



GEORG-AUGUST-UNIVERSITÄT  
GÖTTINGEN

doi:10.53846/goediss-9225

# IMPROVING IOT DEVICE TRANSPARENCY BY MEANS OF PRIVACY LABELS

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

“Doctor rerum naturalium”

der Georg-August-Universität Göttingen

im Promotionsprogramm Computer Science (PCS)  
der Georg-August University School of Science (GAUSS)

vorgelegt von

ALEXANDR RAILEAN

geb. in Orhei, Moldova

Göttingen, 2022



Institute of Computer Science  
Computer Security and Privacy

BETREUUNGSAUSSCHUSS

Prof. Dr.-Ing. Delphine Reinhardt, Institut für Informatik, Georg-August-Universität Göttingen

Prof. Dr. habil. Dr. h.c Simone Fischer-Hübner, Department of Computer Science, Karlstad University

MITGLIEDER DER PRÜFUNGSKOMMISSION

Prof. Dr. Florin Manea, Institut für Informatik, Georg-August-Universität Göttingen

Prof. Dr. Kerstin Strecker, Institut für Informatik, Georg-August-Universität Göttingen

Prof. Dr. Carsten Damm, Institut für Informatik, Georg-August-Universität Göttingen

Prof. Dr. Dieter Hogrefe, Institut für Informatik, Georg-August-Universität Göttingen

Tag der mündlichen Prüfung: 28. März 2022

## ABSTRACT

---

The Internet of Things (IoT) is an umbrella-term that applies to sensors, actuators and other devices that can interact with each other, or with other systems over the Internet. This technology has the potential to improve our quality of life, bringing more convenience, increasing the efficiency of existing systems, or creating new opportunities that did not exist in the past. The growth of IoT is catalyzed by advances in manufacturing techniques, which make it possible to pack more computing power into smaller devices, at a lower cost. This, in turn, accelerates the transition of IoT to the mass-market.

However, this trend has its downsides. As IoT devices grow in number and diversity, large volumes of data can end up under the control of companies that provide such products. The data can potentially be used to infer personal information about users, hence undermine their privacy. The problem is exacerbated by the improved connectivity of modern systems, which facilitates the quick distribution of data around the world, and complicates attempts to “put it back into Pandora’s box” once the data are out.

The General Data Protection Regulation (GDPR) introduces counter-measures to address these privacy issues. One of these measures is *transparency*, which requires that users understand how personal data are handled before they consent to sharing such information. However, the GDPR does not state exactly in what way companies should present this information to users, therefore our research aims to close this gap.

This dissertation takes a cross-disciplinary approach while tackling the problem of IoT transparency, and considers its usability, privacy and legal aspects. It proposes a “privacy facts” label for IoT product boxes, and an online interface that augments the label with search, sort and filtering capabilities. Both, the label and the interface are the result of a human-centered design approach. The thesis presents the rationale behind the design choices, the qualitative and quantitative methods we used to validate these designs with the participants of our studies, as well as the results of these evaluations.

## ZUSAMMENFASSUNG

---

Das Internet der Dinge (engl.: Internet of Things (IoT)) ist ein Oberbegriff für Sensoren, Aktoren und andere Geräte, die über das Internet untereinander oder mit anderen Systemen interagieren können. Diese Technologie hat das Potenzial, unsere Lebensqualität zu verbessern, mehr Komfort zu bieten, die Effizienz bestehender Systeme zu verbessern oder neue Möglichkeiten zu schaffen, die in der Vergangenheit nicht existierten. Fortschritte in der Fertigungstechnik beschleunigen das Wachstum des IoT, wodurch mehr Rechenleistung in kleineren Geräten und zu geringeren Kosten untergebracht werden können. Dadurch wiederum wird der Übergang des IoT in den Massenmarkt beschleunigt.

Dieser Trend hat jedoch auch Nachteile. Mit wachsender Anzahl und Diversität an IoT-Geräten könnte eine größere Menge an Daten von denjenigen Unternehmen kontrolliert werden, die solche Produkte anbieten. Die Daten können potentiell dazu verwendet werden, persönliche Informationen über Nutzerinnen und Nutzer abzuleiten und somit

deren Privatsphäre zu untergraben. Das Problem wird durch die verbesserte Konnektivität moderner Systeme verschärft, welche eine schnelle Verbreitung von Daten in der ganzen Welt ermöglicht. Sind die Daten erst einmal in der Welt verteilt, ist es schwer die Daten wieder "in die Büchse der Pandora zurückzudrängen".

Die Datenschutzgrundverordnung (DSGVO) führt Gegenmaßnahmen ein, um diese Datenschutzprobleme zu adressieren. Eine dieser Maßnahmen ist *Transparenz*, wodurch Nutzerinnen und Nutzer verstehen, wie personenbezogene Daten verarbeitet werden, bevor sie der Weitergabe dieser Informationen zustimmen. In der DSGVO ist jedoch nicht genau festgelegt, wie Unternehmen den Nutzerinnen und Nutzern diese Informationen präsentieren sollten. Unsere Forschung zielt darauf ab, diese Lücke zu schließen.

In dieser Dissertation wird das Problem der IoT-Transparenz mit einem interdisziplinären Ansatz untersucht, bei dem Aspekte der Benutzerfreundlichkeit, Privatsphäre und des rechtlichen Rahmens berücksichtigt werden. Wir schlagen ein Label mit "Privacy-Facts" für die Produktverpackung von IoT-Geräten sowie eine Online-Anwendung vor, die das Label mit Such-, Sortier- und Filterfunktionen ergänzt. Sowohl das Label als auch die Anwendung sind das Ergebnis eines menschenzentriertes Design-Ansatzes. In dieser Arbeit werden die Gründe für die Designentscheidungen, die qualitativen und quantitativen Methoden, welche wir zur Validierung dieser Designs mit den Teilnehmern unserer Studien verwendet haben, sowie die Ergebnisse dieser Evaluierungen vorgestellt.

## ACKNOWLEDGMENTS

---

I am thankful to everyone who contributed to my education at every stage of my life. Compiling a comprehensive list would be a tedious task, and I am certain that I would inadvertently fail to mention someone. I therefore refrain from giving names, because even seemingly tiny influences exerted decades ago by strangers altered my path in ways that brought me where I am today, so the list would be very long.

The rule above is broken on three occasions. A first exception is *Prof. Dr. Delphine Reinhardt*, my supervisor. Like Jupiter in our solar system, she significantly affected my trajectory and shielded me from global cataclysms. The other exception is my co-supervisor, *Prof. Dr. Simone Fischer-Hübner*, who was a great host when I was a visiting researcher at the University of Karlstad, and whose support helped shape this thesis into its current form. Last, but not least, I am thankful to *Harald Zwingelberg*, my supervisor from the Unabhängiges Landeszentrum für Datenschutz (ULD), who not only got me up to speed with the [GDPR](#), but also supported me during my first steps in Germany.

When it comes to collective and impersonal credits, I tip my hat to:

- Tax-payers of the European Union, whose contributions made it possible for this research to initially receive funding from the H2020 Marie Skłodowska-Curie EU project “Privacy&Us” under the grant agreement No 675730.
- The members of this thesis’ examination board.
- The open source contributors whose software I relied on when building prototypes, analyzing data and documenting results.
- Volunteers who participated in the surveys and interviews organized throughout my research, as well as disseminated information about them.
- Colleagues from ULD, the Data Protection Authority of Schleswig-Holstein.
- Everyone involved in the “Privacy&Us” project.
- Anonymous reviewers whose feedback helped refine the papers included in this thesis.
- The random user on Slashdot who mentioned “The design of everyday things” by Donald Norman in a comment for me to stumble upon, thus placing the concept of *usability* on my radar many moons ago.
- People who support the knowledge dissemination ecosystem that helps remove all barriers in the way of science.
- My family and friends.



# CONTENTS

---

1	INTRODUCTORY SUMMARY	1
1.1	Background	1
1.2	Thesis Structure	2
1.3	Motivation	2
1.4	Scope	4
1.4.1	Legal Context	4
1.4.2	Consumer-Oriented Products	5
1.4.3	Neutrality and Transparency	5
1.5	Research Objectives	6
1.5.1	Primary Objectives and Research Questions	6
1.5.2	Secondary Objectives	6
1.6	Research Methods	8
1.6.1	Survey	8
1.6.2	Heuristic Evaluation	9
1.6.3	Prototyping and Wizard of Oz	9
1.6.4	Think-Aloud Task Analysis	9
1.6.5	Interviews	10
1.6.6	Thematic Analysis	10
1.6.7	Statistical Analysis	11
1.6.8	Mathematical Modeling	11
1.7	Main Contributions	11
1.7.1	Answers to Research Questions	12
1.7.2	Replicability	13
1.8	Limitations	19
1.9	Discussion of Practical Considerations	19
1.9.1	Proposed API for Label Management and Usage	19
1.9.2	Widths of the Flows	20
1.9.3	Visualizing How Often Data Are Sent	21
1.9.4	Product Code Life-Cycle	22
1.10	Related Work	23
1.11	Conclusion	25
1.12	Future Steps	25
1.12.1	Retiring	25
1.12.2	Transparent Automated Processing	26
2	SUMMARY OF PUBLICATIONS	27
2.1	<a href="#">P1 Lifecycle</a> → Life-long Privacy in the IoT? Measuring Privacy Attitudes Throughout the Life-Cycle of IoT Devices	27
2.2	<a href="#">P2 LITE</a> → Let there be LITE: Design and Evaluation of a Label for IoT Transparency Enhancement	27
2.3	<a href="#">P3 OnLITE</a> → OnLITE: On-line Label for IoT Transparency Enhancement	28
2.4	<a href="#">P4 Updates</a> → Improving the Transparency of Privacy Terms Updates	28
2.5	Other Contributions	28

BIBLIOGRAPHY	30
3 P1 LIFECYCLE	31
3.1 Introduction	31
3.2 Related Work	32
3.3 Research Goals	33
3.4 Methodology	33
3.4.1 Distribution and audience	34
3.4.2 Self-selection bias	34
3.4.3 Priming concerns	34
3.5 Results	35
3.5.1 Pre-acquisition	35
3.5.2 Set up	37
3.5.3 Usage	38
3.5.4 Maintenance	39
3.5.5 Decommissioning	40
3.6 Testing the hypotheses	41
3.7 Discussion	42
3.7.1 Limitations	43
3.7.2 Recommendations for IoT vendors	43
3.8 Conclusions	44
3.9 Appendix: Questionnaire	44
BIBLIOGRAPHY	47
4 P2 LITE	47
4.1 Introduction	47
4.2 Requirements and Design Space Analysis	47
4.3 Label Design Methodology	49
4.4 Evaluation	50
4.4.1 Recruitment	50
4.4.2 Demographics	50
4.4.3 Experiment Settings	50
4.5 Results	51
4.6 Discussion	56
4.7 Conclusions	56
4.8 Acknowledgments	56
4.9 Appendix: Questionnaire	57
BIBLIOGRAPHY	59
5 P3 ONLITE	59
5.1 Introduction	59
5.2 The Structure of LITE	61
5.3 Requirements and Design Space Analysis	61
5.4 OnLITE Design	61
5.4.1 Usability of Product Codes	64
5.5 Prototype Implementation	65



5.6	Evaluation Methodology . . . . .	65
5.6.1	Experiment Settings . . . . .	66
5.6.2	Recruitment . . . . .	66
5.6.3	Demographics . . . . .	68
5.6.4	Data Analysis . . . . .	68
5.7	Results . . . . .	69
5.7.1	Qualitative . . . . .	69
5.7.2	Quantitative . . . . .	71
5.8	Discussion . . . . .	71
5.8.1	Avoiding Scores . . . . .	72
5.8.2	The Drawback of Sensor Lists . . . . .	72
5.8.3	Limitations . . . . .	73
5.9	Related Work . . . . .	73
5.10	Conclusions . . . . .	74
	<b>BIBLIOGRAPHY</b> . . . . .	75
6	<b>P4 UPDATES</b> . . . . .	75
6.1	Introduction . . . . .	75
6.2	Proposed Approach . . . . .	76
6.3	Formal Notation of Privacy Terms . . . . .	78
6.4	When to Request Consent Again . . . . .	79
6.5	The Information Efficiency Metric . . . . .	80
6.5.1	Table Benefits and Prose Deficiencies . . . . .	81
6.6	Additional Steps Towards Better Update Transparency . . . . .	82
6.6.1	Distinguishing Feature, Security, and Privacy Updates . . . . .	82
6.6.2	The Best Time to Ask Permission . . . . .	82
6.6.3	Inline Differences . . . . .	83
6.7	Discussion . . . . .	83
6.7.1	Reducing Information Asymmetry . . . . .	83
6.7.2	Benefits of a Formal Notation . . . . .	84
6.7.3	Information Efficiency . . . . .	84
6.7.4	Cross-Context Usage . . . . .	85
6.7.5	Listing Non Personally Identifiable Information . . . . .	85
6.8	Related Work . . . . .	85
6.9	Conclusion . . . . .	86
6.10	Appendix: Information Efficiency Calculation Example . . . . .	86
6.11	Appendix: When to Display Consent Prompts . . . . .	88

## ACRONYMS

---

API	Application Programming Interface
ASCII	American Standard Code for Information Interchange
DDoS	Distributed Denial of Service
DNS	Domain Name System
GDPR	General Data Protection Regulation
GUI	Graphical User Interface
HTTP	Hyper-Text Transfer Protocol
ICANN	Internet Corporation for Assigned Names and Numbers
IEEE	Institute of Electrical and Electronics Engineers
IoT	Internet of Things
ISO	International Organization for Standardization
SUS	System Usability Scale
URL	Uniform Resource Locator
UX	User Experience
W3C	World Wide Web Consortium
WCAG	Web Content Accessibility Guidelines

## INTRODUCTORY SUMMARY

---

### 1.1 BACKGROUND

Within the scope of this thesis, we define the Internet of Things (IoT) as *the sum of devices, sensors or actuators, that connect, communicate or transmit information with or between each other through the Internet* (adapted from [1]). Although the term was coined in 1985 [2]<sup>1</sup>, actual IoT systems precede the definition. An early well-documented example is a vending machine that was customized by Carnegie Mellon University students in 1982, to make it possible to remotely query the machine and find out how many drinks were available and whether they were cool or not [4, 5].

At that stage IoT was a tool for academia and technologists with access to expensive computing equipment. However, since then technology has progressed in terms of cost, computing power, connectivity, energy efficiency and miniaturization, therefore enabling IoT to make a transition to the mass market. Today IoT is a ubiquitous technology that has found its way into our clothing, personal gadgets, home appliances, vehicles, factories, cities and critical infrastructure [6, 7, 8, 9, 10, 11, 12].

At the time of this writing, the size of the IoT is estimated at 31.6 billion connected devices, growing from 14 billion in 2016. Due to economies of scale, this trend is expected to continue, as the average cost of a sensor fell from \$1.3 in 2004 to \$0.44 in 2018 and \$0.38 in 2020 [7].

Such a rapid growth of IoT leads to the accumulation of large quantities of data, some of which could be used to identify a person, especially if cross-correlation with other data sets is possible [13, 14, 15, 16]. Moreover, since some data originate from devices that operate in the immediate proximity of a person, e.g., their clothes or their homes, sensitive information about one's lifestyle and health can be directly obtained or inferred.

Thus, IoT growth can have major *privacy implications* [17, 18, 19, 20, 21, 22]. Furthermore, research has shown that privacy is of great concern to users when they reason about IoT devices, hence it has a strong influence on the acceptance of IoT [23, 24, 25]. Failure to address this concern can hinder the adoption of IoT, and consequently humanity could miss some of the benefits this technology offers. This is especially relevant in a context where accurate and timely data from cyber-physical systems can play a vital role in solving our climate crisis, as well as help reduce waste and increase energy efficiency [11, 26].

Legislative measures were taken across the world to address the privacy issues caused by information technologies, including IoT [27, 28, 29, 30]. Some notable examples are the General Data Protection Regulation (GDPR), California's Consumer Privacy Act, or the Brazilian Lei Geral de Proteção de Dados. Although these measures are in place, some service providers and device manufacturers take steps to deliberately confuse users and make it difficult to reason about privacy [31]. An example is the practice of "opaque transparency", where an interface is designed to hide relevant pieces of information

---

<sup>1</sup> Other sources date it to 1999 [3].

behind a series of counter-intuitive clicks [31]. Such a system can be considered transparent in theory, because the information that the law requires *can* be found, but it is opaque in practice, because few users will go to the lengths necessary to find it.

It is thus clear that the protection of users' privacy, whether in the context of IoT or in general, is not a problem that can be solved *solely* through legislative or technical means [32]. This thesis takes an interdisciplinary approach, where we consider the problem of privacy protection from a *legislative, technical* and *usability* perspective.

## 1.2 THESIS STRUCTURE

The first chapter of this document is an introductory summary that defines the context in which we conduct our research and provides background information necessary to understand the thesis. The premises that created the need for our research are discussed in Sec. 1.3. We define the scope and objectives of our work in Sec. 1.4 and Sec. 1.5.1. The research methods we employed are discussed in Sec. 1.6. Our main contributions are highlighted in Sec. 1.7. We go over the limitations of our experiments in Sec. 1.8, and discuss some practical considerations in Sec. 1.9, while a review of related work is presented in Sec. 1.10. We outline the conclusions that our research brought us to in Sec. 1.11, while the open questions that remain to be addressed in the future are listed in Sec. 1.12.

Chapter 2 summarizes each scientific paper included in this collection thesis. It also provides a list of co-authored contributions that are relevant to privacy research.

Each subsequent chapter represents a paper from the collection thesis. For convenience, we refer to each of them via a short mnemonic title that conveys the essence of the paper (e.g., "P1 Lifecycle"), rather than just a number like "P1". The mnemonics and the full titles they correspond to are given in Tab. 1.1 below:

Table 1.1: Paper title mnemonics used throughout this thesis.

<b>Mnemonic</b>	<b>Paper Title and Reference</b>
P1 Lifecycle	Life-long Privacy in the IoT? Measuring Privacy Attitudes Throughout the Life-Cycle of IoT Devices In: Proceedings of the 12th IFIP Summer School on Privacy and Identity Management – the Smart World Revolution (2017)
P2 LITE	Let there be LITE: Design and Evaluation of a Label for IoT Transparency Enhancement In: Proceedings of the 20th ACM International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI Adjunct, 2020)
P3 OnLITE	OnLITE: On-line Label for IoT Transparency Enhancement In: Proceedings of the 25th Nordic Conference on Secure IT Systems (NordSec, 2020)
P4 Updates	Improving the Transparency of Privacy Terms Updates In: Proceedings of the 9th Annual Privacy Forum (APF, 2021)

### 1.3 MOTIVATION

The research documented in this thesis is driven by issues of a social, technological, economic and political nature. In what follows, we explain how these elements are related to each other and to IoT.

On a *social* level, privacy can influence individual and collective behaviour. For example, a person might be treated unfairly by a service provider if non-transparent algorithms evaluated telemetry from their smart devices and decided they were in poor health [20, 33, 34]. These patterns can also manifest at larger scales, applying unfair judgments to entire groups. When such an imbalance persists, marginalized groups are motivated to change the status quo. However, it is not always certain that the changes will be optimal for society as a whole, or even for the marginalized group. For example, in a hypothetical world where IoT causes social imbalance, one solution would be to dismantle and eliminate this technology, and another is to improve it and address the shortcomings. In the latter case, the advantages that IoT brings are preserved, whereas in the former - they are lost, and thus society incurs an opportunity cost. We are aware of the shortcomings of IoT, and this thesis is our contribution towards finding a solution that enables us to retain the benefits and avoid the issues.

On a *technological* level, IoT is a *multiplier* - it can enhance the efficiency of existing processes, as well as open entirely new possibilities that were not viable in the past. For example, ubiquitous sensors can bring benefits such as predictive maintenance alerts [35, 36], reduced power consumption or higher yields in agriculture [11, 12]. However, the multiplier can also be smaller than 1, which would turn gains into losses. A notable example is the fact that IoT has become an integral part of the world's largest distributed denial of service attacks (DDoS) [37, 38]. Although this matter has more to do with security than with privacy, these topics are interdependent, therefore our research also takes security into account.

From an *economics* perspective, other industries can benefit from the "IoT multiplier effect" outlined above. However, there are some negative aspects as well, lost productivity being one of them. It has been quantified that in 2008 it would cost the US economy 781 billion US dollars if all users actually spent the time necessary to fully read every privacy policy they are facing on a yearly basis [39]. The fact that many users do not thoroughly examine each policy suggests that current approaches for displaying privacy terms are inappropriate, which can lead to frustration and apathy among users [40]. Considering that some of those policies are related to IoT devices and services, we understand that IoT also contributes to this problem. Therefore, there is a need for solutions that assist users in managing their privacy more efficiently. For this reason, although our research is IoT-centric, we take into account the possibility of reusing our findings in other contexts, like smartphone apps or web-sites. This way the improvements we bring can have an impact outside of IoT.

The effects listed above determined *policy-makers* to take action and create an environment that fosters privacy research like "Privacy&Us"<sup>2</sup>, where our work originates. It is thus clear that there is a strong commitment to improving the status quo, not only through legislative means, such as the General Data Protection Regulation (GDPR), but also through funding academia.

*Opportunity cost - the potential gains lost when one option is chosen over another.*

<sup>2</sup> This research received funding from the H2020 Marie Skłodowska-Curie EU project "Privacy&Us" under the grant agreement No 675730.

When we started this research, there were no available solutions that are rooted in the GDPR, and are validated by means of usability studies and focused on IoT. As discussed in Sec. 1.10, this still holds true at the time of this writing.

This thesis is our contribution to solving a concrete problem society faces when dealing with IoT privacy, namely usable transparency, i.e., making it easy for users to understand how a system handles their data.

## 1.4 SCOPE

### 1.4.1 Legal Context

*Ex-ante means “before the event”, i.e., before the user decides to share personal data.*

The scope of our research is *ex-ante* transparency, and it is further refined by the GDPR, which is our canonical source of requirements and terminology.

GDPR Art. 5(1) introduces *transparency* as one of the main principles relating to the processing of personal data, along with lawfulness and fairness. The principle is elaborated on in GDPR Art. 12 as follows: “the controller shall take appropriate measures to provide any information referred to in Articles 13 and 14 and any communication under Articles 15 to 22 and 34 relating to processing to the data subject in a *concise, transparent, intelligible and easily accessible form, using clear and plain language*, in particular for any information addressed specifically to a child”. Note, however, that the term “transparency” itself is not defined in this article.

*Recitals provide additional context information to make the rationale behind articles clearer.*

GDPR Recital 58 fills this gap: “the principle of transparency requires that any information addressed to the public or to the data subject be *concise, easily accessible and easy to understand, and that clear and plain language and, additionally, where appropriate, visualisation be used*. [...] This is of particular relevance in situations where the proliferation of actors and the technological complexity of practice make it difficult for the data subject to know and understand whether, by whom and for what purpose personal data relating to him or her are being collected, such as in the case of online advertising” [41].

Note that Art. 12 references Art. 13, 14, 15, 22 and 34, therefore one can only build a complete picture after processing all these entries and the ones they may refer to. We consider Art. 22 out of scope, as it refers to automated processing of data, which we set aside for future work (see Sec. 1.12). We also omit Art. 15<sup>3</sup> and 34, since they are related to ex-post transparency, i.e., what happens after consent for processing data is given. For brevity, we do not go through each of the remaining articles, instead we distill them into the following list<sup>4</sup> of *transparency questions* that users need answers to:

Q<sub>what</sub> *What data are collected?*

Q<sub>who</sub> *Who has access to the data?*

Q<sub>purpose</sub> *What is the purpose of collection?*

Q<sub>access</sub> *How do I view, edit or delete the data?*

Q<sub>where</sub> *Where are the data stored?*

Q<sub>duration</sub> *How long are they kept?*

Q<sub>complaint</sub> *How can I file a complaint?*

<sup>3</sup> Information referenced in GDPR Art. 15 matches the information we present ex-ante, therefore the same designs that we discuss in Ch. 4 and 5 can be used for this purpose without modification.

<sup>4</sup> The list corresponds primarily to legal requirements given in Art. 13, regarding information that has to be presented to data subjects.

Thus, in the context of this thesis we define “transparency” as *the capability to provide answers to the questions above*. Note that making the information available is necessary, *but not sufficient*, because it can be presented in forms that are not easy to understand, or are misleading. Although this can be a consequence of poor design and neglect, sometimes this is done deliberately [31]. Therefore, we aim for “usable transparency”, which takes into account the usability factors outlined in ISO 9241-210, i.e., the efficacy, efficiency of the process of finding these answers, as well as the users’ satisfaction with it [42].

#### 1.4.2 Consumer-Oriented Products

Although the IoT covers a very wide range of systems, starting with personal devices and ending with large-scale deployments in factories or cities [8, 35, 43], our research is focused on consumer-oriented devices. These include personal items such as fitness trackers or smartwatches, as well as household appliances like voice-activated assistants or smart cameras.

As we argued in the previous section, the aim for usable transparency implies the existence of a user, i.e., someone facing the decision of whether to use an IoT device or not. The focus on consumer-oriented products ensures that we operate in a context with a diverse range of IoT devices, and an abundance of potential participants for usability evaluations. Such evaluations are an indispensable part of human-centered design, as discussed in ISO 9241-210 [42].

In addition to that, we identified a gap in the available literature for this market segment when we began our research, which served as further motivation for our work. More details about how this thesis differs from related work are given in Sec. 1.10.

#### 1.4.3 Neutrality and Transparency

Last, but not least, the scope is limited to *informing* users about the way an IoT device handles their data, without attempting to steer them towards specific products, and without tagging products as “good” or “bad”. Due to personality variations, each user has their own needs, goals and expectations, therefore what some perceive as an unacceptable privacy risk, others might find well within their comfort zone [44]. We believe that such a neutral stance makes our proposals more appealing to a wider range of stakeholders, regardless of their privacy views or their position on the political spectrum.

Therefore, despite our own beliefs and general mission of privacy advocacy, we do not consider as failure a buyer’s choice to acquire a device that is rather privacy invasive, as long as they were fully aware of the impact of their decision.

## 1.5 RESEARCH OBJECTIVES

### 1.5.1 Primary Objectives and Research Questions

In this thesis we set out to improve the status quo for usable transparency in IoT. To this end, we have formulated the following research questions:

RQ1 To what extent are users aware of IoT privacy issues?

This is our starting point, where we understand the scope and the magnitude of

Table 1.2: The papers included in this thesis and the research questions they target. Note that for brevity, in this thesis we refer to the papers through shorthand mnemonics defined in Tab. 1.1.

Paper	RQ1	RQ2	RQ3	RQ4
P1 Lifecycle	●			
P2 LITE		●	●	
P3 OnLITE		●	●	
P4 Updates		●	●	●

the problem. We then use the gathered data to draw initial conclusions and plan our next steps. At this stage the idea of a “privacy facts” label for IoT products begins to take shape.

RQ2 *What information* should be presented to users on a label to achieve transparency? Here we refer specifically to transparency, as envisioned by the GDPR. On the one hand, the GDPR is our source of requirements, but on the other - we have to balance it with other constraints, such as limited space on a printed label.

RQ3 *In what way* should this information be presented, such that transparency is usable? This applies on multiple levels: what terminology to use, what layout is more appropriate, which methods of visualization are most effective when it comes to displaying large volumes of data, etc. We also have to ensure that the solution is usable by a wide audience comprised of non-experts.

RQ4 What other *technical and regulatory* means can improve privacy protection? Considering that privacy matters not only at the time an IoT device is acquired, we look into other stages of the IoT device lifecycle, namely the update process.

Tab. 1.2 shows the relationships between the research questions and the papers included in this thesis.

### 1.5.2 Secondary Objectives

We define several additional objectives with the aim of increasing the relevance of this work: foster generativity and reusability, and ensure accessibility. In what follows, we explain what these objectives mean and the rationale behind them. The mapping between each of these objectives and each included paper is summarized in Tab. 1.3.

#### 1.5.2.1 Generativity

This term was proposed in 2006 by Zittrain, it adds a dimension to the way we evaluate a system. Besides describing it with attributes such as “scalable”, or “usable”, we add a new one: *generative* - “the capacity to grow and acquire new capacities based on user-generated contributions” [45].

For example, a typewriter and a computer running a word processor both solve the problem of typing text. However, the computer can be repurposed by users in different ways by means of custom-written software, making it more generative.



Another example is the layered architecture of a network stack. Instead of designing a monolithic system aimed for solving a specific problem, e.g., voice communications, a network can be designed to solve the generic problem of transporting bytes from point A to point B. If documentation is available and users can add their own logic on top of this foundation, they can implement scenarios that were not originally envisioned, e.g., multi-player games, buying and selling shares on the stock-market, online voting, video streaming, online courses, etc.

Zittrain further defines *generativity* as “a technology’s overall capacity to produce unprompted change driven by large, varied and uncoordinated audiences” [45]. We therefore strive towards increasing the generativity of the transparency tools proposed in our research, in order to foster their adoption. Chapters 5 and 6 cover this topic in more detail.

Another argument in favor of aiming for generativity is that the European Interoperability Framework also advocates for this, e.g., it recommends the release of “machine-readable data for use by others to stimulate transparency, fair competition, innovation and a data-driven economy” [46].

### 1.5.2.2 *Reusability*

Although the scope of our work is to solve the transparency problem for consumer-oriented IoT devices, we observe that users are confronted with similar problems in contexts other than IoT. For example, this also happens when users install new apps on their smartphones or create new accounts on online-services. The transparency questions discussed in Sec. 1.4.1 arise throughout all of these interactions. Therefore, users will always ask themselves “what data are collected?” or “for what purpose are they collected?”, regardless of whether they are setting up a device in their smart home or signing up on a web-site.

Taking into account the best practices of usability research, we know that a consistent interface is better for users, because once they familiarize with it, the skill can be reused in other contexts [47]. Therefore, we postulate that a consistent privacy transparency interface is a highly-desired feature, and that it could be applied in scenarios other than IoT (see Sec. 6.7.4).

### 1.5.2.3 *Accessibility*

Accessibility is the practice of designing interfaces that can be used by as many people as possible, such that there are no barriers that hinder any group. For example, an accessible system can be used by the visually impaired or by people with hearing loss, because few or no assumptions were made about the users’ visual or hearing acuity.

The importance of accessibility is emphasized by the fact that it is firmly established in legislation. For example, in the United States, Section 508 of the Rehabilitation Act Amendments of 1998 requires that “technology is accessible to employees and members of the public with disabilities to the extent it does not pose an “undue burden”” [48]. Meanwhile, European directive 2102/2016 states that “Member States shall ensure that public sector bodies take the necessary measures to make their websites and mobile applications more accessible by making them perceivable, operable, understandable and robust” [49].

Table 1.3: Link between the included papers and the secondary objectives defined in Sec. 1.5.2. Note that no entries are given for *P1 Lifecycle*, as it was an exploratory study which eventually lead to these objectives.

Paper	Generativity	Reusability	Accessibility
<i>P1 Lifecycle</i>			
<i>P2 LITE</i>		●	●
<i>P3 OnLITE</i>	●	●	●
<i>P4 Updates</i>	●	●	●

We therefore choose to adhere to these requirements, because we aim for producing transparency solutions that could potentially become a part of a regulation, and thus affect a very large group of people. Although none of the usability tests we conducted were specifically designed to evaluate accessibility, we always follow best practices, such as those outlined in the Web-Content Accessibility Guidelines (WCAG) defined by W3C [50]. More details about the steps we took towards this goal are given in Sec. 4.3, and Sec. 5.5.

## 1.6 RESEARCH METHODS

In this section we describe the methods that were applied throughout our research and explain why they were appropriate for our purposes. The choice of methods is primarily rooted in the human-centered approach to design outlined in ISO 9241-210 [42], thus most of them require some form of personal interaction. Tab. 1.4 at the end of this section summarizes this information.

### 1.6.1 Survey

A survey collects self-reported information about a person’s beliefs, attitudes, perceptions and behaviours. Throughout our research we have conducted both, online and of-line surveys. We have paid attention to wording, sequencing, response options, length and layout, in order to minimize bias [51, 52].

Our first paper was based on an online survey with 110 participants. It gave us a general picture of IoT privacy awareness among the respondents and made it clear that end-users are not well-informed about the technical capabilities of their IoT devices.

We chose to use this method because it provided an inexpensive way to collect feedback from multiple participants in parallel and automatically, by means of a web-site. The studies we conducted subsequently incorporated offline surveys, which were aimed at obtaining more qualitative data from the participants.

We always share the source code of each questionnaire we administered, to facilitate replicability.

### 1.6.2 Heuristic Evaluation

This method was proposed by Nielsen and Molich as a quick and affordable way to find usability issues in a product by getting experts involved in the design process [53]. This

way we can improve a prototype in the early stages of its development, without running a large-scale test which would take more time and resources.

The diversity of backgrounds in our extended research group at “Privacy&Us” ensured that we always have access to legal, user experience (UX) and technical experts. Thus, heuristic evaluation was always a natural choice for us.

### 1.6.3 *Prototyping and Wizard of Oz*

After analyzing the results of *P1 Lifecycle*, we decided to focus our efforts on early-stage intervention. That is, find the earliest point in time when a person can be empowered to make an informed choice of an IoT device. This, in turn, brought us to tangible artifacts: a “privacy facts” label for IoT product boxes, and eventually - a digital interface that accompanies the label.

To this end, we built a series of prototypes of different levels of fidelity, starting with sketches on paper, continuing with printed paper elements which could be moved and recombined arbitrarily, and ending with an interactive implementation that runs on a computer.

We have used the “Wizard of Oz” technique in the very early stages of prototyping to simulate interactivity [54, 55]. The essence of this method is that someone is manipulating the artefact behind the scenes, giving users the illusion that they are facing an actual system, rather than an inert prototype. Therefore, it is a low-cost technique that enables us to iterate rapidly and rule out unpromising ideas.

In subsequent iterations, when the design has reached the next level of maturity, it was quicker and easier to implement the logic of the digital prototype in computer code, thus the interactions were genuine. This also simplified the experiment logistics, since there was no need to have another person that would play the role of the “wizard”.

### 1.6.4 *Think-Aloud Task Analysis*

This technique implies breaking down a process into individual steps and observing how participants go through them [52]. It is a good match for our research, due to the nature of the problem our prototypes aim to solve. As we have discussed earlier in Sec. 1.4.1, we consider a system possesses the attribute of “usable transparency” if users can easily find answers to specific questions such as “what data are collected?”, or “who gets the data?”. Thus, the prototypes are structured such that these answers can be obtained by following a certain workflow. Through task analysis we observe whether users deviate from the flow, and if so - in what way. This gives us valuable information for considering alternative flows and understanding the weaknesses of the implementation at hand.

We chose to follow a think-aloud protocol for the task analyses conducted in our research. This gives us a clearer picture of what the participant is thinking about and how they perceive our interface. This is especially relevant when conducting exploratory research, because participants can sometimes produce ideas and interpretations that we have not thought of ourselves. Furthermore, Ericsson et al. found that “people are reasonably able to speak about and complete a task at the same time without impacting the outcome”, therefore this approach yields valuable data in addition to what a silent task analysis would offer, at no penalty [56].

In addition, we make screen recordings of these sessions when it is technically feasible. This enables us to review the interaction at a later time and observe subtle details that were not obvious during the live sessions. Another benefit is that the facilitator of the experiment does not have to scribble notes continuously, which could make participants more aware of the fact that their behaviour is being observed.

#### 1.6.5 Interviews

This research method is a conversation during which we ask either scripted or unscripted questions, in order to learn more about the participants' opinions and the rationale behind them. Unlike text-based questionnaires, interviews enable us to observe the participants' facial expressions and body language, which provides additional information for evaluating one's level of satisfaction with an interface.

We have conducted in-person interviews when validating our prototypes. We first go through a *structured* phase, during which we ask the same questions, and we do so in the same order, to maintain consistency across sessions. We then proceed to an *unstructured* phase, during which we adjust our questions to the individuality and background of each participant. This makes the interview a good exploratory tool, because we can elicit feedback and collect ideas we have not thought of ourselves, thus guiding our research towards its future iterations.

In our work we have always combined interviews with other methods, such as task analysis or questionnaires.

#### 1.6.6 Thematic Analysis

When qualitative data was collected from our participants, usually in the form of unstructured text or interview transcripts, we applied this method to annotate the data and find common themes brought up by the participants [57, 58]. This enabled us to find shortcomings in our prototypes, e.g., a visualization that was not easy to understand, a button that was barely visible, or terminology that was not clear. This information was then used to iterate our prototype designs and observe if the problems persisted.

In addition, this analysis revealed new issues and trends that we were unaware of initially, thus guiding us towards the next steps of our research and helping us define our priorities.

#### 1.6.7 Statistical Analysis

We rely on statistical analysis to check the soundness of our data, and look for correlations between the attributes of our prototypes and demographic attributes of our participants. Thus, we can make assertions that were found to be statistically significant and are supported by data, as opposed to merely stating opinions.

Besides that, statistics is at the foundation of other methods we relied on, and it plays a key role in choosing sample sizes, for example [59]. Thus, even though we merely followed well-established usability research guidelines, we are aware of the influence of statistics on defining these guidelines.

Table 1.4: Research methods mapped to research questions and the papers included in this thesis.

Research method	P1 Lifecycle	P2 LITE	P3 OnLITE	P4 Updates
	RQ1	RQ2,3	RQ2,3	RQ2,3,4
Survey	●	●	●	
Heuristic evaluation		●	●	
Prototyping		●	●	●
Wizard of Oz			●	
Think-aloud task analysis			●	
Interview		●	●	
Thematic analysis	●	●	●	
Statistical analysis	●	●	●	
Mathematical modeling				●

### 1.6.8 Mathematical Modeling

We devised a mathematical model in *P4 Updates* (see Ch. 6), to quantify and compare various forms of visualizing privacy terms (e.g., prose, tables). To this end, we proposed a formal notation for such terms and explained how it can be used to compute the *information efficiency* of a representation.

This metric is a quick and inexpensive way of comparing different privacy term representations. An efficiency score is computed by plugging numbers into an equation, the result can then be used to rank various candidates, along with other metrics.

## 1.7 MAIN CONTRIBUTIONS

The main outcomes of the research presented in this thesis are:

- A “privacy facts” label design (see Fig. 1.1) that meets the following criteria [60]:
  - Rooted in the [GDPR](#) and complies with it,
  - User-validated and accessible, and
  - Cross-contextual, i.e., it can be reused in non-IoT contexts, such as web-sites.
- An interactive digital extension of the printed label, which is also based on the [GDPR](#), and is cross-contextual, user-validated, and accessible [61] (see Fig. 1.3 through Fig. 1.9).
  - *SUS* scores obtained during the usability evaluation process, which can be used to compare our solution with alternatives developed by other research teams [61].
- Additional tools for processing, querying and analyzing information presented in privacy labels:
  - A formal notation for expressing privacy terms [62],

- A method for calculating the information efficiency of a representation of privacy terms [62],
- A set of rules for avoiding unnecessary pop-ups related to updates or privacy terms [62], and
- API for label data interchange.

The detailed contributions of each included paper are given in Ch. 2.

### 1.7.1 *Answers to Research Questions*

In this section we explain how the contributions relate to the research questions posed earlier.

**RQ1** To what extent are users aware of IoT privacy issues?

Our findings show that users are aware of privacy issues posed by IoT, and yet some of them continue using such devices for various reasons (e.g., if it serves a useful function). We have also found that users often have incorrect mental models about how an IoT device works and what technical capabilities it has, which suggests that they might underestimate the severity or the likelihood of the aforementioned privacy issues. More details about this research question are available in Ch. 3.

**RQ2** *What information* should be presented to users on a label to achieve transparency?

This topic is addressed in Ch. 4 and Ch. 5, where we explain how we transform specific GDPR clauses into requirements for a transparency interface. We do so by taking a “question-oriented” approach, i.e., we ask ourselves “answers to which questions must be provided by the interface?”, which brings us to the list of transparency questions presented in Sec. 1.4.1.

**RQ3** *In what way* should this information be presented, such that transparency is usable?

The usability evaluation of our prototypes shows that several traits are essential for good usability, for example: avoiding the use of technical or legal jargon in favour of straight-forward terminology, representing information in a tabular form, or using short sentences. Specific details about what works and what does not work are given in Ch. 4 and Ch. 5.

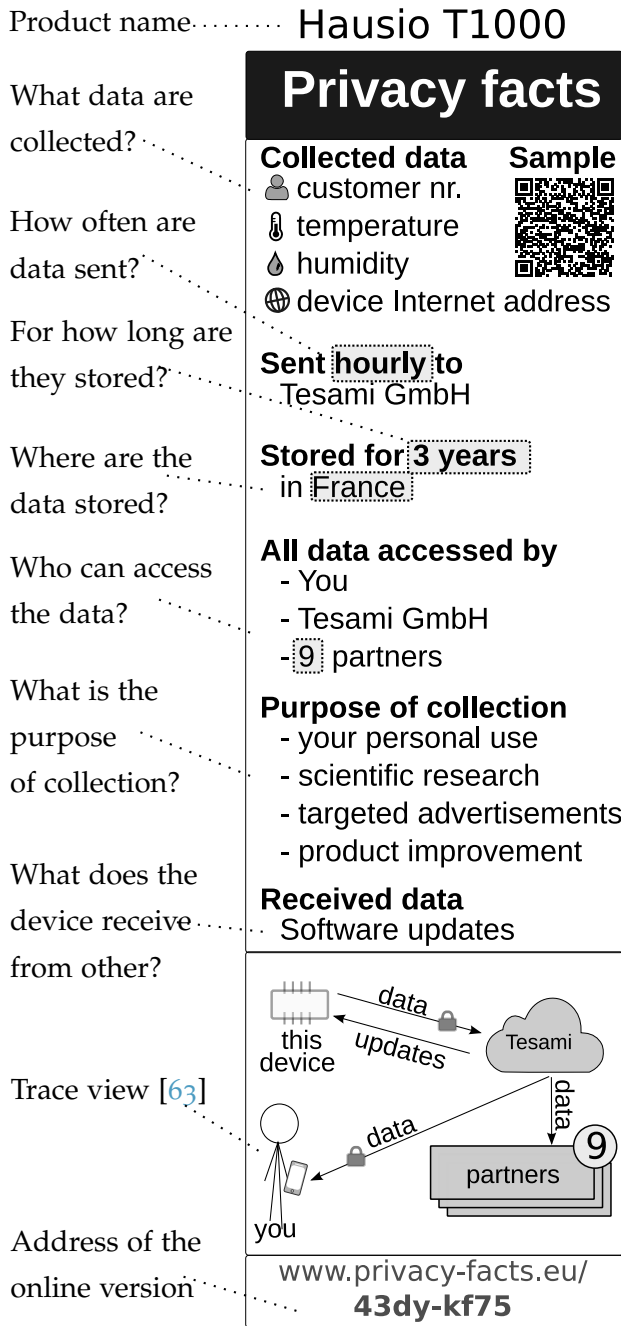
**RQ4** What other *technical and regulatory* means can improve privacy protection?

While the previously mentioned chapters propose solutions that have been explored by others, i.e., labels and online interfaces, in Ch. 6 we introduce some ideas that were not discussed before, to the best of our knowledge. In particular, we propose a standardized mathematical notation of privacy terms, as well a methodology for computing the information efficiency of different representations of such terms. In addition, we also propose an API for the automatic processing of privacy labels.

### 1.7.2 *Replicability*

In accord with good scientific practice, we take the steps that facilitate the replication of our findings by other researchers. To this end, besides the consent forms for GDPR

compliance and the detailed descriptions of our usability tests, we share the source code that makes it easy for others run our interface prototypes and verify our statistical analyses. In addition, these materials can enable others to improve upon our work and take it into directions that we have not envisioned ourselves.



customer number: 481-AHR-1831  
 temperature: 22 C  
 humidity: 34%  
 device Internet address: 93.184.216.34

←

privacy-facts.eu/43dy-kf75



Figure 1.2: (Top) Proposed form-factor for small devices, where space is limited.

(Bottom) False-colour representation of the additional features of the label, where yellow corresponds to an area with a hologram to signalize authenticity, and blue is for embossed areas that are meant to be felt by touch. The dot pattern is Braille for “Privacy facts”.

Figure 1.1: An annotated version of label for an imaginary IoT device [60].

Note that this prototype was not a part of the usability evaluation.






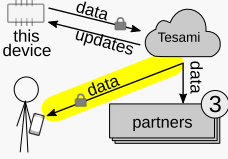
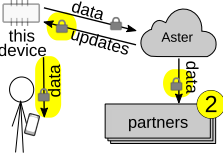
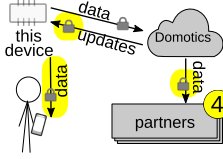
Overview	Who gets the data	Data flows	Data sample	Security	Lifecycle	Contact
<input type="checkbox"/> <input checked="" type="checkbox"/> Show differences <span style="float: right;">+</span>						
Hausio T1000	vs	Casami FX		Domowoj		
						
<b>Collected data</b>						
<ul style="list-style-type: none"> <li>customer nr.</li> <li>temperature</li> <li>humidity</li> <li>device Internet address</li> </ul>		<ul style="list-style-type: none"> <li>customer nr.</li> <li>temperature</li> <li>humidity</li> <li>wind speed</li> </ul>		<ul style="list-style-type: none"> <li>customer nr.</li> <li>temperature</li> <li>UV radiation</li> <li>wind speed</li> </ul>		
<b>Sent</b>						
hourly		daily		daily		
to Tesami GmbH		to Aster SRL		to Domotics s.r.o.		
<b>Stored for</b>						
3 years		6 years		1 year		
in France		in Italy		in the Czech Republic		
<b>All data accessed by</b>						
<ul style="list-style-type: none"> <li>- You</li> <li>- Tesami GmbH</li> <li>- 3 partners</li> </ul>		<ul style="list-style-type: none"> <li>- You</li> <li>- Aster SRL</li> <li>- 2 partners</li> </ul>		<ul style="list-style-type: none"> <li>- You</li> <li>- Domotics s.r.o.</li> <li>- 4 partners</li> </ul>		
<b>Purpose of collection</b>						
<ul style="list-style-type: none"> <li>- your personal use</li> <li>- scientific research</li> <li>- targeted advertisements</li> <li>- product improvement</li> </ul>		<ul style="list-style-type: none"> <li>- your personal use</li> <li>- scientific research</li> </ul>		<ul style="list-style-type: none"> <li>- your personal use</li> <li>- scientific research</li> </ul>		
<b>Received data</b>						
Software updates		Software updates		Software updates		
<b>Data flows</b>						
						
<b>Product information provided by</b>						
<ul style="list-style-type: none"> <li>- vendor</li> <li>- independently verified by Consumix BV (view report)</li> </ul>		<ul style="list-style-type: none"> <li>- vendor</li> <li>- has not been independently verified</li> </ul>		<ul style="list-style-type: none"> <li>- vendor</li> <li>- has not been independently verified</li> </ul>		

Figure 1.3: Screenshot of the main page of OnLITE, displaying the highlighted differences between three IoT devices that are being compared [61].

Overview **Who gets the data** Data flows Data sample Security Lifecycle Contact

See who gets the data, and why Search in table:

Device	Data type	Purpose	Company	Country	Sensitivity
Casami FX	temperature	scientific research	Minerva LTD	Canada	low
Casami FX	humidity	scientific research	Minerva LTD	Canada	low
Domowoj	UV radiation	archive data	Cornix	China	low
Domowoj	customer nr.	scientific research	Minerva LTD	Canada	low
Hausio T1000	customer nr.	targeted ads	Minerva LTD	Canada	low
Hausio T1000	temperature	targeted ads	Minerva LTD	Canada	low
Hausio T1000	humidity	targeted ads	ThirstFirst LTD	USA	low
Hausio T1000	humidity	archive data	Minerva LTD	Canada	low
Hausio T1000	device Internet address	targeted ads	Minerva LTD	Canada	high

Showing 1 to 9 of 9 entries (filtered from 17 total entries)

Figure 1.4: Screenshot of OnLITE, displaying a tabular representation of the data sharing patterns [61].

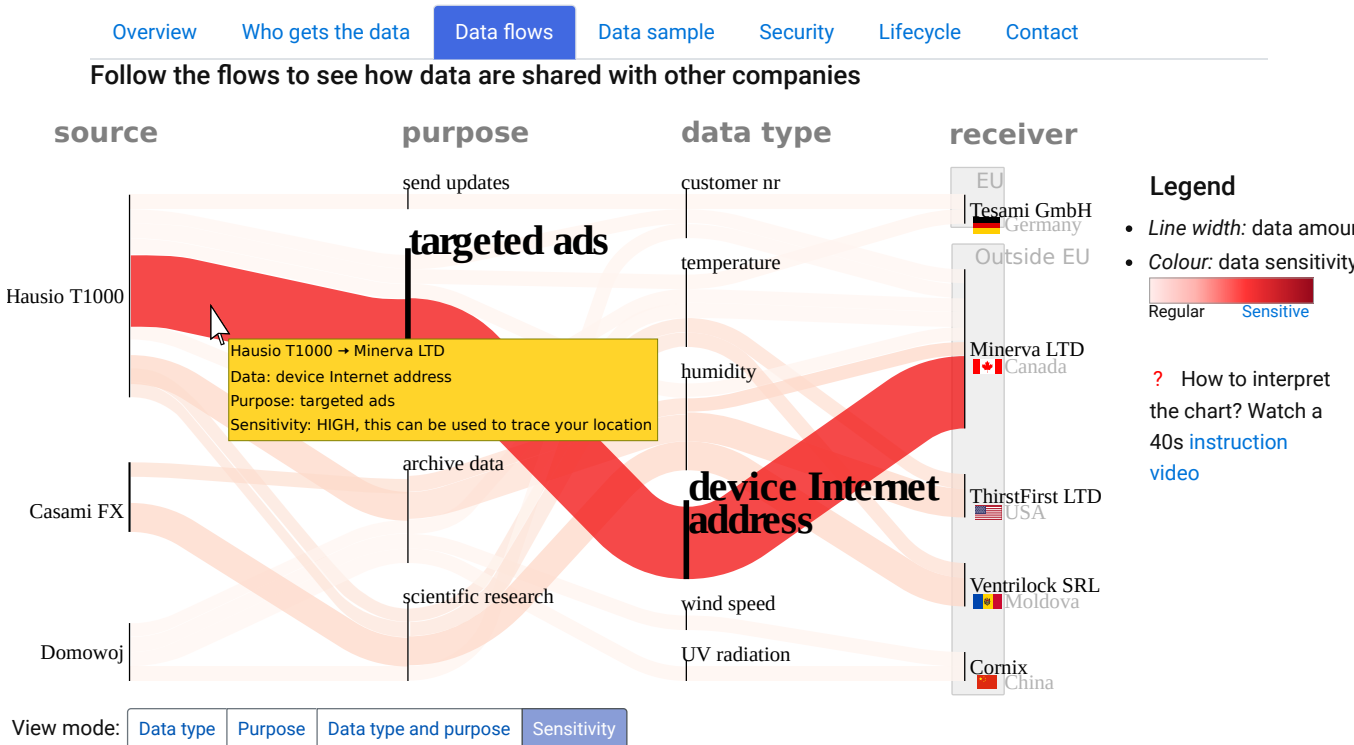


Figure 1.5: Screenshot of OnLITE, displaying a graphical representation of the data sharing patterns, by means of hybrid Sankey diagrams [61, 64].

Overview Who gets the data Data flows **Data sample** Security Lifecycle Contact

This table shows actual samples of data collected by each device

Data	Hausio T1000	Casami FX	Domowoj
customer nr.	481-AHR-1831	mustermann@kiel.de	+43-517987-891
temperature	22 °C	22 °C	22 °C
humidity	34%	34%	-
UV index	-	-	moderate
wind speed	-	2 m/s	2 m/s
device Internet address	93.184.216.34	-	-

Figure 1.6: Screenshot of OnLITE, displaying samples of data collected by each IoT device included in the comparison [61].

Overview Who gets the data Data flows Data sample **Security** Lifecycle Contact

	Hausio T1000	Casami FX	Domowoj
<b>Vulnerabilities</b>			
Reaction time to disclosed vulnerabilities	2 weeks	3 weeks	-
Rewards for reported vulnerabilities	Yes	Yes	No
<b>Communications</b>			
Secure from Internet eavesdroppers	Yes	-	-
Secure from local network eavesdroppers	Yes	Yes	No
<b>Storage</b>			
Stored data are encrypted	N/A, no information is stored on the device	Yes	No

Protected in a way that makes the data unreadable to persons who do not have the password

Figure 1.7: Screenshot of the “Security” tab of OnLITE, where information about the security of an IoT device is summarized, and optionally, compared with other devices [61].

	Hausio T1000	Casami FX	Domowoj
<b>Overview</b> <b>Who gets the data</b> <b>Data flows</b> <b>Data sample</b> <b>Security</b> <b>Lifecycle</b> <b>Contact</b>			
<b>Features grouped by phases of the device lifetime: set-up → usage → maintenance → retiring</b>			
<b>Set up</b> – preparing the device for use			
Unique factory-set password	Yes	Yes	No
Password change required before remote access for the first time	Yes	No	No
<b>Use</b> – typical, daily interactions with the device			
Multiple user accounts	Supported	Supported	No
Separate accounts for children	Supported	Supported	No
Separate account for guests	Supported	No	No
<b>Maintenance</b> – procedures to increase the device longevity and ensure it works well			
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	August 2019	December 2020
Security update period	December 2023	August 2019	December 2020
Long-term support	January 2024 <a href="#">source code release</a>	-	-
<b>Retiring</b> – when the device is sold, sent for repairs, donated or thrown away			
Secure data deletion (wiping)	Yes	No	No

Figure 1.8: Screenshot of the “Lifecycle” tab of OnLITE, where device features are grouped according to the life-cycle phase [25, 61].

	Hausio T1000	Casami FX	Domowoj
<b>Overview</b> <b>Who gets the data</b> <b>Data flows</b> <b>Data sample</b> <b>Security</b> <b>Lifecycle</b> <b>Contact</b>			
<b>Action</b>	<b>Hausio T1000</b>	<b>Casami FX</b>	<b>Domowoj</b>
View, edit or delete collected data by contacting the <i>Data Controller</i>	Tesami GmbH Flachmatuchstr. 42, Kiel, 24148, Germany. info@tesa.mi	Aster SRL Via Macaroni 113, Verona, Italy. contact@casam.it	Domotics s.r.o Bezručova 202, Brno, Czech Republic. gosti@dom.cz
Report privacy-related issues to the <i>Data Protection Officer</i>	dpo@tesa.mi	info@casam.it	rucitel@dom.cz
Lodge a complaint with the <i>supervisory authority</i>	<i>Unabhängiges Landeszentrum für Datenschutz</i> Holstenstraße 98, 24103 Kiel, Germany. mail@datenschutzzentrum.de	<i>Garante per la protezione dei dati personali</i> Piazza di Monte Citorio, Roma, Italy.	<i>Orgánem pro ochranu údajů</i> Svoboda 900, Praha, Czech Republic. pomoc@opou.cz

You can also lodge a complaint with a [supervisory authority in your area](#).

Figure 1.9: Screenshot of the “Contact” tab of OnLITE, where contact details of the data controller are presented [61].

## 1.8 LIMITATIONS

To the best of our knowledge, the work presented in this thesis is subject to several limitations:

1. The sample sizes<sup>5</sup> are sufficient for an exploratory study, but a larger-scale evaluation would be necessary if the proposed solutions were to become a part of a regulation on IoT privacy labels.
2. Although the results of the usability tests we conducted suggest that the proposed designs can work in contexts other than IoT (e.g., smartphone application marketplaces or web-sites), this particular capability has not been evaluated. Therefore, this claim needs to be verified independently if the proposed designs were to be used outside the field of IoT.
3. Throughout our usability tests, the prototypes we evaluated presented privacy facts based on synthetic data sets devised for hypothetical IoT devices that tended to follow the GDPR principle of *data minimization*. For example, a temperature and humidity meter would collect information about temperature, humidity and some metadata like account identifiers. It is possible that real world devices collect much more information, i.e., not just the three aforementioned types of data. Although we have anticipated this possibility and adapted the design to accommodate it, we cannot exclude that the interface would have to be revised further if the true magnitude of the data collection practices would far exceed even our most pessimistic scenarios.

## 1.9 DISCUSSION OF PRACTICAL CONSIDERATIONS

Here we present some additional mechanisms that relate to implementations of privacy label designs and can make them better.

### 1.9.1 Proposed API for Label Management and Usage

An Application Programming Interface (API) facilitates interoperability between systems. In the case of transparency labels, an API can foster generativity (see 1.5.2.1) because it enables third parties to query the data and apply it in ways that go beyond the vision of the original design [45]. For example, privacy researchers can use such an API to automate the analysis of privacy policies and observe whether they get stricter or more relaxed with time, what the most commonly collected types of data are, etc. At the moment such information is not always easy to obtain, because many web-sites prohibit scraping<sup>6</sup>, thus preventing automated extraction and analysis of policies [65]. To close this gap, we propose the use of an HTTP-based API, which would open the data for anyone to use.

The general form of an API request is <base> +[verb+]<device ID>, where the *device id* is a unique identifier of a particular generation of devices, associated with the tuple

<sup>5</sup> We had 110 participants for the online questionnaire in P1 Lifecycle, we interviewed 31 participants for P2 LITE, and 15 participants for P3 OnLITE.

<sup>6</sup> The practice of automatically extracting data from web-sites.

(vendor, device, firmware version). Note that each vendor manages their namespace as they see fit, for example some companies might use a different taxonomy: (vendor, device, country, version).

The `<base>` is a server that handles the data retrieval requests. Ideally, it is managed by an international authority or consortium that everyone trusts, in this document we will assume it is *example.com*. There are systems that rely on similar assumptions, for instance, the Internet Corporation for Assigned Names and Numbers (ICANN) is responsible for several critical aspects required for the Domain Name System (DNS) of the Internet to work. Thus, such an assumption is not unrealistic.

In its shortest form, the `URL` points to a page that visualizes the privacy policy in a human-readable way, i.e., OnLITE. To make the design future-proof, we add the possibility to encode various operation requests into the `URL`, here are some examples:

`<BASE>/BVEJA24K` Retrieve label for device `bveja24k`, the view operation is implied.

`<BASE>/VIEW/BVEJA24K` Same as above, but `view` is explicitly declared. The first notation is preferred, because it is shorter and easier to type by hand.

`<BASE>/COMPARE/BVEJA24K/LKORP3` Compare label between `bveja24k` and `Lkorp3`.

`<BASE>/HISTORY/BVEJA24K` Retrieve a history of changes in the label of the device identified by `bveja24k`.

`<BASE>/UPDATE/BVEJA24K` Retrieve latest label of the device identified by `bveja24k`.  
The IoT device can use this to find if, and how many, changes were applied to the privacy policy.

### 1.9.2 Widths of the Flows

A practical matter that must be addressed before OnLITE can be used by the public is devising a formula that determines the width of each flow in the Sankey diagram. In the example shown in Fig. 1.5, the widths are relative, i.e., they answer the question “what is the volume of a flow relative to other flows?”, without relying on any units. Thus, it makes a comparison easy when multiple devices are shown on the same visualization. However, the drawback is that several print-outs of different devices can contain flows that have the same physical width, but represent different magnitudes.

#### 1.9.2.1 Normalized Encodings

The width is determined by multiplying two values: the frequency of transmission and the payload size. Considering that different IoT devices can encode these values differently, the size of the payload may vary, even when the transmitted information is the same. For example, the value “2” can be transmitted as a one-byte integer, or as a four-byte integer, or as the ASCII string “two” that takes four bytes. This means that a head-to-head comparison will not yield a fair result, therefore the values need to be normalized in a way that ensures the comparison is just. We propose to address this from the standpoint of information theory, and focus on the actual information being transmitted, regardless of the way it is encoded. For example, imagine an IoT device that transmits temperature as an integer data point in the range  $[-50, 60]$  °C. The range consists of 111 discrete values, including 0. Therefore, each value represents  $\lceil \log_2 111 \rceil = 7$

bits. Suppose we have 2 IoT devices that transmit temperature every minute, device A encodes it as a single byte (type “char” in the C programming language), taking 8 bits, while device B encodes it as a signed long of 32 bits. In encoded form, in the period of 1 hour device A will send  $60 \times 8 = 480$  bits, while device B will send  $60 \times 32 = 1920$  bits. Although at first glance device B sends 4 times the amount of data as device A, on our graph the flows will have the same width, because we focus on pure information, hence the side by side comparison is fair, and both devices will be shown as transmitting  $60 \times 7 = 420$  bits. We emphasize that our objective is to compare flows relative to one another, so there is no need to display the units, or the actual value of 420 to end users.

### 1.9.2.2 *Heterogeneous Data Types*

Another point that must be addressed is the visualization of different types of data on the same diagram. For example, if a device transmits humidity as a number between 0 and 100, each data point is worth  $\lceil \log_2 100 \rceil = 7$  bits, so it can be shown next to the previously mentioned flows of temperature data. Thus, even though the flows refer to different intervals (i.e.,  $[-50, 60]$ , and  $[0, 100]$ ) and are measured in different units ( $^{\circ}\text{C}$  vs. %), on the Sankey diagram they can be compared head to head.

### 1.9.2.3 *Representing Continuous Values*

When dealing with floating point values, the approach above is not directly applicable, as the set of real numbers is not countable. A solution to this problem would be to round the float to the nearest integer and treat it as such. Alternatively, one could encode the floats using 32 bits in the IEEE 754 format. However, in this case, the value 1.0 would look 4 times wider on a flow than the value 1 (32 bit vs 7 bit), making the comparison impractical. In this case, the solution could be to treat *all* numeric values as 32 bit floats, because our objective is to visualize magnitudes relative to one another, rather than display the numbers themselves.

### 1.9.2.4 *One Flow for Different Magnitudes*

Another special case occurs when representing volume of data such as photographs, e.g., when an IoT device uploads a photo every minute. How to quantify this information such that it can be paced on the same diagram as a temperature flow, without dwarfing it in scale? If we treat the photo as a matrix of bytes, then a  $800 \times 600$  image will take  $480\,000 \times 8 = 3\,840\,000$  bits, making it difficult to distinguish and reason about smaller magnitudes like temperature and humidity. A solution to this problem could be to apply a logarithmic scale, rather than a linear one, when rendering flows.

## 1.9.3 *Visualizing How Often Data Are Sent*

Now we shift our attention to the remaining component of the calculation, the frequency of transmission. It is critical to reflect this component in the visualization, otherwise a user cannot tell the difference between a device that sends the temperature every hour and one that sends it multiple times per second. In the latter case, the privacy implications can be significant. Considering that devices can be configured to send data at different frequencies, we have to ensure that the same settings are applied in all

cases, i.e., frequencies are expressed in the same unit of time. For example, if device A sends data 10 times/s (i.e., 10 Hz), while device B sends it every minute, they will be represented as 10 Hz and 0.16 Hz respectively.

#### 1.9.3.1 *Visualizing Event-Driven Data*

Another special case is related to information transmitted when an event occurs (e.g., a photo sent when motion is detected), which can happen very often or very seldom. We consider an interactive approach, where the user can adjust the frequency through sliders in the GUI and observe the corresponding data flows. A non-interactive approach is also possible, where the frequency is set to some predefined value for each *category of events* (e.g., motion detection: 5 times per hour). Though this approach is less flexible, it avoids adding complexity to the GUI, while retaining the ability to compare multiple IoT devices side by side, since the same assumptions about frequencies apply to all devices. Moreover, it produces an output that is also suitable for print, which is an advantage.

#### 1.9.4 *Product Code Life-Cycle*

A discussion about the way product codes are managed is necessary to address several practical matters. The first one is how to handle situations in which a label was printed and attached to a box, which then spent several months on a shelf in a warehouse, while new software updates were released. This makes the label obsolete and potentially misleading, if the update changes the way data are handled. To address this issue, we postulate that *a user should be able to use a device on the terms that it was originally released with*. Art. 7 of the GDPR states that consent must be voluntary, which means that a user cannot be forced to accept new terms. Assuming that a company is not open to accepting device returns every time an update is released, this can be accomplished by allowing users to continue running a version of the firmware they already have, for as long as they wish.

##### 1.9.4.1 *Decoupled Feature, Privacy, and Security Updates*

To ensure that running such firmware does not expose users to security risks, we advocate the decoupling of security, privacy and feature updates [62], such that security patches can be applied without re-requesting consent, because nothing is changed in the way the data are handled. While this adds some new requirements to the design of systems, it is technically feasible, because there are systems in use today that already accomplish this. For example, software that uses OpenSSL can take advantage of error fixes by updating the library *libopenssl*, while the remaining logic is unchanged.

What remains to be done is to find ways to encourage companies that provide such decoupled updates. We believe it can be accomplished by means of regulation (make it a requirement, or provide tax breaks to companies that do so), and by exposing consumers to such information, so they could choose products that have this capability. Thus, consumers can potentially influence vendors, by creating more demand for privacy-preserving products.

The life-cycle of a product code is shown in Fig. 1.10. Note how in the lower figure the transitions occur only by accepting new terms, and the two timelines coexist independently. Also note that in this document we focus on how things *should be*, while a



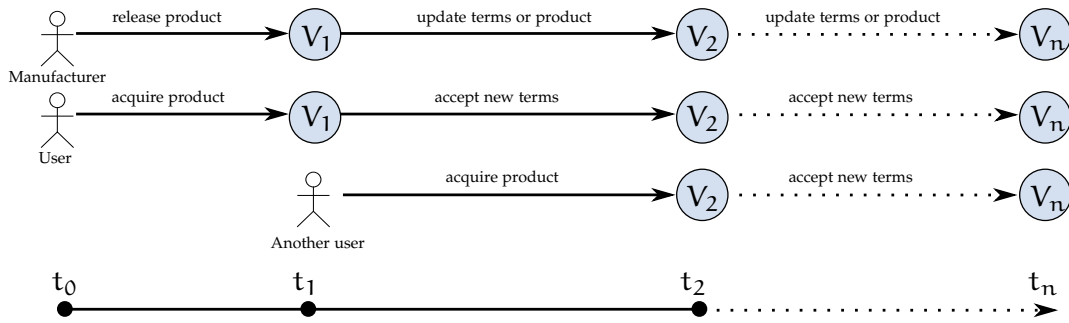


Figure 1.10: The product code lifecycle.

discussion about the engineering and economic challenges of making it possible is outside the scope of our current work. In addition, it should be noted that a new code can be issued in several cases: an update changes how the data are handled by default, or when a new hardware revision is released, even if nothing is changed in the way data are handled by the device. This creates a unified namespace of product codes, and when users refer to code X, they can be certain that they are talking about the same device, running the same hardware and firmware underneath. This is an important feature, because it allows one to distinguish devices even when they look the same and are marketed under the same name, but are different internally. A notable example is the Linksys WRT-54G router that gained popularity because it ran Linux, which made it heavily customizable. It was eventually replaced by an identical-looking device that had the same name and look, but ran different software that had less potential for enhancement by hobbyists.

By taking the latter approach of a unified product code namespace, we also avoid the complexity of dealing with multiple version numbers that users would otherwise have to be aware of (e.g., privacy policy version, firmware version, hardware platform version, etc.). The trade-off is that there could be cases where different product codes will actually refer to the same privacy rules. We don't consider this issue relevant, because in these cases OnLITE will automatically say that no differences were found - so users are not exposed to additional cognitive burdens.

#### 1.9.4.2 *Immutable Codes*

Another requirement that must be taken care of - the product codes should be immutable. That is, once a code was issued, it remains unchanged and any revisions of the privacy policy will produce a new code. This is important for archival and research purposes, but also because it ensures that any URLs that circulated in the past will remain accessible [66, 67], and are guaranteed to point to *the same* privacy policy.

## 1.10 RELATED WORK

Several transparency solutions for IoT devices were conceived by other researchers. In what follows, we adapt the terminology proposed by [68] to summarize<sup>7</sup> the main differences between our own work and the alternatives we found through a literature review:

<sup>7</sup> Additional details are available in Sec. 9 of *P3 OnLITE*, which is the paper where we introduce our digital interface for the privacy label.

**Seal** - A symbol that indicates that a certification was attained.

**Info** - Informational labels present facts about a device, e.g., its supported features, data-handling practices, security capabilities, etc.

**Graded** - Such labels rank IoT devices according to a metric, e.g. the level of privacy protection. Grades can be represented numerically in a range between 1 (worst) and 10 (best), or a “star rating”, where 5 stars indicate a good result, etc.

**Printed** - The design is meant to be used in print, e.g., a sticker on a box.

**Digital** - The design meant to be used in a digital form, e.g., shown on a web-site or viewed via a smartphone app.

**GDPR** - The proposed solution has been designed with the GDPR transparency requirements in mind.

**Validated** - A design was evaluated through usability studies.

We consider that seals of approval cannot be regarded as a viable solution to the IoT transparency problem, because they transmit only one bit of information<sup>8</sup> to the consumer. However, transparency implies the ability to find answers to all the questions listed in Sec. 1.4.1, such as  $Q_{\text{what}}$  (“what data are collected?”). Therefore, seals cannot solve the problem on their own. The same rationale applies to graded labels. A grade places an IoT device on a spectrum, e.g., between 1 and 10, which makes ranking devices easier, but it does not provide answers to the transparency questions. Thus, a grade does not imply transparency, though it could be a part of a larger transparency solution. Moreover, grades can hinder the adoption of a privacy label, because some IoT device manufacturers might resist labeling their products if they only earned a low grade. For this reason, we maintain a position of *neutrality* (as outlined in Sec. 1.4.3), letting end-users decide for themselves what is “good” and what is “bad” after evaluating the facts presented to them.

Another factor that we consider is validation by means of usability studies. If a label becomes part of a regulation, it will affect a very large population of users of various ages, levels of education and cultural backgrounds. In addition, various levels of impairment have to be taken into account, as discussed in Sec. 1.5.2.3. Therefore, it is imperative that the design is tested before going into effect.

To the best of our knowledge, as of this writing, our work is the only research that proposes an IoT privacy label that is rooted in the GDPR and has been validated through user studies.

## 1.11 CONCLUSION

In this thesis we present a possible solution to the problem of usable transparency for IoT devices. We designed it by following the requirements given in the GDPR and expressing

<sup>8</sup> Essentially, they answer the question “Is this product certified?”, which takes a “yes” or “no” answer.

<sup>9</sup> The paper mentions that the design was reviewed by experts, then evaluated with 32 participants. Although the evaluation measured the participants’ “level of satisfaction with each label and perceived trust” [70], no details are given about how this was quantified, nor are specific details of the analysis given.

<sup>10</sup> A prototype of a seal was designed as a recognition symbol for LITE (see Fig. 1.2). However, the seal was not included in the usability evaluation.

Table 1.5: Summary of IoT transparency initiatives, our work is in the lower part of the table

Authors	Seal	Info	Graded	Printed	Digital	GDPR	Validated
Van Diermen [69]		●	●	●	●	●	
Fox et al. [70]		●		●	●	●	● <sup>9</sup>
Shen et al. [71]		●		●			
Naeini et al. [72]		●	●	●	●		●
Bihl [73]	●						
P <sub>2</sub> LITE [60]	● <sup>10</sup>	●		●		●	●
P <sub>3</sub> OnLITE [61]		●		●	●	●	●

them as questions listed in Sec. 1.4.1. We propose a “privacy facts” label and a digital interface that accompanies it. In addition, we describe the usability tests that we conducted to evaluate the efficacy and efficiency of the interfaces, as well as the users’ level of satisfaction with said interfaces. We show the evidence that we rely on when asserting that our proposal is a viable candidate for solving the problem of IoT transparency. Furthermore, we discuss some of the practical aspects that need to be taken into account when considering the real-world application of such solutions.

Our claim is that the proposed design is merely *one* approach to the transparency problem among the many possibilities. Although we have followed best practices of design and relied on prototyping and usability tests to rule out inadequate solutions, our proposal is probably not a global optimum. Therefore the aim of this thesis is to initiate a discussion and suggest a practical starting point, with the assumption that subsequent iterations will bring the research community closer to the optimum.

## 1.12 FUTURE STEPS

In this section we discuss some points that fall outside the scope of this thesis and the IoT life-cycle phases that it targets. Nevertheless, these points are an important component of a holistic privacy-friendly IoT ecosystem.

### 1.12.1 Retiring

Although the main focus of our work is *transparency* during the first stages of an IoT device life-cycle, the picture is only complete if solutions applicable to the *retiring* phase are available as well. This implies that there must be an easy way to wipe data off an IoT device that is about to be sold, gifted or otherwise discarded. Such functionality needs to be implemented and evaluated in terms of usability, to ensure that end-users are aware of such a feature and are confident that their sensitive data are cleared if they have used this feature. In addition to that, it is necessary to consider what technical means can be applied to perform such a wipe operation even when an IoT device is not functioning properly. For example, if a device cannot be powered up or if its screen is broken, the user will be unable to invoke the wipe feature, even if they are fully aware of its existence and are willing to use it. Sending such a device to a repair center or throwing it away would potentially compromise the users’ privacy, because some of their data are still on

the device and can be retrieved with specialized tools [74, 75, 76]. This gap in privacy remains open at the time of this writing, but it needs to be closed.

#### 1.12.2 *Transparent Automated Processing*

The problem of *explainability* can arise when automated processing is employed by systems that deal with personal data [77, 78]. For example, a car insurance company can evaluate one's driving style based on collected data in order to calculate a driving score, then determine the cost of the insurance for the given customer. If one's costs increased, they might want to know why this happened, i.e., ask for an explanation.

However, the calculations may be performed by complex algorithms that may not necessarily be understood by non-experts. In some cases, e.g., when deep learning is involved, introspecting the result may be even more challenging [78].

GDPR Art. 22 explicitly refers to such automated individual decision-making and profiling, therefore, a comprehensive transparency solution should also provide means for end-users to understand how such algorithms work.

## SUMMARY OF PUBLICATIONS

---

This chapter briefly summarizes each paper included in this collection thesis, highlighting the main contributions. The papers are listed in chronological order.

### 2.1 P1 LIFECYCLE → LIFE-LONG PRIVACY IN THE IOT? MEASURING PRIVACY ATTITUDES THROUGHOUT THE LIFE-CYCLE OF IOT DEVICES

In this paper we conducted an online survey with 110 participants, to determine the extent to which users are aware of the privacy implications of IoT, and understand which measures users take to protect their privacy when using IoT.

In addition, we propose a life-cycle model for consumer-oriented IoT devices (see Fig. 3.1), which we then use as a roadmap for our subsequent research, moving further along this timeline.

The results indicate that users are aware of the negative impact IoT has on their privacy, but are nevertheless continuing to use such devices. The findings also suggest that users develop incorrect or incomplete mental models about the capability of IoT devices, hence it is possible that they underestimate the risks or consider that their defensive measures are sufficient.

### 2.2 P2 LITE → LET THERE BE LITE: DESIGN AND EVALUATION OF A LABEL FOR IOT TRANSPARENCY ENHANCEMENT

The findings of *P1 Lifecycle* motivated us to think about the earliest time in the life-cycle of an IoT device when users can make choices that will have a positive impact on their privacy. According to our proposed model, it is the *pre-acquisition phase*, i.e., the moment in time when users are still deciding whether to acquire an IoT device, and which brand and model to go for.

Driven by the idea that the cost of solving a problem is lower if it is detected and addressed earlier, we argue that the cost can be zero if the problem can be avoided altogether. Therefore, helping users choose a device that is a better match for their needs has the potential to solve the problem preemptively, rather than in a reactive way.

The main contribution of this paper is a GDPR-centric “privacy-facts” label for IoT devices. The label summarizes important privacy information in a concise form, and it is suitable for use on product boxes, akin to “nutrition facts” labels. In addition to that, we conducted interviews with 31 participants to measure how well they could interpret the information presented on the label.

The results of the evaluation showed that the participants could interpret the label correctly, they understood its benefits and stated that they would be in a better position to make informed choices if such labels were present on actual IoT products. However, they also wanted more details, which we omitted because of the size constraints of a physical label.

### 2.3 P3 ONLITE → ONLITE: ON-LINE LABEL FOR IOT TRANSPARENCY ENHANCEMENT

The results of the previous paper revealed some gaps in the design of our label, leading us towards the next idea - a digital interface that accompanies the printed label and provides information that would otherwise not fit on a product box.

The main contribution of this paper is an interface prototype which assists users in answering the transparency questions listed in Sec. 1.4.1. The interface augments the printed label with features that become possible when an actual computer is at hand - search, sort, filter and side-by-side comparison of devices. It also comes with an interactive visualization which aims to distill large data-sets into a single image that makes it easy for users to see how their data are used.

The usability of the interface was evaluated by means of think-aloud task analysis and interviews with 15 participants.

### 2.4 P4 UPDATES → IMPROVING THE TRANSPARENCY OF PRIVACY TERMS UPDATES

In this paper we focus on the *maintenance* phase of the IoT device life-cycle, i.e., the phase during which users can make configuration changes or install software updates that can affect their privacy. For example, an update can come with a new privacy policy, which may include changes in the way collected data are handled. If the new terms are presented as lengthy prose, some users may not read them thoroughly [39] before giving consent, thus hindering transparency. We argue that the transparency of updates can be improved if users are presented with an answer to the question of “what has changed in the privacy policy?”, rather than to the question of “what is the new privacy policy?”.

We build upon our previous results and propose the use of our interface to compare different versions of the privacy policy of the same IoT device, rather than entirely different devices.

We also define a structured format for storing privacy policies, as a prerequisite for the transparency label and interface proposed earlier. Such a format would also make it possible to automate the analysis of privacy policies, potentially making privacy-preserving products more prominent and more accessible to users.

Last, but not least, we propose an information efficiency approach to quantify and compare different representations of privacy terms. This method makes it possible to detect visualization methods deliberately designed to be voluminous, but offer little information to users.

### 2.5 OTHER CONTRIBUTIONS

To explore the potential of the transparency mechanisms proposed in this thesis, and to elicit additional requirements for future prototype iterations, I co-authored several peer-reviewed papers about usable privacy for telemetry-based car insurance<sup>1</sup>. Such forms of insurance rely on vehicle telemetry to determine one’s driving style and possibly reduce their insurance costs if they drive safely and obey the rules.

<sup>1</sup> Also referred to as “telematics insurance” or “usage-based [car] insurance” (UBI or UBCI).

The relevance to this thesis is that we regard a car as an IoT device which collects and processes data that can have privacy implications for the drivers. Thus, work on these papers was not only an opportunity to practice user-centered design and gain experience in usability studies, but also a way get early feedback about the reusability of the transparency prototypes discussed in this thesis.

Thus, the following contributions were made:

- Juan Quintero, **Alexandr Railean**, Zinaida Benenson: Acceptance Factors of Car Insurance Innovations: The Case of Usage-Based Insurance. *Journal of Traffic and Logistics Engineering*, Vol. 8 (2020). My contribution to this paper includes the recommendations for telematics insurance providers aimed at improving the usability and privacy of their services. In addition, I helped edit the contents of the paper itself.
- Juan Quintero, **Alexandr Railean**: Users' Privacy Concerns and Attitudes Towards Usage-Based Insurance: an Empirical Approach. *Vehicle Technology and Intelligent Transport Systems* (2022). Both authors have contributed equally and collaborated while designing and conducting the study, analyzing the data and writing this paper.

*Sec. 1.5.2.2 gives the rationale for the reusability requirement.*

Several non peer-reviewed deliverables produced for the European Commission by the Privacy&Us project members included contributions from me:

- Michael Bechinie (editor). User Interface Requirements V1.0. Privacy&Us deliverable D4.1 (2017). I provided Sec. 2.5 of this report.
- Alexandr Railean (editor, contributor), Harald Zwingelberg (editor). Privacy Principles V1.0. Privacy&Us deliverable D5.1 (2017). I provided Sec. 4.4 of this report and helped copy-edit the document.
- Simone Fischer-Hübner, Leonardo A. Martucci (editors). User Interface Designs and Prototypes V1.0. Privacy&Us deliverable D4.2 (2018). I provided Sec. 6 of this report, which discusses early-stage drafts of the prototypes presented in this thesis.
- Ben Wagner (editor). Risk Assessment V1.0. Privacy&Us deliverable D5.2 (2017). I participated in the activities that generated the data this deliverable is based on.

## REFERENCES

- [1] *Internet of Things: Privacy & Security in a Connected World*. Staff report. FTC, 2015. URL: <https://ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf>.
- [2] *Telecommunication Threats in Current Context*. In collab. with Peter Lewis. 2020. URL: <https://youtu.be/Me-SjNDonXY?t=658>.
- [3] Kevin Ashton. *That 'Internet of Things' Thing*. RFID Journal. 2009. URL: <https://www.rfidjournal.com/that-internet-of-things-thing>.
- [4] Tom Lane et al. *CMU SCS Coke Machine*. 1982. URL: [https://www.cs.cmu.edu/~coke/history\\_long.txt](https://www.cs.cmu.edu/~coke/history_long.txt).
- [5] Jordan Teicher. *The Little-Known Story of the First IoT Device*. Industrious. 2018. URL: <https://www.ibm.com/blogs/industries/little-known-story-first-iot-device/>.
- [6] *Analysis of Business Value Creation Enabled by 5G for Manufacturing Industries*. 2021. URL: <https://5gsmart.eu/wp-content/uploads/5G-SMART-D1.2-v1.0.pdf>.
- [7] *Manufacturing Trends Report*. Microsoft, 2019. URL: <https://info.microsoft.com/rs/157-GQE-382/images/EN-US-CNTNT-Report-2019-Manufacturing-Trends.pdf>.
- [8] Min Chen et al. "Smart Clothing: Connecting Human with Clouds and Big Data for Sustainable Health Monitoring." In: *Mobile Networks and Applications* (2016).
- [9] Jie Ding et al. "IoT Connectivity Technologies and Applications: A Survey." In: *IEEE Access* (2020). arXiv: 2002.12646.
- [10] Sichao Liu et al. "An 'Internet of Things' Enabled Dynamic Optimization Method for Smart Vehicles and Logistics Tasks." In: *Journal of Cleaner Production* (2019).
- [11] Juan A. López-Morales, Juan A. Martínez, and Antonio F. Skarmeta. "Improving Energy Efficiency of Irrigation Wells by Using an IoT-Based Platform." In: *Electronics* (2021).
- [12] Toni Perković et al. "Meeting Challenges in IoT: Sensing, Energy Efficiency, and the Implementation." In: *Fourth International Congress on Information and Communication Technology*. 2020.
- [13] Delphine Christin. "Privacy in Mobile Participatory Sensing: Current Trends and Future Challenges." In: *Journal of Systems and Software* (2016).
- [14] M. Kosinski, D. Stillwell, and T. Graepel. "Private Traits and Attributes Are Predictable From Digital Records of Human Behavior." In: *Proceedings of the National Academy of Sciences* (2013).
- [15] Nicholas D. Lane et al. "On the Feasibility of User De-anonymization From Shared Mobile Sensor Data." In: *Proceedings of the 3rd International Workshop on Sensing Applications on Mobile Phones (PhoneSense)*. 2012.
- [16] Arvind Narayanan and Vitaly Shmatikov. "How to Break Anonymity of the Netflix Prize Dataset." In: *arXiv preprint cs/0610105* (2006).



- [17] Xavier Caron et al. "The Internet of Things (IoT) and its Impact on Individual Privacy: An Australian Perspective." In: *Computer Law & Security Review* (2016).
- [18] Mahmoud Elkhodr, Seyed Shahrestani, and Hon Cheung. "A Review of Mobile Location Privacy in the Internet of Things." In: *Proceedings of the 10th International Conference on ICT and Knowledge Engineering*. 2012.
- [19] Robert P. Minch. "Location Privacy in the Era of the Internet of Things and Big Data Analytics." In: *Proceedings of the 48th Hawaii International Conference on System Sciences (HICSS)*. 2015.
- [20] Scott R. Peppet. "Regulating the Internet of Things: First Steps Toward Managing Discrimination, Privacy, Security and Consent." In: *Texas Law Review* (2014).
- [21] Wei Zhou and Selwyn Piramuthu. "Security/Privacy of Wearable Fitness Tracking IoT Devices." In: *Proceedings of the 9th Iberian Conference on Information Systems and Technologies (CISTI)*. 2014.
- [22] Jan Henrik Ziegeldorf, Oscar Garcia Morchon, and Klaus Wehrle. "Privacy in the Internet of Things: Threats and Challenges." In: *Security and Communication Networks* (2014).
- [23] Lisa Diamond et al. "Privacy in the Smart Grid: End-User Concerns and Requirements." In: *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct - MobileHCI '18*. 2018.
- [24] Pascal Kowalczyk. "Consumer Acceptance of Smart Speakers: a Mixed Methods Approach." In: *Journal of Research in Interactive Marketing* (2018).
- [25] Alexandr Railean and Delphine Reinhardt. "Life-Long Privacy in the IoT? Measuring Privacy Attitudes Throughout the Life-Cycle of IoT Devices." In: *Privacy and Identity Management. The Smart Revolution*. IFIP Advances in Information and Communication Technology. 2017.
- [26] Abdul Salam. "Internet of Things for Environmental Sustainability and Climate Change." In: *Internet of Things for Sustainable Community Development*. 2020.
- [27] Parlamentul Republicii Moldova. *Legea Privind Protecția Datelor cu Caracter Personal*. 2011. URL: <http://lex.justice.md/md/340495/>.
- [28] Abigayle Erickson. "Comparative Analysis of the EU's GDPR and Brazil's LGPD: Enforcement Challenges with the LGPD." In: *Brooklyn Journal of International Law* (2018).
- [29] European Parliament and Council of European Union. "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of Natural Persons With Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)." In: *Official Journal of the European Union* (2016).
- [30] W. Gregory Gregory Voss. "The CCPA and the GDPR Are Not the Same: Why You Should Understand Both." In: *CPI Antitrust Chronicle* (2021).
- [31] Soheil Human and Florian Cech. "A Human-Centric Perspective on Digital Consenting: The Case of GAFAM." In: *Human Centred Intelligent Systems*. 2021.

- [32] Lorrie Faith Cranor. "Necessary But Not Sufficient: Standardized Mechanisms for Privacy Notice and Choice." In: *Journal on Telecommunications and High Technology Law* (2012).
- [33] Brian Christian. *The Alignment Problem: Machine Learning and Human Values*. 2020.
- [34] Sandra Wachter. "The GDPR and the Internet of Things: a Three-Step Transparency Model." In: *Law, Innovation and Technology* (2018).
- [35] Federico Civerchia et al. "Industrial Internet of Things Monitoring Solution for Advanced Predictive Maintenance Applications." In: *Journal of Industrial Information Integration* (2017).
- [36] Michele Compare, Piero Baraldi, and Enrico Zio. "Challenges to IoT-Enabled Predictive Maintenance for Industry 4.0." In: *IEEE Internet of Things Journal* (2020).
- [37] *KrebsOnSecurity Hit By Huge New IoT Botnet "Meris"*. 2021. URL: <https://krebsonsecurity.com/2021/09/krebsonsecurity-hit-by-huge-new-iot-botnet-meris/>.
- [38] *KrebsOnSecurity Hit With Record DDoS*. 2016. URL: <https://krebsonsecurity.com/2016/09/krebsonsecurity-hit-with-record-ddos/>.
- [39] Aleecia M McDonald and Lorrie Faith Cranor. "The Cost of Reading Privacy Policies." In: *Journal of Law and Policy for the Information Society* (2008).
- [40] Eszter Hargittai. "'What Can I Really Do?' Explaining the Privacy Paradox with Online Apathy." In: *International Journal of Communication* (2016).
- [41] *GDPR Recital 58 - The Principle of Transparency*. URL: <https://gdpr-info.eu/recitals/no-58/>.
- [42] ISO DIS. "9241-210: 2010. Ergonomics of human system interaction-Part 210: Human-centred design for interactive systems." In: *International Standardization Organization (ISO)*. Switzerland (2009).
- [43] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. "Alexa, Are You Listening?: Privacy Perceptions, Concerns and Privacy-seeking Behaviors with Smart Speakers." In: *Proceedings of the ACM on Human-Computer Interaction (CSCW 2018)*.
- [44] Ponnurangam Kumaraguru and Lorrie Faith Cranor. *Privacy Indexes: A Survey of Westin's Studies*. Tech. rep. Institute for Software Research International, School of Computer Science, Carnegie Mellon University, 2005.
- [45] Jonathan L Zittrain. "The Generative Internet." In: *Communications of the ACM* (2006).
- [46] "Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions - on the European Interoperability Framework." In: *Official Journal of the European Union* (2017).
- [47] Farzaneh Karegar, John Sören Pettersson, and Simone Fischer-Hübner. "The Dilemma of User Engagement in Privacy Notices: Effects of Interaction Modes and Habituation on User Attention." In: *ACM Transactions on Privacy and Security* (2020).
- [48] *U.S. Access Board - Rehabilitation Act*. 1998. URL: <https://www.access-board.gov/law/ra.html>.

- [49] European Parliament and Council of European Union. "Directive (EU) 2016/2102 of the European Parliament and of the Council - of 26 October 2016 - on the Accessibility of the Websites and Mobile Applications of Public Sector Bodies." In: *Official Journal of the European Union* (2016).
- [50] *Web Content Accessibility Guidelines (WCAG)*. 2008. URL: <https://www.w3.org/TR/WCAG20/>.
- [51] Norman M. Bradburn, Seymour Sudman, and Brian Wansink. *Asking Questions: the Definitive Guide to Questionnaire Design— for Market Research, Political Polls, and Social and Health Questionnaires*. 2004.
- [52] Bella Martin and Bruce M. Hanington. *Universal Methods of Design: 100 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions*. 2012.
- [53] Jakob Nielsen and Rolf Molich. "Heuristic Evaluation of User Interfaces." In: *Conference on Human Factors in Computing Systems*. 1990.
- [54] William Buxton. *Sketching User Experiences: Getting the Design Right and the Right Design*. 2007.
- [55] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. *Research Methods in Human-Computer Interaction*. 2010.
- [56] K Anders Ericsson and Herbert A Simon. *Protocol Analysis: Verbal Reports as Data*. 1984.
- [57] Virginia Braun and Victoria Clarke. "Using Thematic Analysis in Psychology." In: *Qualitative Research in Psychology* (2006).
- [58] Ronggui Huang. *RQDA: R-based Qualitative Data Analysis*. 2018. URL: <http://rqda.r-forge.r-project.org>.
- [59] Jeff Sauro and James R. Lewis. *Quantifying the User Experience: Practical Statistics for User Research*. 2012.
- [60] Alexandr Railean and Delphine Reinhardt. "Let There be LITE: Design and Evaluation of a Label for IoT Transparency Enhancement." In: *Proceedings of the 20th ACM International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI Adjunct)*. 2018.
- [61] Alexandr Railean and Delphine Reinhardt. "OnLITE: On-line Label for IoT Transparency Enhancement." In: *Proceedings of the 25th Nordic Conference on Secure IT Systems*. 2020.
- [62] Alexandr Railean and Delphine Reinhardt. "Improving the Transparency of Privacy Terms Updates." In: *Proceedings of the 9th Annual Privacy Forum*. 2021.
- [63] Simone Fischer-Hübner et al. "Transparency, Privacy and Trust—Technology for Tracking and Controlling My Data Disclosures: Does This Work?" In: *IFIP International Conference on Trust Management (IFIPTM)*. 2016.
- [64] R.C. Lupton and J.M. Allwood. "Hybrid Sankey Diagrams: Visual Snaalysis of Multidimensional Data for Understanding Resource Use." In: *Resources, Conservation and Recycling* (2017).
- [65] Murray State University et al. "Legality and Ethics of Web Scraping." In: *Communications of the Association for Information Systems* (2020).

- [66] Ryan Huebsch. *Digital Documents How Dead are Dead Links?* Tech. rep. URL: <http://citeseerx.ist.psu.edu/viewdoc/versions?doi=10.1.1.124.1678>.
- [67] Diomidis Spinellis. "The Decay and Failures of Web References." In: *Communications of the ACM* (2003).
- [68] Shane D. Johnson et al. "The Impact of IoT Security Labelling on Consumer Product Choice and Willingness to Pay." In: *PLOS ONE* (2020).
- [69] Rob van Diermen. "The Internet of Things: a Privacy Label for IoT Products in a Consumer Market." PhD thesis. 2018.
- [70] Grace Fox et al. "Communicating Compliance: Developing a GDPR Privacy Label." In: *Proceedings of the Americas Conference on Information Systems* (2018).
- [71] Yun Shen and Pierre-Antoine Vervier. "IoT Security and Privacy Labels." In: *Privacy Technologies and Policy*. 2019.
- [72] Pardis Emami-Naeini et al. "Exploring How Privacy and Security Factor into IoT Device Purchase Behavior." In: *Proceedings of the 2019 Conference on Human Factors in Computing Systems (CHI)*. 2019.
- [73] Peter Bihl. *A Trustmark for IoT*. ThingsCon, 2017. URL: <https://thewavingcat.com/iot-trustmark/>.
- [74] A. Boztas, A.R.J. Riethoven, and M. Roeloffs. "Smart TV Forensics: Digital Traces on Televisions." In: *Digital Investigation* (2015).
- [75] Haifa Al Hosani et al. "State of the Art in Digital Forensics for Small Scale Digital Devices." In: *2020 11th International Conference on Information and Communication Systems (ICICS)*. 2020.
- [76] Wazir Zada Khan, Mohammed Y. Aalsalem, and Muhammad Khurram Khan. "Communal Acts of IoT Consumers: A Potential Threat to Security and Privacy." In: *IEEE Transactions on Consumer Electronics* (2019).
- [77] Lilian Edwards and Michael Veale. "Slave to the Algorithm? Why a Right to Explanation is Probably Not the Remedy You are Looking for." In: *SSRN Electronic Journal* (2017).
- [78] Zachary C Lipton. "In Machine Learning, the Concept of Interpretability is Both Important and Slippery." In: *ACM Queue* (2018).

## MEASURING PRIVACY ATTITUDES THROUGHOUT THE LIFE-CYCLE OF IOT DEVICES

---

### AUTHORS

Alexandr Railean<sup>1,2</sup> and Delphine Reinhardt<sup>2</sup>

<sup>1</sup>Unabhängiges Landeszentrum für Datenschutz, Kiel, Germany

<sup>2</sup>University of Bonn and Fraunhofer FKIE Bonn, Germany

PUBLISHED IN Proceedings of the 12th IFIP Summer School on Privacy and Identity Management – the Smart World Revolution (2017). DOI: [10.1007/978-3-319-92925-5\\_9](https://doi.org/10.1007/978-3-319-92925-5_9).

**ABSTRACT** The novelty of the Internet of Things (IoT) as a trend has not given society sufficient time to establish a clear view of what IoT is and how it operates. As such, people are likely to be unaware of the privacy implications, thus creating a gap between the belief of what a device does and its actual behaviour. The responses collected in our online survey indicate that participants tend to see IoT as computer-like devices, rather than appliances, though there are some important misconceptions about the way these devices function. We also find that privacy is a primary concern when it comes to IoT adoption. Nevertheless, participants have a propensity to keep using IoT devices even after they find out that the device abuses their trust. Finally, we provide recommendations to IoT vendors, to make their products more transparent in terms of privacy.

### 3.1 INTRODUCTION

The IoT is composed of *devices, sensors or actuators, that connect, communicate or transmit information with or between each other through the Internet* (adapted from [1]). It is rapidly growing, as the number of connected devices per person has increased from 1.84 to 3.3 between 2010 and 2016 [2, 3]. Many IoT devices, such as light bulbs, power switches, air quality monitors, or fitness trackers, are widely available. There is also strong support in the “do it yourself” community: there are 21,714 hits on Github.com, and 49,000 hits on Instructables.com when searching for the term “IoT”. Moreover, some appliance manufacturers aim at increasing the share of their connected products. For instance, Samsung’s CEO stated that all their products will be part of the IoT by 2020 [4]. Governments have also expressed interest in the IoT. For example, the Federal Trade Commission (FTC) issued a privacy and security guide [5] for businesses involved in IoT development, while the European Commission is working on regulations that have provisions for IoT communications [6]. This indicates that IoT is on the path of becoming an indispensable part of our daily lives, based on the current attention of all involved parties, i.e., enterprises, governments, and end users.

However, such products may expose end users and product owners to privacy risks that can occur at the interplay of factors like resource-constrained hardware, poor usability, ubiquitous deployment or the availability of many pools of data. These factors can make the implementation of well-established privacy and security mechanisms difficult.

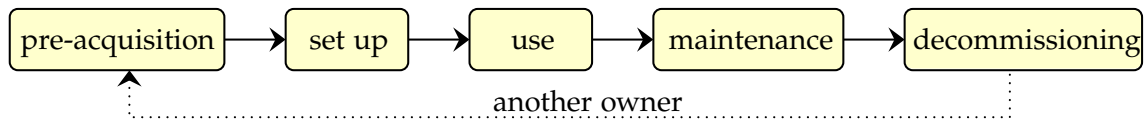
Additionally, users may get little or no feedback about the data collected while interacting with an environment that lacks an interface (e.g. when sensors are seamlessly embedded into walls or furniture). A ubiquitous deployment means that insights about the users can be gathered in locations where they are not expecting data collection. Moreover, linking different data pools having information about the users can facilitate their identification, and hence lead to their deanonymization. For example, studies show that information about a person can be derived by correlating data from disparate sources, such as smartphone sensors [7, 8], social media [9] or online reviews [10]. At the same time, most people are not technically proficient [11], and even those who are often subvert their privacy [12]. This has been shown in the use of social media [13] or instant messengers [14].

This paper starts with a review of related work in Sec. 3.2. We then investigate whether the aforementioned patterns apply to IoT in Sec. 3.3, by means of an online questionnaire introduced in Sec. 3.4. The results, based on the answers of 110 participants, are shared in Sec. 3.5. The answers show that most participants are aware of privacy risks, though they are inclined to keep using a device that infringes on their privacy. Moreover, our results provide an understanding of the reasons behind the adoption of IoT devices by end users, and give a clearer picture of the attention our participants pay to privacy throughout the life-cycle of their IoT devices. We then test our hypotheses in Sec. 3.6. In Sec. 3.7 we discuss the results and limitations of our survey, as well as provide recommendations for IoT vendors. Sec. 3.8 concludes the paper and summarizes our findings. All the materials needed to replicate the survey are given in 3.9.

## 3.2 RELATED WORK

Naeini et al. explore people's preferences regarding IoT data collection and notifications of data collection in [15]. They found that the participants of their study were more open towards data collection in public settings, and less so when data collection occurs in a private environment, if it involves biometric data, or if the data will be stored for long periods of time. They also develop a model that can predict one's data-collection preferences based on three data-points. Other works examine IoT from a legal perspective, a definition of IoT privacy is given in [16], the paper identifies the possible privacy risks related to IoT. Peppet conducts another legal analysis in [17] and discusses how privacy is affected by the difficulty of sensor data de-identification, thus questioning the distinction between personal data and other data. Another raised concern is that some IoT device vendors conflate the notion of "notice" with that of "consent", assuming that informing users about what a technology does is sufficient to indicate that use of technology implies consent ( $S_o$ , please note that the *statements* marked with  $S_n$  will be referred to in Sec. 3.7.2). The analysis also includes a comparison of the packages of several IoT devices with respect to privacy-related information, as well as their privacy policies. An extensive literature review and summary of IoT privacy issues is provided in [18, 19, 20]. Other works are focused on location privacy [21, 22], while [23] focuses on fitness trackers. Volkamer and Renaud discusses the importance of mental models formed by end-users and the role these models play in the trust and acceptance of new technologies in [24]. There are other papers that present IoT life-cycle models, however they take a data-centered approach, examining what happens to the personal data acquired and transmitted by IoT devices [16, 22]. Our work, on the other hand, takes a user-centered

Figure 3.1: IoT device lifecycle



approach, focusing on the different stages of the relationship between users and their IoT devices.

### 3.3 RESEARCH GOALS

To examine the participants' *privacy attitudes* and *user experience* in the context of IoT device ownership, we focus on the following *Research Questions* (RQ):

- RQ<sub>1</sub>: What motivates potential users to acquire IoT devices?
- RQ<sub>2</sub>: Would they continue using a device that infringes on their privacy?
- RQ<sub>3</sub>: Are users aware of the extent to which IoT devices can interact with other equipment they own?

We then map the answers to the corresponding phases of the IoT device life-cycle (defined in Sec. 3.4), and look for user interface friction points that can potentially affect the privacy of end-users. This, in turn, enables us to suggest usability improvements and creates new research questions for the future.

The answers to the research questions help us test the following hypotheses (referred to as H), which are formulated on the basis of autoethnographic observations:

- H<sub>1</sub>: When dealing with IoT devices, most users treat them as *appliances*, rather than *computers*.
- H<sub>2</sub>: Users are inclined to keep IoT devices that infringe on their privacy, if those devices have a high *monetary value*.
- H<sub>3</sub>: Users are inclined to keep IoT devices that infringe on their privacy, if those devices were a *gift from a close person*.

### 3.4 METHODOLOGY

To answer the questions and test the hypotheses, we designed an online questionnaire, which covers the phases of the IoT device life-cycle we consider to have an impact on privacy: pre-acquisition, set-up, usage, maintenance, and decommissioning, as illustrated in Fig. 3.1. Note that we are not concerned with the factors that lead to decommissioning (e.g. resale, recycling, etc), we only focus on the privacy implications due to removal of IoT devices from service, regardless of the cause. In our questionnaire, we take a human-centered perspective and focus on what a person does with the device, rather than on what the device does with the data, in contrast to [16, 22]. We have especially phrased our questions in a way that should elicit what participants *think* about the device and what their *beliefs* about its behaviour are.

Table 3.1: Distribution of points for each considered computer-related skill (Q<sub>30</sub>)

Points	Skills	Points	Skills
2	play video games	5	type complex documents in word processors (e.g. macros, automatic indexes, dynamic fields)
2	view photos and watch videos	10	assemble computers or other electronics from components
2	browse the Internet and send emails	15	I know at least one programming language
2	use a word-processor to type documents		
5	set up email sorting filters		

### 3.4.1 Distribution and audience

We have invited our participants via word of mouth, mailing lists, social media, and survey sharing platforms. Because it appeals to a wide audience, we have particularly taken care that non-experts could understand the goal of our questionnaire. To this end, we have defined and detailed the terminology used and given concrete examples. The introduction also provided key details about how the collected data would be handled, i.e., full anonymity and no disclosure of individual answers.

In total, 193 participants have answered our online questionnaire. Among them, 110 participants have fully filled it out. We have therefore discarded the incomplete ones for computing the following results. The majority of our participants are male (57%), 5% preferred not to disclose their gender. The most represented age category is between 21 and 30 (52%), followed by 31 and 40 (28%), then by 41 and 50 (8%). 45% of the participants have a bachelor degree, 33% have a master degree, 8% have a secondary school level of education, 5% preferred not to disclose information about their education, while 3% have earned a doctorate degree. Geographically, most of our participants are from Eastern Europe (45%), followed by 31% from Western Europe and 14% from North America.

### 3.4.2 Self-selection bias

Since we have initiated the distribution of the survey, it is possible that the recruited participants fit a similar profile, thus biasing the sample. We have therefore asked the participants to indicate the different computer-related skills they have in question Q<sub>30</sub> (see 3.9). We then assign to each skill a number of points according to the distribution presented in Tab. 3.1. The total number of points obtained by a participant finally determines the category they belong to. We categorize participants with a total number of points below 8 as *novice*, between 8 and 20 as *medium*, and greater than 20 as *expert*. Our sample counts 55% rated as expert, 37% are medium and 7% are novice.



### 3.4.3 Priming concerns

To avoid priming participants into a privacy-oriented mindset, the topic of the survey has been announced as “IoT usability”. There was no mention of the term “privacy” in the call for participation, e.g. “You’re invited to participate in an IoT usability survey”. Additionally, privacy-themed questions and answer choices were uniformly distributed among other topics.

## 3.5 RESULTS

Our results are based on the responses of 110 participants and are mapped to phases of our IoT lifecycle model. The first set of questions is aimed at all the participants, whether they own an IoT device or not. We have found that 41% of them do not own IoT devices, whereas the others own smart TVs (38%), smart watches (23%), fitness bracelets (18%), thermostats (12%) and voice assistants (12%) (multiple choices possible). 39% of the participants are planning to purchase new IoT devices in the next 6 months (74% of them already own an IoT device), 30% have no such plans (33% of them own an IoT device), while 27% are not sure about it (47% of them own an IoT device).

### 3.5.1 Pre-acquisition

We have then asked the participants to indicate, in a non-prioritized way, the “reasons to buy Internet-connected appliances” (Q<sub>21</sub>). They have indicated 86 reasons in a free-text field, which we have clustered as follows: automation of routine tasks (38%), better remote control (31%), and new capabilities (31%). Being socially connected (16%) and health improvements (12%) were selected by fewer participants. On the other hand, the participants have given 109 reasons why they would not buy such appliances. The most represented concerns are privacy (34%), security (30%) and cost (12%). Some of the arguments supporting the latter concern being (a) interaction with IoT devices will consume their data plan and inflate the bill, (b) an insecure IoT device that can make purchases can be taken over, allowing hackers to order items for free, (c) the cost of IoT devices is usually greater, due to their novelty, not due to their actual benefits, and (d) these devices become obsolete very fast.

Tab. 3.2 shows what participants would be looking for, if they were purchasing an IoT device. The responses indicate that *convenience* plays a key role. 72% look for ease of use, while 66% seek compatibility with existing devices. We have also seen that privacy is not of particular importance, it ranked 46%, close to “good brand reputation” (48%) and “low price” (47%). Another important highlight is that certifications from organizations like Technischer Überwachungsverein (TÜV) or Federal Communications Commission (FCC) play little role in the choice of IoT devices. Such an attitude may be explained by a greater level of trust in product reviews published on the Internet, or by the fact that brand reputation is sufficient to decide which device to purchase.

Other features mentioned in a free-text field by participants were (a) guaranteed updates period (2 mentions), (b) open hardware/software and firmware access (2 mentions), (c) good security record (3 mentions), (d) wide functionality and customizability (3 mentions). One participant specifically indicated that the privacy policy should be “SHORT and clear” (S<sub>1</sub>).

Table 3.2: Desired IoT features (Q<sub>20</sub>)

Feature	%	Feature	%
ease of use	72	recommendations from friends and others	39
compatibility with my existing devices	66	stylish design	35
good brand reputation	48	availability of technical documentation	35
low price	47	certifications by authorities (e.g. TÜV, FCC)	20
clear privacy policy	46	other (please specify)	8

Table 3.3: IoT benefits that appeal to you personally (Q<sub>23</sub>)

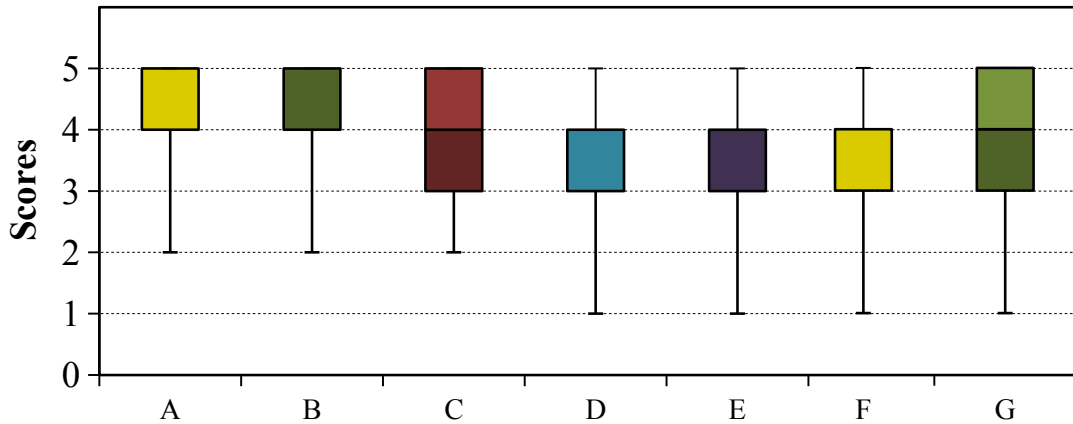
Option	%
automation of routine tasks	59
better remote control	55
new capabilities	52
energy saving	49
easier data management	34
health improvements	30
being connected to friends or family in a new way	26
being connected to strangers or society in general	10
I don't know	10

To learn the reasons why our participants chose to acquire their IoT devices, we have asked them to “[...] indicate the benefits of connected devices that appeal to [them] personally” (Q<sub>23</sub>). Although this question is similar to Q<sub>21</sub>, it enables us to differentiate between benefits participants have heard of in principle, and benefits that they themselves are looking for. The results in Tab. 3.3 show that the responses are similar, the most common and least common reasons follow the same distribution, with a difference in health improvements. 12% chose it as a reason to buy IoT devices, 30% indicated that it is what appealed to them in particular. This observation leads us to the conclusion that in our sample, participants acquire IoT hardware for practical reasons, rather than because it is fashionable to do so.

### 3.5.2 Set up

In this and subsequent sections, we provide the results related to questions that involved participants who own IoT devices. Note that these questions were not displayed to those who indicated that they do not own an IoT device. Therefore the percentages shown are relative to a total of 65 participants. In Q<sub>6</sub>, we have asked participants “how satisfied [they] are with the process of using the device ‘brand’?”, the answers are expressed on a 5-point Likert scale, ranging from “very dissatisfied” (1) to “very satisfied” (5), based on several criteria in Fig. 3.2.

Figure 3.2: Extrema and quartiles of the valid participants' answers to Q<sub>6</sub> based on the following criteria: plugging it in and connecting the cables (A, valid answers: 49), connecting it to [a] network or the Internet (B, 48), configuring the device settings (C, 50), accompanying documentation (D, 46), online materials (e.g. product site, support services) (E, 45), accompanying smartphone application (F, 43), resetting to default settings and wiping all data (G, 37). Invalid answers correspond to participants who skipped the questions or chose not to answer.



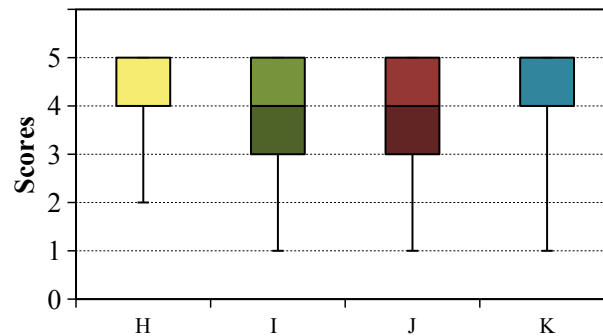
We have found that “satisfied” and “very satisfied” are the most common answers to all the questions, except when it comes to the level of satisfaction with the accompanying documentation, where 42% chose the “neutral” option. A possible explanation is that the manual was never consulted due to lack of need, preference, or lack of interest. Lack of need can be the result of a successful configuration based solely on the clarity of the interface, or the technical experience of the end user. It can also be explained by the fact that the majority of participants rated “online materials (e.g. site, support services)” as “satisfying”, which could indicate that whatever questions they had were addressed online, as such materials are easier or faster to search.

We have further probed this matter by asking participants “when it comes to configuring [the IoT device], how much do [they] agree with the following statements” in Q<sub>9</sub>, and find that 71% agreed and strongly agreed to being able to set up and configure their device without reading the manual (Fig. 3.3). This supports the assumption that *lack of need* is what leads to the documentation being neglected. Such a level of success can have an undesired effect: satisfied end-users can stop tinkering with the device as soon as they accomplish their primary goals, thus missing potentially critical security and privacy tips the documentation could offer. We conclude that important privacy-related controls should be incorporated into the initial setup procedure, to ensure that end-users make informed privacy-related decisions (S<sub>2</sub>).

### 3.5.3 Usage

When asked about continued use of an IoT device that infringes on the owner’s privacy (Q<sub>24</sub>), two of the top three reasons are related to the monetary value of the product, “it was an expensive purchase” and “it is difficult to return it or get a refund” got a combined score of 53%. In contrast, options related to family values are the least convincing reasons to keep it (14%). Other mentioned reasons were: (a) if it provides a unique func-

Figure 3.3: Extrema and quartiles of the valid participants' answers to Q<sub>9</sub> based on the following criteria: configuring the device is easy (H, valid answers: 55), configuring it via a smartphone app is easy (I, 54), configuring it via a web-interface is easy (J, 54), set it up without reading the manual (K, 53).



tion, (b) if it is crucial for daily use, or (c) if the infringement is negligible. *Convenience* is a major factor and its importance is often expressed throughout the collected answers. We have found that *entertainment* scores as high as health-related benefits (20%). This attitude resonates with the “dancing pigs” adage in computer security: “*The user’s going to pick dancing pigs over security every time*” [25]. While studies [26] concluded that a better user interface helps people make wiser security-related decisions, those findings are not necessarily applicable in our context. Our question asks about a participant’s choice *in principle*, which implies that this is a conscious decision they would make, no matter what the interface looked like.

When it comes to discarding an IoT device that infringes on the owner’s privacy (Q<sub>25</sub>), the reasons chosen by participants were: “ethical and moral convictions” (46%), “it is easy to get a refund” (45%), “installing custom firmware voids the warranty” (38%), and “it is easy to re-sell” (32%). Among the reasons indicated in the free-text field, 2 participants mentioned that the decision depends on the magnitude of the infringement.

To get a better understanding of what IoT device owners think about the capabilities of their hardware, we have asked them to indicate “the resources [they] think are exposed to the IoT device” in Q<sub>7</sub>. The distribution of the answers is shown in Tab. 3.4. In 69% of the responses, it is expected that an IoT device can interact with a smartphone, presumably because that is how it is configured and controlled. Other options have been chosen by fewer than 40% of the participants.

We have asked participants “who, in [their] opinion, can use, or otherwise interact with IoT [devices] installed in your home?” in Q<sub>8</sub>. The responses show that 35% of participants consider that hackers are capable of doing so, while 13% think the government can do that as well. These numbers indicate that the efforts of IoT device vendors are insufficient to establish trust and convince the participants that their product is secure (S<sub>3</sub>), as it has been argued in [24]. We have also found, by means of a Kruskal-Wallis test, that expert participants are more likely ( $\chi^2 = 6.857$ ,  $p = 0.032$ )<sup>1</sup> to consider that the government can access their IoT hardware. Note that they do not hold the same opinion about hackers. This may be explained by an expert’s confidence in their own ability to secure a system from typical attackers. On the other hand, their awareness

<sup>1</sup> When  $p \leq 0.05$ , it indicates that the results are not likely to be caused by chance, and that another set of participants would provide similar answers.

Table 3.4: Which of these resources you think are exposed to the IoT device? (Q7)

Option	%
my smartphone	69
other computers on my home network	40
communications between other devices in my home and the Internet	31
purpose-specific data (e.g. temp., humidity)	25
other devices on my home network (e.g. printer)	24
communications between devices in my home	22
other computers on the Internet	15
I don't know	11

Table 3.5: Who can interact with the IoT device? (Q8)

Option	%
me	84
others in my household (e.g. family)	65
the manufacturer	38
hackers	35
the government	13
my neighbors	4

of the fact that state-level actors have much more resources may justify the belief that governments could conduct successful attacks, if they choose so. We have finally asked our participants whether they have “examined the privacy policy” of their IoT device in Q<sub>12</sub>, and find that 22% have done so. To understand whether IoT device adoption is a conscious decision, rather than a forced one (i.e. the IoT-enabled device was purchased because there was no “dumb” analog), we have asked our participants if they “own any appliances, the IoT capabilities of which are not used” (Q<sub>17</sub>). 22% of the participants who own IoT devices always use the IoT features, 5% turn them off explicitly, 5% are aware of the features but are ignoring them, while 2% use various external means to disable them. Among the recorded means, we have found stickers over cameras (two mentions), positioning the device with the camera pointing down (one mention) and using a network router to limit the traffic of particular devices (one mention).

#### 3.5.4 Maintenance

To understand the participants’ attitudes towards software updates, we have asked them “do [they] think IoT devices require software updates?” (Q<sub>4</sub>). 92% consider that IoT devices require software updates, 5% do not know if that is the case, while 3% believe that updates are not necessary. In Tab. 3.6, we present the answers to the question “who should be responsible for updating the IoT device, in your opinion?” (Q<sub>5</sub>). Although 60% of the participants consider that the manufacturer should be responsible for pushing updates to IoT devices (S<sub>4</sub>), two participants indicated that they want to be the ones who decide whether an update is installed or not. This could be the result of prior experience with unwanted updates, that disabled useful features or added undesired ones (S<sub>5</sub>). This could explain why some are aware of the availability of newer versions, but are not installing them (Tab. 3.7).

The results indicate that our participants see IoT devices as computer-like systems that require software updates, rather than “plug in and forget” devices. We emphasize that

Table 3.6: Who should be responsible for updating IoT devices? (Q<sub>5</sub>)

Option	%
the manufacturer	60
me, as the device owner	44
the seller of the device	15
a government agency	1
I don't know	1

Table 3.7: Is your IoT device running fully up-to-date firmware/software? (Q<sub>3</sub>)

Option	%
N/A, I do not own any IoT device	41
yes, it updates itself automatically	27
yes, I update it manually	11
I don't know	10
no, but newer firmware is available	5

the most common expectation is for the updates to be rolled out by the manufacturer. This is an important point to be considered by IoT device designers, because if this expectation will not be met, it is possible that the devices will run outdated firmware, potentially exposing owners to security and privacy risks. The data also reveal a gap between those who expect updates to be automatically installed by the manufacturer (60%) and those who are aware that updates are automatic and are certain that their IoT device uses the latest version (27%). This difference could be explained in different ways, e.g. the IoT devices do not adequately reflect their update availability status (if at all) (S<sub>6</sub>) or end users did not bother to check that. We measure that, using a 5-point Likert scale, by asking participants “How well does the device [...] express what it is currently doing?”, listing several use cases, of which one is “installing an update” (Q<sub>10</sub>). We have found that participants consider this to be expressed clearly (20%) to very clearly (35%), while another 20% have not experienced this use case. Sec. 3.5.5 discusses other implications related to update policies.

### 3.5.5 Decommissioning

To determine whether participants have gone through this procedure and measure their level of satisfaction with it, we have asked them “how satisfied are you with the process of [...] resetting [...] to default settings and wiping all data?” (Q<sub>6</sub>) and “how well does the device express [...] that it is currently resetting itself to default settings and wiping the data?” (Q<sub>10</sub>). We have found that the many of our participants have not had the experience of wiping the data off their IoT device (31%) or have not had the chance to see how this process is reflected in the interface (45%). It should be noted that some of the participants could have chosen the “N/A” option because their IoT device does not provide such a feature or it is not relevant for its function, the survey does not distinguish between these possibilities. Since this use case has been less explored by end users, manufacturers have fewer opportunities to receive feedback about this procedure. Thus, any existing usability shortcomings can possibly remain in the product for a longer period of time. In contrast, use cases related to set up and usage are likely to attract far more attention. We conclude that IoT device manufacturers should not perceive the lack of customer complaints as an indicator of good usability of their product in the decommissioning phase. Instead, they ought to conduct tests targeting this particular scenario (S<sub>7</sub>).

### 3.6 TESTING THE HYPOTHESES

In what follows, we successively test the hypotheses defined in Sec. 3.3, based on the answers given by participants.

3.6.0.1  $H_1$ : *When dealing with IoT devices, most users treat them as appliances, rather than computers.*

On one hand, the arguments detailed in Sec. 3.5.4 suggest that most of the participants consider IoT devices to be computers, rather than appliances, based on their awareness of the fact that such devices require regular updates and have to be secured. However, the analysis in Sec. 3.5.3 indicates that this awareness is limited. For example a smart TV that runs an operating system with network capabilities is exposed to all of the resources listed in Q<sub>7</sub>, yet the participants' responses failed to reflect that. This could mean that some participants' level of confidence exceeds their actual understanding, which can lead to the false belief that the measures taken to protect their privacy are sufficient, when they are not. We cannot definitively support or refute  $H_1$ , because the premise appears to be wrong. It is possible that there exists another model in the spectrum between *computer* and *appliance*, which describes more accurately how IoT devices are perceived. For example, participants may be used to smartphones and tablets, which require updates, but are nevertheless not treated as computers.

3.6.0.2  $H_2$ : *Users are inclined to keep IoT devices that infringe on their privacy, if those devices have a high monetary value.*

The sampled population perceives privacy as a major concern in IoT adoption, but the concern can be overridden if the purchased IoT hardware was expensive, if it has an entertainment or utility value. In these circumstances, a substantial number of participants would continue using an IoT device, even if they are certain that it infringes on their privacy (Q<sub>24</sub>, Q<sub>25</sub>). This can be partially explained by *loss aversion*, thus what matters is whether the owner can get reimbursed easily, regardless of the cost of the IoT device. When a refund is not possible, or if it is a tedious process, an inexpensive device is more likely to be discarded than an expensive one. Thus  $H_2$  is supported, although we have to emphasize that other factors are at play.

3.6.0.3  $H_3$ : *Users are inclined to keep IoT devices that infringe on their privacy, if those devices were a gift from a close person.*

We have also found, by means of a Mann-Whitney U test, that females are more likely to keep using a rogue IoT device ( $U = 1066$ ,  $n = 42$ ,  $p = 0.012$ )<sup>2</sup> if it was a gift from a close person, thus  $H_3$  is partially supported. It is possible that such attitudes are caused by emotional attachment to a person, however there may be other conditions too, e.g. the device has a likeable design, or it stores valuable content, like photographs. These additional factors were not checked by the questionnaire, so they should be investigated separately.

<sup>2</sup> This indicates that the results are not likely to be caused by chance, and that if the same questions were given to other participants, the results would be similar.

### 3.7 DISCUSSION

The answers to Q<sub>7</sub>, “Which of these resources you think are exposed to the IoT device?” discussed in Sec. 3.5.3 could be a reason of concern. For example, in the case of a smart TV, a typical feature is to stream videos from remote sources, which requires some form of communication over networks, such as the Internet. This, in turn, implies that the device has to have an implementation of a network stack and software that leverages it. However, only two participants (rated at a medium skill level) indicated that their smart TV can access both, computers on their home network as well as other computers on the Internet. The same reasoning applies to voice-activated assistants (e.g. “Amazon Echo”). Only one participant correctly identified that their “Echo” can interact with local and remote hosts, which means that some participants are unaware of the fact that this device can transmit information via the Internet. While it is possible that some IoT devices are deliberately constrained by their owners (e.g. using firewalls), this should not be the case for assistants like “Echo”, because they rely on an Internet connection for their basic features. Moreover, configuring Internet access is a required step in the setup phase, which the participants had to go through. This could be explained by the fact that they have an incomplete understanding of the capabilities of their device, or that someone else configured it for them (S<sub>8</sub>). Product designers should consider this, because some of the user categories who could benefit from IoT, such as the elderly, may not be digitally literate, yet they must be aware of the implications of using the IoT device. Either the set-up procedure should be easy enough for anyone, or there should be a separate privacy summary that does not use technical or legal jargon and is easy to understand. We did not anticipate such results, therefore our survey was not crafted in a way that would enable us to determine whether this is a deliberate decision made by manufacturers, or an oversight, thus this matter has to be investigated separately.

Another important aspect is *obsolescence*, which we examine by analogy with smartphones. For example, the most common version of Android today has a market share of 31%, it was released two years ago [27]. The two latest versions, 8.0 and 7.1, have a combined market share of 3.3%. Thus, a substantial number of smartphones are running outdated software. This is one of the reasons why the American Civil Liberties Union (ACLU) filed an FTC complaint over Android security issues [28]. If the same pattern arises in IoT, end-users will be stuck with outdated devices which, at best, can only be secured by applying external technical means (e.g. firewalls) or custom firmware. Neither of these options is novice-friendly. A strategy consumers can adapt is to decommission the device before the support period ends. While this solves *their* problem, the obsolete device will become someone else’s problem. This creates the premises for a “tragedy of the commons” [29], where the cost of security and privacy risks is distributed among all Internet users, instead of affecting IoT vendors or users specifically. Thus, the incentives to continue supporting and updating these devices is weak. This problem should be resolved in the future, otherwise it could hinder IoT adoption (S<sub>9</sub>).

We have found some variation in attitudes, based on technical skills. Experts are more likely to indicate that they use a firewall, encrypted volumes and ad-blockers. They are also better-informed about IoT-related privacy and security news such as those about the Mirai botnet or the German steel factory incident. Note that we chose these topics because they were also covered by the international mainstream press, so non-experts could have heard about them. More surprisingly, the expert participants in our sample



are also more likely to consider that manufacturers should be responsible for deploying IoT updates.

Note that our tests show that gender, age, and location do not have a significant impact on the participants' answers, unless otherwise stated.

### 3.7.1 *Limitations*

We encountered several limitations while running the survey. Firstly, people below the age of 18 were excluded, because of strict EU regulations concerning data collection from minors. However, this population segment could represent a significant portion of IoT technology consumers, thus their opinions should be accounted for. Secondly, we reached out to a technologically proficient audience (only 7% fell into the "novice" category), which is not representative of society in general. The modest number of participants finally gave us some hints about questions worth pursuing, but a study of a larger scale is required to make definitive claims about privacy attitudes.

### 3.7.2 *Recommendations for IoT vendors*

Based on the different statements  $S_0$  to  $S_9$  we highlighted in the paper, we would like to make the following recommendations to IoT manufacturers, to improve their privacy practices:

- $S_0$  Do not conflate "notice" with "consent" (based on [17])
- $S_1$  Write concise privacy policies
- $S_2$  Make privacy-related settings a mandatory part of the set-up phase
- $S_3$  Find ways to address people's security and privacy concerns
- $S_4$  Provide an automatic update feature
- $S_5$  Make the list of version changes public
- $S_6$  Reflect the update availability status clearly
- $S_7$  Include decommissioning in usability tests
- $S_8$  Consider that someone other than the end-user can set up the IoT device
- $S_9$  Planned obsolescence should be more future-oriented

## 3.8 CONCLUSIONS

We have organized an online survey with 110 participants, to explore their privacy attitudes towards IoT devices. The results reveal a generally positive opinion about IoT, despite the awareness of existing privacy and security risks. The challenge is to address these issues before the end-users' skepticism creates a barrier in IoT adoption.

We have found a potential void in the user experience related to the decommissioning of such devices. Most participants have not gone through such a use case and there

is a possibility that they will run into issues when they do so. Device manufacturers should consider this before releasing their products to the market. We have also found that the expected norm is that IoT devices are updated automatically and that it is the responsibility of the manufacturer to ensure the smoothness of the process. IoT device designers should implement such a capability in their product and provide clear information to end users when automatic updates are not available, and it is the user's responsibility to keep the device up to date.

### 3.8.0.1 Acknowledgments

This research is funded by H2020 MSCA ITN Privacy&Us (project no 675730). We would like to thank the survey participants, Harald Zwingelberg and the anonymous peer reviewers for their helpful comments.

## 3.9 APPENDIX: QUESTIONNAIRE

The questions that featured in the survey are shown in Tab. 3.8. The list does not include the provided choices or other accompanying materials, they are available at <https://www.datenschutzzentrum.de/projekte/privacy-us/>. The site also provides the source code needed to replicate the survey and analyze the data.

Note that not all questions were shown to all participants (e.g. those who do not own IoT devices were not asked about their experience with such products). The label 'brand' was replaced with the IoT device name provided by participants in Q<sub>2</sub>. The table also mentions the type of each question, FT: free-text, MS: questions that allowed *several* options to be selected at the same time, MC: questions for which participants had to choose *only one* option out of several, L: Likert scale questions.

Table 3.8: Survey questions

ID	Type	Question
Q <sub>1</sub>	MS	Which of these IoT appliances do you own?
Q <sub>2</sub>	FT	Focus on a specific device (note: here the participant is asked to name a specific device they own)
Q <sub>3</sub>	MC	Is the selected device running fully up-to-date software/firmware?
Q <sub>4</sub>	MC	Do you think IoT devices require software updates?
Q <sub>5</sub>	MS	Who should be responsible for updating the device, in your opinion?
Q <sub>6</sub>	L	How satisfied are you with the process of using the device 'brand'?
Q <sub>7</sub>	MS	Which of these resources you think are exposed to the device 'brand'?
Q <sub>8</sub>	MS	Who, in your opinion, can use, or otherwise interact with a 'brand' installed in your home?
Q <sub>9</sub>	L	When it comes to configuring the device 'brand' how much do you agree with these statements?
Q <sub>10</sub>	L	How well does the device 'brand' express what it is currently doing?
Q <sub>11</sub>	L	How confident are you that the device 'brand' respects your privacy?
Q <sub>12</sub>	MC	Have you examined the privacy policy of 'brand'?

- Q<sub>13</sub> FT What would make the device 'brand' more usable, in your opinion?
- Q<sub>14</sub> FT What are the most important things that you like in 'brand'?
- Q<sub>15</sub> FT What do you dislike the most about your experience with 'brand'?
- Q<sub>16</sub> MC Do you plan to buy any IoT devices in the next 6 months?
- Q<sub>17</sub> MC Do you own any appliances, the IoT capabilities of which are not used?
- Q<sub>18</sub> FT If you answered "yes" above, please list those appliances here. Optionally, indicate the feature.
- Q<sub>19</sub> MC Do you think it is possible that some of your devices or appliances are connected to the Internet without your knowledge?
- Q<sub>20</sub> MS Which qualities would you be looking for if you were buying an IoT device?
- Q<sub>21</sub> FT What are the reasons to buy Internet-connected appliances, in your opinion?
- Q<sub>22</sub> FT What are reasons NOT to buy such appliances, in your opinion?
- Q<sub>23</sub> MS Please indicate the benefits of connected devices that appeal to you personally.
- Q<sub>24</sub> MS You discover that an IoT device infringes on your privacy and you have no capability to change that. Which of these reasons will influence you to KEEP the device?
- Q<sub>25</sub> MS You discover that an IoT device infringes on your privacy and you have no capability to change that. Which of these reasons will influence you to DISCARD the device?
- Q<sub>26</sub> MC If you have a WiFi network at home, which of the options below best describes its security settings
- Q<sub>27</sub> MS Which of these security tools have you got on your computer?
- Q<sub>28</sub> MC What is your age?
- Q<sub>29</sub> MC What is your gender?
- Q<sub>30</sub> MS Please specify the computer-related skills you have.
- Q<sub>31</sub> L Have you heard anything about these in the news?
- Q<sub>32</sub> MC What is the highest level of education that you successfully completed?
- Q<sub>33</sub> MC Which of these best describes your location?
- Q<sub>34</sub> FT If you have any remarks that you would like to make, please use the form below.

## REFERENCES

- [1] *Internet of things: Privacy & Security in a Connected World*. Staff report. FTC, 2015. URL: <https://ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf>.
- [2] Dave Evans. *The Internet of Things - How the Next Evolution of the Internet Is Changing Everything*. Cisco, 2011. URL: [http://www.cisco.com/c/dam/en\\_us/about/ac79/docs/innov/IoT\\_IBSG\\_0411FINAL.pdf](http://www.cisco.com/c/dam/en_us/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf).
- [3] *Trends 17*. Globalwebindex, 2016. URL: <http://insight.globalwebindex.net/hubfs/Reports/Trends-17.pdf>.
- [4] *Samsung: By 2020, All of Our Products Will Be Connected to the Web*. URL: <http://mashable.com/2015/01/05/samsung-internet-of-things>.
- [5] *Careful Connections: Building Security in the Internet of Things*. 2015. URL: <https://ftc.gov/system/files/documents/plain-language/pdf0199-carefulconnections-buildingsecurityinternetofthings.pdf>.
- [6] *Regulation of the European Parliament and of the Council Concerning the Respect for Private Life and the Protection of Personal Data in Electronic Communications and Repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications)*. 2017.
- [7] Delphine Christin. "Privacy in Mobile Participatory Sensing: Current Trends and Future Challenges." In: *Journal of Systems and Software* (2016).
- [8] Nicholas D. Lane et al. "On the Feasibility of User De-anonymization From Shared Mobile Sensor Data." In: *Proceedings of the 3rd International Workshop on Sensing Applications on Mobile Phones (PhoneSense)*. 2012.
- [9] M. Kosinski, D. Stillwell, and T. Graepel. "Private Traits and Attributes Are Predictable From Digital Records of Human Behavior." In: *Proceedings of the National Academy of Sciences* (2013).
- [10] Arvind Narayanan and Vitaly Shmatikov. "How to Break Anonymity of the Netflix Prize Dataset." In: *arXiv preprint cs/0610105* (2006).
- [11] OECD. *Skills Matter*. OECD Skills Studies. 2016. URL: [http://www.oecd-ilibrary.org/education/skills-matter\\_9789264258051-en](http://www.oecd-ilibrary.org/education/skills-matter_9789264258051-en).
- [12] Ruogu Kang et al. "'My Data Just Goes Everywhere' User Mental Models of the Internet and Implications for Privacy and Security." In: *Proceedings of the 11th Symposium on Usable Privacy and Security (SOUPS)*. 2015.
- [13] Susan B. Barnes. "A Privacy Paradox: Social Networking in the United States." In: *First Monday* (2006).
- [14] Alexander De Luca et al. "Expert and Non-Expert Attitudes Towards (Secure) Instant Messaging." In: *Proceedings of the 12th Symposium on Usable Privacy and Security (SOUPS)*. 2016.
- [15] Pardis Emami Naeini et al. "Privacy Expectations and Preferences in an IoT World." In: *Proceedings of the 13th Symposium on Usable Privacy and Security (SOUPS)*. 2017.

- [16] Jan Henrik Ziegeldorf, Oscar Garcia Morchon, and Klaus Wehrle. "Privacy in the Internet of Things: Threats and Challenges." In: *Security and Communication Networks* (2014).
- [17] Scott R. Peppet. "Regulating the Internet of Things: First Steps Toward Managing Discrimination, Privacy, Security and Consent." In: *Texas Law Review* (2014).
- [18] Luigi Atzori, Antonio Iera, and Giacomo Morabito. "The Internet of Things: A Survey." In: *Computer Networks* (2010).
- [19] Xavier Caron et al. "The Internet of Things (IoT) and its Impact on Individual Privacy: An Australian Perspective." In: *Computer Law & Security Review* (2016).
- [20] Diego M. Mendez, Ioannis Papapanagiotou, and Baijian Yang. "Internet of Things: Survey on Security and Privacy." In: *arXiv:1707.01879 [cs]* (2017).
- [21] Mahmoud Elkhodr, Seyed Shahrestani, and Hon Cheung. "A Review of Mobile Location Privacy in the Internet of Things." In: *Proceedings of the 10th International Conference on ICT and Knowledge Engineering*. 2012.
- [22] Robert P. Minch. "Location Privacy in the Era of the Internet of Things and Big Data Analytics." In: *Proceedings of the 48th Hawaii International Conference on System Sciences (HICSS)*. 2015.
- [23] Wei Zhou and Selwyn Piramuthu. "Security/Privacy of Wearable Fitness Tracking IoT Devices." In: *Proceedings of the 9th Iberian Conference on Information Systems and Technologies (CISTI)*. 2014.
- [24] Melanie Volkamer and Karen Renaud. "Mental Models - General Introduction and Review of Their Application to Human-Centred Security." In: *Lecture Notes in Computer Science*. 2013.
- [25] Bruce Schneier. *Secrets and Lies: Digital Security in a Networked World*. 2008.
- [26] Devdatta Akhawe and Adrienne Porter Felt. "Alice in Warningland: A Large-Scale Field Study of Browser Security Warning Effectiveness." In: *Usenix Security*. 2013.
- [27] *Android API Versions*. URL: <https://developer.android.com/about/dashboards/index.html>.
- [28] *ACLU Files FTC Complaint Over Android Smartphone Security*. URL: <https://aclu.org/blog/national-security/aclu-files-ftc-complaint-over-android-smartphone-security>.
- [29] Garrett Hardin. "The Tragedy of the Commons." In: *Journal of Natural Resources Policy Research* (2009).



## LET THERE BE LITE: DESIGN AND EVALUATION OF A LABEL FOR IOT TRANSPARENCY ENHANCEMENT

---

### AUTHORS

Alexandr Railean<sup>1,2</sup> and Delphine Reinhardt<sup>2</sup>

<sup>1</sup>Unabhängiges Landeszentrum für Datenschutz, Kiel, Germany

<sup>2</sup>Institute of Computer Science, Georg-August-Universität Göttingen, Germany

**PUBLISHED IN** Proceedings of the 20th ACM International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI Adjunct, 2020). DOI: [10.1145/3236112.3236126](https://doi.org/10.1145/3236112.3236126).

**ABSTRACT** We present a “privacy facts” label, which aims at helping non-experts understand how an Internet of Things (IoT) device collects and handles data. We describe our design methodology, and detail the results of our user study involving 31 participants, assessing the efficacy of the label. The results suggest that the label was perceived positively by the participants, and is a promising solution to help users in making informed decisions.

### 4.1 INTRODUCTION

The IoT is composed of *devices, sensors or actuators, that connect, communicate or transmit information with or between each other through the Internet* [1]. Ubiquitous use of such technology can have major privacy implications for its users, as well as non-users, who may be unaware of IoT devices in their environment [2, 3]. For example, TV content can be identified from smart energy meter data [4]. Another problem is that users have little awareness of how the data collected by IoT devices are handled [5].

The *General Data Protection Regulation* (GDPR) aims to address some of these risks. It applies to entities that handle personal data of EU citizens, and requires organizations that legally control the data to “take appropriate measures to provide any information [...] relating to processing to the data subject in a *concise, transparent, intelligible and easily accessible form, using clear and plain language* [...]” [6]. In this paper, we map these requirements to a *Label for IoT Transparency Enhancement* (LITE), as shown in Fig. 4.1, that can be distributed with an IoT device, to assist potential buyers in protecting their privacy *before* acquiring the device. This is the earliest point in time, where important privacy-preserving decisions can be made [1]. The main contributions of our work are the label design and the conclusions of the user study we conducted to assess its clarity.

### 4.2 REQUIREMENTS AND DESIGN SPACE ANALYSIS

The primary goal for the label is to be informative, and answer these questions:

- *What* data are collected? (referred to as  $Q_{\text{what}}$ )

Figure 4.1: “Privacy facts” label for IoT devices.



- What is the purpose of collection? ( $Q_{\text{purpose}}$ )
- Where are the data stored? ( $Q_{\text{where}}$ )
- How long are they kept? ( $Q_{\text{duration}}$ )
- Who has access to the data? ( $Q_{\text{who}}$ )

The list is based on the GDPR and the transparency recommendations [7] of the *Article 29 Working Party (WP29)*, an advisory group of representatives from European *Data Protection Authorities (DPA)*. We then extend it with questions derived from autoethnographic observations:

- What do the data look like? ( $Q_{\text{sample}}$ )
- How to access the data? ( $Q_{\text{access}}$ )
- How frequently are the data sent? ( $Q_{\text{freq}}$ )
- Which communications are protected? ( $Q_{\text{sec}}$ )
- What paths do the data follow? ( $Q_{\text{path}}$ )



- What information does the device receive from other sources? ( $Q_{rcv}$ )

In addition, we set these usability requirements: facilitate side-by-side comparison, be compatible with printed and digital media, maintain utility even when shown in gray-scale, be short and simple. Finally, the label has to be future-proof, rather than over-fitted to a particular class of devices.

### 4.3 LABEL DESIGN METHODOLOGY

We structure the design as follows: the *information area* on top is grouped by the questions  $Q_{what}$ ,  $Q_{where}$ ,  $Q_{duration}$ ,  $Q_{who}$ ,  $Q_{purpose}$ . It features a sample of collected data, in the form of a *Quick Response* (QR) code. The QR answers  $Q_{sample}$ , by illustrating a *concrete* set of values, which can improve understanding. For example, a “customer number” can look like “481-AHR-1831”, but it could also take forms that reveal more information, e.g., an email address. To keep the label self-contained, the QR holds human-readable text, as seen in List. 4.31, rather than a link to a site.

The lower part is a *trace view* [8] of the data flows involving the IoT device, it answers  $Q_{sec}$ ,  $Q_{path}$  and  $Q_{rcv}$ . It aims at helping users understand if they operate the IoT device directly, or if it relies on systems outside of their control.

We follow these visual guidelines:

- group related elements [9],
- use indentation to express hierarchy,
- facilitate quick scanning by using bullet points in lists,
- and by emphasizing section titles,
- provide redundant encoding of information via icons,
- use gray-scale, to ensure that LITE is print-friendly and that color-blindness does not hinder readability,
- use additional emphasis to facilitate side-by-side comparison of key parameters,
- keep the number of elements at each level of abstraction below Miller’s “magic number  $7 \pm 2$ ” [10], to reduce the cognitive load when comparing devices.

To make the text accessible to non-experts, we have avoided specialized terms, e.g., “Internet address” instead of “IP address”. We choose words that have a more generic meaning, e.g., “software” instead of “firmware”. We follow the *progressive disclosure* principle and omit low-level information. For instance, we use the padlock icons as security indicators, instead of mentioning algorithms and key lengths. This reduces clutter and removes terms that might not be clear to a novice. Another choice in favour of simplicity is to refrain from listing all the sensors, actuators and connectivity interfaces. Some devices may integrate mechanisms that are not exposed to users, e.g., noise-cancelling headphones may use microphones to improve noise suppression, possibly contradicting one’s mental model of “headphones produce sound, they do not record it”.

Further simplifications are achieved by focusing on *collected*, rather than transmitted data. The GDPR holds organizations accountable for the data they have, rather than the data which may be, in principle, extracted from the metadata of communication

protocols, or derived via post-processing. This also guards against cases where an IoT device is privacy-friendly, while its accompanying smartphone application is not, as it may collect other data using the phone. Given that the data from the device and the smartphone end up on the same online service, they all become “collected data”. As such, it would take a greater effort to conceal potentially abusive privacy practices.

The “purpose” section of the label guards against *purpose creep*, which occurs when collected data are used in ways other than originally declared. When this information is stated upfront, users can decide for themselves if the data are applicable to the purpose.

#### 4.4 EVALUATION

To test the clarity and readability of LITE, we have designed a study that elicits answers to questions about how a mock-up IoT product handles data.

##### 4.4.1 Recruitment

In February 2018, 31 participants were recruited among the students and staff of the University of Karlstad, Sweden. To get a better approximation of non-expert consumers, we have focused our recruitment efforts on areas outside the computer science department. The invitation referred to an “evaluation of a privacy label for IoT (Internet of Things) products” and announced that 6 coupons for the university cafeteria worth 8.5 EUR (10.5 USD), would be randomly distributed after the study. No ethical committee approval was necessary according to the university’s regulations.

##### 4.4.2 Demographics

52% of the participants are female, 48% are male. 58% of the participants are between 18 and 26 years, followed by 27 and 35 years (35%). We measure their *self-reported* technical competence in Q<sub>12</sub> (see [Appendix: Questionnaire](#)), by assigning points to each skill, according to Tab. 4.1. The skill category is determined by the sum of points. As in [1], we have categorized participants with a total number of points below 8 as *novice*, between 8 and 20 as *medium*, and greater than 20 as expert. In our sample, 29% are classified as medium and 23% as novice, the rest are expert.

##### 4.4.3 Experiment Settings

We first gave the participants a consent form, that explains how the information collected during the experiment will be used. Then, we provided a mock-up IoT device and a 128mm × 40mm “privacy facts” label with these instructions: “*You are holding a prototype device produced by Tesami GmbH, it is called “Hausio” and it keeps track of the temperature and humidity in your house. The accompanying “privacy facts” label summarizes how the data are collected and handled. Take as much time as you want to examine the device and the label. When ready, please proceed to the questionnaire*”. We then asked participants to examine the items and fill out our questionnaire, available in Appendix A. Participants were then left alone, having LITE with them all the time. When done, they notified the examiner, who asked follow-up questions and recorded the interview (average duration was 7 minutes).

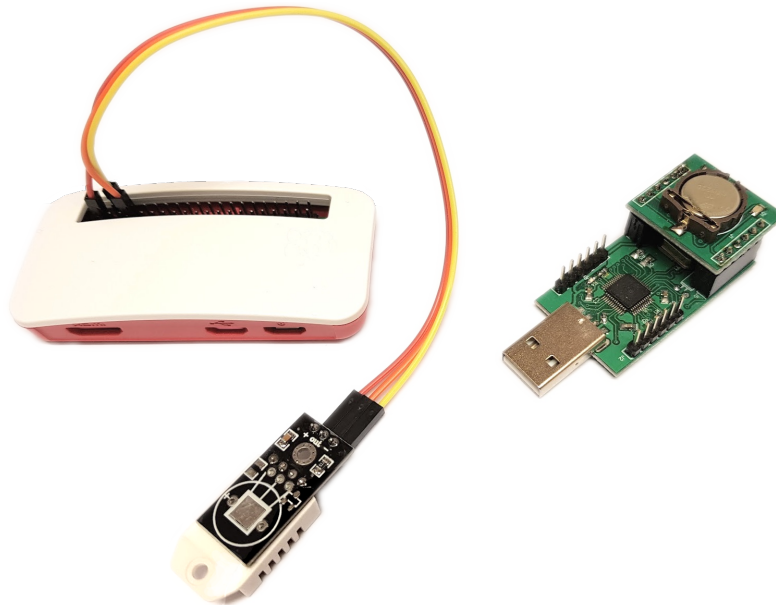
Points	Skills
2	play video games
2	browse the Internet and send emails
2	view photos and watch videos
2	use a word-processor to type documents
5	set up email sorting filters
5	type complex documents in word processors (e.g., macros, automatic indexes, dynamic fields)
10	assemble computers or other electronics from components
15	I know at least one programming language

Table 4.1: Distribution of points for each computer-related skill (Q<sub>12</sub>).

A mock-up device is used to make the experiment more realistic and link LITE to a tangible item. We have used two mock-ups (Fig. 4.2), to check if there is any difference in responses depending on the device. Half of the participants were given a RasPi Zero, the other half got a custom board. We have chosen not to distribute the items in a product box, because it could potentially distract participants from the label, which is the focus of the study.

The transcripts were independently coded by two researchers, who counted the references to label sections, and tagged the participants' interpretation of the "product improvement" purpose listed on the label, as "suspicious" (e.g., intentionally vague, potentially abusive) or "not suspicious".

Figure 4.2: Mock-ups used: RasPi Zero with a DHT-22 sensor (**left**), custom board (**right**).



## 4.5 RESULTS

For a quantifiable evaluation, we count the number of errors in the completed questionnaires, compiled in Fig. 4.3. The score treats any deviation from the correct answer as a separate error. For example, in Q<sub>1</sub> “what purpose are the data collected for?”, the expected answer is to check “my personal use” and “scientific research”, and to write “targeted ads” and “product improvement” in the custom fields. The following deviations would amount to 4 errors: checking another box (1 error), not checking one of the correct ones (1 error) and not filling out correct values in both custom fields (2 errors). The maximum number of errors one can make is 23. Note that Q<sub>5</sub> and Q<sub>6</sub> do not count towards this total, as they are open to interpretation and are exploratory.

We consider the following types of errors: *check incorrect* (i.e., a wrong box is checked), *uncheck correct* (i.e., a correct box is not checked), *custom missing* (i.e., a custom entry field was left empty), *custom incorrect* (i.e., a custom entry field contains an incorrect value).

### 4.5.0.1 Q<sub>1</sub> What purpose are the data collected for?

This entry has the largest number of errors, 87% of the participants made at least one. 54% of these errors are of the *custom missing* type, while none of the other questions have had such errors in their responses.

This could be an artifact of our questionnaire, as most participants have correctly checked the right options from the list, but did not fill in the custom ones, thus taking a penalty of 2 errors. It is also possible that the participants considered that the empty fields were optional, and that it was sufficient to check the correct items that were explicitly listed. Note that questions, which did not require hand-written options besides listed ones, were not subject to this effect.

It is also possible that participants interpreted “marketing offers” (listed) as “targeted advertisements” (had to be written by hand). 26% of the participants have done so, thus taking a penalty of 2 errors. One of the highest error rates was attained by P13, who has forgotten their glasses and used a smart-phone camera as a lens to read the materials.

These “traps” were deliberately placed into the questionnaire, while they increased the error rate, they suggest that LITE works better when used as a *reference*. This also emphasizes the importance of a well-defined vocabulary of terms, as minor inconsistencies lead to errors.

### 4.5.0.2 Q<sub>2</sub> If the data were collected in the year 2045, what will be the last year in which they are still available?

84% of the participants correctly answered “2048”. We expected many off-by-one errors, however only one participant answered “2047”. Another incorrect answer was “2042”, which can be caused by a misinterpretation of the question. In this case, the participant subtracted the given interval, instead of adding it.

### 4.5.0.3 Q<sub>3</sub> What information is collected?

Although the complexity of Q<sub>3</sub> is comparable to Q<sub>1</sub>, the error rate was substantially lower. 65% of the participants have made no errors when answering it. There could be several reasons that explain the difference: the list of collected data features icons, while

the list of purposes does not. However, the questionnaire itself did not include the icons, hence participants *could not have* relied on the graphics to identify the correct entries. Another possibility is that the correct answer required less effort, as there is no need to write custom texts, one simply had to check a subset of the listed options. Finally, the listed options were worded as on the label, thus reducing interpretation issues.

#### 4.5.0.4 Q<sub>4</sub> Which country are the data stored in?

P<sub>13</sub> skipped this question, while others have correctly written “France” in the custom field. It is worth noting that Q<sub>4</sub> did not provide options to choose from, there was only an empty field to write text in. This can explain the high number of “custom missing” errors for Q<sub>1</sub> and their absence in Q<sub>4</sub>. Another possibility is that “France” on the label is highlighted, making it easier to see.

#### 4.5.0.5 Q<sub>5</sub> Who in Tesami GmbH can access the collected data?

Since there is no such information on the label, this question has no exact answer. We use it to see how participants react, expecting no consensus. 36% chose “I don’t know”, 13% ticked all the available options, while 10% answered “not sure”, “everyone?” or “it doesn’t say”. 45% of the participants chose various combinations of the listed options. During the interviews, they would come up with plausible explanations based on the purpose of collection, e.g., “but seeing product improvement and targeted advertisement, you can say it is the marketing staff that will get it” (P<sub>2</sub>).

#### 4.5.0.6 Q<sub>6</sub> Who can access the data while they are transmitted to Tesami?

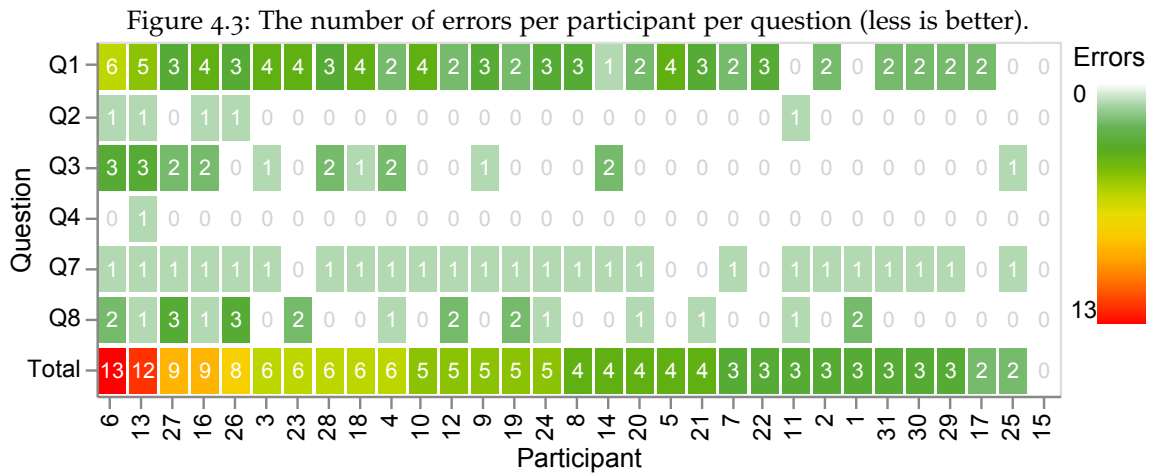
This question is open to interpretation, because the answer depends on one’s assumptions about the system (e.g., type of encryption, network protocols in use). 61% indicated that Tesami can access the data while they are in transit, 16% stated that others in the household can do it, 10% chose “I don’t know”. Contrary to our expectations, only 6% considered that the government can access the data.

#### 4.5.0.7 Q<sub>7</sub> How many organizations can access the data after they were collected?

The expected answer is “10”, comprising Tesami and 9 affiliates. The answers were “9” (42%), “10” (19%), “9..10” or “9 affiliates” (16%), “I don’t know” (10%), while 10% wrote “1”, “1-2” or “1?”. P<sub>16</sub> skipped the question.

It is possible that the answer “9” is an off-by-one error. However, there was only one instance in Q<sub>2</sub>, which could mean that something else has caused this discrepancy. It is also possible that some answered “9” because it is highlighted on the label, so they simply referred to that value.

Some participants have explicitly commented that “it depends on whether you count Tesami or not”, it indicates that they understand the context, but the phrasing of the question made it difficult to settle on one interpretation. Some participants could have made a distinction between “organization” and “affiliate”, hence answering “1”, because the question asked about organizations, not affiliates.



#### 4.5.0.8 Q<sub>8</sub> Which of the following data transmissions are not protected?

This question relies on the interpretation of padlock icons in the trace view. 55% of the participants answered correctly, 23% made one error, 7% chose “I don’t know”. Our analysis rules out the possibility that some participants did not notice the negation in the question, as we have not found answers that are the exact opposite of the correct one.

If participants understand the meaning of the padlock icon, they ought to answer the question correctly, otherwise they would make two errors, one for each use of the icon. The fact that 23% made only one error suggests that they did not understand the principle, or that they understood it, but did not notice the other icon. It is worth noting that a participant who said they usually ignored icons, answered the question correctly (P22). Another one has realized during the interview that they made a mistake in the form (P23).

Other participants’ comments indicate a clear understanding of the role of these icons, e.g., “it’s my own data, and it’s coming to me with some privacy, but my data is going to 9 affiliates without any privacy; isn’t it odd?” (P30).

#### 4.5.0.9 What do you think of when you read “product improvement”?

Contrary to our expectations, this vague purpose statement did not raise suspicions among the participants. All the interpretations were positive, focusing on the product in general, e.g., “making the product better in the future” (P18), or on software updates: “I guess bug-fixes” (P24). One participant has emphasized that they are not concerned by this: “it makes me think of updates for the device perhaps [...] I don’t think that would be something that would feel like a concern to me” (P20). P22 pointed out that there can be different interpretations: “Probably they would associate your preference with your customer number [...] I suppose, I have no idea at this point, this is speculation. It’s quite rough... general, so it depends”.

In their answers to Q<sub>13</sub>, about the advantages and disadvantages of such labels, 68% of the participants consider that the label benefits consumers, e.g., “Yes. I think it is important to be very clear about what information will be gathered, how and by whom it will be used!” (P7), “I do think such kind of labels are essential” (P28). Two participants expressed concerns: “[...] it only informs me, but I cannot control the data or limit it”

Section	Most interesting	Least interesting
Who	12	1
What	9	2
Trace	8	1
Purpose	7	0
Duration	1	3
QR	0	2
Where	0	2

Table 4.2: Most and least interesting sections of the label.

(P1) and “[.] if you only went of the label you might not find loopholes or other things a company could use/abuse” (P2).

In Q<sub>14</sub> we have asked whether participants like or dislike to have such labels. 77% of them answered affirmatively: “I don’t usually look at labels when I buy stuff, but I’d like to have this label” (P8), “Yes, I would like to see as much facts and descriptions as possible, so that I can make a better choice” (P23). None of the surveyed persons disliked the idea of having such labels.

Throughout the interviews, participants expressed satisfaction with the structure of the label and appreciated its contribution to transparency: “it feels like it is more open and more explanatory, they kind of show you their hand, like in poker almost. They don’t try to hide it, they put an emphasis on it so you know about it. I think that is good for the customer” (P2). Others would point out that such information is hard to find: “usually this type of information is buried under a lot of paper” (P7). Some stated that they liked the brevity of the label: “privacy facts should be short, [...] I get so much data just by looking at that, [...] if you make it longer, I will probably not read it” (P10). A common theme was the desire to obtain more information about how the data are used, participants wanted to know who the affiliates were, and what parts of data they were getting. P19 suggested a folding label, like the ones used in medical products, which would allow more information to be provided “under the fold”. Three participants questioned the authenticity of the label: “I need to feel that I trust the label itself” (P17), “labels can lie” (P9). Although such remarks were infrequent, contrary to our expectations, we believe that it is important to support LITE, e.g., via government-endorsed programmes [11]. Two participants expressed preference for a larger label, e.g., “it’s pretty clear, but I would like it bigger” (P5). Some participants stated that they understand the label, but not the full implications: “I believe it is my IP address they’re taking. But I don’t really know how that affects me” (P18).

We have asked participants to point out which parts of the label were most and least interesting to them, mapping each response to an element of the label. A total of 37 “most interesting” mentions were made, and 11 “least interesting” ones (Tab. 4.2).

The answers to our follow-up questions reveal that all of the participants have noticed the QR code, however 10% did not know what it was, while 84% did not scan it, nor intended to. 77% noticed the rectangles that emphasize some parts of the label. In terms of interpretation, all participants stated that they understood the icons, 77% had no diffi-

culties with the text. Although 16% did not know the word “affiliate”, they understood it when the word “partner” was suggested.

#### 4.6 DISCUSSION

Participants wanted to know more details about the way the data are used by each affiliate. The folding label proposed by P19 is an elegant solution, as it keeps the label usable without relying on gadgets or online services.

The results suggest that efficiency can be improved through the use of standardized terms and icons. This would also make the labels consistent across vendors, making comparisons easier, and improve usability, by habituating consumers to these terms.

The fact that none of the participants had suspicions when interpreting “product improvement” (in the “purpose” section) indicates that additional measures are needed to protect consumers. This may be resolved by the introduction of consistent terminology and by legal means.

When it comes to the authenticity of the label itself, our results suggest that most of the participants trusted the information or did not voice their concerns about it.

Statistical analysis of the results did not reveal any correlations between error rates and age, gender, skill level or the mock-up used.

The various errors we measured have a different impact on transparency. For example, the belief that the data are accessed by 9 companies instead of 10 is inaccurate, but still good enough for practical purposes.

For LITE to stay relevant as products evolve, vendors should decouple security and privacy from feature updates. Thus, IoT devices stay current without breaking the terms shown on the label. If users choose to install an update that modifies data collection practices, an updated label can be shown and consent has to be requested again, per GDPR.

#### 4.7 CONCLUSIONS

We have presented a “privacy facts” label for IoT devices and held 31 interviews to test it in practice. This is one out of many possible designs that meet the requirements, in this study we aimed for simplicity. The results are encouraging and they offer pointers for future work. For example, it is clear the creation of a standardized vocabulary and a common set of graphical primitives are important in the long term. Although we have found that participants tend to trust the information in the label, even in the absence of indicators of endorsement by regulators, we believe that such support will improve the viability of LITE.

#### 4.8 ACKNOWLEDGMENTS

This research has received funding from the H2020 Marie Skłodowska-Curie EU project “Privacy&Us” under the grant agreement No 675730. We thank the survey participants, Patrick Murmann for helping us analyze the transcripts, Harald Zwingelberg and the anonymous peer reviewers for their helpful comments.



## 4.9 APPENDIX: QUESTIONNAIRE

Please fill out this questionnaire. Feel free to go back to the label at any time. You can take notes and *use any tools and gadgets* at your disposal. This is not an exam, and there are no wrong answers or grades.

1. What purpose are the data collected for? (multiple choice possible)
 

<input type="checkbox"/> marketing offers <input type="checkbox"/> home automation <input type="checkbox"/> automatic billing <input type="checkbox"/> my personal use <input type="checkbox"/> scientific research <input type="checkbox"/> _____ <input type="checkbox"/> _____ <input type="checkbox"/> I don't know	<input type="checkbox"/> the IoT device <input type="checkbox"/> the government <input type="checkbox"/> _____ <input type="checkbox"/> _____ <input type="checkbox"/> I don't know
--	---
2. If the data were collected in the year 2045, what will be the last year in which they are still available?
 

\_\_\_\_\_                       I don't know
3. What information is collected? (multiple choice possible)
 

<input type="checkbox"/> current time <input type="checkbox"/> device Internet address <input type="checkbox"/> my customer number <input type="checkbox"/> temperature <input type="checkbox"/> my name <input type="checkbox"/> number of computers in my home <input type="checkbox"/> humidity <input type="checkbox"/> my phone number <input type="checkbox"/> _____ <input type="checkbox"/> I don't know	<input type="checkbox"/> data sent from device to Tesami <input type="checkbox"/> updates sent from Tesami to the device <input type="checkbox"/> data sent from Tesami to you <input type="checkbox"/> data sent from Tesami to affiliates <input type="checkbox"/> I don't know
---	---
4. Which country are the data stored in?
 

\_\_\_\_\_                       I don't know
5. Who in Tesami GmbH can access the collected data? (multiple choice possible)
 

<input type="checkbox"/> software engineers <input type="checkbox"/> hardware engineers <input type="checkbox"/> research & development dept. <input type="checkbox"/> marketing staff <input type="checkbox"/> company director <input type="checkbox"/> _____ <input type="checkbox"/> _____ <input type="checkbox"/> I don't know	<input type="checkbox"/> Do you have a QR-scanner program in your smartphone? <input type="radio"/> yes <input type="radio"/> no <input type="radio"/> I don't have a smartphone <input type="radio"/> I don't know <input type="radio"/> prefer not to say
---	---
6. Who can access the data while they are transmitted to Tesami GmbH? (multiple choice possible)
 

<input type="checkbox"/> others in my home <input type="checkbox"/> my neighbors <input type="checkbox"/> Tesami GmbH <input type="checkbox"/> my Internet provider	<input type="checkbox"/> 18..26 <input type="checkbox"/> 45..53 <input type="checkbox"/> 27..35 <input type="checkbox"/> 54 and above <input type="checkbox"/> 36..44 <input type="checkbox"/> prefer not to say
--	--
7. How many organizations can access the data after they were collected?
 

\_\_\_\_\_                       I don't know
8. Which of the following data transmissions are *not* protected? (multiple choice possible)
 

<input type="checkbox"/> play video games <input type="checkbox"/> view photos and watch videos <input type="checkbox"/> browse the Internet and send emails <input type="checkbox"/> use a word-processor to type documents <input type="checkbox"/> set up email sorting filters	<input type="checkbox"/> type complex documents in word processors (e.g., macros, automatic indexes, dynamic fields) <input type="checkbox"/> assemble computers or other electronics from components <input type="checkbox"/> I know at least one programming language
--	---
9. Do you have a QR-scanner program in your smartphone?
 

male                                       other  
 female                                       prefer not to say
10. What is your age?
 

18..26                                       45..53  
 27..35                                       54 and above  
 36..44                                       prefer not to say
11. What is your gender?
 

male                                       other  
 female                                       prefer not to say
12. Please specify the computer-related skills you have
 

<input type="checkbox"/> play video games <input type="checkbox"/> view photos and watch videos <input type="checkbox"/> browse the Internet and send emails <input type="checkbox"/> use a word-processor to type documents <input type="checkbox"/> set up email sorting filters	<input type="checkbox"/> type complex documents in word processors (e.g., macros, automatic indexes, dynamic fields) <input type="checkbox"/> assemble computers or other electronics from components <input type="checkbox"/> I know at least one programming language
--	---
13. Do you see advantages/disadvantages in having such labels on products in the future?
14. Would you like/dislike to have such product labels in the future?

These questions are asked to elicit qualitative data after the survey is filled out:

- Have you encountered any difficulties in understanding the information on the label? If yes, which ones?
- Have you encountered any difficulties in understanding the icons on the label? If yes, which ones?
- Which content has been particularly interesting/not interesting to you?
- What do you understand when reading “personal use” and “product improvement”?
- Have you seen that some of the elements of the label are highlighted? How have you interpreted that emphasis?
- How do you interpret the image in the hand of the human figure?
- Do you know what this figure [QR] is, and what can be done with it?
- What other comments have you got about the “privacy facts” label?

## REFERENCES

- [1] Alexandr Railean and Delphine Reinhardt. "Life-Long Privacy in the IoT? Measuring Privacy Attitudes Throughout the Life-Cycle of IoT Devices." In: *Privacy and Identity Management. The Smart Revolution*. IFIP Advances in Information and Communication Technology. 2017.
- [2] Noura Aleisa and Karen Renaud. "Privacy of the Internet of Things: A Systematic Literature Review." In: *Proceedings of the 50th Hawaii International Conference on System Sciences HICSS (2017)*.
- [3] David De Cremer, Bang Nguyen, and Lyndon Simkin. "The Integrity Challenge of the Internet-of-Things (IoT): on Understanding its Dark Side." In: *Journal of Marketing Management (2017)*.
- [4] Ulrich 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.
- [5] Jessica Vitak et al. "Privacy Attitudes and Data Valuation Among Fitness Tracker Users." In: *Transforming Digital Worlds*. 2018.
- [6] European Parliament and Council of European Union. "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of Natural Persons With Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)." In: *Official Journal of the European Union (2016)*.
- [7] Article 29 Working Party. *Guidelines on Transparency Under Regulation 2016/679*.
- [8] Simone Fischer-Hübner et al. "Transparency, Privacy and Trust—Technology for Tracking and Controlling My Data Disclosures: Does This Work?" In: *IFIP International Conference on Trust Management (IFIPTM)*. 2016.
- [9] Patrick Moore and Chad Fitz. "Gestalt Theory and Instructional Design." In: *Journal of Technical Writing and Communication (1993)*.
- [10] George A Miller. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information." In: *Psychological review (1956)*.
- [11] Abhijit Banerjee and Barry D. Solomon. "Eco-Labeling for Energy Efficiency and Sustainability: A Meta-Evaluation of US Programs." In: *Energy Policy (2003)*.

## ONLITE: ON-LINE LABEL FOR IOT TRANSPARENCY ENHANCEMENT

---

### AUTHORS

Alexandr Railean and Delphine Reinhardt

Institute of Computer Science, Georg-August-Universität Göttingen, Germany

PUBLISHED IN Proceedings of the 25th Nordic Conference on Secure IT Systems (NordSec, 2020). DOI: [10.1007/978-3-030-70852-8\\_14](https://doi.org/10.1007/978-3-030-70852-8_14).

**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.

### 5.1 INTRODUCTION

The number of IoT devices, such as smart appliances, fitness trackers or surveillance cameras, has grown over the last decade [1]. While this brings economic benefits, it also comes with major privacy risks [2]. For example, it has been shown that in some circumstances, individuals can be deanonymized by correlating data sets [3, 4]. Another example is the analysis of smart-meter readings to identify media played on a TV [5]. Such privacy issues can be amplified by factors like device ubiquity, sensor diversity, data collection frequency, and the large volume of collected data [6, 7]. 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 [8, 9, 10]. Another factor that contributes to loss of privacy is the lack of awareness about the technical capabilities of IoT devices [10, 11, 12]. Besides that, users are skeptical of the ways algorithms can infer personal facts about them [13].

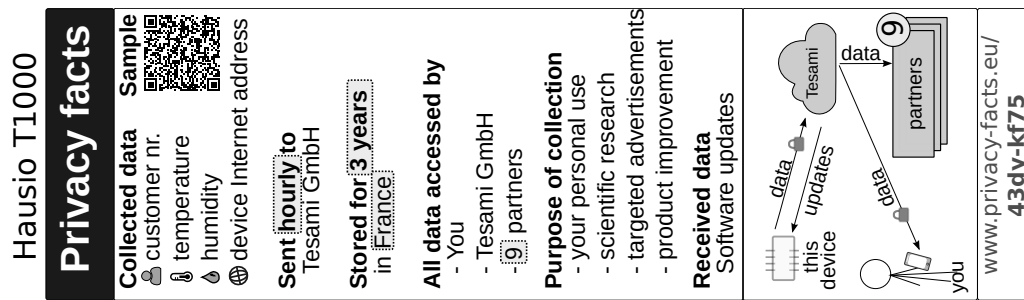


Figure 5.1: LITE label for a hypothetical IoT device called “Hausio T1000” [16].

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 [15].

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, 17, 18]. To address this need, several “privacy facts” labels have been proposed [17, 19, 20, 21], including our own “Label for IoT Transparency Enhancement”, LITE (Fig. 5.1, [16]).

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 [16] 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 Sec. 5.8.1, 5.8.2, 5.9). 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](https://zenodo.org/record/4126346).

## 5.2 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 [22], 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.

## 5.3 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 [16]:

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?

## 5.4 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:

*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. 5.2a).

(a) Overview

Overview Who gets the data Data flows Data sample Security Lifecycle Contact

Show differences

Hausio T1000 vs Casami FX Domowoj

**Collected data**

customer nr.	customer nr.	customer nr.
temperature	temperature	temperature
humidity	humidity	UV radiation
device Internet address	wind speed	wind speed

**Sent**

hourly	daily	daily
to Tesami GmbH	to Aster SRL	to Domotics s.r.o.

**Stored for**

3 years in France	6 years in Italy	1 year in the Czech Republic
-------------------	------------------	------------------------------

(b) Who gets the data

See who gets the data, and why Search in table: ad

Device	Data type	Purpose	Company	Country	Sensitivity
Casami FX	temperature	scientific research	Minerva LTD	Canada	low
Casami FX	humidity	scientific research	Minerva LTD	Canada	low
Domowoj	UV radiation	archive data	Cornix	China	low

Showing 1 to 3 of 3 entries (filtered from 17 total entries)

(c) Security

	Hausio T1000	Casami FX	Domowoj
<b>Vulnerabilities</b>			
Reaction time to disclosed vulnerabilities	2 weeks	3 weeks	-
Rewards for reported vulnerabilities	Yes	Yes	No
<b>Communications</b>			
Secure from Internet eavesdroppers	Yes	-	-
Secure from local network eavesdroppers	Yes	Yes	No
<b>Storage</b>			
Stored data are encrypted	N/A, no information is stored on the device	Yes	No

(d) Contact

Action	Hausio T1000	Casami FX	Domowoj
View, edit or delete collected data by contacting the Data Controller	Tesami GmbH Flachmatuchstr. 42, Lindau Germany. info@tesa.mi	Aster SRL Via Macaroni 113, Verona, Italy. contact@casam.it	Domotics s.r.o Bezručova 202, Brno, Czech Republic. gosti@dom.cz
Report privacy-related issues to the Data Protection Officer	dpo@tesa.mi	info@casam.it	rucitel@dom.cz
Lodge a complaint with the supervisory authority	Unabhängiges Landeszentrum Flachmatuchstr. 42, Lindau Germany. mail@lindau.de	Garante per la protezione dei dati personali Piazza di Monte Citorio, Roma, Italy.	Orgánem pro ochranu údajů Svoboda 900, Praha, Czech Republic. pomoc@opou.cz

You can also lodge a complaint with a [supervisory authority in your area](#).

Figure 5.2: Collage of screenshots of the tabs of OnLITE. The information is provided by vendors themselves, as they are obliged to do so under the GDPR.

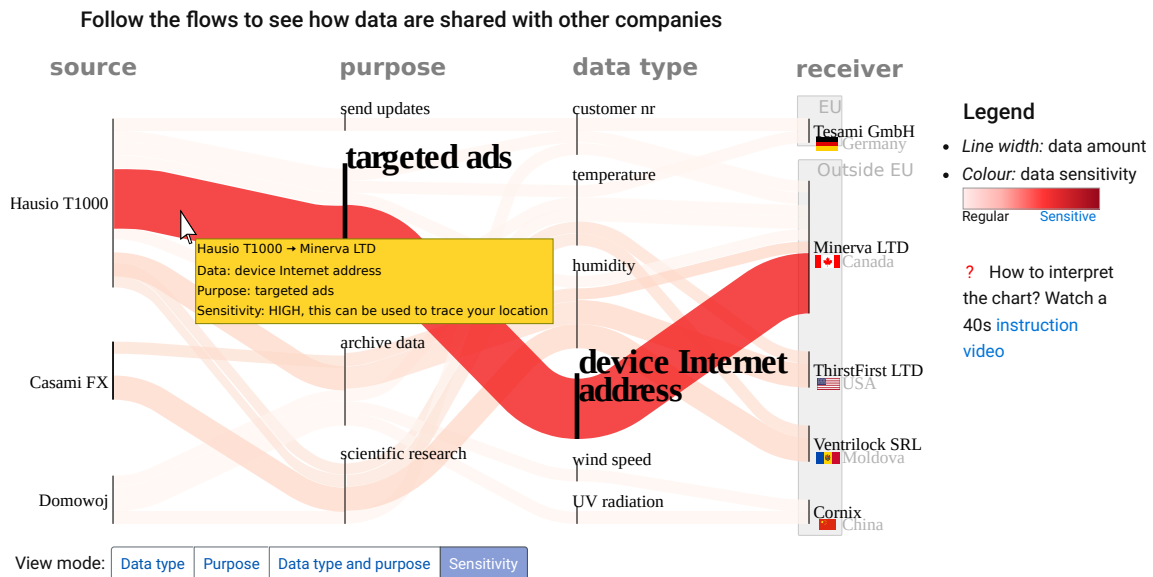


Figure 5.3: The “Data flows” tab shows how data are shared with third parties. The “sensitivity” view highlights special categories of data defined by the GDPR.

*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  $\text{dataSize} \times \text{frequency}$ . 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. 5.3 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 [23]. 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. 5.3). 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”.

*Security* - this tab presents security information (Fig. 5.2c). We have made sure to use common language. For example, “Secure from Internet eavesdroppers”, as opposed to specialized terms [25]. 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 [12] (Fig. 5.4). For example, it informs

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

	Hausio T1000	Casami FX	Domowoj
<b>Set up</b> – preparing the device for use			
Unique factory-set password	Yes	Yes	No
Password change required before remote access for the first time	Yes	No	No
<b>Use</b> – typical, daily interactions with the device			
Multiple user accounts	Supported	Supported	No
Separate accounts for children	Supported	Supported	No
Separate account for guests	Supported	No	No
<b>Maintenance</b> – procedures to increase the device longevity and ensure it works well			
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	-	-
<b>Retiring</b> – when the device is sold, sent for repairs, donated or thrown away			
Secure data deletion (wiping)	Yes	No	No

Figure 5.4: Comparing three IoT devices throughout the phases of their lifecycle.

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. 5.2d). 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.

#### 5.4.1 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., 00 111, 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].



## 5.5 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. 5.2, 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 [15, 27].

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.

## 5.6 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 Sec. 5.3 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% [29]. We further quantified the usability of the GUI using the *System Usability Scale* (SUS) [30], chosen due to its good performance at sample sizes  $\geq 12$  [31], and because scores of similar interfaces can be compared.

### 5.6.1 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''

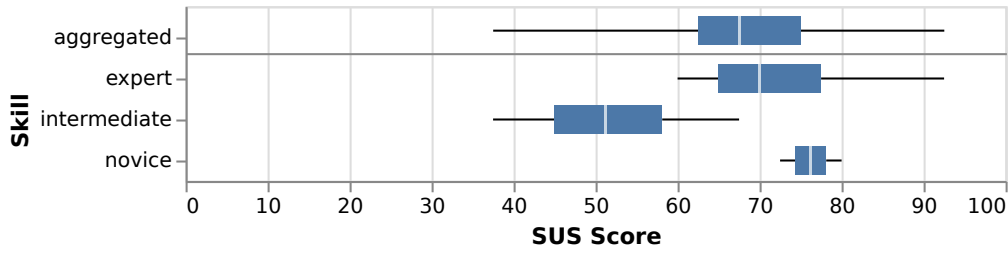


Figure 5.5: SUS scores grouped by skill.

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 128mm × 40mm privacy labels on A6 sheets, each corresponding to a device, as shown in Fig. 5.1. 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 Tab. 5.1 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.

### 5.6.2 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.

### 5.6.3 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 26 years (20%), the rest are between 36 and 44 years (13%). Their self-reported technical competence is computed using the method defined in [12]. In our sample, 60% are expert, 27% are intermediate, and 13% are novice (Tab. 5.2). The group is diverse in terms of academic fields, and includes economists, mathematicians, computer scientists, environmentalists, and

Table 5.1: 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?

Table 5.2: Demographic data and results.

	Age	Sex	Skill	SUS	Time (minutes)		
				score	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

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

#### 5.6.4 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 [32]. 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 [33].

## 5.7 RESULTS

### 5.7.1 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. 5.3). 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. 5.3).

*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. 5.3 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.

### 5.7.2 Quantitative

The SUS results are given in Tab. 5.2 and Fig. 5.5. The mean score of OnLITE is 68, which matches the industry average for web interfaces [34]. Statistical analysis, by means of a t-test<sup>1</sup>, did not reveal any correlation between SUS scores and age or gender. Proto-

<sup>1</sup> 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.

type 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 P<sub>1</sub>, P<sub>3</sub>, P<sub>4</sub> and P<sub>6</sub>, who failed task M. Note that the session durations in Tab. 5.2 are not an indication of invested effort, because we encouraged participants to explore alternatives and elicited additional feedback, even after a task was done.

## 5.8 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 Sec. 5.3. 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 (P<sub>8</sub> and P<sub>14</sub>), 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 [34]. 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 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".

### 5.8.1 *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 [35]. While comparing device privacy ratings via scores is easy for consumers [18, 36], 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 Sec. 5.3. 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.

### 5.8.2 *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” [21], 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 [37], 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).

### 5.8.3 *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 [16]. 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.



## 5.9 RELATED WORK

Several designs were proposed to address IoT privacy and security issues. Some inherit the grid layout and the layered approach of [38]. A taxonomy proposed by [18] 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 [20]. 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 [21]. 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 [19].

Emami-Naeini et al. created a user-validated informational privacy and security label for IoT [17]. A difference is the use of scoring to quantify the level of privacy a device provides, while we have avoided using star ratings (see Sec. 5.8.1). 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 [39]. 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 [40] (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.

## 5.10 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.

## REFERENCES

- [1] *Trends 17*. Globalwebindex, 2016. URL: <http://insight.globalwebindex.net/hubfs/Reports/Trends-17.pdf>.
- [2] Jan Henrik Ziegeldorf, Oscar Garcia Morchon, and Klaus Wehrle. "Privacy in the Internet of Things: Threats and Challenges." In: *Security and Communication Networks* (2014).
- [3] Delphine Christin. "Privacy in Mobile Participatory Sensing: Current Trends and Future Challenges." In: *Journal of Systems and Software* (2016).
- [4] Arvind Narayanan and Vitaly Shmatikov. "How to Break Anonymity of the Netflix Prize Dataset." In: *arXiv preprint cs/0610105* (2006).
- [5] Ulrich 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.
- [6] M. Kosinski, D. Stillwell, and T. Graepel. "Private Traits and Attributes Are Predictable From Digital Records of Human Behavior." In: *Proceedings of the National Academy of Sciences* (2013).
- [7] Nicholas D. Lane et al. "On the Feasibility of User De-anonymization From Shared Mobile Sensor Data." In: *Proceedings of the 3rd International Workshop on Sensing Applications on Mobile Phones (PhoneSense)*. 2012.
- [8] Noura Aleisa and Karen Renaud. "Privacy of the Internet of Things: A Systematic Literature Review." In: *Proceedings of the 50th Hawaii International Conference on System Sciences HICSS* (2017).
- [9] David De Cremer, Bang Nguyen, and Lyndon Simkin. "The Integrity Challenge of the Internet-of-Things (IoT): on Understanding its Dark Side." In: *Journal of Marketing Management* (2017).
- [10] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. "Alexa, Are You Listening?: Privacy Perceptions, Concerns and Privacy-seeking Behaviors with Smart Speakers." In: *Proceedings of the ACM on Human-Computer Interaction (CSCW 2018)*.
- [11] Xinru Page et al. "The Internet of What?: Understanding Differences in Perceptions and Adoption for the Internet of Things." In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*. 2018.
- [12] Alexandr Railean and Delphine Reinhardt. "Life-Long Privacy in the IoT? Measuring Privacy Attitudes Throughout the Life-Cycle of IoT Devices." In: *Privacy and Identity Management. The Smart Revolution*. IFIP Advances in Information and Communication Technology. 2017.
- [13] Serena Zheng et al. "User Perceptions of Smart Home IoT Privacy." In: *Proceedings of the ACM on Human-Computer Interaction*. 2018.
- [14] European Parliament and Council of European Union. "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of Natural Persons With Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)." In: *Official Journal of the European Union* (2016).

- [15] Lorrie Faith Cranor. "Necessary But Not Sufficient: Standardized Mechanisms for Privacy Notice and Choice." In: *Journal on Telecommunications and High Technology Law* (2012).
- [16] Alexandr Railean and Delphine Reinhardt. "Let There be LITE: Design and Evaluation of a Label for IoT Transparency Enhancement." In: *Proceedings of the 20th ACM International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI Adjunct)*. 2018.
- [17] Pardis Emami-Naeini et al. "Exploring How Privacy and Security Factor into IoT Device Purchase Behavior." In: *Proceedings of the 2019 Conference on Human Factors in Computing Systems (CHI)*. 2019.
- [18] Shane Johnson et al. "The Impact of IoT Security Labelling on Consumer Product Choice and Willingness to Pay." In: *preprint*. 2019.
- [19] Grace Fox et al. "Communicating Compliance: Developing a GDPR Privacy Label." In: *Proceedings of the Americas Conference on Information Systems* (2018).
- [20] Rob van Diermen. "The Internet of Things: a Privacy Label for IoT Products in a Consumer Market." PhD thesis. 2018.
- [21] Yun Shen and Pierre-Antoine Vervier. "IoT Security and Privacy Labels." In: *Privacy Technologies and Policy*. 2019.
- [22] Simone Fischer-Hübner et al. "Transparency, Privacy and Trust—Technology for Tracking and Controlling My Data Disclosures: Does This Work?" In: *IFIP International Conference on Trust Management (IFIPTM)*. 2016.
- [23] Mary Theofanos et al. *Usability Testing of Ten-print Fingerprint Capture*. Tech. rep. National Institute of Standards and Technology, 2007.
- [24] R.C. Lupton and J.M. Allwood. "Hybrid Sankey Diagrams: Visual Snalysis of Multidimensional Data for Understanding Resource Use." In: *Resources, Conservation and Recycling* (2017).
- [25] Bruce Schneier. *Click Here to Kill Everybody*. 2018.
- [26] George A Miller. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information." In: *Psychological review* (1956).
- [27] Bert Bos. *Data Privacy Vocabulary*. W3C Recommendation. W3C, 2019. URL: <https://www.w3.org/ns/dpv>.
- [28] Jakob Nielsen and Rolf Molich. "Heuristic Evaluation of User Interfaces." In: *Conference on Human Factors in Computing Systems*. 1990.
- [29] Laura Faulkner. "Beyond the Five-user Assumption: Benefits of Increased Sample Sizes in Usability Testing." In: *Behavior Research Methods, Instruments, & Computers*. 2003.
- [30] John Brooke. "SUS - a Quick and Dirty Usability Scale." In: *Usability Evaluation in Industry*. 1986.
- [31] Thomas S Tullis and Jacqueline N Stetson. "A Comparison of Questionnaires for Assessing Website Usability." In: *Usability Professional Association Conference*. 2004.
- [32] Virginia Braun and Victoria Clarke. "Using Thematic Analysis in Psychology." In: *Qualitative Research in Psychology* (2006).

- [33] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. "Reliability and Inter-rater Reliability in Qualitative Research." In: *Proceedings of the ACM on Human-Computer Interaction* (2019).
- [34] Aaron Bangor, Philip T. Kortum, and James T. Miller. "An Empirical Evaluation of the System Usability Scale." In: *International Journal of HCI*. 2008.
- [35] Douglas Engelbart. *Augmenting Human Intellect: A Conceptual Framework*. Tech. rep. 1962. URL: <https://www.dougelbart.org/content/view/138/000/>.
- [36] Pardis Emami-Naeini et al. "The Influence of Friends and Experts on Privacy Decision Making in IoT Scenarios." In: *Proceedings of the ACM on Human-Computer Interaction*. 2018.
- [37] Abe Davis et al. "The Visual Microphone: Passive Recovery of Sound From Video." In: *ACM Transactions on Graphics*. 2014.
- [38] Patrick Gage Kelley et al. "A Nutrition Label for Privacy." In: *Proceedings of the 5th Symposium on Usable Privacy and Security (SOUPS)*. 2009.
- [39] Peter Bihl. *A Trustmark for IoT*. ThingsCon, 2017. URL: <https://thewavingcat.com/iot-trustmark/>.
- [40] ETSI. "Cyber Security for Consumer IoT: Baseline Requirements." In: *European Standard 303 645*. 2020.

## IMPROVING THE TRANSPARENCY OF PRIVACY TERMS UPDATES (OPINION PAPER)

---

### AUTHORS

Alexandr Railean and Delphine Reinhardt

Institute of Computer Science, Georg-August-Universität Göttingen, Germany

PUBLISHED IN Proceedings of the 9th Annual Privacy Forum (APF, 2021).

DOI: [10.1007/978-3-030-76663-4\\_4](https://doi.org/10.1007/978-3-030-76663-4_4).

**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.

### 6.1 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 [1, 2]. As a result, many systems remain insecure, while users are frustrated and may lose interest in the maintenance of their systems [1, 2]. 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"[3, 4]. Prior research has shown that the current level of transparency is inadequate and that in many cases end users cannot exercise their rights [5, 6]. Users face problems such as excessive length of privacy policies, complex language, vagueness, lack of choices, and fatigue [7]. 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 [8]. 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 [9, 10]. This suggests that end-users are in a vulnerable position and that their privacy is undermined.

Data type	Purpose	Company	Country	Duration	Frequency
🌡 temperature	research	Minerva LTD	🇨🇦 Canada	1y	daily
💧 humidity	marketing	ThirstFirst LTD	🇺🇸 USA	1y	hourly

Figure 6.1: The “who gets the data” table, adapted from [11]. Note that the table can be configured to show personal and non-personal data (see Sec. 6.7.5 for details).

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. 6.2 provides a high-level overview of our approach. Sec. 6.3 introduces a formal notation of privacy terms, which is then used in Sec. 6.4 to determine when update prompts can be omitted. In Sec. 6.5, we argue that our proposed way of expressing updated privacy terms is more efficient than prose typically used for this purpose. Sec. 6.6 describes additional steps that can be taken to further improve transparency. In Sec. 6.7 we discuss the implications of applying our approach, while Sec. 6.8 reviews related work. We make concluding remarks in Sec. 6.9.

## 6.2 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 [3]. 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. 6.1, 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?

The table in Fig. 6.1 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. 6.2 [11]. Although the aforementioned transparency questions are

Hausio T1000 v1.1	vs	Hausio T1000 v1.2
<b>Collected data</b>		
customer nr.		customer nr.
temperature		temperature
humidity		humidity
device Internet address		device Internet address
		wind speed
<b>Sent</b>		
hourly		daily
to Tesami GmbH		to Tesami GmbH
<b>Stored for</b>		
3 years		6 years
in France		in France

Figure 6.2: Comparing two versions of the same device side by side, while highlighting differences (adapted from [11]).

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 [11].

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 [11]. 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, 12]. Such an effect occurs after just two exposures, and grows with repeated exposure [13, 14]. 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.

### 6.3 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. 6.2. For example, privacy is affected if the *retention* period changes from "1 month" to "10

years”, or if the collection *frequency*<sup>1</sup> changes from “once per day” to “twice per second” [15]. Thus, our notation aims to capture these parameters, using the following symbols:

**Data type**  $\Delta$  type of collected data

**Purpose**  $\Pi$  purpose of collection

**Time**  $T$  the retention period

**Company**  $C$  a company that gets the data

**Location**  $\Lambda$  location of said company

**Frequency**  $\Phi$  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**  $\Theta$  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.,  $K = \{\Theta_1, \Theta_2, \Theta_3, \dots, \Theta_i\}$ .

Thus, when a user gives consent, we formally represent that in an expanded form as:  $K = \{(\Delta_1, \Pi_1, C_1, \Lambda_1, T_1, \Phi_1), \dots, (\Delta_i, \Pi_i, C_i, \Lambda_i, T_i, \Phi_i)\}$ . 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  $\Lambda$  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.

#### 6.4 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. 6.1, 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. 6.2). To this end, we propose the concept of a *privacy protection gradient*, which differentiates areas by level of privacy protection mechanisms in place.

<sup>1</sup> 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 $\Theta_i$ 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 6.1: Primary filters. If any rule is matched, a consent prompt is unnecessary.

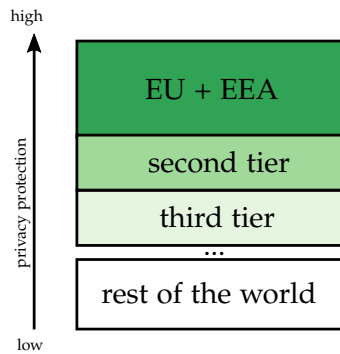


Figure 6.3: Privacy protection levels in different political, economic or strategic unions.

In this hypothetical example (Fig. 6.3), 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 Proteção de Dados, modeled after the GDPR [16], California and its Consumer Privacy Act [17], 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  $\Lambda$  is written as  $\Lambda^\pi$ . Thus, if the old location of the data was in an area less secure than the new location, we express that as  $\Lambda_{old}^\pi < \Lambda_{new}^\pi$ .

Rule	Logic	Formal notation	Intuition
5	Go up or sideways on the “privacy gradient”	$\Theta_i \wedge_{\text{old}}^\pi \leq \Theta_i \wedge_{\text{new}}^\pi$	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 $\Theta_i$ is identical)

Table 6.2: Secondary filter, subject to discussion, can be deactivated by users.

Such secondary filters can be controversial. For example, there was an attempt to use the GDPR to silence journalists in Romania [18], therefore some 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. 6.6), or choose other criteria for computing  $\wedge^\pi$ , such as the democracy index<sup>2</sup>.

## 6.5 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 [6, 8], 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 [19]. In what follows, we present an example calculation, using the notation proposed in Sec. 6.3.

Recall that each term of a privacy policy is a tuple expressed as  $\Theta = (\Delta, \Pi, C, \wedge, T, \Phi)$ . For example,  $\wedge$  represents one of the world’s 193 countries<sup>3</sup>. 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.

Before this information can be communicated, we must *encode* it [20]. 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  $\eta = \frac{\text{info}_{\text{useful}}}{\text{info}_{\text{total}}} \times 100 = \frac{8}{40} \times 100 \approx 20\%$ . 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  $\Theta$ , by relying on existing terminology that defines types of data, purposes of collection and retention periods [21, 22, 23], reaching

<sup>2</sup> [eiu.com/topic/democracy-index](http://eiu.com/topic/democracy-index)

<sup>3</sup> According to the UN [un.org/en/member-states](http://un.org/en/member-states)

a total of **155 bits**. The complete calculation is omitted for brevity, but is available in 6.10.

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 [24].

### 6.5.1 *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 [9]. (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 [6, 9]. (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. 6.1). Moreover, if a user does not need certain columns, they can hide them.

## 6.6 ADDITIONAL STEPS TOWARDS BETTER UPDATE TRANSPARENCY

### 6.6.1 *Distinguishing Feature, Security, and Privacy Updates*

Sometimes updates can force users into a “take it or leave it” dilemma [7]. 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 [2]. 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

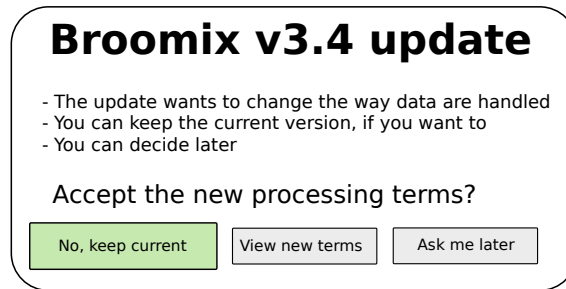


Figure 6.4: Hypothetical interface where users indicate whether they want to update without changing the terms. Note the default option is the most conservative, and that there is no “accept terms” option, because the user needs to understand them before accepting, otherwise it would not be an *informed* 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. 6.2).

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. 6.4), since we know exactly what terms were previously accepted.

### 6.6.2 *The Best Time to Ask Permission*

Another improvement in the way privacy updates are handled is to consider the best time<sup>4</sup> 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.

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. 6.6.1, 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.

### 6.6.3 *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<sup>5</sup> between the new and old terms is beneath a threshold. The

<sup>4</sup> Here we mean it in the sense of the Greek word “kairos”, which refers to an opportune moment, not to chronology.

<sup>5</sup>  $|K_{old} \Delta K_{new}|$ , i.e., the cardinality of the symmetric difference between the old and new terms.

## Broomix v3.4 update

- The update wants to change the way data are handled
- You can keep the current version, if you want to
- You can decide later

What has changed:

	Data type	Purpose	Company	Country	Duration
new	🌡 temperature	research	Minerva LTD	🇨🇦 Canada	1y
	💧 humidity	marketing	ThirstFirst LTD	🇺🇸 USA	<del>1y</del> 3y

Accept the new processing terms?

No, keep current

Accept new terms

Ask me later

Figure 6.5: Hypothetical update featuring an inline consent prompt.

### Settings

I give consent automatically when:

- New terms are more strict than the ones I already agreed to
- The data moves to a country with better privacy protections

Figure 6.6: Hypothetical interface where users indicate whether they want to give consent automatically in some cases.

sweet spot remains to be established experimentally, but a good default value could be Miller’s “magic number  $7 \pm 2$ ” [25]. For example, in Fig. 6.5 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. 6.6) 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”.

## 6.7 DISCUSSION

### 6.7.1 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 [26].

### 6.7.2 *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 [27, 28]. 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, 30].

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 [11], 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.

### 6.7.3 *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.

### 6.7.4 *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 [31]. 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.

### 6.7.5 *Listing Non Personally Identifiable Information*

Given that the proposed design grew out of IoT-centric research, Fig. 6.1 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 il-

lustrative 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 [32].

## 6.8 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 [2, 33]. 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 [34], machine learning, and neural networks [27, 28, 35].

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 [27, 28]. 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 [35]. 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 chat-bot 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 [36]. Their

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

Table 6.3: 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.

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.

## 6.9 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.

## 6.10 APPENDIX: INFORMATION EFFICIENCY CALCULATION EXAMPLE

We extend the material from Sec. 6.5 by providing another example. Consider the last term of the tuple,  $\Phi$ , 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  $\Theta$ , using terminology adapted from several sources: Platform for Privacy Preferences (P3P) [23], Data Privacy Vocabulary (DPV) [22], and Apple developer guidelines [21], summarized in Tab. 6.3. 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.

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.



$\Delta$	$\Pi$	$C$	$\Lambda$	$T$	$\Phi$
<u>temperature</u>	<u>research</u>	<u>Minerva LTD</u>	<u>Canada</u>	<u>1 year</u>	<u>daily</u>
161 data types	31 purposes	20 symbols per company <small>39-symbol alphabet (a..z, 0..9, \t, \n, \space) 6 bit/symbol</small>	193 countries	8 units + n <small>year month week day hour minute second</small>	8 frequencies <small>0.255 many times per second every second every minute hourly daily weekly monthly on demand</small>
bits <b>8</b>	<b>5</b>	<b>20x6=120</b>	<b>8</b>	<b>3+8=11</b>	<b>3</b>

Figure 6.7: Annotated calculations that explains how the amount of information in each privacy term is computed, yielding a total of 155 bits.

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 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. 6.1, 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:  $\eta = \frac{\text{info}_{\text{useful}}}{\text{info}_{\text{total}}} \times 100 = \frac{155 \times 2}{294 \times 2} \times 100 \approx 53\%$ .

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. 6.1: “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  $\eta = \frac{\text{info}_{\text{useful}}}{\text{info}_{\text{total}}} \times 100 = \frac{155 \times 2}{453 \times 6} \times 100 \approx 11\%$ .

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 [6, 8, 10].

## 6.11 APPENDIX: WHEN TO DISPLAY CONSENT PROMPTS

The following pseudo-code illustrates the logic defined in Sec. 6.4 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] Michael Fagan, Mohammad Maifi Hasan Khan, and Ross Buck. "A Study of Users' Experiences and Beliefs About Software Update Messages." In: *Computers in Human Behavior* (2015).
- [2] Kami E. Vaniea, Emilee Rader, and Rick Wash. "Betrayed by Updates: How Negative Experiences Affect Future Security." In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 2014.
- [3] European Parliament and Council of European Union. "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of Natural Persons With Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)." In: *Official Journal of the European Union* (2016).
- [4] *GDPR Recital 58 - The Principle of Transparency*. URL: <https://gdpr-info.eu/recitals/no-58/>.
- [5] Fred H. Cate. "The Limits of Notice and Choice." In: *IEEE Security & Privacy Magazine* (2010).
- [6] Ehimare Okoyomon et al. "On The Ridiculousness of Notice and Consent: Contradictions in App Privacy Policies." In: *Workshop on Technology and Consumer Protection* (2019).
- [7] Florian Schaub et al. "A Design Space for Effective Privacy Notices." In: *Eleventh Symposium On Usable Privacy and Security (SOUPS)*. 2015.
- [8] Aleecia M McDonald and Lorrie Faith Cranor. "The Cost of Reading Privacy Policies." In: *Journal of Law and Policy for the Information Society* (2008).
- [9] Christoph Bösch et al. "Tales from the Dark Side." In: *Proceedings on Privacy Enhancing Technologies* (2016).
- [10] Soheil Human and Florian Cech. "A Human-Centric Perspective on Digital Consenting: The Case of GAFAM." In: *Human Centred Intelligent Systems*. 2021.
- [11] Alexandr Railean and Delphine Reinhardt. "OnLITE: On-line Label for IoT Transparency Enhancement." In: *Proceedings of the 25th Nordic Conference on Secure IT Systems*. 2020.
- [12] Steven Michael Casey. *Set Phasers on Stun: And Other True Tales of Design, Technology, and Human Error*. 1993.
- [13] Bonnie Brinton 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.
- [14] Bonnie Brinton 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.
- [15] Ulrich 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] Abigayle Erickson. "Comparative Analysis of the EU's GDPR and Brazil's LGPD: Enforcement Challenges with the LGPD." In: *Brooklyn Journal of International Law* (2018).
- [17] W. Gregory Gregory Voss. "The CCPA and the GDPR Are Not the Same: Why You Should Understand Both." In: *CPI Antitrust Chronicle* (2021).
- [18] Nikolaj Nielsen. *EU Warns Romania not to Abuse GDPR Against Press*. EUobserver. 2018. URL: <https://euobserver.com/justice/143356>.
- [19] Jef Raskin. *The Humane Interface: New Directions for Designing Interactive Systems*. 2011.
- [20] C. E. Shannon. "A Mathematical Theory of Communication." In: *Bell System Technical Journal* (1948).
- [21] *App privacy details on the App Store*. Apple Developer. URL: <https://developer.apple.com/app-store/app-privacy-details/>.
- [22] Bert Bos. *Data Privacy Vocabulary*. W3C Recommendation. W3C, 2019. URL: <https://www.w3.org/ns/dpv>.
- [23] Lorrie Faith Cranor. *Web Privacy with P3P*. 2002.
- [24] ISO DIS. "9241-210: 2010. Ergonomics of human system interaction-Part 210: Human-centred design for interactive systems." In: *International Standardization Organization (ISO)*. Switzerland (2009).
- [25] George A Miller. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information." In: *Psychological review* (1956).
- [26] Philipp Morgner, Felix Freiling, and Zinaida Benenson. "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).
- [27] Hamza Harkous et al. "Polisis: Automated Analysis and Presentation of Privacy Policies Using Deep Learning." In: *27th USENIX Security Symposium* (2018).
- [28] Hamza Harkous and Kassem Fawaz. "PriBots: Conversational Privacy with Chatbots." In: *12th Symposium on Usable Privacy and Security* (2016).
- [29] Amir Rosenfeld, Richard Zemel, and John K. Tsotsos. "The Elephant in the Room." In: *arXiv:1808.03305 [cs]* (2018).
- [30] J. Su, D. V. Vargas, and K. Sakurai. "One Pixel Attack for Fooling Deep Neural Networks." In: *IEEE Transactions on Evolutionary Computation* (2019).
- [31] Jakob Nielsen. *Jakob's Law of Internet User Experience*. URL: <https://www.nngroup.com/videos/jakobs-law-internet-ux/>.
- [32] Philipp Morgner et al. "Privacy Implications of Room Climate Data." In: *Computer Security – ESORICS*. 2017.
- [33] Kami Vaniea and Yasmeeen Rashidi. "Tales of Software Updates: The Process of Updating Software." In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 2016.
- [34] Norman 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 (SOUPS)* (2014).

- [35] Razieh 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.
- [36] Travis D. Breaux, Hanan Hibshi, and Ashwini Rao. "Eddy, a Formal Language for Specifying and Analyzing Data Flow Specifications for Conflicting Privacy Requirements." In: *Requirements Engineering* (2014).