requirements and their enforcement in the

¹WP29 stands for Working Party 29, a European advisory board, now European Data Protection Board (EDPB)

underlying system. Furthermore, some of these proposals are equipped with auditing tools which facilitate, for DPA or DS, verifying that no violations of a privacy policy have occurred. Unfortunately, these solutions are not widely used.

In this work, we systematically analyse 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. In doing so, we have identified that there exist three main ways to express privacy policies: natural language, graphical and machine-readable; which we call the *dimensions* of privacy policies. Each of these dimensions have arisen (independently) 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 restricted to different communities that have been working in the same issue separately and with different objectives. Unsurprisingly, each dimension mainly provides benefits to the specific community it was defined in. Therefore, we take the insights of our study to define how the different dimensions of privacy policies can complement each other. This synergy can include the benefits of each dimension and minimize their limitations. Hence, we propose a new type of privacy policies which we denote *multifaceted privacy policies* combining aspects from all dimensions. More concretely, in bringing the above ideas to the forefront, our contributions are:

1. An in-depth study of the existing dimensions of privacy policies:
 - (a) Privacy policies expressed in natural language (Section 2). We denote these policies *natural language privacy policies*. Natural language privacy policies are necessary for legal compliance, and are required for DC to conduct lawful collection and processing of personal data.
 - (b) The expression of privacy policies graphically, *i.e.*, using visual aids such as icons and pictograms (Section 3). We denote these policies *graphical privacy policies*. Graphical privacy policies can be suitable for conveying intelligible information.
 - (c) Privacy policies that can be automatically processed by machines (Section 4). We denote this type of privacy policies *machine-readable privacy policies*. Machine-readability can be useful to provide tools to assist users and auditors in their tasks.

For each dimension, we provide an overview of: i) its content; ii) the available tools; iii) its benefits; and iv) its limitations.

2. A categorization of existing works in each dimension according to a privacy taxonomy, and the specificities of each dimension.
3. A discussion on the intrinsic limitations of mono-dimensional solutions (Section 5). The discussion is based on the insights extracted from the systematic study of existing means of expression of privacy policies.
4. A multifaceted approach to express privacy policies, which overcomes the limitations of each dimension by combining their benefits (Section 6).

The glossary provides terms necessary for the comprehension of the paper in Appendix A.

2 Natural language privacy policies

Most legislations now require notices expressed in natural language to inform DS about the collection and processing of their personal data:² the use of natural language is necessary to ensure that the policy has a legal value. The ways these documents can be authored — *i.e.*, drafted automatically or written manually — and the manners to assist their authoring can vary greatly. There are many ways to express privacy policies in natural language. In what follows, we present the content expressed by natural language privacy policies in Section 2.1, the tools used to assist their authoring and to analyze existing natural language privacy policies in Section 2.2, the benefits in Section 2.3, and the limitations in Section 2.4.

2.1 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, or mobile applications etc. Because these privacy policies are expressed in natural language, they are not restricted in terms of content. This content can be categorized according to different taxonomies. In the following, we start by providing a high-level overview of existing taxonomies, and succinctly present our own.

2.1.1 Overview of existing taxonomies

In [91] Solove's introduces one of the first, and perhaps the most known, taxonomy of privacy. This taxonomy focuses on activities that invade privacy, and distinguishes four categories. The first category is *information collection*, and encompasses surveillance and interrogation. The second category is *information processing* and comprises aggregation, identification, insecurity, secondary use, and exclusion. The third category is *information dissemination*, it covers breach of confidentiality, disclosure, exposure, increased accessibility, blackmail, appropriation and distortion. The last category is *invasions*, and includes intrusion and decisional interference. This taxonomy focuses on privacy harms, which makes it unsuitable for classifying the content of natural language privacy policies: we look for a classification of personal data management.

Paul *et al.* [73] introduced an evaluation framework to help DS assess how “privacy friendly” the privacy policies of IoT devices are. Concretely, they provide a scoring system based on the content of the privacy policy. The framework includes categories such as *Right to object*, *Right to access*, *Right to erase*, *Period of storage*, etc. As before, this work cannot be used to classify the content of natural language privacy policies, as it focuses on helping DS to take informed decisions based on the assessment of their privacy policies.

Wilson *et al.* [103] proposed a taxonomy tailored for privacy policies. It is 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 to the privacy policy”, *Other* “Additional sub-labels for introductory or general

²For instance, the GDPR mentions a list of information to provide where personal data are collected in Art. 13 and 14.

³We denote *item* a piece of information provided in a natural language privacy policy. See Appendix B.

text, contact information, and practices not covered by the other categories”, and *Choice Control* “Choices and control options available to users”. This taxonomy is appropriate for our purposes for two main reasons: 1) it was devised according to existing natural language privacy policies and therefore reflects their content; and 2) it encompasses most requirements of the current legislations, such as the GDPR but also the Fair Information Practice Principles (FIPPs) [36] in some cases.⁴ This taxonomy was also used in Polisis [45] as discussed in Section 2.2.

2.1.2 Presentation of our taxonomy

We use a slight variation of Wilson *et al.*’s taxonomy — this variation does not change the content of the taxonomy but accommodates it to the purposes of our study. Concretely, the differences are: i) We use *DS rights* to denote both *Access*, *Edit*, *Delete* and *Choice Control* as they relate to DS rights in the sense of the GDPR — see Chapter III of the GDPR. and subsume *Specific Audiences* and *Do Not Track* under *Other* as they are both small items in terms of size; and ii) 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.

Table 1 summarizes the chosen taxonomy, and shows what legal requirements appear explicitly in the GDPR and the FIPPs. We only study the requirements of the GDPR and FIPPs as they are the two main laws that determine the content required when informing DS of data collection and processing.⁵ 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) that represent widely-accepted principles concerning fair information practices. We do not consider other widely-known regulations because they have a restricted scope (such as health data for HIPAA [99], or children for COPPA [35]). In what follows we examine the aforementioned items in detail, illustrated with examples from existing natural language privacy policies such as Facebook [34], Twitter [97], Dropbox [31], Netflix [68], and Google [43].

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*. Common types of data collected can be the name of the DS, an email address, geolocation, messages, etc; or the social graph, more specifically to social networks. As an example, Facebook collects “Networks and connections. [...] information about the people, Pages, accounts, hashtags and groups you are connected to and how you interact with them across our Products[...]. We also collect contact information [...] (such as an address book [...]).” Cookies — small pieces of data sent from a website and stored on the DS’s computer by the DS’s web browser for various purposes — 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 in a natural language privacy policy. 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. For instance, Twitter’s privacy policy states: “Location Information: We require information about your signup and current location, which we get from signals such as your IP address or device settings, to securely and reliably set up and maintain your account and to provide our services to you.”

⁴We discuss below the relevance of legislations with respect to the content of policies.

⁵See Appendix B for the details.

	Description	GDPR	FIPPs
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 regarding their personal data, <i>e.g.</i> , right to access, rectify port or erasure.	●	○
Data Retention	Duration of data storage	●	○
Data Security	Modalities of protection of data, <i>e.g.</i> , encrypted communication and storage.	●	●
Policy Change	Modalities of notification for policy changes.	●	○
Other	Other items such as identity of DC, information related to DNT, 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 purpose of processing often comes along the type of data collected. It is possible to find among the purposes, *marketing* and *advertising*, which are prevalent in natural language privacy policies. Analytics is often mentioned as a purpose to improve the functioning of services, to provide a better overview of what is actively used or not in a service, or to automatically retrieve malfunctions. Data can also be collected for security reasons: to remove illegal or harmful content, or to prevent payment fraud. Certain services collect data for research, and this broad purpose can be exempt of some constraints, notably for the definition of a more concrete research purpose.⁶ DC can conduct data collection to operate a service: for instance Facebook mentions “Provide, personalize and improve our Products” as a purpose of processing, and this is a reason often put forward for data collection and processing.

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 mean.

Informing about the type of data, the purpose of processing, the recipients and the means of collection is required by both the GDPR and the FIPPs.

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, *i.e.* type of data and purpose. For instance, Dropbox declares in its privacy policy: “Dropbox uses certain trusted third parties (such as providers of customer support and IT services) to help us provide, improve, protect and promote our Services. These third parties will only access your information to perform tasks on our behalf in compliance with this Privacy Policy, and we’ll remain responsible for their handling of your information per our instructions. For a list of trusted third parties that we use to process your personal information, please see our FAQ.”

Informing about third party collection is required by both the GDPR and the FIPPs. Note that *recipient* in the sense of the GDPR encompasses first and third party collection.

⁶See Recital 159 of the GDPR.

Legal basis Legal basis (or legal ground) is regularly found as a complement of the purpose of processing.⁷ 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,⁸ 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 of the obtention of consent [38]. Other legal basis can be found in natural 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,⁹ and major stakeholders generally consider cumulatively either all of them — such as Facebook which combines all possible legal bases — or a large subset — *e.g.*, Netflix's policy considers all of them except public interest.

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”.

DS Rights DS can often 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.¹⁰ As an example, Google's privacy policies mentions: “You can export a copy of your information or delete it from your Google Account at any time”. DS rights can be seen more restrictively as possibilities opt-in or opt-out.¹¹ Thus natural language privacy policies present how to subscribe or unsubscribe to specific services.

Informing about DS rights is required by the GDPR.

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 a variable one — *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.

Data Security DC regularly explain in their policies how data is stored, if its communication is secured or its storage encrypted. As an example, Netflix's privacy policy claims: “Security: We use reasonable administrative, logical, physical and managerial measures to safeguard your personal information against loss, theft and unauthorized access, use and modification. These measures are designed to provide a level of security appropriate to the risks of processing your personal information.”

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”.¹²

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.

⁷Article 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)

⁸“Freely given, specific, informed and unambiguous indication of the data subject's agreement” in Recital 32.

⁹See Art. 6.

¹⁰These rights are explicitly mentioned in the GDPR.

¹¹Opt-out is now illegal in Europe

¹²See Recital 39 of the GDPR.

This item is not required by the GDPR nor by the FIPPs.

Other We subsume the identity and contact of DC, requirements towards specific audiences such children, and Do Not Track (DNT) 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,¹³ children have specific considerations. As a result, it is possible to find a dedicated section in many natural language privacy policies, even if it is only to mention that personal data of children under thirteen is not collected without parental consent.

Identity of the DC is required by the GDPR and the FIPPs, but not the rest of this item (although both legislations specifically address data collected from children).

2.2 Tools

A variety of solutions exist to assist in the authoring of natural language privacy policies, ranging from the least to the most automated ones. We denote these tools *authoring tools*. We distinguish, in this section, *templates*, *generators*, and *retrievers*. While natural language does not limit the expressiveness of privacy policies, authoring tools are often tailored to websites and mobile app owners, and are constrained in terms of content. In addition to the authoring tools presented above, we denote *analysis tool* the piece of software able to parse or to analyze natural language privacy policies, in order to produce a machine-readable or a graphical version of a policy. This section does not aim at providing an exhaustive list of the available tools.

Templates Tools such as Docracy [30], Termsfeed [96], SEQ Legal [94], and 3DCart [2] provide a *fill-in-the-gap* form, where the author writes appropriate terms in the fields. The redundancy is not taken into account, and no verification can be made regarding the validity of the terms written. It is not possible, for instance, to check if the email of a service owner is valid. They usually do not allow the possibility to describe the type of data collected, nor its use: their expressive power often cannot provide all the items required for legal compliance — at least in EU. Their level of automation is low, and the result is easily prone to errors.

Generators These tools differ from *templates* by inputting information in a form using software components. The author inputs information only once. Most *generators* do not allow incorrect data: email addresses without @ are highlighted, and the author do not have the possibility to go further in the process and generate the policy. *Generators* also give the option of expressing the same policy according to different legal contexts. Legislation varies between countries. A *generator* can overcome this issue, and propose text corresponding to the applicable legal framework. For instance, `privacypolicies.com` offers clauses specific to the GDPR or CalOPPA (the California Online Privacy Protection Act) for an additional cost, if required. We can distinguish *light* from *detailed generators*. Privacy Policy Generator [78], Privacy Policy Online [27], or GetTerms [41] are examples of the former, and `PrivacyPolicies.com` [80] and `FreePrivacyPolicy.com` [39] of the latter. On the one hand, *light generators* usually have a restricted set of parameters (name of the company, URL if it is a website, use of cookies or not, advertisers). On the other hand, *detailed generators* usually offer a choice between a policy tailored to websites or mobile applications, with a more exhaustive list of items [57]. *Generators* provide more granularity and tailoring for cookies, third-parties and various laws such as CalOPPA or COPPA (the Children’s Online Privacy Protection Act), but often lack granularity when it comes to the type of data collected.

¹³In particular Recitals 38 and 58 of the GDPR, and the Children’s Online Privacy Protection Act (COPPA).

Retrievers Finally, *retrievers*, such as those offered by Miao [66], Apolinarski *et al.* [8] and Yu *et al.* [107], automatically extract relevant information from code of mobile application, using static code analysis or user behavior analysis ([8] analyze sharing behavior when using online collaboration tools for instance). These prototypes work on Android applications, in which personal data management are structured around the concept of permissions [42]. 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 thus could be difficult to implement in other ecosystems. Furthermore, they cannot automatically retrieve certain information, such as the purpose of collection or the retention time. *Retrievers* reach the highest level of automation among authoring tools.

Analysis tools An early work has been conducted by Brodie *et al.* [18]. They developed a *policy workbench*: SPARCLE, to parse natural language privacy policies in order to produce an XACML format of the policy. It has been devised to provide verification through compliance auditing of the enforcement logs. Costante *et al.* [22] built an Information Extraction system: their tool uses classical NLP techniques to extract the type of data collected by websites. More recently, Polisis [45] was proposed to annotate natural language privacy policies. Polisis is a machine learning framework which summarizes any policy provided in simple text. The way they present information is more exhaustive than the two previous examples: their taxonomy has been presented at the beginning of Section 2.1. They achieve an accuracy of 88.4% when assigning icons to a privacy policy.

Although natural language is versatile, authoring tools often propose a restricted set of items: they do not make use of the full potential of natural language. However, this restriction is not a limitation as long as the items required by law can be expressed. Analysis tools, for their part, cannot convey information in natural language privacy policies with certainty.

2.3 Benefits

The main benefit of natural language privacy policies is their legal value. Most legislations require DC to provide a lawful statement¹⁴ detailing the processing of personal data, and natural language privacy policies often aim to fulfill this obligation.

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. 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, such as the GDPR. However, a document holding legal value is not necessarily compliant with the law. For instance, lawyers or DPA 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. In other words, legal compliance is twofold: with respect to information requirements, and with respect to the actual processing.

¹⁴For instance, DC have to inform of the following in order to be compliant with the GDPR: their identity and contact, the type of data collected, its purpose and its legal basis for processing, the recipient of data, the third-parties involved, the retention time, and the rights of the DS (see Appendix B).

Value produced by authoring tools The content produced by most authoring tools do not have legal value. Most authoring 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” [49], but advocates for a professional legal consultancy “Nothing can substitute a professional legal consultancy in the drafting of your privacy policy” [50], 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.

2.4 Limitations

Ambiguity Natural language privacy policies can be ambiguous [83], as they may be interpreted in different ways. Reidenberg *et al.* [83, 85] 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 languages can be interpreted in different ways. Ambiguity has a direct impact on the understanding, the enforcement, and the auditability of privacy policies.

Understanding McDonald and Cranor [64] 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 of precise semantics, making it difficult to decide how data must be processed by the underlying system. Likewise, Natural language 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 by *templates*, *generators*, and *retrievers*. These tools are often tailored to specific solutions, such as website or mobile applications, therefore restricting their scope. Analysis tools were also devised, but 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.

3 Graphical privacy policies


In the previous section, we analysed 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 ambiguous. Privacy policies can also be expressed with graphical representations, that we denote *graphical privacy policies*. Graphical privacy policies cover icons sets as well as solutions providing additional information, such as warnings or judgments, sometimes combined with

simple text. 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 [104]. We review in this section these means to express privacy policies graphically. In particular, we categorise each work based on: i) the elements in the taxonomy presented in Section 2 that it captures; ii) its features, whether it is made of icons, complementary text, or of something else; and iii) the intended audience of the language, *e.g.*, DS or DC. Table 2 in Section 3.4 summarizes our study.

3.1 Content

Based on their content, graphical privacy policies can be divided in two main types: *icons* and *rating solutions*. Graphical privacy policies based on icons intend to express the content of privacy policies, for DC as for DS policies. These icons try to cover all the items of the taxonomy we introduced in Section 2. Other graphical privacy policies provide rating information concerning certain aspects of privacy policies such as 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. However, they highlight important information for DS. In the following, we describe the content of graphical privacy policies according to the elements of the taxonomy in Section 2, their type (icons or rating) and whether they are intended to be used by DS or DC.

3.1.1 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. In practice, graphical privacy policies have a restricted set of icons, and express specific items (seen in Section 2.1). Sets of icons initially emerged from the academia, but they were quickly adopted by privacy advocates. It is also possible to find a solution coming from the business sector.

Academia Rundle [86] introduced the first set of icons in 2006. It includes icons for selling, and second-use of data. However, the type of data cannot be specified.

The PrimeLife project ([1, Chapter 15]) also proposed a set of privacy icons, that they tested in order to determine if they were understood. The icons presented hereafter (an excerpt is presented in Figure 2) can express retention time and sharing to third-parties, but the exhaustive set of icons is not available.

Kelley *et al.* [55] represented policies in a table such as nutrition labels observed on food packaging (see Figure 3). They developed a privacy nutrition label based on P3P,¹⁵ with the goal of providing efficient and well-organized privacy information. They present the icons in a table such as nutrition labels observed on food packaging. Their solution focused on the type of data and the purpose of collection, with third-parties sharing. However, the information presented is quite precise and fine-grained.

König and Schallaboeck [58] created Privicons, a minimalist set of icons to inform mail correspondents of how the data should be handled (see Figure 4). The icons should be sent along

¹⁵P3P is an obsolete set of specifications for websites to declare their DC policies. See Section 4.








	You agree not to use this data for marketing purposes.
	You agree not to trade or sell this data.
	You agree to submit to a third-party audit program on data use; if government has requested access to my data, you agree to involve my governmental ombudsman.
	You agree to make available to me the data that you have on me without my having to pay for it/at a minimal charge.
	You allow me to address inaccuracies in the data and request its removal.
	You agree to take reasonable steps to keep my data secure.
	You agree to arrange with X organization to help resolve any disputes we have over your treatment of this data. [The seal / name of the entity follows.]

Figure 1: Rundle set of privacy icons



Figure 2: Excerpt of Primelife icons

an email thus indicating their privacy policy: whether it can be shared, to whom, whether it can be printed etc. The set does not consider many items in the sense defined in Section 2.

Egelman *et al.* [32] developed a set of icons for the IoT, with the help of Intel first, then refined with crowdsourcing. The final set of icons (see Figure 5) focuses only on the type of data collected — voice, gesture, image — and its purposes — detection of gender, emotion, language.

A different example of a graphical policy is Polisis by Harkous *et al.* [45]. Polisis can represent the natural language privacy policies as a combination of icons (see Figure 6b), highlights of the corresponding paragraphs in the natural language privacy policies, and a flow diagram (see Figure 6a). Icons are not the core feature of Polisis, but they help DS navigate through the policy. Because Polisis expresses information retrieved from natural language privacy policies, the content expressed is diverse. This possibility is largely due to the support of natural language, which, even if simple, compensates for the lack of granularities of its icons.

Privacy advocacy Many privacy advocates contributed to this area and provided numerous sets of icons, such as Mehlau [65] who developed a set of 30 privacy icons, describing the type



Figure 3: Privacy Nutrition Label

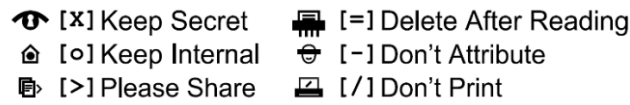


Figure 4: Privicons

of data, third-parties, the purpose of processing and the retention time (see Figure 7). On the one hand many items are missing or incomplete: his proposal only has three different categories of purposes possible, whereas real use cases often have more than that. But on the other hand some choices were pioneer: the icons expressing retention time sufficient as they can represent various durations.

Aaron [3] came up with a smaller set (see Figure 8), that can only express three types of data, whether data may be disclosed to third-parties or not, whether it may be sold, and three types of 'ownership' (user, company or shared control).¹⁶ The goal was to promote privacy commons rather than to propose a comprehensive set.

Raskin [82] developed a set of icons for Mozilla. The type of data is not considered, and only the retention time, third-party use, ad networks, and law enforcement are considered (see Figure 9). They also introduced the concepts of what they call statutory¹⁷ and transparent¹⁸

¹⁶What they denote as ownership is rather a notion of control and of its generated content.

¹⁷Defined in [82] as "This means that when an organization gets a phone call, letter, or other legally insufficient request for your data, they don't comply because the law requires the government to take additional steps before getting your data. These organizations require the government to comply, at a minimum, with the legal process provided by the law before getting users' data."

¹⁸Defined in [82] as "These organizations might provide your data to a government that asks for it without following the legally required process, but always follows a publicly-documented and consistent process."

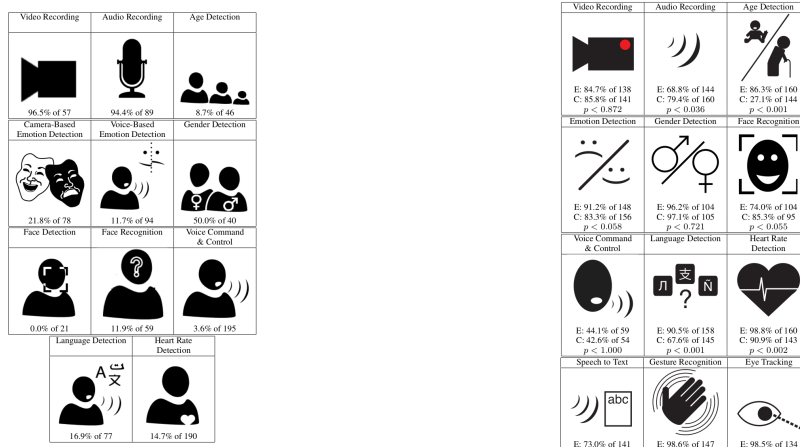


Figure 5: "Is this thing on?" icons

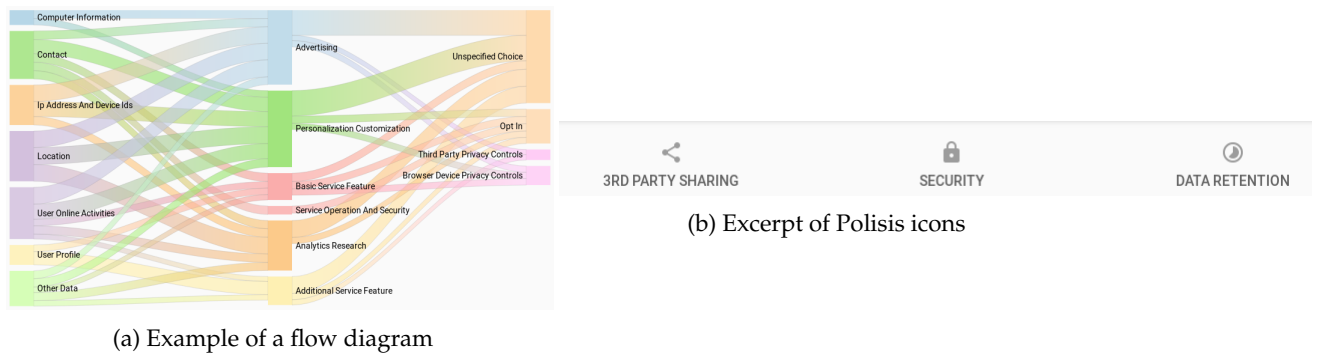


Figure 6: Polisis

processes.

Recently, a set of privacy icons was designed by Privacy Tech [79]. This set considers many types of data, as well as advanced representations of sharing, such as adequacy transfer (see Figure 10). It considers many common items, but not the rights of DS nor policy change — these items may be seen as less relevant for data transparency, even though mandatory under the GDPR.

Business sector A notable example of privacy icons are the android permissions [42], created by Google. They present icons combined with simple natural language (see Figure 11). For each application installed on a mobile phone running Android, the permission manager presents a short graphical policy. However, 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 application's natural language privacy policies in order to find more information.

Iconset for Data-Privacy Declarations v0.1

Let's simple declare what data is how used, stored, given away or deleted.

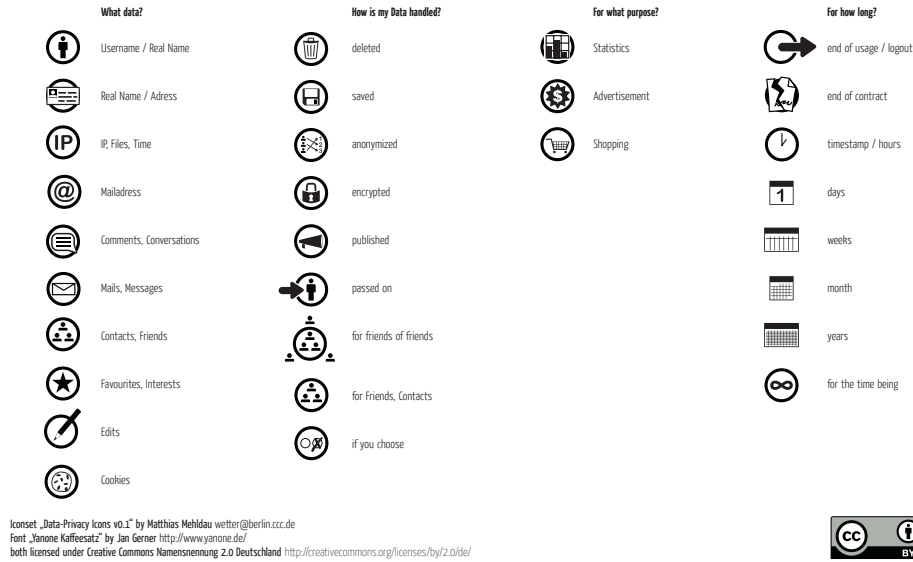


Figure 7: Mehldau's set of icons



Figure 8: Privacy Commons icons set



Figure 9: Raskin's set for Mozilla

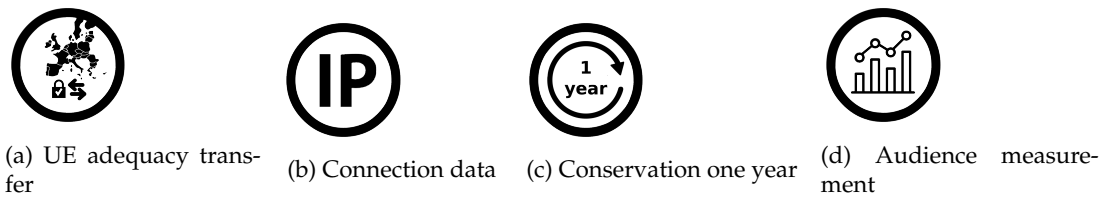


Figure 10: Excerpt of the Privacy Tech icons

3.1.2 Rating solutions

Certain graphical privacy policies do not consist of icons but provide other graphical representations instead. These solutions chose to present extra information related to privacy policies,

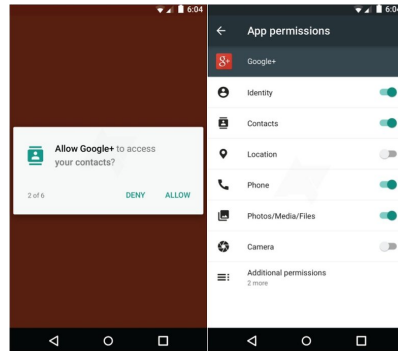


Figure 11: Example of android permissions

often a judgement of the risk level associated to a DC policy, or a comparison between DC and DS policies. We denote them *rating solutions*. Both academics and privacy advocates contributed to the solutions.

Academia Van den Berg and Van der Hof [101] devised a wheel whose spokes show how data are handled (see Figure 12). Their solution highlights fairness of processing rather than transparency: it issues a judgment on the processing, but shows little information with respect to what data are collected, by whom, and for what purpose.



Figure 12: Privacy wheel

Sweeney *et al.* [93] proposes a simplified interface for access requirement to medical data called Datatags — further developed in [13] (see Figure 13). To each piece of data can be associated a tag presenting the risks, the security features associated, and the credentials required to access it. Here, only data security is considered (security features and access credentials).

Hagan describes in [44] a user-centered privacy policy design project. The proposals include for instance a Visual Data Privacy Diagram to visualize data flow intuitively, Multi-character stories to present concrete situations (see Figure 14), and Context-Specific Alert for a selection of common questions regarding location data. The proposition cannot and do not intend to express items defined in Section 2, but attempts to increase DS awareness about consequences of data processing.

Privacy Bird [25] was an assistant developed by the CyLab of CMU for P3P, helping users to make decisions by comparing their privacy preferences with a website's privacy policy. It

Tag Type	Description	Security Features	Access Credentials
Blue	Public	Clear storage, Clear transmit	Open
Green	Controlled public	Clear storage, Clear transmit	Email- or OAuth Verified Registration
Yellow	Accountable	Clear storage, Encrypted transmit	Password, Registered, Approval, Click-through DUA
Orange	More accountable	Encrypted storage, Encrypted transmit	Password, Registered, Approval, Signed DUA
Red	Fully accountable	Encrypted storage, Encrypted transmit	Two-factor authentication, Approval, Signed DUA
Crimson	Maximally restricted	Multi-encrypted storage, Encrypted transmit	Two-factor authentication, Approval, Signed DUA

Figure 13: Final version of the Datatags

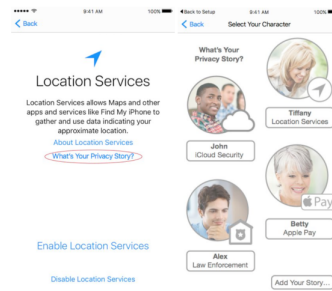


Figure 4. Privacy terms communicated through archetypal user stories, about privacy preferences, scenarios, and consequences.

Figure 14: Multi-character stories

consists of a coloured bird, where the colour indicates the matching (green for a match between the DS policy and the website's DC policy, red for conflict, yellow for uncertain, grey when disabled) (see Figure 15).



(a) Matching policies

(b) Conflicting policies

(c) Uncertain

(d) Disabled

Figure 15: Privacy Bird

In [71], Pardo & Le Métayer present a web interface 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?” (see Figure 16). DS simply need to click on “Analyze” to automatically obtain the answer to the questions. Additionally, DS may introduce risk assumptions in order to specify possible misbehaviors that the collecting parties can perform. In Section 4, we describe in detail the underlying privacy language and the automatic risk analysis.

Privacy advocacy The ToS;DR initiative helps DS understanding the risks associated to a DC policy [95]. 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

PILOT Privacy Policy

Enter the PILOT privacy policy you would like to analyse:

Parket may collect data of type and use it for until

21 / 03 / 2019

This data may be transferred to:

ParketWW which may use it for until

Risk Analysis

Risk Assumptions

Choose the assumptions for the model:

☐ ParketWW may transfer personal data to disregarding the associated DS policies.

☐ CarInsure has strong interest in using data for .

Risk Questions

Click on **Verify!** to get answers to the questions below. The answer depends on the PILOT policy and the assumptions you have chosen.

- Can Parket receive my data?
- Can ParketWW receive my data?
- Can CarInsurance receive my data?
- Can Parket use my data for other purpose than commercial offers?
- Can ParketWW use my data for other purpose than commercial offers?
- Can CarInsure use my data for profiling?

Figure 16: Input Forms of Risk Analysis Web Application.

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 assigned automatically by pondering the average scores (see Figure 17).

- Class A** are the best terms of services: they treat you fairly, respect your rights and will not abuse your data.
- Class B** The terms of services are fair towards the user but they could be improved.
- Class C** The terms of service are okay but some issues need your consideration.
- Class D** The terms of service are very uneven or there are some important issues that need your attention.
- Class E** The terms of service raise very serious concerns.
- No Class Yet** We haven't sufficiently reviewed the terms yet.

Figure 17: Terms of Services; Didn't Read

3.2 Tools

Tools for representing graphical privacy policies (*graphical tools* in the sequel) are tailored to the classical web, and are often found as add-ons for web browsers.

Privacy Bird [25] is one of the first graphical tool (see Figure 15). It is represented as an add-on for Internet Explorer, restricted to Microsoft Windows. A dedicated website [25] provides an explanatory tour as well as a feature named *privacy finder*. Privacy finder displays the search

results of a search engine, combined with the analysis of Privacy Bird. DS could rank the results according to the matching between their DS policy and the websites DC policies.

Another example of graphical tool is ToS;DR [95], started in 2011 by the Chaos Communication Camp. ToS;DR is an add-on for both Firefox and Chrome. ToS;DR ranks policies based on crowdsourced analyses by a community, directly within the web browser. The project is still active as of early 2019.

The most recent example of a graphical tool is Polisis [45]. A website has been built to present the tool [76], and add-ons for Chrome [74] and Firefox [75] are available (the add-ons redirect to the corresponding part of the website).

Only one tool is made of icons, unlike the others which represent rating solutions. The add-on “Disconnect Privacy Icons” [29], in collaboration with TRUSTe, which evolved from Raskin’s set of icons for Mozilla (see Figure 9), 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.

Schaub *et al.* [88] introduced a Design Space for Privacy Notices to foster understanding of *privacy notices*, which can be understood as graphical privacy policies in our sense. They propose guidelines for developing graphical policies. They distinguish the *timing* (at setup, just in time, context-dependent, periodic, persistent, and on-demand), the *channel* (primary, secondary, and public), the *modality* (visual, auditory, haptic, and machine-readable), and the *control* brought by notices (blocking, non-blocking, and decoupled). They argue that each privacy notice should be thought in these terms, and that an appropriate use combined with user studies would participate in a better understanding from DS.

3.3 Benefits

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

Designed for lay-user understandability Many solutions coming from privacy advocates ([65], [82], and [3] for instance) aim to provide intelligible information to lay-users: “In order for privacy policies to have meaning for actual people” [81]. These solutions were built with the will to popularize natural language privacy policies, and were designed to be understood quickly and to take into account simplicity. It is also the case for academic solutions such as the privacy labels [55], 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: “Our only requirement was that English be the participant’s native language”. Graphical privacy policies can also provide intelligible notices for scientists and health physicians using sensitive datasets, such as the DataTags [93] — further developed in [13]. 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.

Measuring understanding Attempts were made to analyse what icons were recognisable and to measure their reliability. Egelman *et al.* [32] crowdsourced privacy indicators for the Internet of Things. In their study, they found out that some icons are well-recognized (for example, the camera symbol was recognized by more than 95% of participants as representing *video recording*), while others not (only 3.6% recognized the *voice command & control* icon). The PrimeLife project also proposed a set of privacy icons ([1, Chapter 15]), that they tested in order to determine if they were understood. They concluded that clear icons with few details were

preferred. Kelley *et al.* [56] 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*.¹⁹

Legal Design Legal design can be defined as “[the] application of design-thinking (processes by which design concepts are developed by designers) principles to the practice of law, to make legal systems, products, services and processes more useful, useable, understandable and engaging for all” [90]. For instance, the set of privacy icons from Privacy Tech [79] was designed with the GDPR in mind, to raise awareness among DS (see Figure 10). In a blog article, McCartney [63] discusses the Legal Design Lab lead by Margaret Hagan. She states that legal design — which includes graphical privacy policies *de facto* — could help people understanding complex legal issues. With respect to privacy, Hagan [44] proposed different tools to help lay-users understanding (see description in Section 3.1.2). Each proposition is an approach to convey privacy notices in a way that would provide better interaction with DS for a better understanding.

3.4 Limitations

Ambiguity Though accessible to lay-users, graphical privacy policies may be interpreted in different ways, thus leading to ambiguities (see Section 3.1). 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. Nothing has been done to produce a reasonably recognized set of icons for privacy — *e.g.* validated by a user study — despite the attempts of [32] and [1, Chapter 15] to see what were the most recognizable icons, and of [56] to provide a graphical policy where results could be found accurately.

Incompleteness Graphical privacy policies are limited by their restricted scope. As seen in Section 3.1, 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, and two aspects in particular are never mentioned in graphical privacy policies (legal basis and policy change).

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. Not all items of the taxonomy are considered: those not appearing in any of the surveyed works are omitted.

¹⁹Note that they also test the speed of retrieval, as well as comparisons between DC policies in addition to information retrieval.

4 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.* [54], 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 (cf. [54, 100]). Here we review the different ways in which privacy languages are used to express machine-readable privacy policies. In particular, we categorise each work based on: i) the elements in the taxonomy presented in Section 2 that it captures; ii) the type of enforcement mechanism it uses and

	Icons	Simple text	Rating	DC	DS	1 st party	3 rd party	DS rights	Data retention	Data security
Privacy Bird [25]			✓	✓	✓	—	—	—	—	—
Rundle [86]	✓	✓			✓	●	○	●	○	●
Mehldau [65]	✓			✓		●	●	○	●	●
Privacy Commons [3]	✓			✓		●	●	○	○	○
Privacy Nutrition Label [55]	✓	✓		✓		●	●	○	○	○
Primelife [1, Chapter 15]	✓			✓		—	—	—	—	—
Raskin [82]	✓			✓		●	●	○	●	○
Privicons [58]	✓				✓	○	●	○	○	○
Privacy wheel [101]		✓	✓	✓		○	●	○	○	●
Android permissions [42]	✓	✓		✓		●	○	○	○	○
“Is this thing on?”[32]	✓			✓		●	○	○	○	○
Datatags [13]		✓	✓	✓		○	○	○	○	●
Hagan [44]		✓	✓	✓		●	●	●	●	○
Polisis [45]	✓	✓		✓		● _a	● _a	● _a	● _a	● _a
Privacy Tech [79]	✓			✓		●	●	○	●	●
ToS DR [95]		✓	✓	✓		● _a	● _a	● _a	● _a	● _a
	Features			Type		Content				

Table 2: Framework of comprehension of graphical policies

We use _a to denote that a solution extensively uses natural language in combination with graphical representations.

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 2. 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 some items of the taxonomy are omitted since no solution includes them.

whether it has been implemented; iii) additional tools for policy analysis or comparison; iv) the intended audience of the language, *e.g.*, DS or DC; and v) whether it is intended to be directly used by lay-users. Table 3 in Section 4.4 summarizes our study.

4.1 Content

In this section, we describe the different types of content that machine-readable privacy policies include. This content is determined by the syntax of the privacy language. Many languages are defined using XML or JSON that 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 2), the format used to express the policies and their target audience.

Access control languages such as XACML [7] and RBAC [87] 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 [61] include conditions depending on geolocation information, *e.g.*, “Alice can only access data from Lyon”. However, none of these languages captures concepts such as retention time, purpose or transfers that are in the privacy policy taxonomy described in Section 2. In other words, access control languages cannot impose any usage constraints after data has been accessed.

Usage control (UCON) [72, 77] appeared as an extension of access control to express how the data may be used after being accessed. To this end, it introduces *obligations*, which are actions to be executed after data has been received — *e.g.*, “do not transfer data item *i* to Company X” or “remove data on 28/01/2019”. These obligations make it possible to express items such as retention time, purpose and allowed data transfers. The Obligation Specification Language (OSL) [46] is an example of a fully-fledged UCON language together with an enforcement mechanism through Digital Right Management systems (DRMs) [37].

Neither access control nor UCON were developed with the idea of expressing privacy policies in mind. For instance, these languages do not offer mechanisms to describe DS policies. They are mostly used by DC to define their policies. New policy languages focused on expressing privacy policies appeared to address this problem.

Several languages dedicated to privacy policies have been proposed. A pioneer project in this area was the “Platform for Privacy Preferences” (P3P) [23]. P3P was conceived as a policy language for websites. It allows clients to declare their privacy preferences, and online service providers (mostly websites) to inform how they use customers’ data. P3P policies are specified in XML format, and include notions such as 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 [59, 10, 6], where its syntax has been extended — for instance, E-P3P [10] extends P3P’s syntax with obligations *à la* UCON. After P3P appeared, new languages with similar syntax have been proposed such as the “Enterprise Policy Authorization Language” (EPAL) [9], “An Accountability Policy Language” (A-PPL) [11], “Customer Profile Exchange” (CPExchange) [17], “Privacy Rights Markup Language” (PRML) [108], “Purpose-to-Use” (P2U) [51] and “Layered Privacy Language” (LPL) [40]. None of them add new features to the content of policies, but enhancements in terms of usability or enforcement (see Section 4.2). For instance, in LPL and PRML it is mandatory to include natural language explanations of policies, and EPAL offers automatic policy comparison.

Another line of work is that of formal privacy languages (*formal languages* in the sequel). S4P [15], SIMPL [60], QPDL [100], CI [69], PrivacyAPIs [62], PrivacyLFP [28] and PILOT [71] are languages which have their syntax and semantics defined by means of mathematical definitions. More precisely, they use formal languages such as *Linear Temporal Logic* [48], *First-Order Logic* [48] or *Authorization Logic* [4]. However, not all these formal languages have the same focus. S4P, SIMPL and PILOT are focused on expressing DS and DC policies. Thus, they do not differ much in content from the languages mentioned in the paragraphs 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 languages mentioned above as they are based on logical languages. For instance, PILOT makes it possible to include spatio-temporal conditions which allow DS and DC to describe when, where and by which devices data may be collected. On the other hand, CI, PrivacyAPIs and PrivacyLFP are focused on encoding privacy regulations such as HIPAA [99], COPPA [35] or GLBA [98]. 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.* [14] 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”. Finally, QPDL is a meta-language to reason about privacy languages. While privacy policies can be expressed in QPDL, it is not its intended use. The language was conceived as a framework to formally reason about different policy languages, *e.g.*, to compare the expressive power of different languages.

Jeeves [106] is a programming language with built-in support for a limited form of privacy policies. It allows programmers to declaratively specify confidentiality conditions based on the execution context. For instance, in a double-blind conference management system, paper authors can only be seen by organisers or the authors itself until the review process is completed.

4.2 Tools

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

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, *e.g.*, DS and DC. There are several ways to express semantics formally. For instance, SIMPL, S4P and CI use trace semantics, *i.e.*, they defined what are the allowed sequences of events (traces) given a set of privacy policies. PILOT uses small step operational semantics that define what events may be executed given the state of the system and the privacy policies of DS and DC. Jeeves, which is defined as a fully-fledged programming language with support for privacy policies, has its semantics formalised using lambda calculus [19]. Rei has its semantics defined as a set of logical rules in Prolog [20]. Though precise and unambiguous, in most cases formal semantics are not directly executable — there is a gap between a formal definition 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 [47].

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 have in common that they 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 the type of the requested 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. Usually, data requests include more information, *e.g.*, P3P data requests include data type, purpose of usage, requesting user, and the action to be performed (*e.g.*, read, write, delete, etc.). Most languages do not have mechanisms to enforce that data will be used according to the policies — *e.g.*, checking whether data are deleted before the retention time, or used for the specified purposes — but there are some exceptions. LPL erases automatically data from the central repository after the retention time has elapsed. UCON-based languages, such as OSL, use DRM to guarantee that obligations are enforced. A common factor of all these languages is that their request evaluation engines have been implemented and are ready to be deployed.

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. Comparison is necessary to make it possible to mechanize consent. If the policy of a DC is more restrictive than that of a DS, then DS privacy preferences are satisfied. Examples of such languages include EPAL, P3P and PILOT. In fact, the graphical tool Privacy Bird (mentioned in Section 3.2) uses P3P's comparison algorithm to provide visual feedback to DS. CI, SIMPL and S4P follow a different approach. They define how restrictive a policy is, based on its semantics.²⁰ Languages that do not differentiate between DS and DC policies — such as RBAC, EPAL, A-PPL, or OSL — tend not to define 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 [12] 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?” or “Can my data be collected by Company Z?”. 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: for instance, it can determine who can access patients medical files depending on their content or role.

4.3 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 4.2, all policy languages have the

²⁰Using 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.

means to guarantee that data are accessed according to the policies. Languages based on UCON or formal languages often provide stronger guarantees as they define how data are processed by all the parties after data collection. For example, they ensure that data are only used for purposes in the policies or that data are only transferred to allowed entities. Languages based on request evaluation offer weaker guarantees as they only protect access to the data, but not how the receiving party must process the data. Nevertheless, due to their simplicity and ease of implementation, request evaluation languages are more widespread. Typically, every party holding personal data must implement the request evaluation engine. The implementation of formal languages tends to be more complicated. Normally, they require tracking actions applied on 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 enable the possibility of auditing whether data are being handled according to their respective privacy policies. This functionality is of great value for DPA. Auditing mechanisms are typically implemented as logs that record 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. Ensuring the integrity of the logs is an orthogonal issue which is crucial for the legal validity of the auditing mechanism [16, 89].

Correctness The lack of ambiguity in policy languages makes it possible to precisely reason about their correctness, *i.e.*, that data are handled as stated in the privacy policies. This is specially true for formal languages. Their mathematical machinery — such as formal 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, using PrivacyAPIs it was possible to find unexpected ambiguities in HIPAA. These ambiguities were also found by commenters four years after it was enacted [62]. 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. Therefore, formal languages should include auditing mechanisms, as the languages mentioned in the previous paragraph.

Automation Machine-readable privacy policies enable the possibility of automating certain procedures such as information communication and consent management. Automatic information communication facilitates transparency by making DS more aware of how their data is being handled — notably in ubiquitous systems where passive data collection is the norm. For instance, Das *et al.* [26] propose Personalized Privacy Assistants for the IoT. These assistants can inform DS of surrounding IoT devices thanks to the machine-readability of the information communicated. Automatic consent management can empower DS — *e.g.*, by mitigating the burden of choice [92] — if managed in a protective way, and facilitate the retrieving of an informed consent for DC. Cunche *et al.* [24] devise a generic framework to manage informed consent in the IoT, using DS and DC policies based on PILOT semantics [71]. Automatic communication of privacy policies also makes possible a negotiation of privacy choices: DC and DS can interact more quickly by means of machines.

4.4 Limitations

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

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 XPref, SIMPL, LPL or PILOT take into account readability requirements — they include a natural language version of each policy. However, it is questionable whether they can actually be understood. To put things into perspective, the OECD [70] conducted a study which shows 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 OECD 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. Table 3 highlights the languages which address this issue in the column *usability*.

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. P3P can claim to have had an impact on the civil society, albeit minor. It has been an official set of specifications of the W3C supported by the web browser Internet Explorer.²¹ It is to be noted that other languages were published as specifications by companies [17, 7] and can therefore be considered as having had some recognition. On the other hand, most formal languages lack a practical scalable implementation which makes it difficult to use in practice. All in all, usability, implementation and widespread recognition is 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 often results in a lack of adoption. To provide a visual and global overview of machine-readable policies, we present a framework to review them (see Table 3). Not all items of the taxonomy are considered: those not appearing in any of the surveyed works are omitted.

5 Insights

As we have seen in the previous sections, each dimension has a number of limitations and benefits. In this section, we show that each dimension is tailored to a specific audience, and that this specificity 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 5.1, we highlight the benefits of each dimension for their particular audiences, and argue that a single dimension cannot comply with

²¹The Electronic Privacy Information Center (EPIC) wrote a report about it, even though it was to highlights its defects.

	Usability	Syntax	Enforcement	Implemented	Tools	DS	DC	Time	Space	1 st party	3 rd party	DS rights	Data security	Data Retention
P3P [23]		XML	Informal	✓	Comparison	✓	✓	●	○	●	○	○	○	●
CPEXchange [17]		XML	Informal				✓	○	○	●	○	○	●	●
PRML [108]	✓	XML	Informal				✓	○	○	●	○	○	●	○
APPEL [59]	✓	XML	Informal	✓		✓	✓	●	○	●	○	○	○	○
E-P3P [10]		XML	Formal			✓	✓	●	○	●	○	○	●	○
Rei [53]		Formal	Formal		Analysis		✓	●	○	●	○	○	○	○
Xpref [6]	✓	XML	Informal	✓		✓		●	○	●	○	○	○	○
XACML [7]		XML	Informal	✓			✓	○	○	○	○	○	○	○
EPAL [9]		XML	Informal	✓	Comparison		✓	○	○	●	○	○	○	○
CI [14]		Formal	Formal			✓	✓	●	○	○	●	○	○	○
SIMPL [60]	✓	Formal	Formal			✓	✓	○	○	●	●	●	○	○
S4P [15]	✓	Formal	Formal			✓	✓	○	○	●	●	○	○	○
Jeeves [106]	✓	Formal	Formal	✓			✓	○	○	○	○	○	○	○
P2U [51]	✓	XML	Informal			✓		○	○	●	○	○	○	○
QPD [100]	✓	Formal	Formal			✓	✓	○	○	●	●	○	○	○
RBAC [87]		XML	Informal	✓			✓	○	○	○	○	○	○	○
OSL [46]		Formal	Formal	✓			✓	○	○	○	○	○	○	○
GeoXACML [61]		XML	Informal	✓			✓	○	○	○	○	○	○	○
A-PPL [11]		XML	Informal				✓	○	○	○	○	○	○	○
LPL [40]	✓	XML	Informal	✓		✓	✓	○	○	○	○	○	○	○
PrivacyAPIs [62]		Formal	Formal		Analysis	✓	✓	○	○	○	○	○	○	○
PrivacyLFP [28]		Formal	Formal			✓	✓	○	○	○	○	○	○	○
PILOT [71]	✓	Formal	Formal		Analysis	✓	✓	○	○	○	○	○	○	○
Features						Audience		Conditions		Content				

Table 3: Framework of comprehension of privacy languages.

Features Machine readable privacy policies specific features:

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 This column specifies type of available tools for the language.

Audience Whether the language can describe a DS or a DC policy, see Appendix B.

Conditions Whether the languages supports conditional rules describing when (time) and/or where (space) data may be collected.

Content Whether the language can express the items described in Section 2.1. We use ● to denote that the item is explicitly included in the language; ○ to denote that the item is partially supported, *e.g.*, may be encoded through conditions or obligations; and ○ the item is not present in the language and cannot be encoded. Note some items of the taxonomy are omitted since no solution includes them.

every requirement. In Section 5.2, we put in perspective the works which attempt to overcome the limitations of mono-dimensional solutions.


5.1 Limitations of mono-dimensional solutions

Limitations in one dimension correspond to benefits in others. This is not a surprise, the different dimensions target different audiences and have different goals. Therefore, to each dimension corresponds benefits for an audience — *e.g.*, DS, DC, DPA, and external auditors — and limitations for another type of audience. In the following, we describe the limitations of each dimension for audiences outside the dimension.

Natural language privacy policies aim at defining the terms for data collection and processing. They must be precise enough so that, given a set of facts, a lawyer or a DPA can determine whether the privacy policy is consistent with what is properly enforced. In other words: they are required to check the compliance with the law and with the processing conducted by DC (see Section 2.3). Details specific to natural language privacy policies are often used by lawyers to check that the policy complies with privacy protection regulations — such as the GDPR — or they refer to functionalities of the system — *e.g.* logging or cookie management. In general, lay-users may not have the knowledge to fully understand these details, which makes it less accessible for them. For instance, the item *legal basis* privacy policies may be difficult to understand by DS. Likewise, natural language privacy policies do not include low level details related to the enforcement of the policies by a machine. In fact, those details are often unnecessary for the purposes of law enforcement and make it difficult for lawyers — who may lack the technical knowledge to understand the details — to use privacy policies. As a result, natural language privacy policies are mostly specified by DC, and they address DS and DPA.

Graphical privacy policies aim at providing a simplified version of the policy to lay-users. They are useful for a better understanding. As a consequence, they are used by DC, and target DS. These policies do not contain details related to the legal aspects of the policy, nor aim to be automatically enforced by a machine. Graphical privacy policies are primarily used as complements to natural language privacy policies. Furthermore, their consistency with natural language privacy policies should be checked by DPAs in order to avoid misleading DS.

Machine-readable privacy policies, on the other hand, 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, consequently, they are unsuitable for lawyers and lay-users. Machine-readable privacy policies can be automatically enforced, and they enable robust auditing. Therefore, they are normally used by DC as part of their data collection and processing systems. Additionally, when they support DC and DC policies, they may be used for automatic policy comparison and consent management.

Illustrative example In order to illustrate the differences in the details that the different dimensions capture, we use an example of privacy policies regarding retention time. In Section 3 we presented the icon  (from Privacy Tech Icons) that denotes that data is deleted after 1 year. Now we show an excerpt of Facebook's privacy policy related to retention time.

[...] 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 a very concrete piece of data which is produced after 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 now 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 Listing 1 shows the above policy in a machine readable language (APPEL-P3P).

```
...
<PURPOSE>
  <log-search-query />
</PURPOSE>
<RETENTION>
  <stated-purpose/>
  <EXTENSION>
    <retention-time days=182>
      xmlns="https://www.example.com/P3P/retention-time/" />
    </EXTENSION>
  </RETENTION>
...
```

Listing 1: Example of APPEL Policy.

This policy includes details that are not present in the natural language policy above. For instance, the format of the policy (XML) can be seen as including technical details about the underlying enforcement of the policy. Also, the policy includes additional technical parameters such as `xmlns` which is required so that the computer can retrieve the set of possible values for the element in the policy (the XML namespace), or the fact that retention time must be specified in days. Additionally, the enforcement mechanism of the policy would precisely define when the data is removed. For example, a background process checks on a daily basis whether the retention time has expired and, if so, deletes the data; or perhaps this process is executed once a year — both would comply with the privacy policies. The graphical and natural language versions of the policy only state for how long data will *not* be removed. These details are necessary for a machine to enforce the policy, but they have little or no value for lay-users and lawyers — in fact, some of these details negatively impact the understanding of the policy.

Intrinsic character of these limitations As a result, a single dimension cannot cover all the benefits required — *i.e.* legal validity, understandability by all parties, and enforceability. This is due to the tension between the details of a privacy policy and i) its suitability for lawyers, ii) lay-users, and iii) its automatic enforcement by machines. Concretely, there are details that only have meaning in one dimension and are irrelevant in others. Natural language privacy policies include details related to compliance with data protection regulations, which are unnecessary for the machine-readable dimension. 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.

5.2 Overcoming limitations

As argued in Section 5.1, a privacy policy expressed in only one dimension cannot satisfy requirements for lay-users, lawyers, and auditors. The current section aims to show that it is possible, however, to combine different dimensions to overcome their respective limitations. In many cases, limitations in one dimension can be addressed by other dimensions. For instance, natural language privacy policies may use graphical policies to enhance readability. Similarly, machine-readable privacy policies can use natural language privacy policies in order

for the latter to be automatically enforced by DC. Furthermore, analysis tools provided with machine-readable privacy policies can be combined with graphical privacy policies to enhance the presentation of risks to DS. In other words: these three dimensions must be seen as complementary ways to express privacy policies, and they should be used together in order to meet the requirements described in Section 1.

Several initiatives are already proposing cross-dimensions solutions, such as Harkous *et al.* [45] for natural language privacy policies and graphical privacy policies: they combine the accessibility of icons and simple text with the legal value of a natural language privacy policy. Policies can be more easily understood thanks to the results of the automatic analysis of natural language privacy policies. Le Maître [60] proposed a combination of natural language privacy policies and machine-readable privacy policies: a machine-readable formal privacy language, enforceable, but close enough to natural language to be readable. PILOT [71] also combines machine-readable privacy policies and natural language privacy policies by providing a natural language user interface for DS to input their machine-readable privacy policies. Finally, Kelley *et al.* [55] added a graphical representation on top of P3P — a machine-readable privacy policy — resulting in both intelligible and enforceable privacy policies. These examples show that the combination of two dimensions makes it possible to take advantage of each dimension, without losing benefits.

Other initiatives attempted to dissociate the dimensions within the same solution. For instance, Van den Berg and Van der Hof [101] implemented a privacy wheel whose spokes present how data handled. Those spokes have different layers: a first graphical layer aimed at DS, with the possibility to access two other layers of information comprising more details (“legalistic” information). Another prominent example is the multi-layer approach of Abrams and Crompton [5]. They proposed a first layer as a short notice (of whether data are collected), a second layer as a condensed notice of the data practices in a common graphic format, and finally a third layer, also called full notice, aimed at lawyers. However, these last two initiatives do not consider machine-readability — and therefore enforceability of privacy policies.

6 Recommendations

We saw in Section 5 that no existing solution is satisfactory and encompasses the requirements for legal compliance, understandability, and enforceability. To mitigate this issue, we propose an approach to define policies covering the three dimensions studied in this paper: natural language, graphical, and machine-readable. We denote these policies *multifaceted privacy policies*. We use the term *facet* to refer to a dimension of privacy policies when more than one is considered. The natural language facet can be verified for legal compliance, *i.e.*, lawyers should be able to check whether all information required by law is present. The graphical facet should be carefully designed to provide clear and concise information, and more information should be easily accessible. The machine-readable facet, endowed with a well-defined semantics, allows for the enforcement of the policy, and, in some cases, provides tools for analysis and comparison of privacy policies (see Section 4.2). Facets should be consistent: a graphical representation will surely omit details, but by no means it should mislead DS in their understanding of the privacy policy.

The natural language facet is a requirement, as it is the only one holding legal value. It can be decomposed according to the taxonomy in Section 2, and each element of the taxonomy can itself be further decomposed in simple *clauses*. A clause is a statement that would not make sense if further decomposed, *e.g.*, “we collect username”, or “collected data is stored until it is no longer necessary to provide our services”.

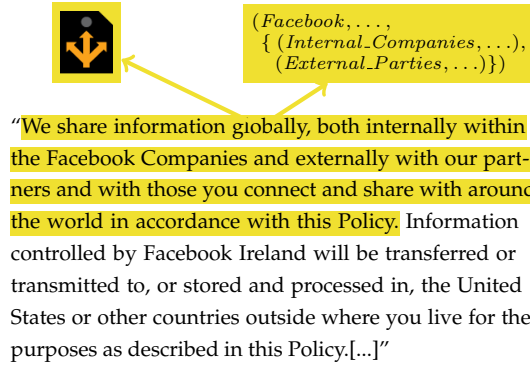


Figure 18: Multifaceted Privacy Policy Overview.

Figure 18 shows an example of multifaceted privacy policy for the clause highlighted in yellow. The clause refers to data transfers. Hence the multifaceted policy includes an icon from the Mozilla Raskin set of icons and a PILOT machine-readable policy as both can express data transfers. Note that the categorization of the solutions in each dimension according to privacy policy taxonomy presented in this paper can be of great use for the design of multifaceted privacy policies.

Ideally, multifaceted privacy policies should represent every clause in the graphical and in the machine-readable facets, in addition to the natural language facet, but it may not be necessary to include all facets for all clauses in practice. For example, the second clause in Figure 18: “Information controlled by Facebook Ireland will be transferred or transmitted to, or stored and processed in, the United States or other countries outside where you live for the purposes as described in this Policy.” corresponds to a specific detail which can be omitted in the graphical facet as it might confuse lay-users.

7 Conclusion

In this paper, we have systematically studied the different ways to express privacy policies: in natural language, with graphical representations, and using machine-readable means. We have categorised the existing works in each dimension according to a taxonomy of privacy policies, as well as their particular features. Additionally, we have studied the benefits and limitations of each dimension, and we have shown that the limitations of one dimension can be addressed by the benefits of the other dimensions. Finally, we have proposed a novel approach to express privacy policies combining the three dimensions, denoted multifaceted privacy policies, which overcomes the limitations of each dimension by bringing together their benefits.

We envision this work as an effort to bring separate approaches together — from the legal domain, design, and computer science — and to provide a big picture of how transparency can be ensured through privacy policies.

References

- [1] Privacy and Identity Management for Life. URL <http://link.springer.com/10.1007/978-3-642-20317-6>.
- [2] 3DCart. Create an online store with 3dcart store builder. URL <https://www.3dcart.com/personalized-policy.html>.
- [3] Helton Aaron. Privacy Commons Icon Set .:aaron.helton:. URL <https://web-beta.archive.org/web/20090601215200/http://aaronhelton.wordpress.com/2009/02/20/privacy-commons-icon-set>.
- [4] 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.
- [5] Marty Abrams and Malcolm Crompton. Multi-layered privacy notices: A better way. 2(1):1–4. URL <http://www.iispartners.com/downloads/Multi-LayeredNoticesPaperPublishedinPrivacyLawBulletinVol2No1June2005PLP2.1-orderform.pdf>.
- [6] 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. URL <http://dl.acm.org/citation.cfm?id=775241>.
- [7] 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. URL <http://courses.cs.vt.edu/cs5204/fall05-kafura/Papers/Security/XACML-Specification.pdf>.
- [8] Wolfgang ApolinarSKI, Marcus Handte, and Pedro Jose Marron. Automating the Generation of Privacy Policies for Context-Sharing Applications. pages 73–80. IEEE. ISBN 978-1-4673-6654-0. doi: 10.1109/IE.2015.18. URL <http://ieeexplore.ieee.org/document/7194273/>.
- [9] Paul Ashley, Satoshi Hada, GÃ¼nter Karjoth, Calvin Powers, and Matthias Schunter. Enterprise privacy authorization language (EPAL). .
- [10] 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, . URL <http://dl.acm.org/citation.cfm?id=644538>.
- [11] Monir Azraoui, Kaoutar Elkhyaoui, 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.
- [12] Christel Baier and Joost-Pieter Katoen. *Principles of Model Checking*. MIT Press, 2008. ISBN 978-0-262-02649-9.

- [13] 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. URL <http://ieeexplore.ieee.org/abstract/document/7527746/>.
- [14] 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. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1624011.
- [15] Moritz Y. Becker, Alexander Malkis, and Laurent Bussard. S4P: A generic language for specifying privacy preferences and policies. URL <http://www.msr-waypoint.com/pubs/122108/main.pdf>.
- [16] Giampaolo Bella, Rosario Giustolisi, and Salvatore Riccobene. Enforcing privacy in e-commerce by balancing anonymity and trust. 30(8):705–718. ISSN 01674048. doi: 10.1016/j.cose.2011.08.005. URL <http://linkinghub.elsevier.com/retrieve/pii/S0167404811001052>.
- [17] Kathy Bohrer and Bobby Holland. *Customer Profile Exchange (Cpexchange) Specification*. URL <http://mail.ctiforum.com/standard/standard/www.cpexchange.org/cpexchangevl0F.pdf>.
- [18] 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. ISBN 978-1-59593-448-2. doi: 10/b3tswp. URL <http://portal.acm.org/citation.cfm?doid=1143120.1143123>.
- [19] Alonzo Church. An unsolvable problem of elementary number theory. *American Journal of Mathematics*, 58(2):345–363, 1936. ISSN 00029327, 10806377.
- [20] William F. Clocksin and Christopher S. Mellish. *Programming in Prolog (4. ed.)*. Springer, 1994. ISBN 978-3-540-58350-9.
- [21] CNIL. The CNIL's restricted committee imposes a financial penalty of 50 Million euros against GOOGLE LLC. URL <https://www.cnil.fr/en/cnils-restricted-committee-imposes-financial-penalty-50-million-euros-against-google-llc>.
- [22] 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. URL <http://dl.acm.org/citation.cfm?id=2381979>.
- [23] Lorrie Cranor, Marc Langheinrich, Massimo Marchiori, Martin Presler-Marshall, and Joseph Reagle. The platform for privacy preferences 1.0 (P3P1. 0) specification. 16. URL <https://elearn.inf.tu-dresden.de/hades/teleseminare/wise0405/Act.%208%20Models%20Languages%20Pierangela/Materials/P3P.pdf>.
- [24] Mathieu Cunche, Daniel Le MÃ©tayer, and Victor Morel. 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*, 2019. To appear.

- [25] CyLab Usable Privacy and Security Laboratory. Privacy Bird. URL <http://www.privacybird.org/>.
- [26] Anupam Das, Martin Degeling, Daniel Smullen, and Norman Sadeh. Personalized Privacy Assistants for the Internet of Things. 2018. doi: 10.1109/MPRV.2018.03367733.
- [27] Daniel DelPercio. Privacy Policy Online (2011). URL <http://www.PrivacyPolicyOnline.com>.
- [28] 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.
- [29] Disconnect. Privacy Icons. URL <https://web.archive.org/web/20160304013156/https://disconnect.me/icons>.
- [30] Docracy. An open source privacy policy for mobile apps. URL <https://web.archive.org/web/20171124185357/https://blog.docracy.com/post/27931026976/an-open-source-privacy-policy-for-mobile-apps>.
- [31] Dropbox. Dropbox Privacy Policy. Effective: 25 May 2018. URL <https://www.dropbox.com/privacy>.
- [32] Serge Egelman, Raghudeep Kannavara, and Richard Chow. Is This Thing On?: Crowdsourcing Privacy Indicators for Ubiquitous Sensing Platforms. pages 1669–1678. ACM Press. ISBN 978-1-4503-3145-6. doi: 10.1145/2702123.2702251. URL <http://dl.acm.org/citation.cfm?doid=2702123.2702251>.
- [33] European Parliament. General Data Protection Regulation.
- [34] Facebook. Facebook Data Policy. Date of last revision: April 19, 2018. URL <https://www.facebook.com/privacy/explanation>.
- [35] Federal Trade Commission. Children’s Online Privacy Protection Rule; Final Rule, . URL https://www.ftc.gov/system/files/documents/federal_register_notices/2013/01/2012-31341.pdf.
- [36] Federal Trade Commission. FAIR INFORMATION PRACTICES IN THE ELECTRONIC MARKETPLACE. page 208, .
- [37] 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.
- [38] Forbrukerrådet. Deceived by Design. URL <https://fil.forbrukerradet.no/wp-content/uploads/2018/06/2018-06-27-deceived-by-design-final.pdf>.
- [39] FreePrivacyPolicies.com. Free Privacy Policy Generator & Template with GDPR - Free Privacy Policy. URL <https://www.freeprivacypolicy.com/>.

- [40] 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.
- [41] GetTerms. Getterms.io. URL <http://getterms.io/>.
- [42] Google. Android Permissions overview, . URL <https://developer.android.com/guide/topics/permissions/overview>.
- [43] Google. Privacy Policy — Privacy & Terms — Google. Effective 22 January 2019, . URL <https://policies.google.com/privacy?gl=en&hl=en-GB>.
- [44] Margaret D. Hagan. User-Centered Privacy Communication Design. URL <https://www.usenix.org/system/files/conference/soups2016/wfnp16-paper-hagan.pdf>.
- [45] Hamza Harkous, Kassem Fawaz, R  mi Lebre  t, Florian Schaub, Kang G. Shin, and Karl Aberer. Polisis: Automated Analysis and Presentation of Privacy Policies Using Deep Learning. URL <http://arxiv.org/abs/1802.02561>.
- [46] 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.
- [47] Gerard J. Holzmann. *The SPIN Model Checker - Primer and Reference Manual*. Addison-Wesley, 2004. ISBN 978-0-321-22862-8.
- [48] Michael Huth and Mark Dermot Ryan. *Logic in computer science - modelling and reasoning about systems (2. ed.)*. Cambridge University Press, 2004.
- [49] Iubenda. Features — Compliance Solutions, . URL <https://www.iubenda.com/en/features>.
- [50] Iubenda. Terms of service, . URL <https://www.iubenda.comhttps://www.iubenda.com/en/user/tos>.
- [51] Johnson Iyilade and Julita Vassileva. P2U: A Privacy Policy Specification Language for Secondary Data Sharing and Usage. pages 18–22. IEEE. ISBN 978-1-4799-5103-1. doi: 10.1109/SPW.2014.12. URL <http://ieeexplore.ieee.org/document/6957279/>.
- [52] 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. URL <http://dl.acm.org/citation.cfm?id=985752>.
- [53] Lalana Kagal. Rei. URL <http://ebiquity.umbc.edu/get/a/publication/57.pdf>.
- [54] Saffija Kasem-Madani and Michael Meier. Security and privacy policy languages: A survey, categorization and gap identification. URL <https://arxiv.org/abs/1512.00201>.
- [55] 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, . URL <http://dl.acm.org/citation.cfm?id=1572538>.

- [56] Patrick Gage Kelley, Lucian Cisca, 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, . URL <http://dl.acm.org/citation.cfm?id=1753561>.
- [57] Jan Kolter and Günther Pernul. Generating User-Understandable Privacy Preferences. pages 299–306. IEEE. ISBN 978-1-4244-3572-2. doi: 10.1109/ARES.2009.89. URL <http://ieeexplore.ieee.org/document/5066486/>.
- [58] Ulrich Kö nig and Jan Schallaboeck. Privacy preferences for E-Mail messages. URL <https://tools.ietf.org/html/koenig-privicons-03.txt>.
- [59] Marc Langheinrich, Lorrie Cranor, and Massimo Marchiori. Appel: A p3p preference exchange language. URL <https://www.w3.org/TR/P3P-preferences/>.
- [60] Daniel Le Mâtayer. A formal privacy management framework. In *International Workshop on Formal Aspects in Security and Trust*, pages 162–176. Springer. URL http://link.springer.com/chapter/10.1007/978-3-642-01465-9_11.
- [61] Andreas Matheus and J Herrmann. Geospatial Extensible Access Control Markup Language (GeoXACML). *Open Geospatial Consortium Inc.* OGC, 2008.
- [62] 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.
- [63] Allison McCartney. How Lawyers Can Benefit From Visual Content. URL <https://visually.ly/blog/lawyers-visual-content/>.
- [64] Aleecia M. McDonald and Lorrie Faith Cranor. The cost of reading privacy policies. 4:543. URL <http://heinonline.org/hol-cgi-bin/get.pdf.cgi?handle=hein.journals/isjlp4§ion=27>.
- [65] Matthias Mehldau. Icons of privacy (original). URL <https://netzpolitik.org/wp-upload/data-privacy-icons-v01.pdf>.
- [66] Daniela Yidan Miao. PrivacyInformer: An Automated Privacy Description Generator for the MIT App Inventor. URL <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1029.2434&rep=rep1&type=pdf>.
- [67] National Telecommunications and Information Administration. Short Form Notice Code of Conduct to Promote Transparency in Mobile Apps Practices. URL https://www.ntia.doc.gov/files/ntia/publications/july_25_code_draft.pdf.
- [68] Netflix. Netflix Privacy Statement. Effective date: 11 May 2018. URL <https://help.netflix.com/en/legal/privacy>.
- [69] Helen Nissenbaum. Privacy as contextual integrity. 79:119. URL <http://heinonline.org/hol-cgi-bin/get.pdf.cgi?handle=hein.journals/washlr79§ion=16>.
- [70] Organisation for Economic Co-operation and Development. Skills matter: Further results from the survey of adult skills. OCLC: ocn953634518.

- [71] Raúl Pardo and Daniel Le Métayer. PILOT: A Privacy Policy Language to Enhance Informed Consent. In *Proceedings of the 33rd Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy, DBSec*, 2019. To appear.
- [72] Jaehong Park and Ravi S. Sandhu. The UCON_{ABC} Usage Control Model. *ACM Trans. Inf. Syst. Secur.*, 7(1):128–174, 2004.
- [73] 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 Kutylowski, editors, *ICT Systems Security and Privacy Protection*, volume 529, pages 156–169. Springer International Publishing. 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.
- [74] Polisis. Chrome Polisis, . URL <https://chrome.google.com/webstore/detail/polisis/bkddolggokpghlbhbkflbbhhjghjdojck>.
- [75] Polisis. Firefox Polisis, . URL <https://addons.mozilla.org/en-US/firefox/addon/polisis/>.
- [76] Polisis. Polisis, . URL <https://www.priobot.org/polisis>.
- [77] Alexander Pretschner, Manuel Hilty, and David A. Basin. Distributed Usage Control. *Commun. ACM*, 49(9):39–44, 2006.
- [78] Privacy Policy Generator. Privacy Policy Generator. URL <https://privacypolicygenerator.info/>.
- [79] Privacy Tech. Privacy icons. URL <https://www.privacytech.fr/privacy-icons/>.
- [80] PrivacyPolicies.com. Privacy Policy Generator: Free, GDPR, CalOPPA - PrivacyPolicies.com. URL <https://www.privacypolicies.com/>.
- [81] Aza Raskin. Making Privacy Policies not Suck, . URL <http://www.azarask.in/blog/post/making-privacy-policies-not-suck/>.
- [82] Aza Raskin. Privacy Icons - MozillaWiki, . URL https://wiki.mozilla.org/Privacy_Icons.
- [83] Joel R. Reidenberg, Jaspreet Bhatia, Travis Breaux, and Thomas B. Norton. Automated comparisons of ambiguity in privacy policies and the impact of regulation. . URL http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2715164.
- [84] 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, . ISSN 0047-2530, 1537-5366. doi: 10/gdcdzm. URL <https://www.journals.uchicago.edu/doi/10.1086/688669>.
- [85] 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, . URL http://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/berktech30§ion=6.

- [86] Mary Rundle. International Data Protection and Digital Identity Management Tools, presentation at IGF 2006.
- [87] 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.
- [88] 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. URL <https://www.usenix.org/conference/soups2015/proceedings/presentation/schaub>.
- [89] Bruce Schneier and John Kelsey. Cryptographic Support for Secure Logs on Untrusted Machines. page 11.
- [90] Meera Sivanathan. What is legal design? — Q&A with Meera Sivanathan (Legal Designer) — The Legal Forecast. URL <https://thelegalforecast.com/what-is-legal-design-qa-with-meera-sivanathan-legal-designer/>.
- [91] Daniel J. Solove. A taxonomy of privacy. 154:477. URL http://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/pnlr154§ion=20.
- [92] Cass R. Sunstein. Choosing Not to Choose. ISSN 1556-5068. doi: 10/gftmr3. URL <http://www.ssrn.com/abstract=2377364>.
- [93] Latanya Sweeney, Merc  Crosas, and Michael Bar-Sinai. Sharing sensitive data with confidence: The datatags system. URL <http://techscience.org/a/2015101601/>.
- [94] Alasdair Taylor. Privacy policy. URL <https://seqlegal.com/free-legal-documents/privacy-policy>.
- [95] Terms of Service; Didn't Read. Terms of Service Classification. URL <https://tosdr.org/classification.html>.
- [96] Termsfeed. Generic Privacy Policy template. URL <https://www.termsfeed.com/assets/pdf/privacy-policy-template.pdf>.
- [97] Twitter. Twitter Privacy Policy. Effective: May 25, 2018. URL https://cdn.cms-twdigitalassets.com/content/dam/legal-twitter/site-assets/privacy-page-gdpr/pdfs/PP.Q22018_April.EN.pdf.
- [98] United States Congress. Gramm—Leach—Bliley Act, .
- [99] United States Congress. Health Insurance Portability and Accountability Act, . URL <https://www.hhs.gov/sites/default/files/privacysummary.pdf>.
- [100] Jasper van de Ven and Frank Dylla. Qualitative Privacy Description Language. In *Annual Privacy Forum*, pages 171–189. Springer. URL http://link.springer.com/chapter/10.1007/978-3-319-44760-5_11.
- [101] Bibi Van den Berg and Simone Van der Hof. What happens to my data? A novel approach to informing users of data processing practices. doi: 10.5210/fm.v17i7.4010. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2100417.

- [102] Shomir Wilson, Florian Schaub, Aswarth Abhilash Dara, Sushain Cherivirala, Sebastian Zimmeck, Mads Schaarup Andersen, Pedro Giovanni Leon, Eduard Hovy, and Norman Sadeh. Demystifying privacy policies with language technologies: Progress and challenges. .
- [103] 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)*, . URL <http://www.aclweb.org/anthology/P/P16/P16-1126.pdf>.
- [104] WP29. Guidelines on transparency under Regulation 2016/679, .
- [105] WP29. Opinion 8/2014 on the Recent Developments on the Internet of Things. .
- [106] Jean Yang, Kuat Yessenov, and Armando Solar-Lezama. A language for automatically enforcing privacy policies. page 85. ACM Press. ISBN 978-1-4503-1083-3. doi: 10.1145/2103656.2103669. URL <http://dl.acm.org/citation.cfm?doid=2103656.2103669>.
- [107] Le Yu, Tao Zhang, Xiapu Luo, and Lei Xue. AutoPPG: Towards Automatic Generation of Privacy Policy for Android Applications. pages 39–50. ACM Press. ISBN 978-1-4503-3819-6. doi: 10.1145/2808117.2808125. URL <http://dl.acm.org/citation.cfm?doid=2808117.2808125>.
- [108] Zero-knowledge. Privacy Rights Markup Language Specification.

A Glossary

To avoid any ambiguities, we highlight terms considered as important for the reading. As it is possible to find redundant terms in the literature, and different concepts expressed under the same word, we present a glossary to help the reader.

Personal data According to the GDPR, “personal data means any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person”.²² Personal data can identify someone directly and uniquely — *e.g.* a social security number — with less precision — *e.g.* a pseudonym — or by combination with other information — *e.g.* metadata left by online behavior.

Data subject According to the GDPR, a data subject is “an identified or identifiable natural person”.²³ We name the data subject *DS* in this document.

Data controller According to the GDPR, “[data] controller means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law”.²⁴ We name the data controller *DC* in this document.

Privacy policy A privacy policy is a statement made by *DS* or *DC* to declare respectively their requirements and commitments in terms of personal data management. A privacy policy can be expressed in different ways: in natural language, graphically, or in machine-readable way.

Natural language privacy policies A *natural language privacy policy* is a privacy policy expressed in natural language.

Graphical privacy policies A *graphical privacy policy* is a privacy policy expressed graphically.

Machine-readable privacy policies A *machine-readable privacy policy* is a privacy policy expressed in a format readable by machines. This format is usually derived from a privacy policy language with a well-defined syntax and also in some cases a formal semantics.

Privacy policy language A *privacy policy language* is a language used to define privacy policies. It can describe *DC* as well as *DS* policies.

DC policy A *DC policy* is the privacy policy of a data controller. It is a commitment of the *DC* regarding its processing of personal data.

DS policy A *DS policy* is the privacy policy of a data subject. It defines the requirements of the *DS* concerning the processing of this data by *DC*.

Item An *item* is a piece of information provided in a privacy policy. Appendix B lists the items required by the GDPR.

²²See Article 4 of the GDPR.

²³See Article 4 of the GDPR, and the definition of *personal data* in this Glossary.

²⁴See Article 4 of the GDPR.

B Definition of policies

This section provides guidelines for the definition of *DS* and *DC policies* in line with the GDPR.

DC policy A *DC policy* should be understood as a privacy policy expressing the requests of DC related to the collection of DS personal data. It can be express in different ways, *i.e.* not only as a text but also as icons or in a formal language. Either way, the policy should at least provide the following items to answer the transparency requirements of the GDPR:

- the identity of the DC and its contact
- the type of data collected
- its purpose
- the legal basis for the processing
- the recipient of data
- the 3rd parties involved
- the retention time
- the rights of the DS

These requirements should be complemented with the following items to fully empower DS:

- the frequency of collection
- the location of the device
- its range of collection
- the beneficiary of the processing
- the risks associated
- what can be inferred from the collection and processing

On the one hand, location of device, frequency and range of collection are tailored to ubiquitous environments such as the IoT, and can provide better insights about devices. On the other hand, beneficiaries, risks and inferable information can summarize important information at a glance, and thereby provide more useful and impacting information.

DS policy A *DS policy* should be understood as a privacy policy expressing the requirements of DS related to the disclosure of their personal data. Among the settings DS should be able to adjust, the most obvious is their consent for data collection. And if the answer is positive, it has to be according to the following items:

- the type of data ²⁵
- the purpose of data collection
- 3rd party dissemination
- the retention time
- the data controller
- other requirements (anonymization, encryption, context ...)

²⁵Considering the difficulties a DS can have with technical terms, it should be possible to have a different granularity for the type of data, *e.g.* *unique identifiers* instead of MAC address and UID, geolocation etc



**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