





OnLITE: On-line Label for IoT Transparency Enhancement

Alexandr Railean^(✉)  and Delphine Reinhardt 

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

Abstract. We present a privacy transparency tool, which helps non-expert consumers understand and compare how *Internet of Things* (IoT) devices handle data. The need for such tools arises with the growing number of IoT products and the privacy implications of their use. This research is further motivated by legal acts, such as the *General Data Protection Regulation* (GDPR), which mandates the communication of privacy practices in a clear language. Our solution summarizes key privacy facts and visualizes information flows in a way that facilitates quick assessments, even for large data sets. We followed an interdisciplinary iterative design process that combines input from legal and usability experts, as well as feedback from 15 participants of our think-aloud task analysis study. In addition to explaining the rationale behind the design and evaluation methodology, we compare our solution, implemented as a graphical user interface, with existing ones. The results show that participants consider the interface straightforward and useful. Our solution encourages them to think critically about privacy and question some of the manufacturers' claims. Participants also reported that they would be glad if such tools were widely available, to further improve privacy awareness. Besides, our solution can be a part of an evidence-based standardization process, enabling policy-makers to further promote privacy.

Keywords: Internet of Things · IoT · Privacy · Usability · GDPR

1 Introduction

The number of IoT devices, such as smart appliances, fitness trackers or surveillance cameras, has grown over the last decade [37]. While this brings economic benefits, it also comes with major privacy risks [40]. For example, it has been shown that in some circumstances, individuals can be deanonymized by correlating data sets [6, 27]. Another example is the analysis of smart-meter readings to identify media played on a TV [18]. Such privacy issues can be amplified by factors like device ubiquity, sensor diversity, data collection frequency, and the large volume of collected data [21, 22]. Moreover, the risks to privacy do not only target users of IoT devices, but also bystanders who are uninformed about the presence of such devices in their surroundings [1, 9, 23]. Another factor that contributes to loss of privacy is the lack of awareness about the technical

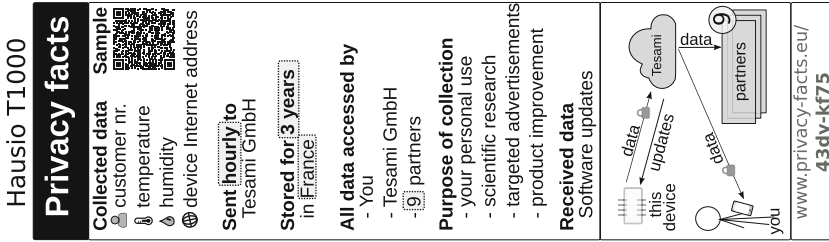


Fig. 1. LITE label for a hypothetical IoT device called “Hausio T1000” [32].

capabilities of IoT devices [23, 29, 33]. Besides that, users are skeptical of the ways algorithms can infer personal facts about them [39].

The GDPR aims to improve privacy, by requiring organizations that control personal data to explain how the data are handled “in a concise, transparent, intelligible and easily accessible form, using clear and plain language” [14]. The regulation creates a context in which privacy tools can gain more traction than in markets that lack enforcement or rely on self-regulation [7].

Despite the introduction of the GDPR, solutions to support IoT transparency have not been sufficiently researched yet. In addition to the legal requirements, demand for such solutions also comes from potential users, who explicitly expressed interest in transparency information or stated that it would influence their purchase decisions [10, 19, 23]. To address this need, several “privacy facts” labels have been proposed [10, 17, 31, 35], including our own “Label for IoT Transparency Enhancement”, LITE (Fig. 1, [32]).

LITE implements the GDPR transparency requirements to inform and help potential buyers protect their privacy, *before* deciding to acquire an IoT device. It provides answers to questions such as “what information is collected?” or “who gets the data?”. The answers are presented in a concise way, allowing IoT products to be compared side by side. The results of the usability study conducted in [32] show that participants could interpret the contents of LITE correctly and found it useful. However, they wanted extra details, that did not fit into the label due to size constraints.

In this paper, we present OnLITE, a *Graphical User Interface* (GUI) that extends LITE and addresses its shortcomings. Although LITE was the only user-validated GDPR-based label at the time we started this research, we also considered other designs (see Sect. 8.1, 8.2, 9). We follow ISO-9241, a human-centered, multi-disciplinary, iterative design approach when developing OnLITE. Compared to LITE, the new design shows more information and provides search, sort, and comparison features, as well as visualizations that distill large data sets into concise representations that can be reviewed at a glance. Its goal is to make the ways in which IoT devices handle data more transparent, informing users *before* and *after* the purchase (e.g. when updates are released). Our other contributions are the insights derived from the user validation of OnLITE, based on think-aloud task analysis with 15 participants. We also share evaluation scores

that can be used to compare OnLITE with similar interfaces. To foster replicability, we provide the source code of the prototype, our statistical calculations, and other supplementary materials at zenodo.org/record/4126346.

2 The Structure of LITE

The original label is divided into sections that provide information about collected data, destination and frequency of transmission, duration of storage, third-parties that access data, purpose of collection, and received data. The label also contains a “trace view” - a high-level graphical representation of the data flows [16], as well as a *quick-response* (QR) code with actual data samples.

This design has been revised to include a web address with a unique product number, which is also a part of the QR code payload. This change enables users to retrieve the digital version of the label, either by typing the address manually, or by using a specialized program that will scan and interpret the QR code.

3 Requirements and Design Space Analysis

The primary goal of OnLITE is to implement GDPR transparency by assisting consumers in making informed decisions when choosing IoT devices. It uses the same terminology and structure as LITE. Each element of the paper version, can be directly mapped to a section of OnLITE. The second goal is to enhance LITE with search and sort capabilities, and provide details that do not fit on the printed label. Our third goal is to facilitate comparisons, by showing labels side by side, and highlighting differences. This applies not only to different devices, but also to software updates of the same device, released after its purchase. Next, OnLITE must provide practical information to novices, even after brief use. We aim for a design that works on desktops and mobile devices. In addition, accessing OnLITE should take little effort once the physical label is at hand. We also strive for a generic design that can be applied outside of IoT (e.g. smartphone apps).

The information architecture of OnLITE is rooted in the GDPR and is centered around questions about data collection practices [32]:

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. What do the data look like?
7. How to access the data?
8. How often are the data sent?
9. Which communications are protected?
10. What paths do the data follow?
11. What does the device receive from other sources?

4 OnLITE Design

Based on our analysis, we propose the following design for OnLITE. For brevity, we do not describe the intermediate stages of the prototype, only the last iteration is presented. The interface consists of the following tabs:

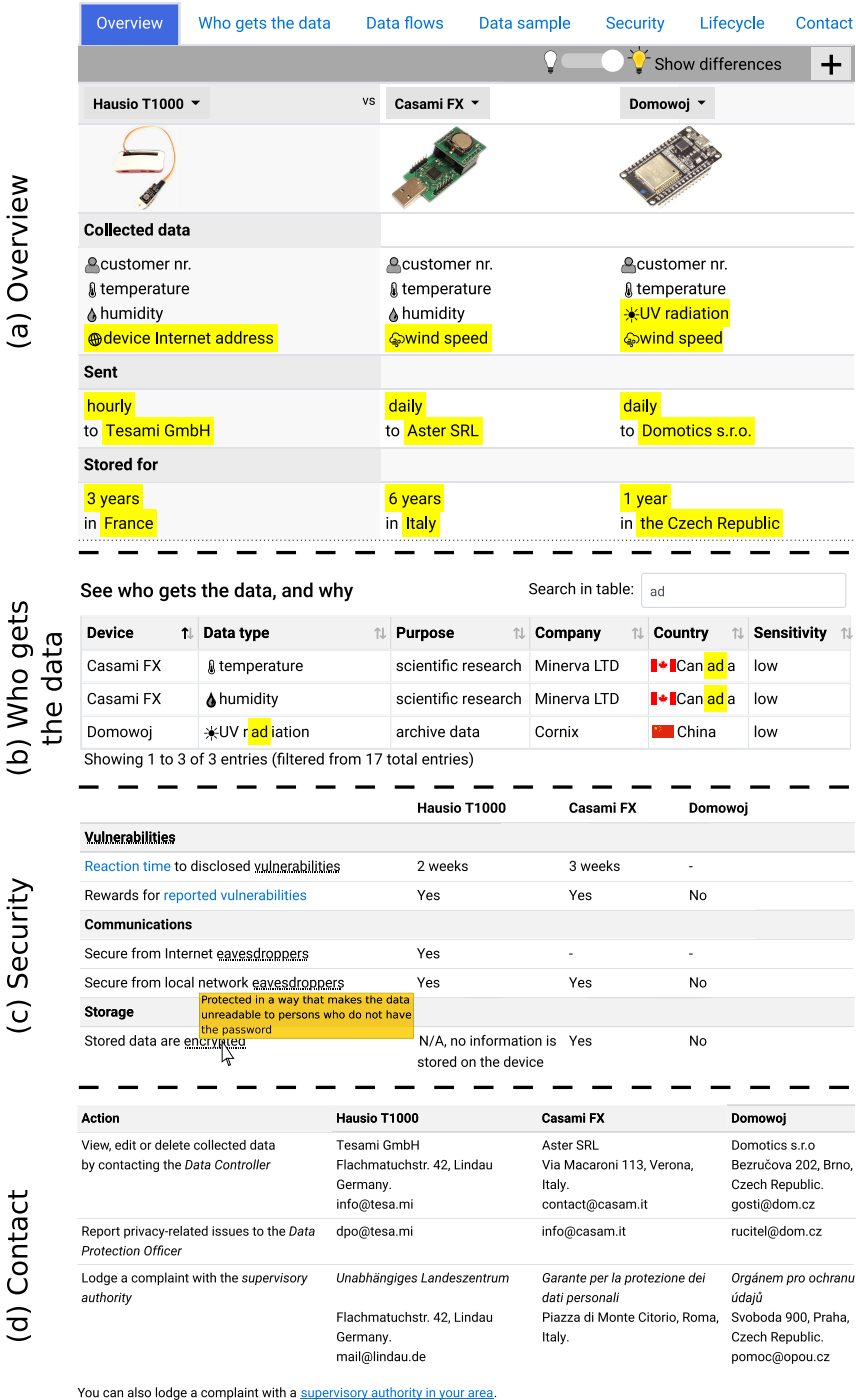


Fig. 2. Collage of screenshots of the tabs of OnLITE. The information is provided by vendors themselves, as the they are obliged to do so under the GDPR.

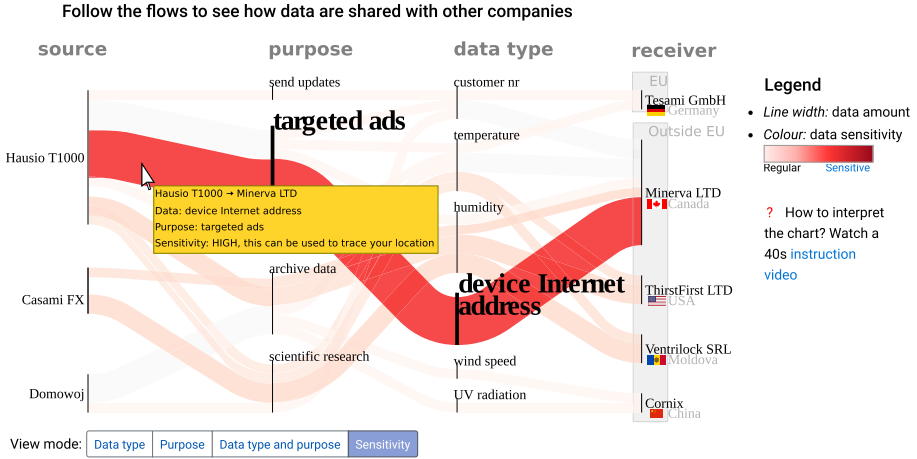


Fig. 3. The “Data flows” tab shows how data are shared with third parties. The “sensitivity” view highlights special categories of data defined by the GDPR.

Overview - the starting page provides the same information as LITE, plus a photo of the device. When several devices are compared, they are shown side by side, and optionally, the differences between devices can be highlighted (Fig. 2a).

Who gets the data - this tab contains a table with the columns: data type, purpose of collection, company, country, and sensitivity. When multiple devices are compared, a “device” column is added. The table can be sorted by each column. A search function is available, it highlights the matching text and only displays rows that contain the searched string, thus reducing the total amount of information shown on the screen.

Data flows are a graphical complement of the previous table, they facilitate a quick comparison of relative data flow sizes, making outliers more prominent. Flow widths are computed as $dataSize \times frequency$. This is a simplified model that is sufficient to test the interpretability of the image; devising a more elaborate formula is outside the scope of this paper. Several visualizations are available, each will group the flows in different ways. Colours are used to differentiate data types or devices, while the view shown in Fig. 3 offers a quantified measure of the sensitivity of each data transfer, highlighting special categories of data defined by the GDPR. The image features a legend and a link to a video that guides the user in interpreting the image. Theofanos et al. found that instruction videos are effective in helping users understand how to use a system [36]. We use Sankey diagrams [24] to distill multidimensional data into a compact view, give a sense of scale of the data flows and reveal the relationships between flow attributes (Fig. 3). Such diagrams can also be interpreted in grayscale.

Data sample - this tab shows actual samples of collected data, revealing aspects that would otherwise go unnoticed. For example, two devices can collect a “customer number”, however, one of them can use an email address, while the other could use a more privacy-preserving identifier, such as “481-AHR-1831”.

Features grouped by phases of the device lifetime: set-up → usage → maintenance → retiring

	Hausio T1000	Casami FX	Domowoj
Set up – <i>preparing the device for use</i>			
Unique factory-set password	Yes	Yes	No
Password change required before remote access for the first time	Yes	No	No
Use – <i>typical, daily interactions with the device</i>			
Multiple user accounts	Supported	Supported	No
Separate accounts for children	Supported	Supported	No
Separate account for guests	Supported	No	No
Maintenance – <i>procedures to increase the device longevity and ensure it works well</i>			
Automatic updates	Yes	Yes	No
Manual approval of updates	Optional	No	No
Update availability indication	In smartphone app	Mailing list	No
Feature update period	August 2020	March 2020	June 2020
Security update period	December 2023	March 2020	June 2020
Long-term support	January 2024	-	-
Retiring – <i>when the device is sold, sent for repairs, donated or thrown away</i>			
Secure data deletion (<u>wiping</u>)	Yes	No	No

Fig. 4. Comparing three IoT devices throughout the phases of their lifecycle.

Security - this tab presents security information (Fig. 2c). We have made sure to use common language. For example, “Secure from Internet eavesdroppers”, as opposed to specialized terms [34]. Low-level details, such as encryption algorithms or key lengths can be revealed by clicking on “More technical details”.

Lifecycle - this tab structures the attributes of the IoT device around the phases of its lifecycle: set up, use, maintenance, and retiring [33] (Fig. 4). For example, it informs consumers whether unique passwords are factory-set, what the duration of the support period is, or whether automatic updates are available.

Contact - according to the GDPR, a consumer has to be informed about several points of contact: the data controller, the *Data Protection Officer* (DPO), and the *Data Protection Authority* (DPA). This tab groups the contact details based on the action that prompted the need for contact: view, edit or delete data, report a privacy issue to the DPO, or lodge a complaint with the DPA (Fig. 2d). The structure is based on the feedback from a DPA representative, who stated that consumers often contact the DPA right away, expecting that appealing to the highest authority will address a problem faster. This creates unnecessary workload and causes delays, because a DPA can only step in if the DPO was contacted, but did not respond within a certain period of time.

4.1 Usability of Product Codes

These codes enable users to switch from the printed label to OnLITE. To make it a smooth transition, we use the Base58 character set, which excludes look-alike symbols, e.g., 00 I11, to avoid ambiguities. We split the code in two chunks, to make it easier to keep in short-term memory when writing down or sharing orally [26].

5 Prototype Implementation

We developed a web-based prototype, using standard graphical widgets such as tables, buttons or tabs, to ensure compatibility with accessibility tools and enable users to leverage their experience with GUIs. We refrain from using colour as the sole channel to convey a message, to ensure the interface preserves its efficacy even if viewed in grayscale. We use tables, such as in Fig. 2, as the main way of visualizing information, to make it easier to compare IoT devices side by side.

Non-specialized terms are preferred. When they cannot be avoided, tooltips provide extra details. Text is further simplified by avoiding paragraphs. The information consists of keywords grouped in tables; sentences are an exception, the longest one is 12 words long. While defining a dictionary of terms was outside the scope of our work, we encourage the reuse of terminology from projects such as P3P or SPECIAL [3, 7].

To further enhance accessibility, we leverage semantic HTML markup. Interactivity is used to indicate what parts of the interface are clickable, and highlight certain elements when the mouse is above them. The GUI is touch-friendly.

Progressive disclosure is used to show the most important information first. The start page offers a concise privacy facts summary, while exploring other parts of the GUI provides more details.

6 Evaluation Methodology

To test the readability, clarity, and usability of OnLITE, we first applied heuristic evaluation, reviewing early prototypes with usability and legal experts [28]. We presented various elements of the interface to 14 experts, of which 7 had repeated exposure to the complete UI. These sessions prompted us to shorten texts, replace specialized terms with general ones, add more information, and simplify the controls. For brevity, we omit ideas that did not make it into the final version, and the intermediate iterations.

We then conducted a task analysis study with 15 participants, who had to think aloud while carrying out tasks under the observation of a facilitator. The tasks are derived from the GDPR transparency questions listed in Sect. 3 and are aimed at evaluating whether the presented information can be interpreted correctly. After interviewing the first group of five people, the interface was revised and a new iteration was produced for the next group. We iterated until we reached the point of feedback saturation and no new insights were gained.

The incremental nature of the changes between versions means that participants using v2 were looking at a slightly evolved v1, and so on with v3 and v2. Thus, we regard this study as one with a sample of 15 (rather than 3 smaller ones with a sample of 5), which yields a minimum of 90% of usability issues found and a mean of 97% [15]. We further quantified the usability of the GUI using the *System Usability Scale* (SUS) [5], chosen due to its good performance at sample sizes ≥ 12 [38], and because scores of similar interfaces can be compared.

6.1 Experiment Settings

The experiment protocol was approved by our Ethics Committee. After signing an informed consent form, the participant is seated at a laptop equipped with a mouse, touchpad and trackpoint. The GUI is viewed in Firefox v66, running full-screen on a 13.3" 1366 × 768 display. We chose a laptop due to availability of tools for debugging and video recording, and because we could hide all toolbars and menus of the operating system, such that participants only see OnLITE. These instructions were given in written form, and then orally summarized, to set the focus on our UI as the primary interaction goal: The aim of this experiment is to evaluate an interface that provides privacy information about devices, enabling you to review their privacy practices and make informed decisions when choosing products. We ask you to analyze the privacy facts of several smart temperature and humidity meters using this interface. Please think aloud and comment your actions and decisions. Remember, that we are testing the interface, not you! There are no wrong actions or incorrect assumptions, do not worry about making mistakes or hurting our feelings, your "raw thoughts" are what we need. An assistant will help if you get stuck, but try to do everything on your own. The participant also gets three 128 mm × 40 mm privacy labels on A6 sheets, each corresponding to a device, as shown in Fig. 1. The labels are centered, such that if they stand side by side, there is spacing between them, as it would be in the case of real product boxes. Audio and screen recordings are made for later analysis. The facilitator sits next to the participant, and gives them a task from Table 1 at a time, observing and taking notes, reminding them to think aloud, if needed. After going through the tasks, the facilitator steps out so the participant can fill out a questionnaire that collects demographic data and includes a SUS form. When the participant is done, they call the facilitator and the evaluation proceeds to the last phase, where several open-ended questions are discussed.

Interviews lasted between 42 and 76 min, the median duration being 57 min.

6.2 Recruitment

We recruited 15 participants from a German language study group at the University of Kiel, Germany, offering an optional 10€ (USD 11) cash reward. The selection criteria were fluency in English and a minimum age of 18 years. The interviews were carried out between April and June 2019.

6.3 Demographics

Among our participants, 53% are male, 40% are female, 7% did not disclose their gender. 67% of the participants are between 27 and 35 years, followed by 18 and

Table 1. The tasks of the experiment. The entries A-F were given sequentially because they depend on one another. Tasks G-N were randomized, to avoid order effects. The entries O-V are open-ended questions that were asked at the end of the session.

Task	Description
A	Retrieve the privacy facts of the device <i>Hausio T1000</i>
B	Which partner companies get data collected by this device?
C	What partner company gets the largest amount of data?
D	Compare <i>Hausio T-1000</i> with the other two devices
E	Remove the device <i>Domowoj</i> from the comparison
F	Add it back to the comparison table
G	Which device shares data that might have the greatest impact on your privacy?
H	What data are used by partner companies for targeted ads?
I	Which device uses a form of customer numbers that protects the owners' identities better?
J	Which device can securely erase all the data before the owner gives the device away?
K	If you suspected that the device <i>Casami FX</i> was not protecting your data correctly, whom would you contact?
L	Which collected data is stored outside of the European Union?
M	Who provided the information about each of the devices?
N	In what way are these devices different?
O	Which tab gave you the best assistance in comparing these devices?
P	To what extent did the graphical data flows support you in comparing the devices?
Q	Which of the flow views you found most informative?
R	What conclusions do you draw from the "verified by an independent auditor" marker?
S	What other information or features, if any, would you like this interface to provide?
T	What parts of the interface were not clear to you?
U	Which of the shown devices is the best choice for the given task, in your opinion?
V	What other comments have you got about the system?

26 years (20%), the rest are between 36 and 44 years (13%). Their self-reported technical competence is computed using the method defined in [33]. In our sample, 60% are expert, 27% are intermediate, and 13% are novice (Table 2). The group is diverse in terms of academic fields, and includes economists, mathematicians, computer scientists, environmentalists, and lawyers. Our sample included participants from all of the continents except Australia and Antarctica.

Although we did not collect demographic details about our heuristic evaluators, their ages are between 30 and 65 years. Note that they belong to an older age category than the participants of our study. Since their age is not determi-

Table 2. Demographic data and results.

	Age	Sex	Skill	SUS score	Time (minutes)		
					Tasks	Interv.	Total
P1	27..35	F	expert	92.5	40	13	53
P2	27..35	M	expert	90	43	24	67
P3	18..26	F	expert	60	40	16	56
P4	27..35	F	interm.	67.5	42	15	57
P5	27..35	F	interm.	55	36	19	55
P6	36..44	M	expert	72.5	39	15	54
P7	18..26	M	novice	80	30	12	42
P8	27..35	F	interm.	37.5	42	18	60
P9	18..26	F	expert	65	39	25	64
P10	27..35	M	expert	70	49	11	60
P11	27..35	-	expert	77.5	55	21	76
P12	36..44	M	expert	67.5	27	26	53
P13	27..35	M	expert	65	47	12	59
P14	27..35	M	interm.	47.5	38	20	58
P15	27..35	M	novice	72.5	28	23	51

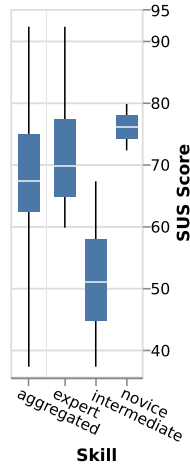


Fig. 5. SUS scores grouped by skill.

nant to their evaluation, we have applied the concept of data minimisation and hence not collected it.

6.4 Data Analysis

To understand the strengths and weaknesses of the prototype, we reviewed the screen recordings, observing the actions and comments of each sample of five participants. The interface was refined, and tested with the next sample.

The interviews were transcribed and processed through thematic analysis, to reveal common interaction patterns and themes [4]. We did not rely on several coders to independently encode transcripts, as the codes are only a step in the process of UI refinement, rather than the end product of our research [25].

7 Results

7.1 Qualitative

The qualitative feedback was used to refine the prototype and is therefore reflected in its latest iteration. We now share the highlights of thematic analysis.

Expectation of clickability was one of the main reasons for design changes. Participants clicked on static UI elements, expecting them to provide tooltips, e.g.: “I wanted it to show me the details of this line, but I cannot, I don’t know what is wrong <clicks on flows again>” (P3). The most common click targets were sections of the “Overview” tab and the graphical flows (Fig. 3). This prompted us to make these elements clickable to reduce friction and provide interactivity where users expect it.

Manual comparisons were another common pattern. Some participants counted how often each company occurs in a table, to understand which of

them gets the most data: “I counted ... the number of times they appear” (P4). It is more efficient to use the sorting feature, or rely on the graphical flows and look for the widest curve. Though the manual approach is effective, most participants prefer the more efficient methods once they discover them: “I think this one, <points to thickest flow> Minerva from Canada, because of the line width” (P5).

Time to understand how flows work was needed by many participants. They said it was not immediately clear how the graphical flows should be interpreted, and that it took them a while to grasp: “I needed more time to understand them” (P1), “The graphic is also just fine, I just needed a couple more seconds to understand the idea” (P2). In the subsequent prototype iterations, we added a 40s video that explained the logic behind the diagrams, as suggested by P5: “maybe a tutorial on how to interpret the charts of the data flow”. The video had a positive impact on user satisfaction and comprehension, e.g., “<watches video> ok, now it’s much more clear” (P15), and most participants watched it entirely, without being prompted to do so.

Flows are comprehensive and useful, as stated by many participants: “The data flow gives a lot of information as well, and it’s visual” (P6), “It’s visual, it has colors and it’s easy to use” (P11), “The faster way for me was looking at the data flow, it was more concise” (P12), “I think the graphical representation was really good for making a conclusion about the similarities and dissimilarities between the 3 devices” (P13), and “[flow] is really complete and very dense in information, not too dense” (P15).

Verified information about IoT devices is often referred to as a strong influence on a purchase decision: “it sounds more trustable if there is an independent verification, not just the vendor. They just want to convince you they have the best option, that is not necessarily the case” (P6), another participant said “I’ll choose the independently verified one, because things should be verified” (P7).

The authority void came up when we asked participants about an authority, whose independent verification of product information they would trust. Most referred to the government: “anything related to the government” (P6), “I will trust the EU” (P15); and failed to name a specific organization: “I don’t know, the international society of web developers, anything similar to that, the board of trust of... I don’t know” (P6).

The most useful tab is “Overview”, as indicated by most participants: “I could easily see the things written in each column and I saw that [show differences] switch” (P4), “definitely the first one, because it had this option to show differences” (P6), and “It gives information about what parameters are collected and also how long this info is stored. It is the most helpful. If you want more details, you go to other tabs” (P2).

Extra information mentioned by participants, when asked what else they would want to see in OnLITE: price (3 mentions), reviews (3 mentions). Each of the following was referred to twice: how many people bought the device, detailed technical specifications, more device photos and videos, device user guide, and the physical size of the device. P7 wished for telephone numbers, so they could

talk to a person in emergency cases. Others would say the interface is complete, for example: “To be honest, I don’t know, because it looks very complete” (P6), “I think the interface has a lot of information, I really couldn’t think of anything else to add” (P5), “I cannot think of any more to add to this” (P9).

The “Contact” tab is well-structured. Participants understood it and correctly identified the address they would have to write to when solving a particular type of problem: “I think it is this one, because it is just for reporting privacy related issues” (P3).

An educational opportunity arises when reasoning about an IoT device and drawing incorrect conclusions. For example, “I won’t be very stressed ... if the information about the temperature in my apartment ... would be read by someone else. I mean, what can they do? ... As long as they don’t have the key from my apartment, they can’t do anything, I think” (P2). In this case, privacy tools can provide tips like “temperature data can tell whether anyone is at home”, which might improve awareness about the privacy implications of sharing seemingly harmless data (e.g. yellow area in Fig. 3).

Data samples are useful, as shown by the participants’ ability to reason about different forms of customer numbers: “I think the first one is better, because it is just a sequence of numbers and letters” (P1), “The first one for sure!” (P6). This information prompted some participants to think of workarounds, such as “this could be resolved with an email address that is not important to you” (P2).

Privacy profiles are a personalized formula for computing a sensitivity score, which determines the colour of each data flow in the sensitivity view. Profiles can be created and shared by trusted authorities, or the users themselves. This idea was mentioned during heuristic evaluation and in the interviews: “maybe a multiple choice at the start ... where they can decide which kind of data is sensitive for them ... the data will be presented in that way” (P12). OnLITE determines sensitivity by referring to Art. 9 of the GDPR, which defines “special categories of data”, such as religious beliefs or sexual orientation. Note that the flow colours in Fig. 3 are not necessarily aligned with the GDPR, they were hand-tuned for experimental purposes, to see if the participants would notice the difference and how they would interpret it.

Critical thinking is an attitude that OnLITE helps foster, encouraging participants to reflect on the information shown to them. In some cases, they doubt that certain types of data are required for serving the declared purpose: “truth be told, I don’t understand why they need to store the device Internet address” (P2), or “why would a temperature measuring device have this feature? This, I don’t understand” (P11). In other cases, they would question the data retention period: “6 years, that’s a long time for such a small purpose, I can’t say it is reasonable” (P15). We consider this an important effect, as it guides participants towards questioning the status quo, as opposed to telling them what to believe.

7.2 Quantitative

The SUS results are given in Table 2 and Fig. 5. The mean score of OnLITE is 68, which matches the industry average for web interfaces [2]. Statistical analysis, by

means of a t-test¹, did not reveal any correlation between SUS scores and age or gender. Prototype iterations have no significant difference in scores either, which we attribute to the incremental nature of the changes between versions. We have not found significant differences between expert and non-expert participants' SUS scores. This suggests that the observed variations can be attributed to individual preferences rather than the level of technical skill. While the low power of the t-test with such a sample size cannot rule out differences between groups, it would have revealed major and obvious effects, if they existed.

All participants completed all the tasks, except P1, P3, P4 and P6, who failed task M. Note that the session durations in Table 2 are not an indication of invested effort, because we encouraged participants to explore alternatives and elicited additional feedback, even after a task was done.

8 Discussion

Our results show that participants can understand and use the presented information. The data also reveal a void when it comes to an authority that regulates such labels. All participants agreed they would trust a label that came from “the government” or “a reputable international organization”, however none gave a specific name. We believe the EU could be in a unique position to fill this gap, given that it is an international body, and that the GDPR is now in effect.

Sankey diagrams effectively visualize data sharing flows towards partner companies. They appealed to some of our participants and enabled them to make rapid judgments about which IoT device they prefer. However, some found them difficult to read at first. Thus, it is important to ensure that information is also conveyed in another form. Adding an instructional video that explains how the diagrams work had a positive impact on comprehension, and most participants watched the entire video without being nudged to do it. We believe that repeated exposure to OnLITE or the act of observing others reading the diagrams can further decrease the perceived effort.

“Overview” was chosen as the most informative tab by all participants, suggesting that it summarizes well the answers to the transparency questions in Sect. 3. We consider it a good choice for a starting page, as this way OnLITE conveys useful information to users, even if they do not explore other tabs.

Based on participants' positive feedback, we expected higher SUS scores. While this can be explained by two outliers who drove the score down (P8 and P14), it is also possible that OnLITE can be improved, or that a privacy-focused GUI is simply not appealing to users. They may not find the topic of privacy exciting, or the GUI could be perceived as a nuisance that stands in the way of using an IoT device that they are enthusiastic about. According to Bangor et al., the average SUS score varies depending on the type of system [2]. To the best of our knowledge, no SUS scores of similar transparency tools are available at the moment, so we cannot say with confidence whether or not “IoT transparency

¹ We chose this test because it is suitable for a sample size of 15, and because we have a normal distribution of scores, verified by means of a Shapiro-Wilk normality test.

tools” constitute a separate UI category with its specific average score. Sankey diagrams may be another reason why some scores were low. Even though the participants completed the tasks by finding answers in other tabs, we always insisted that they interpret the diagrams too. Thus, the diagram could have been seen as an “unnecessary effort”.

8.1 Avoiding Scores

Our design only conveys facts and avoids judgment. Instead of telling consumers “what is better”, we summarize information, so they can decide for themselves. This is inspired by the concept of *intelligence amplification*, where humans are assisted in various ways, yet remain central in the decision-making process [12]. While comparing device privacy ratings via scores is easy for consumers [11, 19], such grading schemes have limitations. (1) Privacy does not map to a linear scale, unlike measurable physical quantities. (2) There is no scoring method that all stakeholders agree with yet. (3) Transparency requires an understanding of the answers to the questions listed in Sect. 3. Some of that information is qualitative in nature and cannot be expressed numerically. (4) Scores can hinder adoption. It is possible that a substantial portion of current IoT devices would get a low privacy score, potentially prompting manufacturers to use their lobbying power to limit a label’s standardization. Thus, a gradual introduction of scores could be appropriate. While we have chosen not to use scores, we do not exclude doing so in the future, when the raised issues are addressed.

8.2 The Drawback of Sensor Lists

In contrast to Shen et al., who consider it “critical to enumerate all the sensors that are used by an IoT device” [35], we argue that a better approach is to show what information is *collected*, regardless of whether it was retrieved from sensors, inferred, or obtained through correlation with other data. Sensor lists can (1) obfuscate true intentions, while creating a false sense of security. For example, a device that is equipped with a camera and does *not* have a microphone can reasonably be considered as a “device that cannot record my voice”. However, it is possible to extract an audio signal from video [8], thus companies can claim compliance, while engaging in unethical practices. (2) Such lists take valuable space, potentially drawing attention away from other details. (3) Products can contain sensors that are only used internally (e.g., a thermometer is needed to prevent overheating), and listing them could confuse users. (4) Sometimes a sensor can be physically present, but remain unused (e.g., due to economies of scale, keeping it may be cheaper than making a product version without it).

8.3 Limitations

Our tests did not include participants above the age of 44 and we had few novice participants. Although we may have overlooked issues that could occur

with some groups, the interface is derived from a design that was evaluated with 31 participants of a wider range of ages and skills [32]. We also believe that heuristic evaluation further compensates this limitation, especially when most of the experts were at least in their forties. Another limitation is that we only tested the GUI on a laptop. We might have missed some issues that arise on touch-only devices with smaller screens. Finally, our evaluation did not explore what happens with repeated exposure to the GUI.

9 Related Work

Several designs were proposed to address IoT privacy and security issues. Some inherit the grid layout and the layered approach of [20]. A taxonomy proposed by [19] places privacy labels into one of three categories: *graded* labels that quantify security or privacy; *seals of approval* which show that a certification was attained, and *informational* labels that communicate facts about a device.

Van Diermen designed a graded and informational label for IoT, accompanied by an electronic interface [31]. The design is inspired by the EU energy efficiency label; it includes details about the support period, a list of processed data types and the available communication technologies, like Wi-Fi or Bluetooth. An extended version of the label provides information about security and the purpose of collection. However, this design has not been subjected to usability tests.

Shen et al. propose two informational labels for IoT [35]. Unlike in the case of LITE, more technical details are provided, e.g., a complete list of sensors and communication interfaces. This label employs a “traffic light” colour-scheme. For example, if encryption is not supported, the corresponding line will have a red marker. The design has not undergone a usability evaluation.

Grace et al. designed an informational privacy label and UI based on the GDPR. The details include a list of collected data, the purpose of collection, contact information and a list of rights that the user has. Although it has been user-validated by means of a focus group, it is not tailored for IoT devices [17].

Emami-Naeini et al. created a user-validated informational privacy and security label for IoT [10]. A difference is the use of scoring to quantify the level of privacy a device provides, while we have avoided using star ratings (see Sect. 8.1). Moreover, their design is not GDPR-centric, so it does not offer some specific information, like the location of the data, or the contact details of a DPA.

Bihl proposes a *trustmark for IoT*, a self-assessed, voluntary seal of approval [30]. Several regulators, e.g., Traficom (Finland) and the National Cyber Security Centre (UK) issue seals for IoT devices that meet a certain standard of security. The seals are derived from ETSI guidelines that dictate what security measures IoT devices should employ [13] (similar to the *security* tab of OnLITE). However, the seals do not convey privacy-related details, nor mandate the way this information ought to be visualized. Thus, they are not directly comparable to OnLITE.

10 Conclusions

We have proposed OnLITE, an on-line label for IoT transparency enhancement. The design has been examined through heuristic evaluation by legal and usability experts, and tested by 15 participants in a think-aloud task analysis study. The results indicate that the prototype conveys privacy facts in a way that can be understood by non-experts and experts alike. The participants find the interface useful, and are in favour of its wider availability. Our findings also suggest that the credibility of such a transparency tool could be higher, if it were regulated by governments or a reputable international organization.

Acknowledgments. We thank the participants of our study and our colleagues at the DPA of Schleswig-Holstein and USECON GmbH, as well as the open source contributors whose software we relied on. This research has received funding from the H2020 Marie Skłodowska-Curie EU project “Privacy&Us” under the grant agreement No 675730.

References

1. Aleisa, N., et al.: Privacy of the Internet of Things: a systematic literature review. In: International Conference on System Sciences (2017)
2. Bangor, A., et al.: An empirical evaluation of the system usability scale. *Int. J. HCI* **24**, 574–594 (2008)
3. Bos, B.: Data Privacy Vocabulary. W3C Recommendation, July 2019
4. Braun, V., et al.: Using thematic analysis in psychology. *Qual. Res. Psychol.* **3**, 77–101 (2006)
5. Brooke, J.: SUS - a quick and dirty usability scale. In: Usability Evaluation in Industry (1986)
6. Christin, D.: Privacy in mobile participatory sensing: current trends and future challenges. *J. Syst. Softw.* **116**, 57–68 (2016)
7. Cranor, L.F.: Necessary but not sufficient: standardized mechanisms for privacy notice and choice. *JTHTL* **10**, 36 (2012)
8. Davis, A., et al.: The visual microphone: passive recovery of sound from video. *ACM Trans. Graph.* **33**, 1–10 (2014)
9. De Cremer, D., et al.: The integrity challenge of the IoT. *J. Mark. Manag.* **33**, 145–158 (2017)
10. Emami-Naeini, P., et al.: Exploring how privacy and security factor into IoT device purchase behavior. In: CHI (2019)
11. Emami-Naeini, P., et al.: The influence of friends and experts on privacy decision making in IoT scenarios. In: ACM on HCI (2018)
12. Engelbart, D.: Augmenting human intellect. Technical report (1962)
13. ETSI: Cyber Security for Consumer IoT: Baseline Requirements. European Standard 303 645 (2020)
14. European Parliament and Council of European Union: Regulation 2016/679 of 27 April 2016. Official Journal of the European Union (2016)
15. Faulkner, L.: Beyond the five-user assumption. *Behav. Res. Methods Instrum. Comput.* **35**, 379–383 (2003)

16. Fischer-Hübner, S., Angulo, J., Karegar, F., Pulls, T.: Transparency, privacy and trust – technology for tracking and controlling my data disclosures: does this work? In: Habib, S.M.M., Vassileva, J., Mauw, S., Mühlhäuser, M. (eds.) IFIPTM 2016. IAICT, vol. 473, pp. 3–14. Springer, Cham (2016)
17. Fox, G., et al.: Communicating compliance: developing a GDPR privacy label. In: AMCIS (2018)
18. Greveler, U., et al.: Multimedia content identification through smart meter power usage profiles. In: IKE (2012)
19. Johnson, S., et al.: The impact of IoT security labelling on consumer product choice and willingness to pay. *PLoS ONE* **15**(1), e0227800 (2019)
20. Kelley, P.G., et al.: A nutrition label for privacy. In: SOUPS (2009)
21. Kosinski, M., et al.: Private traits and attributes are predictable from digital records of human behavior. *PNAS* **110**, 5802–5805 (2013)
22. Lane, N.D., et al.: On the feasibility of user de-anonymization from shared mobile sensor data. In: PhoneSense (2012)
23. Lau, J., et al.: Alexa, are you listening? In: ACM on HCI (2018)
24. Lupton, R., et al.: Hybrid sankey diagrams. In: Resources, Conservation and Recycling (2017)
25. Mcdonald, N., et al.: Reliability and inter-rater reliability in qualitative research. In: ACM on HCI (2019)
26. Miller, G.A.: The magical number 7 ± 2 . *Psychol. Rev.* **63**, 81 (1956)
27. Narayanan, A., et al.: How to Break Anonymity of the Netflix Prize Dataset. arXiv preprint (2006)
28. Nielsen, J., et al.: Heuristic evaluation of user interfaces. In: CHI (1990)
29. Page, X., et al.: The internet of what? *IMWUT* **2**, 1–22 (2018)
30. Bihl, P.: A trustmark for IoT. Technical report, ThingsCon (2017)
31. van Diermen, R.: A privacy label for IoT products. Ph.D. thesis (2018)
32. Railean, A., et al.: Let there be LITE. In: MobileHCI (2018)
33. Railean, A., et al.: Life-long privacy in the IoT? In: IFIP PIM (2017)
34. Schneier, B.: Click Here to Kill Everybody (2018)
35. Shen, Y., Vervier, P.-A.: IoT security and privacy labels. In: Naldi, M., Italiano, G.F., Rannenber, K., Medina, M., Bourka, A. (eds.) APF 2019. LNCS, vol. 11498, pp. 136–147. Springer, Cham (2019)
36. Theofanos, M., et al.: Usability testing of ten-print fingerprint capture. Technical report, National Institute of Standards and Technology (2007)
37. Trends-17. Technical report, Globalwebindex (2016)
38. Tullis, T.S., et al.: A comparison of questionnaires for assessing website usability. In Usability Professional Association Conference (2004)
39. Zheng, S., et al.: User perceptions of smart home IoT privacy. In: ACM on HCI (2018)
40. Ziegeldorf, J.H., et al.: Privacy in the IoT. In: Security and Communication Networks (2014)