Alexandr Railean, Delphine Reinhardt: Improving the Transparency of Privacy Terms Updates. In: Proceedings of the 9th Annual Privacy Forum (APF), pp 70-86, June 2021.

Improving the Transparency of Privacy Terms Updates*

Opinion paper

Alexandr Railean[1] (✉) 0000-0002-7472-2108 and Delphine Reinhardt[1] 0000-0001-6802-2108

Institute of Computer Science, Georg-August-Universität Göttingen
{arailea,reinhardt}@cs.uni-goettingen.de

Abstract. Updates are an essential part of most information systems. However, they may also serve as a means to deploy undesired features or behaviours that potentially undermine users' privacy. In this opinion paper, we propose a way to increase *update transparency*, empowering users to easily answer the question "what has changed with regards to my privacy?", when faced with an update prompt. This is done by leveraging a formal notation of privacy terms and a set of rules that dictate when privacy-related prompts can be omitted, to reduce fatigue. A design that concisely visualizes changes between data handling practices of different software versions or configurations is also presented. We argue that it is an efficient way to display information of such nature and provide the method and calculations to support our assertion.

Keywords: IoT, privacy, usability, updates, transparency, GDPR.

1 Introduction

Although updates are an inherent part of the lifecycle of most information systems, the update process is affected by a number of technical and usability issues, which can be seen in contexts ranging from mobile and desktop applications, to embedded systems and Internet of Things (IoT) appliances [13], [34]. As a result, many systems remain insecure, while users are frustrated and may lose interest in the maintenance of their systems [13], [34]. Among these update-related issues, we focus on *transparency*, discussed in Art. 12(1) of the General Data Protection Regulation (GDPR), which requires that information addressed to users should be "*concise, easily accessible and easy to understand, and expressed in clear and plain language*", such that they can figure out "*whether, by whom and for what purpose* personal data are collected"[12], [14]. Prior research has shown that the

* The final publication is available at Springer via https://link.springer.com/chapter/10.1007/978-3-030-76663-4_4

current level of transparency is inadequate and that in many cases end users cannot exercise their rights [8], [26]. Users face problems such as excessive length of privacy policies, complex language, vagueness, lack of choices, and fatigue [31]. The need for improvements is also motivated by estimations that show that the expectation for users to fully read and understand privacy policies is not realistic, as it would take circa 201 hours for a typical American user to read the privacy policies they are exposed to in the course of a year [19]. Moreover, even when users read policies, they are often confronted with "opaque transparency" - a practice of deliberately designing user experiences in a way that obfuscates important information [5], [18]. This suggests that end-users are in a vulnerable position and that their privacy is undermined.

In this paper we focus on the scenario in which a user is notified about an update for an IoT device they own, prompting them to consider potential privacy implications of installing the update. We propose a set of measures that simplify this analysis, and posit that a net gain in transparency can be attained by (1) avoiding unnecessary prompts, (2) showing less information, (3) displaying it in a common form, and by (4) decoupling feature, security and privacy updates. As a result, end-users can increase awareness of how data collection may affect their privacy, and thus be in a better position to make informed decisions.

In what follows, we elaborate on each of the points above. Sec. 2 provides a high-level overview of our approach. Sec. 3 introduces a formal notation of privacy terms, which is then used in Sec. 4 to determine when update prompts can be omitted. In Sec. 5, we argue that our proposed way of expressing updated privacy terms is more efficient than prose typically used for this purpose. Sec. 6 describes additional steps that can be taken to further improve transparency. In Sec. 7 we discuss the implications of applying our approach, while Sec. 8 reviews related work. We make concluding remarks in Sec. 9.

2 Proposed Approach

Art. 6(1a) and Art. 7 of the GDPR require informed and freely given consent before the collection of personal data, unless exemptions from Art. 6(1) apply. This is also required when something changes in the way personal data are handled since consent was previously granted [12]. In this paper we explore a scenario where instead of flooding users with information, we show them a minimal subset of facts that are sufficient to make a rough, but actionable assessment. Further refinement can be accomplished by investing more time in the evaluation, should the user wish so.

We assert that this minimal subset of information is a "who gets the data" table shown in Fig. 1, because it is easy to interpret, and it can be used to quickly derive answers to these questions related to transparency:

1. *What* data are collected?
2. *What is the purpose* of collection?
3. *Where* are the data stored?
4. *How long* are they kept?
5. *Who* has access to the data?
6. *How often* are the data sent?

Data type	Purpose	Company	Country	Duration	Frequency
temperature	research	Minerva LTD	Canada	1y	daily
humidity	marketing	ThirstFirst LTD	USA	1y	hourly

Fig. 1. The "who gets the data" table, adapted from [27]. Note that the table can be configured to show personal and non-personal data (see Sec. 7.5 for details).

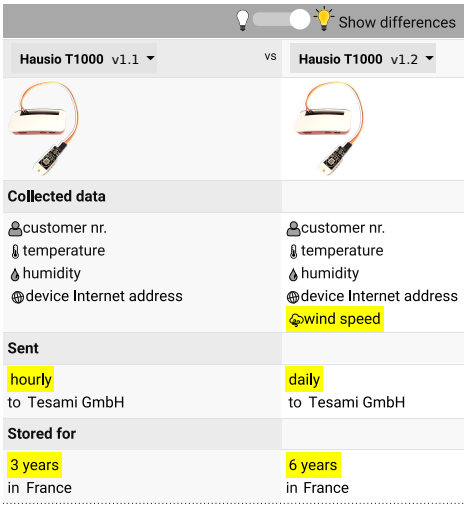

Fig. 2. Comparing two versions of the same device side by side, while highlighting differences (adapted from [27]).

The table in Fig. 1 was originally conceived as a component of an Online Interface for IoT Transparency Enhancement (OnLITE), which summarizes data collection practices and privacy information, and makes it easy to compare different IoT devices side by side, as shown in Fig. 2 [27]. Although the aforementioned transparency questions are not directly expressed in the legal requirements, they are derived from Art. 13 of the GDPR, and the results of our previously conducted usability evaluation showed that such a formulation is clear to non-experts [27].

In this work we take the idea further, applying OnLITE when an update is available, enabling users to compare an IoT device, a program, or a web-site against *another version of itself*. Thus, we leverage a design that we evaluated and which received positive feedback from our participants [27]. Considering that the privacy impact variations between updates are expected to be minimal, we have reasons to believe that the proposed UI will focus the users' attention on the few things that have changed, making it more difficult for companies to deploy features that are potentially privacy-abusive.

In the context of consent prompts for updated terms, the earliest time when we can take steps to protect a user's privacy is *before* displaying the prompt. It

has been established that exposing a person to frequent stimuli leads to fatigue, making them more likely to dismiss potentially important interactions [7], [31]. Such an effect occurs after just two exposures, and grows with repeated exposure [1], [2]. Conversely, decreasing the total number of exposures can reduce fatigue. Thus, we have to understand in what circumstances consent prompts can be omitted without undermining users' privacy. To this end, we propose a notation of privacy terms, and then use it to formally define these circumstances.

3 Formal Notation of Privacy Terms

There are multiple factors that can influence a user's privacy. We take a GDPR-centric approach and focus on the items targeted by the transparency questions listed in Sec. 2. For example, privacy is affected if the *retention* period changes from "1 month" to "10 years", or if the collection *frequency*[1] changes from "once per day" to "twice per second" [15]. Thus, our notation aims to capture these parameters, using the following symbols:

Data type Δ type of collected data
Purpose Π purpose of collection
Time T the retention period
Company C a company that gets the data
Location Λ location of said company
Frequency Φ how often the data are transmitted

These symbols are then encapsulated into structures of a higher level of abstraction, such that they are easier to write down and reason about:

Term Θ a tuple of the form $(\Delta, \Pi, C, \Lambda, T, \Phi)$, indicating agreement to sharing a type of data, for a specific purpose, with a company located in a particular country, for the given duration of time, shared at a certain regularity.
Consent K a set of terms accepted by the user, e.g., $K = \{\Theta_1, \Theta_2, \Theta_3, ..\Theta_i\}$.

Thus, when a user gives consent, we formally represent that in an expanded form as: $K = \{(\Delta_1, \Pi_1, C_1, \Lambda_1, T_1, \Phi_1), \ldots, (\Delta_i, \Pi_i, C_i, \Lambda_i, T_i, \Phi_i)\}$. Here is a practical example with some actual values: $K = \{$
$(temperature, research, MinervaLTD, Canada, 1y, daily)$,
$(humidity, marketing, ThirstFirstLTD, USA, 1y, hourly)\}$.

This notation facilitates the automatic processing of privacy terms by software and enables us to define a formal set of rules that govern when consent *must* be requested again, and when it can be omitted.

Note that in the example above Λ is a country, but it could also be a less granular value such as "within EU" or "outside EU". At this stage we only argue that a location component must be present in the tuple, without having a strong preference towards one option or the other. Finding the optimal approach is outside the scope of this opinion paper.

[1] Art. 13 of the GDPR does not require showing information about how often the data are transferred. We include it, because increasing sampling rates can lead to privacy implications, especially when correlation with other data-sets is possible.

Rule	Logic	Formal notation	Intuition
1	Strict subsets	$K_{new} \subset K_{old}$	I agree to fewer (i.e., more stringent) terms than before
2	Equal sets	$K_{new} = K_{old}$	I still agree to identical terms
3	Shorter duration	$\Theta_i T_{new} \leq \Theta_i T_{old}$	If I agreed to sharing it for 5 years, I agree with sharing it for 3 years (assuming everything else in Θ_i is the same)
4	Reduced frequency	$\Theta_i \Phi_{new} \leq \Theta_i \Phi_{old}$	If I agreed to sharing it every minute, I agree with sharing it every hour (i.e., less often)

Table 1. Primary filters. If any rule is matched, a consent prompt is unnecessary.

4 When to Request Consent Again

In what follows, we propose a set of rules that act as filters, if at least one of them *is matched*, it means that consent *must not* be requested from the user again. Please refer to Tab. 1, where we denote previously accepted terms with K_{old}, and the new terms that the software wants the user to accept with K_{new}.

We can also apply additional filters, based on the privacy protections offered in different parts of the world (for an example, refer to Tab. 2). To this end, we propose the concept of a *privacy protection gradient*, which differentiates areas by level of privacy protection mechanisms in place.

In this hypothetical example (Fig. 3), we consider the EU and the European Economic Area (EEA) as the region with the highest level of protection, because the GDPR directly applies here. It is followed by a "second tier", which includes countries considered to provide an adequate level of data protection, per Art. 45 of the GDPR. As of this writing, the list includes Andorra, Argentina, Canada, Faroe Islands, Guernsey, Israel, Isle of Man, Japan, Jersey, New Zealand, Switzerland, Uruguay and South Korea. A hypothetical "third tier" could include countries or states that are said to have legislation comparable to the GDPR (e.g., Brazil with the Lei Geral de Proteçao de Dados, modeled after the GDPR [11], California and its Consumer Privacy Act [36], etc.), followed by the rest of the world, assumed to provide the weakest protections. Note that this is only a simplified model that enables us to reason about the "privacy gradient". Finding the optimal number of tiers and assigning each country to a tier is outside the scope of this paper.

We postulate that "moving up" along the gradient increases privacy, and thus can happen without re-requesting consent. In contrast, moving in the opposite direction would potentially weaken a user's privacy, hence such a transition would require consent to be obtained again.

In our formal notation, the level of protection applicable to a location Λ is written as Λ^{π}. Thus, if the old location of the data was in an area less secure than the new location, we express that as $\Lambda^{\pi}_{old} < \Lambda^{\pi}_{new}$.

Such secondary filters can be controversial. For example, there was an attempt to use the GDPR to silence journalists in Romania [24], therefore some

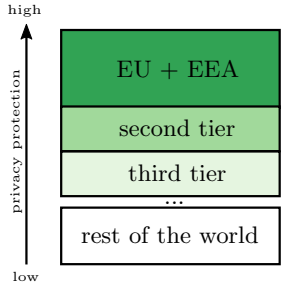

Fig. 3. Privacy protection levels in different political, economic or strategic unions.

Rule	Logic	Formal notation	Intuition
5	Go up or sideways on the "privacy gradient"	$\Theta_i \Lambda^{\pi}_{old} \leq \Theta_i \Lambda^{\pi}_{new}$	Moving from an area with fewer and weaker protections to an area with more and stronger protections, or to an area with comparable protections (assuming everything else in Θ_i is identical)

Table 2. Secondary filter, subject to discussion, can be deactivated by users.

users might rank the privacy protection levels of this EU member differently, while others would prefer to consider the EU as a single entity. A compromise solution might be to let users choose beforehand whether they want to treat such changes as major or minor ones (an example is shown in Fig. 6), or choose other criteria for computing Λ^{π}, such as the democracy index[2].

5 The Information Efficiency Metric

Since one of the ways in which users' privacy is undermined is through exposure to lengthy privacy policies that are not likely to be read [19], [26], one step towards improving the status quo is to reduce the volume of data users have to analyze when making decisions that can affect privacy. Therefore, we need a way to quantify this volume, in order to objectively compare different representations of privacy terms.

One way to accomplish this is by computing information efficiency, i.e., the ratio between "total" and "useful" information [28]. In what follows, we present an example calculation, using the notation proposed in Sec. 3.

Recall that each term of a privacy policy is a tuple expressed as $\Theta = (\Delta, \Pi, C, \Lambda, T, \Phi)$. For example, Λ represents one of the world's 193 countries[3]. Therefore, when specifying a country, we choose one of 193 discrete values, i.e., we produce $\lceil log_2 193 \rceil = 8$ bits of *useful* information.

[2] eiu.com/topic/democracy-index
[3] According the to UN un.org/en/member-states

Before this information can be communicated, we must *encode* it [32]. Assume we use an alphabet of 26 letters and that our text is case-insensitive, thus each letter is worth $\lceil log_2 26 \rceil = 5$ bits. Therefore, if we want to encode "Portugal", we need 8 letters, i.e., $8 \times 5 = 40$ bits. Now we calculate the efficiency of our encoding as $\eta = \frac{info_{useful}}{info_{total}} \times 100 = \frac{5}{40} \times 100 \approx 13\%$. This result can be roughly quadrupled by using the ISO 2-letter country code, "PT", instead of the full name. Thus, the ratio makes it obvious that one of the encodings incurs an overhead of circa 80%, prompting a search for better alternatives.

We then quantify the other elements of Θ, by relying on existing terminology that defines types of data, purposes of collection and retention periods [3], [4], [9], reaching a total of **155 bits**. The complete calculation is omitted for brevity, but is available in Appendix A.

We propose using this metric as a standard practice applied to rule out inefficient representations, because they are likely to lead to poor usability.

Although a high information efficiency is desired, we must consider metrics like the time and the mental effort necessary to interpret the message. For example, replacing country names with flags, or using icons to instead of text to represent data types will improve efficiency, but it might not work well with all users, or it could affect screen readers and automated translation software. Therefore, when reasoning about ways to represent privacy policies, information efficiency should be counter-balanced with a human-centered design process, taking aesthetics, and user satisfaction into account [10].

5.1 Table Benefits and Prose Deficiencies

While our calculations show that expressing privacy terms as a table is more efficient than as prose, we posit that tables may also have the *highest information efficiency* among options. This is due to the fact that tables omit "glue text", which improves the flow of prose, but also constitutes the bulk of the message.

In addition, a tabular layout for privacy terms comes with the following benefits. (1) It is easier to skim through because it is a fixed structure consisting of similar elements. In contrast, prose would have to be read entirely, otherwise users cannot be sure there is no abusive or unfair clause [5]. (2) Tables are easier to translate (even automatically), because they use predefined values, whereas prose is open to interpretation and can be confusing even to native speakers of the language [5], [26]. (3) Sorting, grouping and filtering works well with tables, but not with prose. (4) A table does not have to be processed entirely to be useful. For example, the number of rows can be a powerful signal when comparing something that shares data with 3 vs. 150 partners. (5) Tables pave the way for high permission granularity, where users can accept only specific terms while rejecting others. (6) Consequently, this makes possible the automated processing of terms, e.g., by means of trusted AI assistants that act on the user's behalf. (7) No extra training for users is necessary if the table is extended with new columns (e.g., a "condition" column could represent opt-in and opt-out logic, which is not reflected in the example in Fig. 1). Moreover, if a user does not need certain columns, they can hide them.

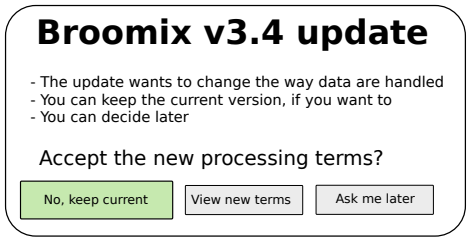

Fig. 4. Hypothetical interface where users indicate whether they want to update without changing the terms. Note the default option is the most conservative, and that there is no "accept terms" option, because the user needs to understand them before accepting, otherwise it would not be an *informed* consent. Clicking "view new terms" opens a "classic OnLITE" page where the current and new versions are compared side by side, with an option to highlight differences (refer to Fig. 2).

6 Additional Steps Towards Better Update Transparency

6.1 Distinguishing Feature, Security, and Privacy Updates

Sometimes updates can force users into a "take it or leave it" dilemma [31]. This creates an asymmetry in which vendors can force users into accepting new terms, because otherwise users will not get continued service or remain exposed to security risks. As others suggested, software can be designed in a way that decouples security updates from regular ones [34]. In the same fashion, we advocate the additional decoupling of updates that change the way personal data are handled. If such a level of granularity is achieved, consent forms can be shown less often, thus making it more likely that users will pay attention to one when they see it. In addition, this would mean that end users can exercise the rights enshrined in the GDPR, choosing not to accept the new terms (since consent must be voluntary) and thus continuing to use the software on previously accepted terms. In other words, the "I take it, but I keep the old terms" option becomes possible (as shown in Fig. 4), since we know exactly what terms were previously accepted.

6.2 The Best Time to Ask Permission

Another improvement in the way privacy updates are handled is to consider the best time[4] to display a consent prompt. Usually this happens when it is convenient for the software (e.g., at system boot, at program start-up or at regular intervals), without regard for the users' preferences. In these circumstances, a consent prompt is likely to interfere with a user's primary task, causing them to either accept the update in order to dismiss the prompt as quickly as possible, or postpone it. Either way, the damage is done - the user was interrupted.

[4] Here we mean it in the sense of the Greek word "kairos", which refers to an opportune moment, not to chronology.

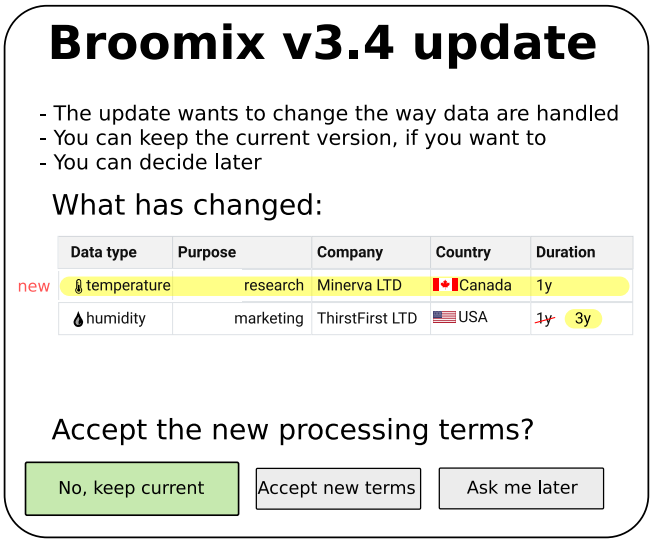

Fig. 5. Hypothetical update featuring an inline consent prompt.

Some operating systems let users decide when to apply updates. While this is done out of reliability considerations (the system must be plugged in, or there must be sufficient battery power left), it can also be done to avoid unnecessary distractions. The operating system could group updates based on their type, as discussed in Sec. 6.1, thus reducing potential interference with users' tasks. Alternatively, it can apply some heuristics to determine whether the user is actively involved in a task, and only display these non-disruptive prompts when the system is idle.

6.3 Inline Differences

We propose an "inline difference" prompt, which does not refer to the new terms in a separate window, but displays them in the prompt itself. This is only applicable when the number of differences[5] between the new and old terms is beneath a threshold. The sweet spot remains to be established experimentally, but a good default value could be Miller's "magic number 7±2" [20]. For example, in Fig. 5 you can see that only 2 differences exist between K_{old} and K_{new}, thus they can be displayed inline.

Depending on the user's preferences, a consent prompt may be shown only in a subset of cases. This can be configured in the interface (Fig. 6) or defined when an event occurs: the prompt is always shown the first time, and it contains a checkbox that says "ask me again whenever the data moves within the EU".

[5] $|K_{old} \Delta K_{new}|$, i.e., the cardinality of the symmetric difference between the old and new terms.

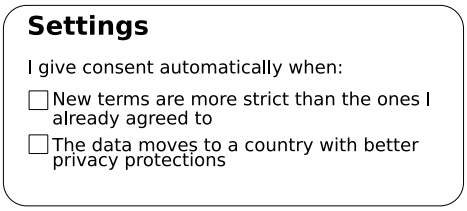

Fig. 6. Hypothetical interface where users indicate whether they want to give consent automatically in some cases.

7 Discussion

7.1 Reducing Information Asymmetry

Applying the measures outlined in this paper can reduce the information asymmetry between consumers and companies, making data processing practices more transparent and accessible to end users. This can enable users to make decisions based on criteria they may not have been aware of otherwise, and thus reward products that are more privacy-friendly. This, in turn, can incentivise vendors to become more transparent [21].

7.2 Benefits of a Formal Notation

Although the analysis of a privacy policy can be carried out by means of natural language processing and artificial intelligence (AI) tools, such approaches can have accuracy issues and are technically more complex [16], [17]. Moreover, even if human-level general intelligence were available, it is not unreasonable to assume that the AI will have to deal with ambiguities, contradictions or incomplete data, just like humans do when confronted with complex texts. It is also possible that vendors engaged in "opaque transparency" will explore adversarial approaches to deceive such software, akin to methods that trick a program into identifying a deer as an airplane by manipulating a specific pixel [29], [33].

We argue that this problem can be addressed in a simpler way - by mandating vendors to provide the data in a structured format. As we have shown earlier, this information would be easy for humans to comprehend [27], and it would also facilitate automated processing of such data using conventional means. Another benefit is that legal liability can be assigned to the vendor, leaving no wiggle room that would otherwise be created by potentially inaccurate interpretations generated by an AI. Other potential legal ramifications of applying the granular consent notation proposed in this paper will be discussed in our future work.

7.3 Information Efficiency

Another benefit of a formal notation is that it makes it possible to quantify the information efficiency of a representation of privacy terms. The metric is

easy to compute and can serve as an early indication of "opaque transparency". Although this method does not answer the question "*how* to do better?", it is still useful because (1) it tells us how well we are doing on a scale from 0 to 100, (2) it can be used to measure improvement during iterative prototyping, and (3) it can be used to objectively compare completely different designs.

7.4 Cross-Context Usage

A unified way of visualizing privacy terms is a major benefit, because end users can leverage their prior experience and apply it in other contexts [23]. For example, once a user familiarizes with the layout of a "who gets the data" table, they can recognize it in a smartphone application marketplace, on web-sites, on IoT devices, and other interfaces.

In such circumstances, one's ability to query the data set can become a general, rather than a specialized skill. This, in turn, can make users more perceptive to the subject of privacy and better equipped to reason about it.

7.5 Listing Non Personally Identifiable Information

Given that the proposed design grew out of IoT-centric research, Fig. 1 contains examples such as temperature or humidity, which do not constitute personal data, at least not without cross-correlating with other data sets. This information is presented for illustrative purposes, and ultimately it is a matter of policy or user preference, whether the table will display strictly personal data, or all collected data in general.

The benefit of listing all types of collected data is that a consumer can make a better judgment. For example, logging room temperature on an hourly basis is less sensitive than doing it every minute. In the latter case, the higher sampling rate can be used to infer whether the room is occupied or empty, how many persons are inside, and whether they sit, stand, or move around [22].

8 Related Work

Several works by Vaniea et al. analyse user behaviour in the context of updates. They found that sometimes prior experience determines users to intentionally ignore updates, in an attempt to avoid negative consequences, such as loss of functionality or undesired changes in the interface. They provide guidelines for improving the update experience through simple steps, such as explaining the changes the update brings or offering a rollback capability. They also advocate the separation of feature and security updates [34], [35]. In our paper, we apply some of these ideas to the context of update transparency. We describe a formal method and a UI design for effectively explaining how the changes in an update can influence a user's privacy. In addition, we argue in favour of decoupling privacy updates from other types of updates, with the purpose of reducing unnecessary interruptions.

We also consider relevant the literature related to summarizing privacy policies, because it is a more general form of the "what are the terms I have to accept?" problem users face when dealing with updates. So far this has been attempted through a combination of crowd-sourcing [30], machine learning, and neural networks [16], [17], [25].

Harkous et al. trained a neural network that analyzes, annotates and summarizes a policy, such that a user would not have to read it entirely. In addition, they provide a chat bot that answers questions about the policy in a natural language [16], [17]. While such a mode of interaction reduces the amount of information one has to read at once, a drawback is that some facts will not be revealed unless a user asks about them. Thus, *unknown unknowns* can only be found by stumbling upon them when reading the entire text, hence one cannot rely solely on a dialogue with the bot. Nokhbeh Zaeem et al. propose another automated tool for generating a concise summary of a policy and assign a privacy score to the product or service in question [25]. As in the case of the chat-bot, this approach reduces the volume of text a user has to read, but it is subject to the same limitations as other AI-based methods - a guarantee that the summary is 100% accurate is not provided, which also raises the question of legal liability. In contrast, we propose practical methods of reducing the total volume of text, rather than transforming it and showing a derivative form to the users. Further, the simplicity of our approach makes it immune to adversarial formulations that can trick an AI into misinterpreting a text.

Nevertheless, we believe that our works can complement each other. A chatbot and a summary screen will be more accurate when they rely on data structured like our "who gets the data" table (versus relying on free-form prose), while the issues of interpretation accuracy and legal liability are also resolved.

Breaux et al. propose a formal language for defining privacy terms. Their notation aims at helping requirements engineers and software developers detect potential contradictions in a policy, especially when the software relies on external services [6]. Their notation differs from the one we describe in this opinion paper in several ways: our proposal is GDPR-centric, hence we include some additional information, e.g., location of collected data. Further, our notation and the logic built upon it is aimed at a wider audience, not only developers.

9 Conclusion

We have described a series of measures that can improve the transparency of updates with respect to data collection practices. The measures rely on a simplified formal notation for privacy terms and heuristics that can be used to reduce the frequency of displaying update prompts. We argue how this approach can reduce habituation effects and we also provide an information efficiency metric that can be used to determine whether privacy terms (or the differences between terms brought by an update) can be expressed in a more concise form. By applying these measures, we believe that the information asymmetry between users and

companies can be reduced, putting users in a better position to make informed decisions with respect to their privacy.

Acknowledgments This research is a continuation of an activity that has originally received funding from the H2020 Marie Skłodowska-Curie EU project "Privacy&Us" under the grant agreement No 675730.

Appendix A Information Efficiency Calculation Example

We extend the material from Sec. 5 by providing another example. Consider the last term of the tuple, Φ, which represents the frequency with which data are sent. Suppose that in this case we express it as a choice among these options: {*multiple times per second, every second, every minute, hourly, daily, weekly, monthly, on-demand*}. Given that the set has 8 options to choose from, it means that a choice of a specific element yields $\lceil log_2 8 \rceil = 3$ bits of useful information.

Following the same principle, we quantify each component of a privacy term Θ, using terminology adapted from several sources: Platform for Privacy Preferences (P3P) [9], Data Privacy Vocabulary (DPV) [4], and Apple developer guidelines [3], summarized in Tab. 3. Note that different vocabularies provide a different level of granularity, for example, DPV distinguishes between 161 types of data, while P3P only 16. Since devising a vocabulary is outside the scope of this paper, we err on the safe side and take the maximum values (highlighted in bold) among the considered examples.

	DPV		**P3P**		**Apple**	
	items	bits	items	bits	items	bits
Data type	**161**	8	17	5	32	5
Purpose	**31**	5	16	4	6	3
Duration	-	-	**5**	3	-	-

Table 3. Summary of discrete choices to indicate the type of collected data, purpose of collection and retention period, using notation proposed by DPV, P3P and Apple developer guidelines.

After substituting each component, we get: $\Theta = 8{+}5{+}20{\times}6{+}8{+}11{+}3 = 155$ bits. Therefore, the pure information required to express a term is 155 bits, this is how much we would transmit, if we could upload it directly into the conscience of a person. However, some overhead is added because the information is encoded into words, or other forms that have to be perceived by end users.

We argue that the tabular representation is a highly efficient way of encoding privacy terms. This assertion is supported by the following calculation. Suppose that the notation consists of 26 small letters of the Latin alphabet, 10 digits, the `SPACE`, `TAB` and `NEWLINE` symbols. The notation has a total of 39 characters, which means that a single character is worth $\lceil log_2 39 \rceil = 6$ bits. In addition, the

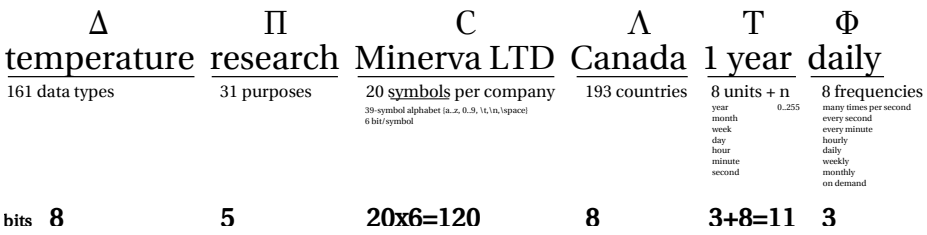

Fig. 7. Annotated calculations that explains how the amount of information in each privacy term is computed, yielding a total of 155 bits.

following conventions apply: a company name is assumed to be a string of 20 characters, thus it is worth up to $20 \times 6 = 120$ bits.

We now apply this encoding to Tab. 1, ignoring the data type icons and the country flags for simplicity. Each line is 49 characters long, yielding $49 \times 6 = 294$ bits. At this stage we can compute the efficiency of this representation: $\eta = \frac{info_{useful}}{info_{total}} \times 100 = \frac{155 \times 2}{294 \times 2} \times 100 \approx 53\%$.

Armed with this number, we can consider various ways to improve efficiency and measure their impact. For example, we can remove the country names and leave only their flags, or use two-letter ISO codes instead of full names. Entries can also be grouped, e.g., all terms related to temperature can skip the word "temperature" in all but the first entry. In addition, search and filter functionality can be used to hide all the rows except the ones the user wants to focus on, thus reducing the total amount of displayed information. With such an efficiency metric at hand, one can argue in favour of one design over another, supporting the choice with hard data.

In addition, we can use the same metric to compare entirely different notations. For example, consider this hypothetical prose version of the terms expressed in Fig. 1: *"We care about your privacy, therefore our smart indoor temperature and humidity meter only collects and shares your data with 2 companies. Temperature data are shared on a daily basis with Minerva LTD, located in Canada. The data are retained for a period of 1 year and are used for research purposes. Humidity is shared on an hourly basis with ThirstFirst LTD, and retained by them for 1 year, in the USA. Humidity data are used for marketing purposes"*. It is 453 characters long, and for the sake of simplicity let us assume that it also uses an alphabet of 39 symbols: 26 lower case Latin letters, 10 digits, space, comma, period. As in the previous case, each symbol is worth 6 bits, therefore $\eta = \frac{info_{useful}}{info_{total}} \times 100 = \frac{155 \times 2}{453 \times 6} \times 100 \approx 11\%$.

The prose version is clearly a step down from an efficiency of 53%! While we acknowledge that this synthetic version of a prose policy could have been shorter, such laconic policies are not the norm [18], [19], [26].

Appendix B When to Display Consent Prompts

The following pseudo-code illustrates the logic defined in Sec. 4 in action:

```
def is_consent_necessary():
    """Returns True if consent needs to be requested again,
    otherwise False"""
    for rule in rules:
        if rule matched:
            return False  # No need to ask for consent

    # if we got this far, re-asking for consent is required
    return True
```

A more granular approach enables us to tell whether a primary or a secondary filter matched, allowing more control (e.g. the GUI can display different prompts, depending on the magnitude of the difference):

```
def is_consent_necessary_granular():
    """Returns a tuple consisting of (necessary, reason),
    where necessary is True or False, while reason is
    one of {MAJOR, MINOR, NONE}."""
    for rule in primary_rules:
        if rule matched:
            # a primary rule was fired, no need to ask
            # consent again. E.g. some terms were removed
            # or made more strict
            return False, MAJOR

    for rule in secondary_rules:
        if rule matched:
            # a smaller change, we don't necessarily need
            # to ask consent again, but we might have to,
            # depending on the user's preferences. E.g.,
            # switch to another EU country, or moving up to
            # a "stronger privacy" place
            return False, MINOR

    # if we got this far, re-asking for consent is required
    return True, NONE
```

References

[1] B. B. Anderson et al. "How Polymorphic Warnings Reduce Habituation in the Brain: Insights from an fMRI Study". In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 2015.

[2] B. B. Anderson et al. "Users Aren't (Necessarily) Lazy: Using NeuroIS to Explain Habituation to Security Warnings". In: *Proceedings of the 35th International Conference on Information Systems*. 2014.

[3] *App privacy details on the App Store*. Apple Developer. URL: https://developer.apple.com/app-store/app-privacy-details/.
[4] B. Bos. *Data Privacy Vocabulary*. W3C Recommendation. W3C, 2019. URL: https://www.w3.org/ns/dpv.
[5] C. Bösch et al. "Tales from the Dark Side". In: *Proceedings on Privacy Enhancing Technologies* (2016).
[6] T. D. Breaux et al. "Eddy, a Formal Language for Specifying and Analyzing Data Flow Specifications for Conflicting Privacy Requirements". In: *Requirements Engineering* (2014).
[7] S. M. Casey. *Set Phasers on Stun: And Other True Tales of Design, Technology, and Human Error*. 1993.
[8] F. H. Cate. "The Limits of Notice and Choice". In: *IEEE Security & Privacy Magazine* (2010).
[9] L. F. Cranor. *Web Privacy with P3P*. 2002.
[10] I. DIS. *9241-210: 2010. Ergonomics of Human System Interaction-Part 210: Human-Centred Design for Interactive Systems*. 2009.
[11] A. Erickson. "Comparative Analysis of the EU's GDPR and Brazil's LGPD: Enforcement Challenges with the LGPD". In: *Brook. J. Int'l L.* (2018).
[12] European Parliament and Council of European Union. "Regulation 2016/679 of the European Parliament and of the Council". In: *Official Journal of the European Union* (2016).
[13] M. Fagan et al. "A Study of Users' Experiences and Beliefs About Software Update Messages". In: *Computers in Human Behavior* (2015).
[14] *GDPR Recital 58 - The Principle of Transparency*. URL: https://gdpr-info.eu/recitals/no-58/.
[15] U. Greveler et al. "Multimedia Content Identification Through Smart Meter Power Usage Profiles". In: *Proceedings of the International Conference on Information and Knowledge Engineering (IKE)*. 2012.
[16] H. Harkous et al. "Polisis: Automated Analysis and Presentation of Privacy Policies Using Deep Learning". In: *27th USENIX Security Symposium* (2018).
[17] H. Harkous et al. "PriBots: Conversational Privacy with Chatbots". In: *12th Symposium on Usable Privacy and Security* (2016).
[18] S. Human et al. "A Human-Centric Perspective on Digital Consenting: The Case of GAFAM". In: *Human Centred Intelligent Systems*. 2021.
[19] A. M. Mcdonald et al. "The Cost of Reading Privacy Policies". In: *Journal of Law and Policy for the Information Society* (2008).
[20] G. A. Miller. "The Magical Number 7 ± 2". In: *Psychological review*. 1956.
[21] P. Morgner et al. "Opinion: Security Lifetime Labels – Overcoming Information Asymmetry in Security of IoT Consumer Products". In: *Proceedings of the 11th ACM Conference on Security & Privacy in Wireless and Mobile Networks* (2018).
[22] P. Morgner et al. "Privacy Implications of Room Climate Data". In: *Computer Security – ESORICS*. 2017.

[23] J. Nielsen. *Jakob's Law of Internet User Experience*. URL: https://www.nngroup.com/videos/jakobs-law-internet-ux/.

[24] N. Nielsen. *EU warns Romania not to abuse GDPR against press*. EUobserver. 2018. URL: https://euobserver.com/justice/143356.

[25] R. Nokhbeh Zaeem et al. "PrivacyCheck v2: A Tool that Recaps Privacy Policies for You". In: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 2020.

[26] E. Okoyomon et al. "On The Ridiculousness of Notice and Consent: Contradictions in App Privacy Policies". In: *Workshop on Technology and Consumer Protection* (2019).

[27] A. Railean et al. "OnLITE: On-line Label for IoT Transparency Enhancement". In: *Proceedings of the 25th Nordic Conference on Secure IT Systems*. 2020.

[28] J. Raskin. *The Humane Interface: New Directions for Designing Interactive Systems*. 2011.

[29] A. Rosenfeld et al. "The Elephant in the Room". In: *arXiv:1808.03305 [cs]* (2018).

[30] N. Sadeh et al. "Towards Usable Privacy Policies: Semi-automatically Extracting Data Practices From Websites' Privacy Policies". In: *Poster Proceedings of the 10th Symposium On Usable Privacy and Security* (2014).

[31] F. Schaub et al. "A Design Space for Effective Privacy Notices". In: *Proceedings of the 11th Symposium On Usable Privacy and Security*. 2015.

[32] C. E. Shannon. "A Mathematical Theory of Communication". In: *Bell System Technical Journal* (1948).

[33] J. Su et al. "One Pixel Attack for Fooling Deep Neural Networks". In: *IEEE Transactions on Evolutionary Computation* (2019).

[34] K. Vaniea et al. "Betrayed by Updates". In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 2014.

[35] K. Vaniea et al. "Tales of Software Updates". In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 2016.

[36] W. G. G. Voss. "The CCPA and the GDPR Are Not the Same: Why You Should Understand Both". In: *CPI Antitrust Chronicle* (2021).