Controlling Data Streams of IoT Devices in Smart Homes from a User Perspective

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

Chathurangi Ishara Wickramasinghe
aus Leubsdorf

Göttingen, February 2024

Betreuungsausschuss

Prof. Dr. Jens Grabowski,
Institut für Informatik, Georg-August-Universität Göttingen

Prof. Dr. Patrick Harms,
Fakultät Elektrotechnik, Feinwerktechnik, Informationstechnik, Technische Hochschule Nürnberg

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

Mitglieder der Prüfungskommission

Referent:	Prof. Dr. Jens Grabowski, Institut für Informatik, Georg-August-Universität Göttingen
Korreferent:	Prof. Dr. Patrick Harms, Fakultät Elektrotechnik, Feinwerktechnik, Informationstechnik, Technische Hochschule Nürnberg
Korreferent:	Prof. Dr. Timo Jakobi, Fakultät Elektrotechnik, Feinwerktechnik, Informationstechnik, Technische Hochschule Nürnberg

Weitere Mitglieder der Prüfungskommission

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

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

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

Tag der mündlichen Prüfung
22.03.2024

Abstract

Increasingly, humans are surrounded by intelligent devices, such as smart devices, in the *Internet of Things* (IoT) context. These smart devices are used in various environments, such as smart environments as well as homes. On the one hand, these devices allow users to carry out daily routines comfortably at home. On the other hand, these devices simultaneously collect a vast amount of data types to execute the functionalities of these devices. Especially in home environments, the control of these data is essential as these devices are installed in private environments and spaces like homes. Also, sensitive and personal data are collected, processed and shared, especially in these private environments. Laws such as the *General Data Protection Regulation* (GDPR) require more transparency and user-centric data controls. Furthermore, several recent research projects, such as the *Security Assistance Manager* (SAM) from Fraunhofer Institute, are planning to develop standards for data privacy, -security and -control for smart home devices.

This work aims to present user-centric data stream controls for smart home devices to provide users with more transparency and control of data streams, including collected, processed and shared data types in this context. Furthermore, we also present a proposal for providers and users of smart home devices, which allows them to focus on addressing the GDPR requirements and increasing user acceptance of those devices.

We conducted two user-centric experiments to derive the user-centric data stream controls, called *User-Centric-Control-Points* (UCCPs), and their setting options to achieve this goal. The results of the questionnaire-based user-centric experiments allowed us to derive six UCCPs for smart home devices. Subsequently, we further analyzed six existing smart home devices and to what extent the derived UCCPs are supplied by those devices. Based on this analysis, we derived a proposal called quality check for providers and users of smart home devices. This proposal enables providers to evaluate and classify their devices regarding the provision of the derived UCCPs to increase user acceptance and implement the required GDPR requirements. Moreover, our quality check solution allows users to view and set supplied UCCPs during purchase and setup. Additionally, our solution allows users to gain more transparency on which data types are collected, processed, and shared for which smart home device functionalities. Our proposal in this work can also be deployed in environments other than smart homes.

Zusammenfassung

Zunehmend werden die Menschen von intelligenten Geräten (Smart Devices) umgeben. Diese Geräte werden in verschiedenen Umgebungen eingesetzt, unter anderem auch im privaten Haushalt. Auf der einen Seite ermöglichen diese Geräte eine komfortable Erledigung der täglichen Routinen im Haushalt. Auf der anderen Seite sammeln und verarbeiten die Geräte eine Vielzahl von Datentypen, um die Funktionalitäten der Geräte zu gewährleisten. Besonders im eigenen Haushalt ist die Kontrolle dieser Daten essenziell, da in dieser Umgebung persönliche und besonders sensible Daten gesammelt werden. Gesetze wie z.B. *Datenschutz-Grundverordnung* (DSGVO) verlangen mehr Transparenz und nutzerorientierte Datenkontrollen. Diesbezüglich gibt es verschiedene Forschungsprojekte wie z.B. *Security Assistance Manager* (SAM) von Fraunhofer Institut, um Standards für Datenschutz, Datensicherheit und Datenkontrolle für intelligente Haushaltsgeräte (Smart Home Devices) zu entwickeln.

Ziel dieser Arbeit ist es, nutzerorientierte Datenstromkontrollen für die Nutzung der intelligenten Haushaltsgeräte vorzustellen, um so den Nutzern mehr Transparenz und Kontrolle über Datenströme zu ermöglichen. Weiterhin stellen wir im Rahmen dieser Arbeit einen Lösungsvorschlag für Anbieter und Nutzer der intelligenten Geräte vor, um die Anforderungen der DSGVO umzusetzen und die Nutzerakzeptanz für diese Geräte zu erhöhen.

Um dieses Ziel zu erreichen, haben wir zwei nutzerorientierte Experimente in Form von Umfragen durchgeführt, um daraus nutzerorientierte Datenstromkontrollen, genannt *User-Centric-Control-Points* (UCCPs), inklusiv Einstellungsmöglichkeiten abzuleiten. Aus den Umfragen leiteten wir sechs UCCPs für intelligente Haushaltsgeräte ab. Danach haben wir sechs auf dem Markt vorhandene intelligente Haushaltsgeräte analysiert, inwiefern die abgeleiteten UCCPs von diesen Geräten ermöglicht werden. Auf Basis dieser Analyse haben wir eine Lösung, quality check, für Anbieter und Nutzer der intelligenten Geräte entwickelt und vorgestellt. Unsere Lösung ermöglicht es den Anbietern, ihre intelligente (Haushalts-)Geräte bzgl. der UCCPs zu überprüfen und zu klassifizieren, um auf diese Weise die Nutzerakzeptanz zu erhöhen sowie die Anforderungen der DSGVO umzusetzen. Außerdem bietet unsere Lösung den Nutzern die Möglichkeit, mehr Transparenz zu erlangen, welche Datentypen vom Gerät für welche Funktionalität gesammelt, verarbeitet und geteilt werden. Weiterhin sollen Nutzer dadurch die Möglichkeit haben, sich bereits beim Kauf die möglichen UCCPs anzuschauen, die bei der Inbetriebnahme eingestellt werden können. Unsere vorgestellte Lösung lässt sich auch in anderen Umgebungen als in privaten Haushalten einsetzen.

Acknowledgements

At this point, I would like to thank several persons who supported me during my PhD studies. First of all, I would like to thank my supervisors, Prof. Dr. Jens Grabowski, Prof. Dr. Patrick Harms and Prof. Dr. Hogrefe. My special thanks go to Prof. Dr. Graboswki and Prof. Dr. Harms, who guided me to the end of my PhD work. I am infinitely grateful for the human nature of these two professors, who showed me that professors can also be like gods, and that's when I experienced and saw what the word doctoral supervisor means. I would especially like to thank Prof. Dr. Harms, who gives students a chance and motivates them to achieve their goals. I have rarely come across such great people as professors.

I additionally, want to thank Prof. Dr. Ina Kersten and Dr. Dirk Kamin, who supported me during a tough time during my PhD studies besides my job. I really appreciate how they support me during this tough time. Prof. Dr. Kersten will always be a special person for me, and I will be grateful to her for the rest of my life. I would also like to thank Prof. Dr. Baum, who supported me with advice during my PhD studies. Moreover, I would also like to thank Prof. Dr. Kerstin Strecker, Prof. Dr. Florin Manea, Prof. Dr. Timo Jakobi and Prof. Dr. Bäumel-Courth for their support and time.

Many thanks go to my colleagues from the research group: Dr. Johannes Erbel, Dr. Philip Makedonski, Zaheed Ahmed, Lindrit Kquiku, Patrick Kühtreiber, Luca Hernandez-Acosta, Dr. Jan Tolsdorf, Dr. Alexander Railean, Dr. Alexander Richter. My colleagues always gave me valuable feedback and hints to improve my work. I would also like to thank Prof. Dr. Delphine Reinhardt for the initial chance and sparring, as well as Birgit Schubauer and Daniel Franke for reviewing my scientific papers and giving me valuable feedback. I would especially like to thank Dr. Philip Makedonski and Zaheed Ahmed for reviewing the thesis and giving me valuable feedback, which helped me a lot to improve my work and achieve the best result. I also thank Suran Jayathilaka for sparring and my friends, who motivated me and accepted that I had to work on my doctoral thesis during my free time.

Finally, I want to thank my golden parents, Ganga and Philip Benjamin, who supported me during my PhD studies and my whole life. They always believed in me and supported me in every step of my life. They both always had a supporting hand and tips to strengthen me to overcome any challenges I faced until today. I am incredibly grateful to have such wonderful parents in my life. My thanks also go to my whole family, who supported me during my PhD. I can't express my inner happiness in words. By completing my PhD, I am fulfilling one of my main wishes in life, which I have had since I moved to Germany at the age of 14. I would also like to dedicate this work to my grandmother, Karunawathee Samaraweera, who was my second mother and supported me as long as she lived.

Contents

List of Figures

List of Tables

Acronyms

AES *Advanced Encryption Standard*. 44

App *Application*. xiii, 44, 49, 114

AWS *Amazon Web Services*. 44

DSC *Data Sharing Information Categories* . xv, 38, 40

DSGVO *Datenschutz-Grundverordnung*. v

E2E *End-to-End*. 9

FIT *Fraunhofer Institute for Applied Information Technology*. 17

GDPR *General Data Protection Regulation*. iii, 1, 2, 11, 17, 77, 78, 79, 80, 86

ICT *Information and Communications Technology*. 47

IoT *Internet of Things*. iii, 1, 4, 7, 9, 10, 13, 14, 15, 17, 19, 21, 22, 73, 77, 79, 80, 81, 86

ITU *International Telecommunication Union*. 7

OCF *Open Connectivity Foundation*. 9

RBAC *Role Based Access Control*. 14, 73

RQs *research questions*. xv, 3

RSA *Rivest–Shamir–Adleman*. 44

SAM *Security Assistance Manager*. iii, v, 2

SVM *Support Vector Machines*. 78

TLS *Transport Layer Security*. 44

UCCPs *User-Centric-Control-Points*. iii, v, xv, 2, 3, 5, 9, 10, 11, 13, 16, 17, 19, 20, 34, 35, 36, 37, 41, 43, 45, 47, 48, 50, 51, 52, 53, 55, 56, 57, 60, 64, 65, 68, 69, 70, 71, 72, 73, 74, 75, 77, 78, 79, 80, 81, 82, 83, 85, 86

1. Introduction

After *Internet of Things* (IoT)'s first introduction in 1985 and its first use solving a problem, [1,2], it took around 20 years until it was integrated into our daily lives as smart devices carrying out the daily routines [1,3], making industrial and social processes more convenient, automatic and efficient. In 2023, there were already about 15 billion connected devices worldwide [4], and the predictions say that about 30 billion devices will be connected via the Internet by 2030 [4]. Smart home environments are one of the main environments in the IoT context in which such connected devices are integrated. All these smart home devices collect a vast amount of data types [5]. The questions in this context are whether the device owners know which data types are collected, shared with whom, and why. Until today, from a user perspective, there are no standards and opportunities to control the data streams, including data collection, processing, and disclosure, on the device or data type level. Furthermore, those data types are disclosed to third parties without users' consent. Disclosing data without users' consent includes several perspectives: (1) voluntary disclosure due to incomprehensible data stream settings, (2) lack of transparency about data streams, (3) data breach due to improper securing and (4) data type disclosure to gain advantages [5–8]. The rapid technological growth causes different smart home devices more and more surround us without standards for user-centric data stream controls [7]. Especially in smart home environments, such user-centric data stream controls gain an essential role because those smart devices are installed in personal spaces, such as at home [9]. The lack of those controls in those devices has a significant impact on data security and privacy [8,10,11] and includes the violation of norms of distribution according to the concept of contextual integrity from Nissenbaum [12]. Legislation is dealing with these data-related security and privacy issues in the IoT context worldwide, as we can see that laws like the *General Data Protection Regulation* (GDPR) with different Rights like "Processing of special categories of personal data", "Right for Access", "Right to be forgotten" (Art. 5, 9, 12, 15, 17 and 19), are still calling for privacy-preserving solutions with more user control and involvement [13–15]. Furthermore, existing works, such as Kounoudes and Kapitsaki [16], outline that there are still open research areas in addressing GDPR requirements regarding data stream controls, security and privacy issues from a user perspective in the IoT context [16]. The mentioned open research areas in this context are informed consent, data-related risk analysis, data minimization strategies, and context-aware privacy modelling considering users' perceptions [16]. Not addressing the issues mentioned above and not allowing users to control the data streams in the IoT context leads to less acceptance of the users [5] and less application of such technological progress. Less acceptance and application will lead to humanity not

making use of these technologies, which does not allow us to benefit from technological development and contribute to the efficient and effective development of our society, economy and environment [5].
Furthermore, a current survey by the Fraunhofer Institute [17] recently started a project in order to derive a *Security Assistance Manager* (SAM), which follows the aim to improve the current poor, smart home device standards in data protection and security areas by supplying users overviews regarding collected data types and their further usage [17]. This also outlines the need to address the open research aims in this area. Hence, this thesis focuses on deriving data stream controls from a user perspective, called *User-Centric-Control-Points* (UCCPs), which allows us to address the existing research gaps regarding data-related security and privacy issues in the smart home context. Additionally, we also present a UCCP-based proposal as a standard for device classification and configuration, including a progress report for controlling data streams of smart home devices. In our thesis, we also outline how the device provider can use the UCCP-based proposal to increase user acceptance regarding their devices by providing the users with transparency regarding collected and controllable data types while purchasing and using the corresponding devices.

1.1. Motivation

The data streams of smart home devices must be transparent and controllable for the users. Today, humans are dealing with a vast amount of different smart home devices to make their daily routines comfortable. Simultaneously, the users are also concerned about the data streams while using these devices. Controlling data streams from a user perspective has the main advantage that our society will accept those devices as controllable and use them in their daily lives with less or no concerns. Additionally, will help to address the open issues in the context of different laws, like the GDPR. From a user perspective, controlling the data streams of different devices interacting with other devices in the smart home environment is challenging. First, the users must gain transparency regarding the data streams of their devices. Second, the devices should supply UCCPs for users to control the data streams. In order to address these steps, we introduce in this thesis UCCPs and a UCCP-based proposal, which allows users to gain transparency regarding the data types of the data streams and to control the data streams of corresponding devices. Furthermore, we outline how the users can set the supplied UCCPs and receive a progress report regarding applied settings while using those devices. Our proposal also supplies device providers with a solution that can be applied to make their devices more acceptable to the users by allowing users to gain transparency regarding collected and controllable data types in the purchasing process.

1.2. Scope of the Thesis

This thesis focuses on improving the UCCPs in the smart home context. For this purpose, we present UCCPs to control the data streams of smart home devices. In this context, the main hypothesis is that controlling the data streams in smart homes from a user perspective is possible. To evaluate this hypothesis, we focus on answering the following four *research questions* (RQs):

- **RQ 1:** To what extent are users aware of the data streams of smart home devices?
- **RQ 2:** What are the UCCPs for smart home devices?
- **RQ 3:** To what extent do the existing smart home devices supply UCCPs?
- **RQ 4:** How to consider the UCCPs while using smart home devices?

Tab. 1.1 presents in which chapter we address each of the RQs.

Chapters	**RQ 1**	**RQ 2**	**RQ 3**	**RQ 4**
Chapter 4	●	●		
Chapter 5			●	●
Chapter 6				●

Table 1.1.: Overview of the addressed RQs in the respective chapters

1.3. Goals and Contributions

With this thesis, we address the existing research gaps and contribute to the following points:

- A contribution regarding the **transparency of data streams of smart home devices** to improve users' understanding and perspectives (Sec. 4.1 and Chapter 5).
- A **set of UCCPs** for devices, allowing users to control the data streams in the smart home context (Section 4.2).

- A combination of the contributions mentioned above into a **proposal** (Chapter 6), **which is applicable to different smart home devices by the device providers and users**. This combination allows, on the one side, users to holistically control the data streams of their devices while using the advantages of technological progress in the IoT context and helps device providers, on the other hand, to improve user acceptance regarding their devices.

1.4. Impact

During the preparation of this thesis, the following papers have been published in peer-reviewed conference proceedings in order to contribute to the existing research gaps in this context:

- C. I. Wickramasinghe and D. Reinhardt, "A Survey-based Exploration of Users' Awareness and their Willingness to Protect their Data with Smart Objects," *in Proceedings of the 14th IFIP Summer School on Privacy and Identity Management Data for Better Living - AI and Privacy*, vol. 576, pp. 427 – 446, 2020. [Online]. Available: `https://doi.org/10.1007/978-3-030-42504-3_27`
- C. Wickramasinghe and D. Reinhardt, "A User-Centric Privacy-Preserving Approach to Control Data Collection, Storage, and Disclosure in Own Smart Home Environments," in *International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services (MobiQuitous)*, pp. 190–206, 2021. [Online]. Available: `https://doi.org/10.1007/978-3-030-94822-1_11`
- C. I. Wickramasinghe, "Best-Practice-Based Framework for User-Centric Privacy-Preserving Solutions in Smart Home Environments," in *International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services (MobiQuitous)*, pp. 101–120, 2022. [Online]. Available: `https://doi.org/10.1007/978-3-031-34776-4_6`

1.5. Structure of the Thesis

We structured the thesis as follows: After giving an introduction in this Chapter, in Chapter 2, we describe the relevant background information, including definitions regarding IoT architecture, smart home categories, and the relevant results of the previous research work in this context.

The subsequent Chapter 3 presents the existing works in the context of this thesis topic. While putting our work in the context of existing works, we also outline in this Chapter the current research gaps regarding the control points for data streams in smart homes from a user perspective. We subdivided this Chapter into three subsections, in which we first describe our methodology of the related work analysis (Sec. 3.1), then present the results of the literature review (Sec. 3.2 and 3.3), and subsequently, we summarize and outline the research gaps in this context (Sec. 3.4).

Chapter 4 presents the results of our two user-centric experiments (Sec. 4.1) first and subsequently, the derived UCCPs for smart home devices (Sec. 4.2).

The subsequent Chapter (Chapter 5) presents the results of our smart home device evaluation. In our evaluation, we outline to which extent the derived UCCPs from Chapter 4 are supplied by the existing devices.

In Chapter 6, we present a proposal for device classification and configuration based on derived UCCPs including a progress report. In this Chapter, we propose a solution for smart home device providers and users.

We discuss our results in Chapter 7 while presenting the strengths of our results (Sec. 7.1) and outlining the limitations of our work (Sec. 7.2). In Chapter 8, we conclude the thesis, including suggestions for future research work.

2. Foundations

In this chapter, we introduce the foundations of this thesis, which includes relevant terminologies and basic concepts in the context of this research topic. First, we introduce terms related to *Internet of Things* (IoT) architecture, smart homes, their components and related controls. Second, we introduce the relevant service-oriented categories in smart home environments and the relevant results from previous research work.

2.1. IoT Architecture

According to the German Federal Ministry of Economics and Climate Protection, IoT makes our daily usable devices, such as thermostats, bulbs, cameras, cleaners, washing machines, and toasters, more intelligent through integrated programmable memories, sensors and communication capabilities via the Internet [18]. Furthermore, The *International Telecommunication Union* (ITU) defined IoT as "...a global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving, interoperable information and communication technologies." [19]. In this thesis, we define **IoT** as the technology which allows physical devices of our daily life with embedded sensors to communicate and data transformation between each other via the Internet (derived from [18, 19]).

In general, the IoT ecosystem consists of three layers: (1) IoT device, (2) service and (3) enterprise layer [20]. In Fig. 2.1, we present the IoT ecosystem from [20]. While the IoT device layer comprises the physical smart devices and gateways, the service layer consists of IoT platform provider services, such as data flow, processing and storage. The enterprise layer includes different applications and service management components [7, 20].

The integration of IoT devices in homes is called **smart homes**, according to the German Federal Ministry of Economics and Climate Protection [18]. According to Statista, the market volume in this area will reach more than 200 billion euros by 2028 from today's ca. 140 billion euros expected revenue [21]. The integrated IoT devices in the smart home context are defined as **smart home devices** and, in the course of this thesis, called devices. Some examples of those devices are smart bulbs, door locks, thermostats, washing machines, cleaning robots, cameras, and voice assistants [3, 9, 22]. Smart homes allow us to improve the comfort and security of our daily lives by supporting humans in their daily routines [18] with devices allowing users to control, monitor and regulate different functionalities [3, 9, 22]. An example of a smart home environment is presented in Fig. 2.2. In this way, it should

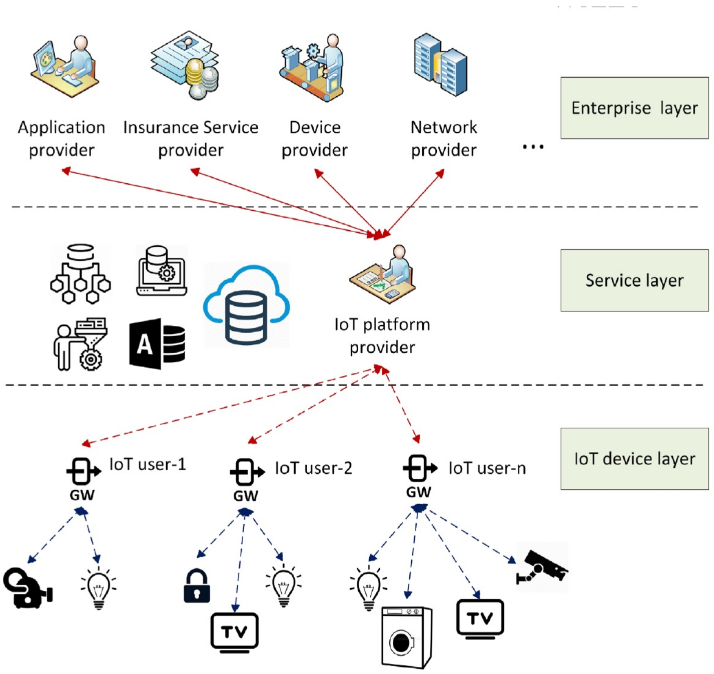


Figure 2.1.: IoT ecosystem with three layers [20]

be possible for our washing machine to independently find the cheapest electricity rate during the day and run the washing programs at the time with the most affordable electricity rate [18]. The **device functionality** includes the scope of the services or operations of any device [9, 24, 25], such as smart robots, bulbs or heaters, supply functionalities like cleaning, automatic lighting and heating own house. The device functionality can be divided into two categories. The two categories are main and optional functionality. The **main functionality** of a device includes the scope of the services or operations of any device to supply users with the essential functionality of the corresponding device, such as a smart camera monitoring the house. In comparison, the **optional functionality** of a device consists of the scope of the services or operations of any device to supply users with the additional functionality of the corresponding device, such as talking via a smart camera and remote control while monitoring the house. Each device collects different data types to allow users to use the corresponding device functionalities [5, 9, 24, 25]. A **data type** is a piece of individual information collected by a unique device, such as fingerprint and name. The data types necessary for the main functionality of the devices are called **mandatory data types**. The data types for the optional functionality of the device are called **optional data types**. One data type includes different data points. A **data point** is an individual value of a data type collected by a device, such as the measured values of a smart heater.

While using those devices in the smart home context, the users can apply data stream

Figure 2.2.: An example for smart home environments based on [23]

controls in two main layers of the IoT ecosystem, IoT and the service layer. We define the **data streams** as collecting, processing, and sharing different data types while using devices. Furthermore, we define the **data stream controls** in the context of this thesis as mechanisms which can be applied to control the data streams. Regarding smart homes, data stream controls include the definition that a device's collected, processed and shared data types can be monitored, examined and limited. In order to allow users to apply *User-Centric-Control-Points* (UCCPs) in smart homes in the IoT and service layers, those UCCPs must be integrated into a framework allowing interoperability between different IoT devices and systems. One of those standards is the ISO/IEC 30118-1 [26].
ISO/IEC 30118-1 is a framework based on the *Open Connectivity Foundation* (OCF), a common standard for interoperability between different IoT devices and systems [26]. The framework also includes architectures, protocols, interfaces and services for IoT implementation, which enables seamless communication and data exchange between various devices [26] regardless of the **device providers**, called providers in the context of this thesis. Fig. 2.3 presents the overview of the ISO/IEC 30118-1 architecture. The OCF framework is based on open standards and protocols, including solutions for establishing connections and managing information flow between the different IoT devices [26, 27]. Furthermore, this framework is scalable and supports a wide range of IoT devices and use cases with simple integration [26, 27]. In addition, the OCF framework also contains robust security mechanisms, such as *End-to-End* (E2E) encryption, authentication, and

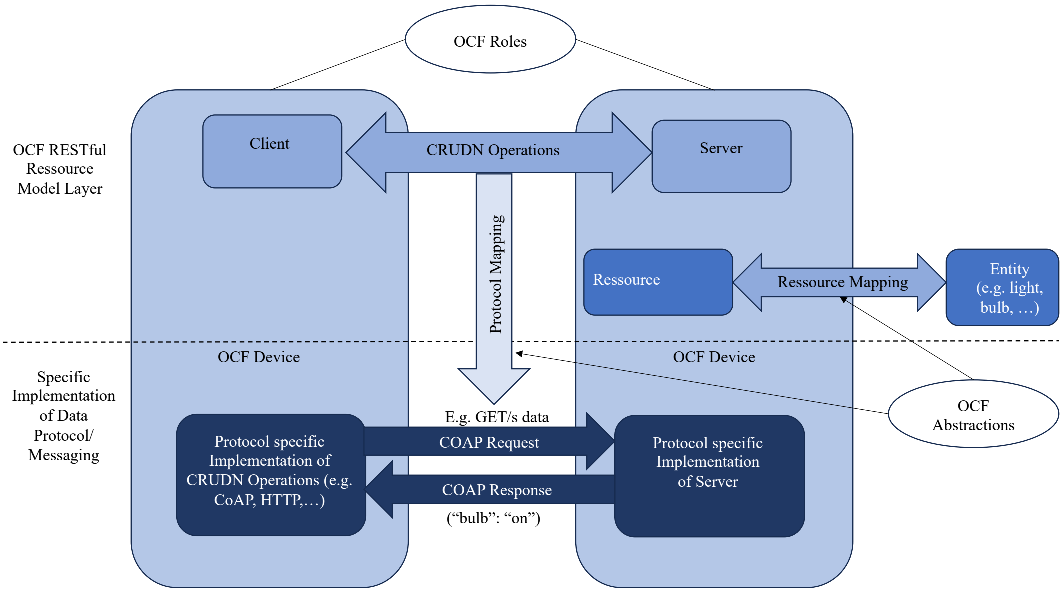

OCF Roles
OCF RESTful Ressource Model Layer
Client
CRUDN Operations
Server
Protocol Mapping
Ressource
Ressource Mapping
Entity (e.g. light, bulb, …)
OCF Device
OCF Device
Specific Implementation of Data Protocol/ Messaging
OCF Abstractions
E.g. GET/s data
COAP Request
Protocol specific Implementation of CRUDN Operations (e.g. CoAP, HTTP,…)
COAP Response
("bulb": "on")
Protocol specific Implementation of Server

Figure 2.3.: ISO/IEC 30118-1 framework based on [26]

authorization mechanisms, to mitigate risks of data breaches and cyber-attacks [26, 27]. The OCF is an industry foundation that integrates the latest trends and technologies in the IoT ecosystem [26]. Since the OCF allows the providers to certify their devices according to this standard [26, 27], the best way of integrating the UCCPs for the devices is to integrate such control mechanisms in the ISO/IEC 30118-1 framework.

2.2. Smart Home Categories

In this section, we present the considered service-oriented smart home categories while investigating the RQs of this thesis. According to N. Paulsen and S. Klöß from Bitkom [28], ca. 41% of Germans use devices in their homes, and according to the forecasts, the utilization trend will increase further in the future. Moreover, these statistics also cluster the applied devices in service-oriented categories. The categories are [21]:

1. Security,
2. Comfort and lightning,
3. Energy management,
4. Household appliances,

5. Networking and control,
6. Entertainment

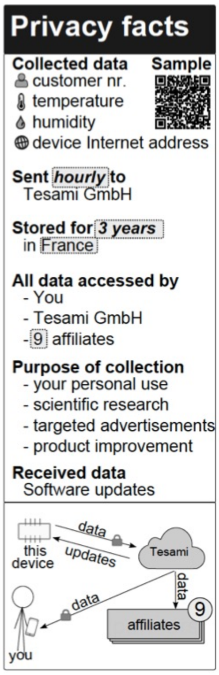
We consider the abovementioned six categories as the main categories of devices in our research. Different devices can be clustered into the six service-oriented categories. To ensure representativeness, we consider one device in each category, which we derive based on various comparative tests and statistics, such as [29–31]. The considered devices in each category are listed below and derived based on current statistics [29–31] to consider popular and most applied devices in the six considered categories. The considered statistics from [29] are presented by Bitkom e. V., the industrial association for the German information and telecommunications sector and the analyses from [30] are presented by Chip, the largest consumer portal in Germany, carrying out trusted product comparisons. In addition, the statistics from [31] are presented by Statista, a German online platform presenting different market and industry statistics. We consider the previously mentioned statistics as primary sources for our evaluation to consider trustworthy and current comparative analyses in this context. The considered devices in each category are:

1. Security: Smart cameras [29, 31]
2. Comfort and lightning: Smart bulbs [29, 31]
3. Energy management: Smart thermostats [29, 31]
4. Household appliances: Smart robot cleaner [29, 31]
5. Networking and control: Smart control hub [30]
6. Entertainment: Smart voice assistant [31]

In Sec. 5, we will consider these derived smart home categories and the focused devices to analyse the functionality, collected data types, and supplied UCCPs.

2.3. Privacy Facts Label

To address the *General Data Protection Regulation* (GDPR) requirements, one of the previous works proposed a *Privacy Facts Label* in their work [5]. In [5], they presented a label based on usability tests for devices to increase transparency regarding data collection and disclosure while using those devices. The proposed label is derived for an imaginary device to present the logic of *Privacy Facts Label*. The Fig. 2.4 presents the original *Privacy Facts Label*.



Privacy facts

Collected data Sample
customer nr.
temperature
humidity
device Internet address

Sent *hourly* to
Tesami GmbH

Stored for *3 years*
in France

All data accessed by
- You
- Tesami GmbH
- 9 affiliates

Purpose of collection
- your personal use
- scientific research
- targeted advertisements
- product improvement

Received data
Software updates

this device
data
updates
Tesami
data
9
affiliates
data
you

Figure 2.4.: *Privacy Facts Label* [5]

The label from [5] outlines the following information regarding the corresponding device.

- Collected data
- Collected data communication frequency
- Shared data
- Shared data recipients
- Purpose of data collection
- Received data
- Picture regarding data transformation path

In this thesis, we recommend integrating the derived UCCP-based proposal into an adjusted version of the original *Privacy Facts Label* (see Sec. 7.1.2).

3. Related Work

In our research work, we propose UCCPs and a UCCP-based proposal for device providers and users for smart home environments. In this Chapter, we present the related work in this research area. In order to outline the related work and present the research gaps, we conducted a literature review on the topic of UCCPs of devices. First, we briefly present the literature review methodology in Sec. 3.1. Second, we present the overview of already covered topics in Sec. 3.2 and 3.3. Subsequently, in Sec. 3.4, we outline the research gaps, delimit our work from existing works and summarize the findings of our literature review.

3.1. Literature Review Methodology

In order to review the relevant works in the context of our work, we consider in our literature review works focusing on the topics: (1) control mechanisms for data stream in the IoT context and (2) user perception regarding IoT data control issues. We used the following three-stage process in our literature review according to [32]:

1. **Step 1**: Identify relevant articles in leading journals and conferences
2. **Step 2**: Review the citations of the identified articles from Step 1 with a backward search
3. **Step 3**: Apply a forward search to identify further articles citing the articles from Step 2

In our search phase, we used scientific work indexing services and databases to gather an overview of existing and relevant works in the context of our research topic. We used the following database repositories for the identification of the relevant articles: Springer Link repository, Web of Science, IEEE Xplore, ACM Digital Library, ScienceDirect, Google Scholar, AIS Electronic Library, USENIX Papers Search. To identify the relevant papers, we used the following keywords as our search criteria:

- *article* including (IoT or smart homes) AND (data stream control or data controls) AND (data disclosure control or data collection control) AND (user-centric) AND (privacy or security or privacy controls).

We considered only peer-reviewed articles published in English with full-text availability in our systematic analysis. We clustered the literature review results according to the above-mentioned focus topics, and the results are presented in the following subsections.

3.2. Technical Solutions for Controlling Data Collection and Sharing

This section presents the existing technical solutions to control data collection and sharing in the IoT context. Different solutions have been proposed in this context. Existing works present solutions for avoiding the misuse of IoT devices and collected data by attackers for burglaries [33–36]. While Yu et al. [36] present an approach with a set of new security policies for detecting abnormal behaviour of each device, the approach from Hussain et al. [33] implements a strong password authentication policy in their smart home automation system. Additionally, Jia et al. [34] introduce a solution including a new context-based permission system allowing users to decide whether the abnormal action will be performed based on collected context information. The work from Perera et al. [35] proposes a Privacy-by-Design framework including a set of guidelines to evaluate IoT platforms. Summarized, the proposed guidelines are (1) minimization of the number of data sources, storage, acquisition, retention period, knowledge discovery; (2) data anonymization and encrypted data communication as well as data processing; (3) time period-, category-, chain-, knowledge discovery- and geography-based data aggregation; (4) pre- and post-information regarding information disclosure; (5) publication of the source code, data flow diagrams, certifications, applied standards and fulfilled compliance for code reviews. Furthermore, the existing works from Barhamgi et al. [37], Chakravorty et al. [38], Huang et al. [25, 39] and Ouaddah et al. [40] present different mechanisms including privacy preserved access control and *Role Based Access Control* (RBAC) with k-anonymity technique for IoT environments. The functionality of the frameworks is described based on the collected data from smart healthcare sensors, wearables and a few home and hotel automation systems. The authentication protocols presented in those frameworks allow users to share event-based data with specially defined roles by the device owner, and examples of such defined roles can be doctor, partner, colleague [24, 25, 37–40]. Especially, the proposed framework in [24, 37] contains a trade-off data sharing decision component allowing users to make data disclosure decisions after balancing privacy risks and potential benefits while using their smart healthcare devices. Moreover, solutions from Abutair et al., Alanezi and Mishra [41, 42] present further standalone privacy-preserving solutions, including security and privacy-preserving IoT architecture and privacy negotiation mechanisms ensuring privacy-preserving data storage and allowing end users to limit the data access. In addition, further approaches propose authentication protocols, context-based permission systems, privacy-preserving policies and data encryption methods to protect the collected sensor data of the users [34, 43–53]. Furthermore, in this context some technical standalone solutions are presented, which help users to differentiate between sensitive and non-sensitive data in the context of devices disclosing data, for instance the works from M. Keshavarz and M. Anwar [54] and A. Subahi and G. Theodorakopoulos [55]. The framework in [54] presents an active learning technique, which classifies the collected data of the devices as sensitive and non-sensitive based on users' privacy preferences. In comparison, the IoT-app inspector presented in [55] labels the data packets of the network traffic as sensitive or non-sensitive based on packet sizes and content, like for example user credentials, usernames, or user locations, to avoid privacy attacks. In further works [56–60], a few machine-learning solutions are presented, which deliver users data disclosure and privacy recommendations in different contexts. Sadeh et al. [56] propose a machine learning mechanism helping users to indicate disclosure preferences in a location-sharing system. In comparison, Fang and LeFevre [57] present a machine learning-based privacy wizard configuring privacy settings after asking users questions in the social network context. Moreover, the work from Knijnenburg and Jin [58] outline that receiving privacy recommendations by an assisted system is desired by users. Xie et al. propose in their work [59] a prediction algorithm allowing users to configure privacy settings in the location-sharing context and present that the data consumers and the context influence their location-related privacy preferences. Furthermore, the work from Pallapa et al. [60] propose a context-aware privacy-preserving solution in the mobile context. This approach derives privacy preferences based on historical interactions of the users among each other and applies those in new situations [60]. Khan et al. outline in their work [61] a solution to improve privacy concerns in the context of ownership change during device usage. Additionally, Zimmermann et al. [62] present a smart home configurator, which informs users regarding privacy and security implications to increase transparency and reduce the lack of clarity during decision-making in the purchase process. However, the approach from [62] does not allow users to control the data collection, storage, and disclosure process while using devices.

3.3. Surveys regarding Privacy Issues and Sensitivity Perception in IoT

Compared to Sec. 3.2, this section includes existing works, including user surveys, which investigate the perception and opinions of device users regarding privacy, security and data stream control issues in the IoT context. Besides outlining that users' primary motivation for using devices lying in convenience, Zheng et al. and Zeng et al. recommend in their works [8, 63] developers to design applications and devices allowing users to access and control the collected data according to users' privacy needs. Furthermore, the user survey from Martin and Nissenbaum outlines in their work [64] that the usage of the collected data is more relevant for the users than the sensitivity of the corresponding data. Additionally, a few large-scale surveys [65–67] outline that users' privacy issues regarding IoT devices highly depend on the context. Some of the user studies [68, 69] also gave hints for developers and device providers on how to address users' needs regarding privacy and security issues to increase user acceptance of devices like smart toys and watches. The hints include recommendations like visual recording indicators on devices, data storage controls, local data processing mechanisms, and transparency of the collected data [68, 69]. The works from Zeng et al. and Zimmermann et al. [8, 70] deliver hints regarding UCCP integrated solutions based on analysing users' mental and threat privacy models. To sum up, the recommendations include the following aspects: (1) providing transparency regarding privacy consequences in an understandable way and (2) limiting data recipients while focusing on presenting UCCPs for privacy- and data-preservation for the context-based data in smart home environments [8, 70]. Moreover, further works are carried out to analyse the sensitivity perception of the users in different contexts, for instance [71–77], because the sensitivity perception influences users' privacy awareness as mentioned in [71, 77–80]. While Milne et al. [71] present a typology which categorises 52 general information types into six clusters and outlines the perceived risk associated with different data types, Rumbold et al. [72] present a structured approach allowing ethical assessment to understand the sensitivity concerns regarding the healthcare data usage. They also outline that this approach only presents a rough guide because the sensitivity issues radically differ from person to person and highly depend on the context [72]. Moreover, further surveys are carried out to investigate the sensitivity perceptions, privacy, and security issues of users from different countries [73–76]. Schomakers et al. [76] present the sensitivity perception of 40 different data types in the context of internet usage. In their paper, they also compare the sensitivity perception of persons from Germany, Brazil, and the USA and outline that culture (on certain points), risk affinity, and education level influence the users' sensitivity perception. Additionally, Fietkiewicz and Ilhan [74] present in their paper the users' sensitivity perceptions and privacy concerns regarding the data collected by fitness tracking technologies based on an online survey with 590 participants from the EU and USA. The results outline that the perceived data sensitivity and privacy concerns are data type-dependent when using fitness-tracking applications. In addition, Al-Ameen et al. [73] outline in their paper, based on a qualitative user study with 32 Bangladesh participants from urban areas, that the users' privacy and security perceptions in the context of collecting sensitive information are influenced by the users' knowledge and their technical efficacy. Furthermore, Balapour et al. [77] investigated the effects of privacy perceptions on the security perceptions of mobile application users. They used two categories of mobile applications, as mobile applications are collecting less and more sensitive information. In their results, they outline that the sensitivity of the information collected by the mobile applications influences the users' privacy perceptions in that way that the relationship between the perceived effectiveness of privacy policies and perceived risk is stronger when using those mobile applications that collect more sensitive information [77]. Moreover, Kulyk et al. [75] compare the privacy and security issues of Germans, Romanians, and Spanish people with 575 participants in the context of smart home and health environments and outline that there are differences in risk awareness between the countries and very few participants mentioned specific harms arising from disclosing collected data types in the respective context.

3.4. Research Gaps and Summary

The related work outlines that more and more research works focus on the research area of privacy- and security preservation in the IoT and smart home context. The existing works include different approaches and standalone solutions which address privacy and security preservation regarding data streams in the smart home context and other contexts. In Sec. 3.2, listed approaches provide different frameworks and standalone mechanisms to address the security and privacy issues identified by different user surveys, as presented in surveys from Sec. 3.3. However, the considerations show us that the proposed solutions include the following lacks and gaps regarding UCCPs for IoT devices in the smart home context. To the best of our knowledge, the existing gaps are that proposed (technical) solutions include (1) less user involvement, (2) no clear hints regarding users' readiness and willingness to be involved in their privacy protection, (3) no clear UCCPs and control mechanisms for devices, for users and providers and (4) no possibilities to apply data minimization, aggregation strategies as well as data access and usage limitations. Additionally, none of the previous works focus on supplying solutions for providers to evaluate their devices according to supplied UCCPs. So far, the providers do not share those data transparently with the users to allow users to apply those UCCPs during the setup process and before using the devices.

Furthermore, laws like GDPR, especially Art. 9, 12, 15, 17, 19 and 22, are still calling for user-centric solutions allowing users to have transparency and control regarding the data streams while using devices [13–15]. Recently, *Fraunhofer Institute for Applied Information Technology* (FIT) started a user-centric project with at least two years to investigate solutions for smart homes to allow users to protect their data in the smart home context consciously [17]. This project also underlines that we need UCCP-integrated mechanisms in order to (1) address the data protection-, privacy- and security-related issues in the smart home context and (2) increase users' acceptance of the devices.

In our work, we address the above-mentioned research gaps and propose a quality check mechanism including UCCPs derived from the end user perspectives for devices. The device providers can use this quality check mechanism to evaluate and classify their devices in terms of supplied UCCPs. On the one hand, users can review the device classification in the setup process and set the supplied UCCPs based on the quality check results presented by the provider before using those devices. Addressing these gaps will help providers gain more and more user acceptance regarding their devices, improve their devices according to UCCPs, and address the GDPR requirements. In this way, our society can also benefit from technological progress. On the other hand, the proposed quality check mechanism allows users to control the data streams while using different devices. Additionally, the proposed quality check mechanism also allows us to address the GDPR requirements, especially Art. 5, 9, 12, 13, 14, 15 and 19, giving users more transparency and control in the data collection and sharing process in the smart home context [13].

4. Awareness of Data Streams and User-Centric Data Stream Controls regarding Smart Home Devices

This chapter first introduces the results of our user-centric experiments. Second, based on the user-centric experiment results, it introduces the derived UCCPs in the smart home context. The contents of this chapter have already been published in scientific papers [6, 7, 81] within the framework of this work. We first present in Sec. 4.1 the outcome of our user-centric experiments regarding the UCCPs. Then we present in Sec. 4.2 the derived UCCPs, which is the main output of this thesis to address the research gaps outlined in Sec. 3.4.

4.1. Users' Awareness regarding Data Streams of Smart Home Devices

In this section, we describe the two user-centric experiments carried out to understand users' awareness regarding data streams in the IoT and smart home context and to derive UCCPs for devices. First, we describe the setup and the commonalities of both user-centric experiments. Second, we present each experiment's specifications and associated results in separate sections.

4.1.1. User-Centric Experiment Setup

In both user-centric experiments, we carried out an online questionnaire-based survey, which was available in English to reach a representative number of participants from different countries. The questionnaires were distributed on several platforms like LinkedIn, Xing, IoT Subreddit, SurveyCircle, several companies' community platforms and a panel provider (ISO 26362 certified regarding experiment 2) to reach frequent Internet users. The questionnaire-based surveys included several matrix, multiple choice, and open-ended questions and required approximately fifteen minutes to be answered. A few volunteer participants tested the questionnaires before starting the final surveys. Based on the pretests, the survey questions were modified in advance to eliminate language errors and misunderstandings regarding the statements and questions. Both user-centric experiments help us to answer the research questions, RQ 1 and RQ 2 (from Sec. 1.2). The survey questions are presented in the corresponding sections with the results.

4.1.2. User-Centric Experiment 1: Awareness and Control Willingness

In the first user-centric experiment, we focused on answering the RQ 1 and RQ 2 and gaining user input to derive UCCPs. In order to answer RQ 1 and RQ 2, we asked our participants to answer 22 questions, and no incentives were given to the participants. First, the questionnaire gathered insights regarding participants' knowledge and experience with devices. Second, it addressed the participant's awareness of collected and disclosed data by devices and corresponding privacy risks. Third, it focused on participants' willingness to gain more transparency and control regarding collected and shared data by devices. Fourthly and finally, it gathered participants' requirements and motivation to apply UCCP-integrated solutions while using devices. The results of the user-centric experiment 1 have already been published in scientific paper [6]. The questions of this questionnaire are presented in Appendix A. During our analysis of this experiment's results, we tested the following five hypotheses:

- $H_{1.1}$: The users want to control the data collected by devices.
- $H_{1.2}$: The users want to get information regarding arising privacy risks from the disclosure of the collected data types.
- $H_{1.3}$: The users are willing to have an overview of the collected data types.
- $H_{1.4}$: The users want to determine who is allowed to access the disclosed data types.
- $H_{1.5}$: The users want to determine for which purpose the disclosed data types are used.

We applied statistical tests that fulfilled the predictions, Mann-Whitney, Fisher's Exact, multiple linear regression, and correlation tests, as well as group-wise comparisons. The testing allows us to (1) confirm the hypotheses mentioned above, (2) derive UCCPs, and (3) get more insights regarding the RQ 1 and RQ 2. The Mann-Whitney tests are applied to conduct the group-wise analysis regarding participants' awareness of data collection. Additionally, multiple linear regression and correlation tests are applied to investigate participants' willingness regarding information and privacy-preserving measurement usage. In order to confirm the above-mentioned hypotheses, we applied Fisher's Exact tests. In the following, we present the experiment results.

Demographics: In total, 229 participants completed the questionnaire. After discarding the invalid data entries based on invalid answers regarding Q_{13}, Q_{20} and time stamps, 209 valid data entries are considered in the analysis. 69% of our participants are male (Q_{19}) and 58% are between 26 and 50 years old (Q_{18}). Further, 25% are over 51 and 16% under 26 years. The majority of the participants are German with 74% followed by US Americans with 7%, British citizens with 5%, Sri Lankans with 5% and the remaining 9% are distributed among 15 other nationalities of the world (Q_{20}). In total, around 79% of the participants indicated their annual income range, which ranges between *less than 25.000 euro* and *more than 100.000 euro* (Q_{21}) and about 34% indicated that they annually earn *between 40.000* *and 75.000 euro.*

Knowledge: Around 93% of our participants indicated in Q_1 that they have already heard about IoT. To gain more insights, we asked our participants in Q_2 in which context they have heard about IoT. They mentioned the following contexts: (1) smart home (ca. 27%), (2) industry 4.0 (ca. 20%), (3) smart/intelligent things (ca. 19%), (4) smart city (ca. 19%), and (5) smart factory (ca. 13%). In an additional free text box, a few participants mentioned further contexts, like smart vehicles, smart clothes, wearables, smart meters, smart grids, smart supply chains, smart campuses, smart agriculture, robotic machines, smart logistics, smart health devices, and predictive maintenance. Moreover, about 89% of the participants indicated in Q_3 that they know or use devices and the frequently given answers were: (1) 12% controlling home technology apps, (2) 10% smart voice control devices, like Amazon Echo, (3) 8% smart health devices, (4) 8% smart door/window locks, (5) 7.5% smart bulbs, (6) 7.2% smart fridge, (7) 7% augmented/virtual reality glasses, (8) 6% smart washing machine, (9) 5.7% smart alarm clock, (10) 4% smart toothbrush, (11) 3% smart grid apps and (12) 2.7% smart scale. In addition, in Q_4, 71% mentioned using specific smartphone apps for this purpose and 11% used associated web interfaces. Furthermore, in Q_5, about 70% of the participants mentioned that they use IoT devices connected to the Internet frequently, and among these participants, 76% use them at least once per day, and 24% only use them occasionally. Analysis of the answers based on cross tables grouped by gender and age groups outlines that male participants and participants between 26 and 50 years significantly use IoT devices more frequently than the others. The answers regarding Q_5 allow us to derive the following three user categories, which we consider in our further analysis.

1. Frequent users: Using smart devices more than once a day,
2. Average users: Using smart devices at most once a day or less,
3. Non-users: Not using any smart devices.

In Q_{11}, we asked our participants to rate the following statement: *In a few years, I believe that it will be difficult to live without using smart objects*, by using a 5-point Likert scale [1]. The answers outline that 87% of participants agreed it would be difficult to live without device utilization. While most of our participants indicated that they appreciate the advantages of devices in Q_{12}, 20% mentioned that they do not see any advantages that devices offer. In the following, we list the seven most frequently mentioned advantages by the participants: (1) support in the execution of routine tasks, (2) increased comfort and convenience, (3) reduce error rates, (4) adjust according to own lifestyle, (5) record interesting personal insights, (6) find out optimization potentials and (7) carry out specific things/tasks automatically.

Awareness regarding Data Streams: While about 93% of our participants believe that the devices collect data about themselves and their environments (Q_6), only 58% indicated that they know the data types collected by those devices (Q_7). Our participants indicated

[1] The value of 1 corresponds to a strong disagreement, while a value of 5 to a strong agreement.

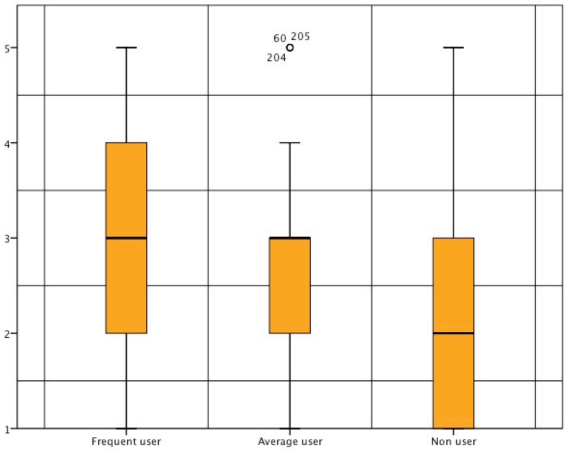


Figure 4.1.: Boxplots regarding Q_7: *I believe that I know the information collected by smart objects*, grouped by user profiles (Q_5) [6]
(1 = strong disagreement, while 5 = strong agreement)

in Q_8 location (29%), health (25%), browsing (24%) and personal data (19%), like bank details as collected data types by those devices. Considering the derived user categories shows that frequent and average users seem to be more aware of the data collection than the non-users ("...frequent users vs non-users: $p-value = 0.003 < 0.05$, r = 0.248, average users vs non-users: $p-value = 0.048 < 0.05$, r = 0.195, Mann-Whitney test" [6]). Fig. 4.1 underlines the results mentioned above, while the outliers represent the answers that deviate from the most frequently mentioned answers of most participants. Each outlier in Fig. 4.1 represents the record of the corresponding anonymous participant.

Our results outline that a minority of our participants (24%) believe that they know the third parties accessing the collected data types by devices (Q_9), and 72% mentioned that they do not know the third parties accessing the collected data types. Asking the participants to mention the third parties they know who access the collected data types by those devices in a free text box, delivered the following answers: "retail companies (like Amazon, Apple), service providers (like Google), cyber security firms, social media companies (like Facebook), several smart object/telco providers, institutes/companies using data for statistics and analyses, (health) insurance companies, hospitals, doctors, manufacturers of the heating systems/cars, banks and government departments" [6]. In Q_{14}, we asked our participants to indicate the potential privacy issues and risks in the context of IoT by using the given free text box. To summarize their statements, they mentioned that the IoT devices collect and share a vast amount of data types with third parties, while those devices make our daily lives and routines easier. They also mentioned that those shared data types are used for individualized services or offers or to create (more or less) detailed personal profiles and to manipulate the device owners. Most of our participants (93%) indicated in Q_{15} that they believe devices endanger their privacy. About 38% mentioned in Q_{10} that they carry out special measures to protect their privacy while using those devices, and 48% declined it. In this question, the participants were also asked to mention the special measures they take to preserve their privacy. The mentioned answers were: (1) 35% switching off the devices to avoid data collection, (2) 57% checking the privacy settings and adjusting or disabling the device or device functionalities, (3) 6% using local servers rather than the cloud connections. Additionally, 2% of our participants mentioned that they do not know any measurements to protect their privacy.

Willingness regarding Data Stream Control: In order to analyse whether our participants are willing to control the data streams, data collection, processing and disclosure in Q_{16}, the participants had to rate the given statements using a 5-point Likert scale. Fig. 4.2 presents the distribution of the participants' answers, where the outliers present the divergent answers of the corresponding anonymous participants. While in $Q_{16.1}$ about 94% of our participants indicated that they want to have more transparency regarding collected data types about themselves and their surroundings by devices, 96% mentioned $Q_{16.2}$ that they are willing to have an overview of all the collected data types by used devices. Additionally, the answers regarding $Q_{16.3}$ show that 94% of our participants want to see a summary of the collected data types over a given period, like daily, weekly, and monthly. About 84% indicated in $Q_{16.4}$ that they want more information regarding the collected data types in their smart home environments in real time. Moreover, about 92% of the participants want more information regarding associated privacy risks caused by disclosure of the collected data types to increase the transparency regarding collected and shared data types ($Q_{16.5}$). Additionally, about 87% of the participants mentioned in $Q_{16.6}$ that they also want more information regarding arising personal and social advantages caused by the data disclosure. Our analysis also shows that the participants want more information regarding associated privacy risks and advantages caused by simultaneously sharing the collected data types. This observation is confirmed by the positive correlation between the statements from $Q_{16.5}$ and $Q_{16.6}$ ("r = 0.608, significant at the 0.01 level - 2-tailed" [6]). Furthermore, a majority of our participants (97%) in $Q_{16.7}$ and in $Q_{16.8}$ mentioned that they want to control the data types collected and shared by devices. Our analysis also shows that there are no statistically significant differences between the answers given by the participants clustered in the derived user categories from Q_5 (Mann-Whitney test: $p-values > 0.05$) and the participants applying special measures to preserve their privacy in Q_{10} (multiple linear regression test: $p-values > 0.05$).

The analysis of the statements from $Q_{16.9}$ and $Q_{16.10}$ show that a majority want to determine which third parties can access the collected data types ($Q_{16.9}$: 95%) and for which purpose ($Q_{16.10}$: 95%). On the other hand, 86% of our participants mentioned in $Q_{16.11}$ that they want to spend time on auditing the collected data types, while 74% of our participants indicated in $Q_{16.12}$ that they are willing to be supported by an automated system taking privacy decisions while taking such decisions. In $Q_{16.13}$, 96% of our sample indicated that they want clear policies with the provider regarding data collection in the smart environment

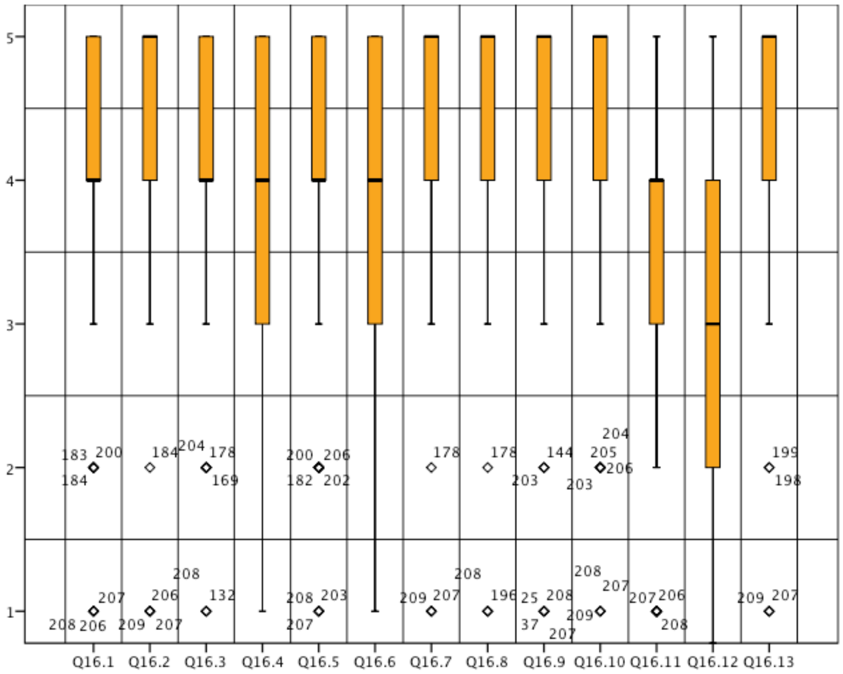


Figure 4.2.: Boxplots regarding Q_{16} statements [6]
(1 = strong disagreement, while 5 = strong agreement)

context. Furthermore, our participants mentioned in Q_{17} the motivating factors to use devices while having the opportunity to control the collected and shared data types. The following listing includes the most mentioned answers: “Having control about the usage of collected data about me (32%) followed by feeling myself secured and protected (29%) as well as avoiding to draw a digital biography (22%) and having information about the data consumer of my data (15%)" [6]. In the free text box, two participants mentioned no motivating factors for using devices while having data collection and disclosure controls. In addition, two other participants indicated that they do not use devices because of the lack of security, and one participant also mentioned that s/he would never want to spend time validating or examining the collected or disclosed data types.

Results regarding the Hypotheses: Our analysis of participants’ answers, especially regarding Q_{16}, allows us to investigate the five hypotheses mentioned above in Sec. 4.1.2. Our hypotheses allow us to investigate whether most participants, who want more transparency regarding collected data types and arising privacy risks and advantages from shared data types, are willing to consider that information while controlling the data collection and sharing process. In order to investigate our derived hypotheses and to find out whether there is a significant dependency between receiving more transparency regarding data collection and sharing and considering that information in controlling the data-sharing process, we applied the Fisher's Exact test. The results of the Fisher's Exact tests confirm all the considered hypotheses ($p = 0.00 < 0.05$) and show that the participants want to control the collected data types by devices (H_1) and also want to have an overview of the collected data types (H_3). Furthermore, the test results also show that the participants want to receive more information regarding associated privacy risks and personal and social advantages arising from data disclosure (H_2). While having control over collected and shared data types, the participants also want to determine who can access the shared data types (H_4) and for which purpose the shared data types are used (H_5).

4.1.3. User-Centric Experiment 2: Sensitivity Perception regarding Data Types of Data Streams

Subsequently to the first experiment, we carried out the second user-centric experiment to investigate the sensitivity perception of the participants in the context of collected and shared data types in smart home environments in order to improve the UCCP setting options and our answers regarding RQ 1 and RQ 2, from Sec. 1.2. In this user-centric experiment, we carried out a questionnaire-based survey with 11 questions (see Appendix B) after the approval of the University's ethics committee and data protection officer. In the analysis of the results, we only consider the answers regarding the questions: Q_1, Q_{2a}, Q_{2b}, Q_3, Q_5, Q_6, Q_7, Q_8 and Q_9. The questions Q_4, Q_{10} and Q_{11} are not considered in this survey analysis since these questions do not deliver further relevant insights to the considered topic in terms of content. In total, our questionnaire gathered different information: It mainly gathered participant's sensitivity perception regarding collected and shared data types in the smart home context and participant's perception regarding linkability of the collected data types in the considered context. Additionally, our questionnaire also gathered further insights regarding participants' sensitivity in the online and mobile context compared to the smart home context in order to investigate context and data type dependency of participants' sensitivity perception. We asked our participants to use a Likert scale while indicating their sensitivity perception in the corresponding questions. The scenarios considered in the questions and corresponding statements were derived from current surveys [82–84] about expected usage of devices in the coming years as well as most commonly used devices, allowing the comparability to the online and mobile scenarios, in which the same data types are collected. In this online questionnaire, we considered the six categories in the smart home context from Sec. 2.2. Those are (1) security, (2) comfort & lightning, (3) energy management, (4) household appliances, (5) networking & control and (6) entertainment. The considered devices in this experiment from these categories collect different data types, like energy consumption data, purchase habits, physical measurement data, sports data, biometrical data (face ID, fingerprint, voice print), private life data, data about lifestyle, availability, and current location at home, sleeping habits, diagnoses, and medical history data. We considered in this experiment the following comprehensible devices collecting the above-mentioned data types: smart door locks, speakers, scales, wearables, heaters, and smoke detectors. The results regarding sensitivity perception from user-centric experiment 2 have partially been published in the scientific paper [81]. We also investigated the following hypotheses during our analysis:

- $H_{2.1}$: The type of the data collected influences users' sensitivity perception in the smart home context.
- $H_{2.2}$: The users' sensitivity perception in online and mobile contexts differs from the sensitivity perception in the smart home context.

We applied different statistical tests associated with dependency analysis that fulfil the preconditions, Friedman, Wilcoxon signed-rank, Kruskal-Wallis H, and pairwise comparison (Bonferroni corrected) tests, to investigate the above-mentioned hypotheses and get more insights regarding UCCP setting improvements and RQ 1 and RQ 2. The Friedman and Wilcoxon signed-rank tests are applied to analyse the differences in sensitivity perception between the contexts of online, mobile, and smart homes. Additionally, the Kruskal-Wallis H and pairwise comparison (Bonferroni corrected) tests are used to conduct the group-wise analysis regarding differences in the participants' sensitivity perception. The questions of the questionnaire are presented with the results of the experiment in the following sections.

Demographics: In total, 810 participants completed the questionnaire. After discarding the invalid data entries, based on invalid answers regarding the mathematical question (Q_5), age (Q_6), gender (Q_7), countries (Q_8), 765 valid data entries were considered in the analysis. 58% of our 765 participants are male (Q_7) and 48% of the participants are older than 45 years (Q_6). Further, 27% are between 30 and 45 and 11% between 18 and 30 years old. In Q_8, 41% mentioned that they have resided in Germany most of the time during the past 15 years, 28% mentioned in the United Kingdom (UK), and 13% each in Switzerland and Austria. The remaining 5% of the participants mentioned countries, Canada, France, Italy, Sri Lanka, United States, Argentina, Denmark, Iran, Ireland, Israel, Japan, Kosovo, Moldova, Netherlands, Oman, Slovakia, Thailand and Turkey. In our group-wise analysis, we consider Germany, UK, Switzerland, and Austria as frequently mentioned countries (n = 729).

Smart Home Ownership: In our sample, 37% of our participants indicated in Q_9 that they own several devices. At the same time, 29% mentioned owning one device, and 13% mentioned that they are considering getting devices. Another 20% mentioned that they do not want to get any devices, and 1% have abstained from answering. Based on the responses in Q_9 (n = 760), we derived four user categories, which we considered in the group-wise analysis to analyse the differences regarding the sensitivity perception between those distributions:

1. Non-user: I do not want to get any devices,
2. Planning user: I am thinking of getting devices,

3. Less familiar user: I own one device,
4. Familiar user: I own a couple of devices.

Sensitivity Perception regarding Data Types of Data Streams in the Smart Home Context: Our analysis regarding Q_{2a} shows that the participants believe that most of the collected data types in the smart home context are directly linkable to themselves. The distribution of participants' answers is presented in Fig. 4.3, confirming these observations. The statements of the Q_{2a} are presented in Tab. 4.1. The results also outline that almost half of the participants think that the following data types are directly linkable to themselves:

- Biometrical data, like fingerprint or face IDs from smart door locks ($Q_{2a.1}$: 59%)
- Private details like home address from smart watches ($Q_{2a.4}$: 55%)
- Diagnoses and medical history from smart wearables ($Q_{2a.7}$: 45%)

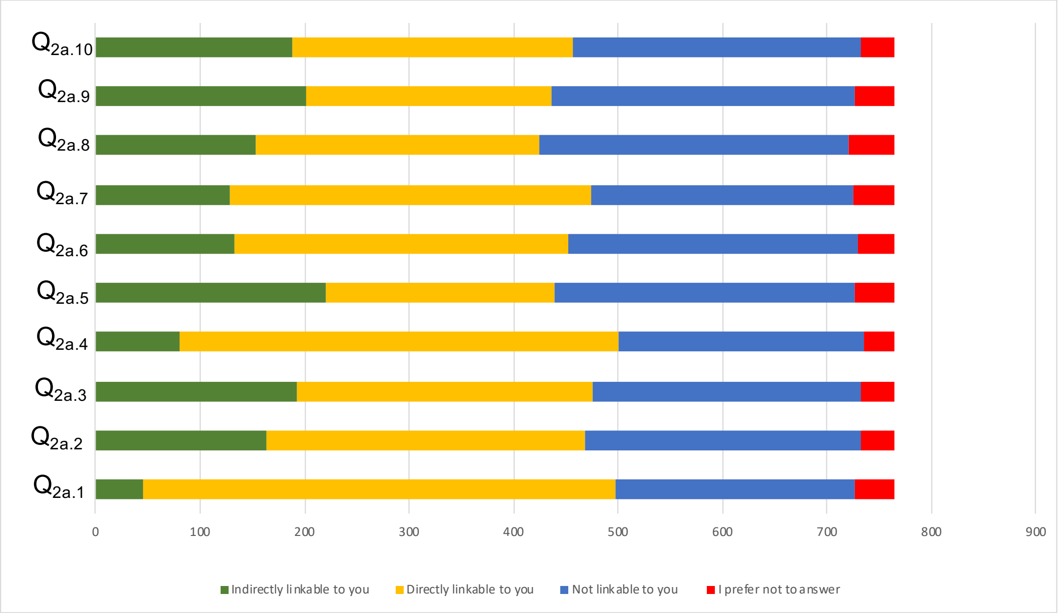


Figure 4.3.: Participants' perception regarding linkability of the data type from the statements of the Q_{2a} from Tab. 4.1 to own self

Further analysis regarding the statements $Q_{2b.1}$ - $Q_{2b.10}$ shows that the participant's sensitivity perception differs according to the collected data type in the smart home context, which we analysed in detail in the following. The statements of Q_{2b} are presented in Tab. 4.1, and the distribution of the participant's answers is presented in Fig. 4.4. The quantitative analysis confirms that there is a difference between participants' sensitivity perception depending

Numeration	Statement
$Q_{2a.1}$ and $Q_{2b.1}$	Biometrical data, like fingerprint or face IDs from your smart door lock.
$Q_{2a.2}$ and $Q_{2b.2}$	Voiceprint from your smart speakers.
$Q_{2a.3}$ and $Q_{2b.3}$	Physical measurements, like weight and height from your smart scale.
$Q_{2a.4}$ and $Q_{2b.4}$	Private details like home address from your smartwatch.
$Q_{2a.5}$ and $Q_{2b.5}$	Purchasing habits from your smart fridge.
$Q_{2a.6}$ and $Q_{2b.6}$	Personal preferences, like sexual preferences, political affiliation, and religion from your smart speaker.
$Q_{2a.7}$ and $Q_{2b.7}$	Diagnoses and medical history from your smart wearable.
$Q_{2a.8}$ and $Q_{2b.8}$	Information about your family and environments from your smart speaker.
$Q_{2a.9}$ and $Q_{2b.9}$	Lifestyle information, like energy consumption data, sleeping habits, doing sports from your smart heaters and watch.
$Q_{2a.10}$ and $Q_{2b.10}$	Information about your availability at home from your smart smoke detectors.

Table 4.1.: Statements from the survey questions: Q_{2a} and Q_{2b}

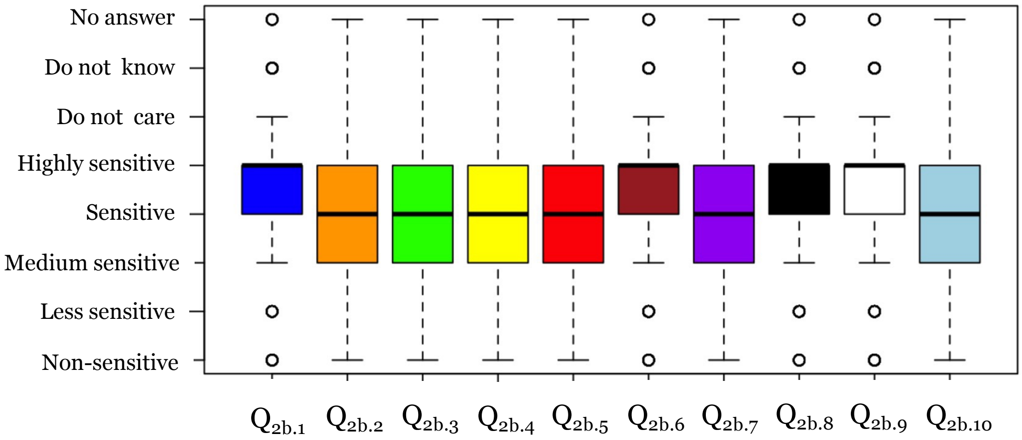


Figure 4.4.: Distribution of participants' responses regarding the statements $Q_{2b.1}$ - $Q_{2b.10}$

on the collected data type in the smart home context (x^2 = 267.62, p-value < 0.05, strong effect size: r > 0.5). Note that the mean rank comparison between the statements shows that participants' sensitivity perception is highest regarding collected biometrical data in the smart home context. Clustering the participants' responses, who rated *sensitive* and *highly sensitive* regarding the statements from Q_{2b}, shows that the collected biometrical data ($Q_{2b.1}$), and personal preferences ($Q_{2b.2}$) have the highest sensitivity perception, but also that at least 50% of our participants have a higher sensitivity perception regarding all the collected data types in the smart home context. The details regarding these observations are presented in Tab. 4.2. Furthermore, the group-wise analyses show that there are significant differences between the derived user categories (Q_9), age groups (Q_6) and few differences between considered countries, Germany, UK, Switzerland, and Austria (Q_8). Tab. F.1, F.2 and F.3 in Appendix F present the details regarding significant differences of the group-wise analyses based on the Kruskal-Wallis H tests. The additional mean rank analyses show that non-users, participants above 45 years and participants from Germany have a higher sensitivity perception regarding biometrical data.

Statements	Percentage of the Participants with sensitive or highly sensitive
$Q_{2b.1}$: Biometraical Data	71%
$Q_{2b.2}$: Voiceprint	53%
$Q_{2b.3}$: Physical Measurements	50%
$Q_{2b.4}$: Private details	52%
$Q_{2b.5}$: Purchasing habits	57%
$Q_{2b.6}$: Personal preferences	66%
$Q_{2b.7}$: Diagnoses and medical history	60%
$Q_{2b.8}$: Information about family and environments	64%
$Q_{2b.9}$: Lifestyle information	62%
$Q_{2b.10}$: Information about your availability at home	57%

Table 4.2.: Participants' sensitivity perception regarding the collected data types in Q_{2b}, clustered *sensitive* and *highly sensitive*

Sensitivity Perception regarding Data Streams in the Smart Home Context Compared to Online and Mobile Contexts: The responses regarding Q_1 and Q_3 allowed us to investigate the differences in the perceived sensitivity between the considered contexts when the same data type is collected (Q_1: online vs. smart homes and Q_3: mobile vs. smart homes). While Tab. 4.3 presents the statements of Q_1 and Q_3, Fig. 4.5 provides an overview of the distribution of participants' responses regarding the corresponding statements from Q_1 and Q_3. In Fig. 4.5, two boxplots have the same colour within the corresponding question when the collected data type is the same, and only the considered contexts are different. The boxplots in Fig. 4.5 outline that there are differences in participants' sensitivity perception

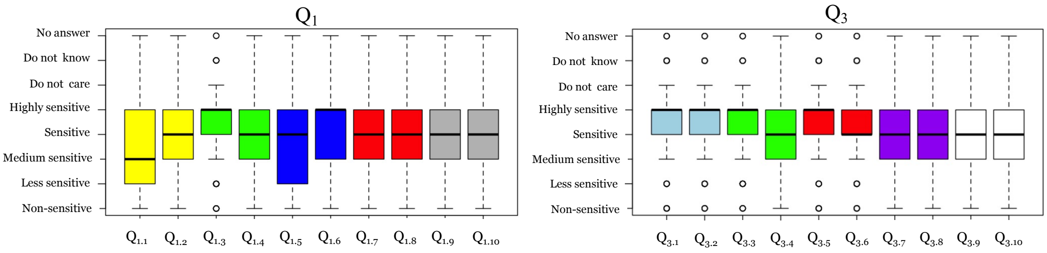


Figure 4.5.: Distribution of participants' responses regarding the statements Q_1 and Q_3 from Tab. 4.3

collecting the same data types in the respective context comparisons in most of the statements from Q_1 and Q_3. In the following, we analyse these observations in detail. The additional quantitative analysis confirms that there are differences between participants' sensitivity perception depending on whether the respective data types are collected by an online service, mobile application, or device (Q_1: x^2 = 298.02, p-value < 0.05, strong effect size: r > 0.5, Q_3: x^2 = 283.59, p-value < 0.05, strong effect size: r > 0.5). Comparing the contexts, online and smart home, collecting the same data type shows significant differences in sensitivity perception regarding considered data types except collected face ID as presented in Tab. 4.4. In contrast, the context comparison between mobile and smart home only shows significant differences in sensitivity perception regarding two considered data types as outlined in Tab. 4.5. The mean rank comparison and relative frequencies of the participants, who rated *sensitive* and *highly sensitive* regarding all the statements from Q_1 and Q_3, underline the observation that the sensitivity perception prevails differently depending on the collected data type in the respective context. Fig. 4.6 presents the relative frequency comparisons regarding Q_1 and Q_3 and outlines that the sensitivity perception in Q_1 is highest when (1) purchasing habits are collected by an online service ($Q_{1.3}$) and (2) in Q_3 fingerprint data is collected by a mobile application ($Q_{3.1}$) rather than by the corresponding devices in the smart home context.

In addition, the analyses show that there are significant differences between the user categories (Q_9), age groups (Q_6) and few differences between considered countries, Germany, UK, Switzerland, and Austria (Q_8). Tab. G.1 to G.6 in Appendix G present the details regarding significant differences of the group-wise analyses based on the Kruskal-Wallis H tests. The analysis regarding the contexts of mobile and smart homes outlines that there is no significant difference between the user categories regarding the collected fingerprint data by a mobile banking application ($Q_{3.1}$). The pairwise comparison of the statement couples also underlines the significant differences between user categories, age groups and countries regarding Q_1 and Q_3. Additionally, the mean rank comparison of all the statement couples regarding Q_1 and Q_3 outlines that in most cases, the non-users have a higher sensitivity perception in the smart home context compared to the online and mobile

Numeration	Statement
$Q_{1.1}$	The online service of the energy supplier collects energy consumption data of your heaters.
$Q_{1.2}$	The smart heaters collect energy consumption data of your heaters.
$Q_{1.3}$	The online purchasing service collects data on your purchasing habits.
$Q_{1.4}$	The smart fridge collects data on your purchasing habits.
$Q_{1.5}$	The online medical service collects data on your weight.
$Q_{1.6}$	The smart scale collects data on your weight.
$Q_{1.7}$	The online medical service collects data how often you do sports.
$Q_{1.8}$	The smart watch collects data how often you do sports.
$Q_{1.9}$	The online banking service collects your face ID.
$Q_{1.10}$	The smart door lock collects your face ID.
$Q_{3.1}$	The mobile banking app collects your fingerprint.
$Q_{3.2}$	The smart door lock collects your fingerprint.
$Q_{3.3}$	The mobile home security app collects information about your availability at home.
$Q_{3.4}$	The smart smoke detectors collect information about your availability at home.
$Q_{3.5}$	The mobile social media apps collect information about your private life, like relationship status, age, name, address, details about family members. etc..
$Q_{3.6}$	The smart speaker collects information about your private life, like relationship status, age, name, address, details about family members. etc..
$Q_{3.7}$	The mobile location app collects data on current location at home.
$Q_{3.8}$	The smart bulbs collect data on current location at home.
$Q_{3.9}$	The mobile sleeping app collects data on your sleeping habits.
$Q_{3.10}$	The smart bed sensors collect data on your sleeping habits.

Table 4.3.: Statements from the Survey Questions Q_1 and Q_3

Q_1: The following data types are collected in online services to which you have signed up and in your smart home environment that you control and Q_3: The following data types are collected in mobile applications to which you have signed up and in your smart home environment that you control. Please indicate your sensitivity perception in Q_1 and Q_3 using the scale provided regardless of the usage purposes and whether the data are collected and processed locally or in a cloud.

Statements	Collected data type	Significant difference	$p-value$
$Q_{1.1}$ vs $Q_{1.2}$	energy consumption data	✓	*4.006e-06*
$Q_{1.3}$ vs $Q_{1.4}$	purchasing habits	✓	*0.003357*
$Q_{1.5}$ vs $Q_{1.6}$	weight data	✓	*5.351e-12*
$Q_{1.7}$ vs $Q_{1.8}$	sports data	✓	*0.03613*
$Q_{1.9}$ vs $Q_{1.10}$	face ID	-	-

Table 4.4.: Significant differences in sensitivity perception comparing the contexts online and smart home collecting the same data type (Q_1) ("✓" = significant difference and "-" = no significant difference)

Statements	Collected data type	Significant difference	$p-value$
$Q_{3.1}$ vs $Q_{.2}$	fingerprint	-	-
$Q_{3.3}$ vs $Q_{3.4}$	availability at home	✓	*2.2e-16*
$Q_{3.5}$ vs $Q_{3.6}$	Information about private life	-	-
$Q_{3.7}$ vs $Q_{3.8}$	current location at home	✓	*6.66e-08*
$Q_{3.9}$ vs $Q_{3.10}$	sleeping habits	-	-

Table 4.5.: Significant differences in sensitivity perception comparing the contexts mobile and smart home collecting the same data type (Q_3) ("✓" = significant difference and "-" = no significant difference)

context. In comparison, the mean ranks also outline that users above 45 years have a higher sensitivity perception in the online and mobile context compared to the smart home context. **Results regarding the Hypotheses:** To sum up, the results mentioned above regarding Q_2 support the $H_{2.1}$ that the perception of the sensitivity is influenced by collected and shared data type in the smart home context. Additionally, the analysis regarding Q_1 and Q_3 confirms the $H_{2.2}$ partially. In the case of the comparison between online and smart home contexts, the context influences the participant's sensitivity perception. The comparison between mobile and smart home contexts led to the observation that the sensitivity perception could be equal in the mobile and smart home contexts if the collected data type did not differ. A possible explanation for this observation could be that users tend to see parallelisms between the mobile application and device usage in smart homes, which influences their sensitivity perception in respective contexts. Moreover, besides the context and the data type, user categories (Q_9) and age groups (Q_6) can be considered as influencing factors for participants' sensitivity perception in smart home, online and mobile contexts.

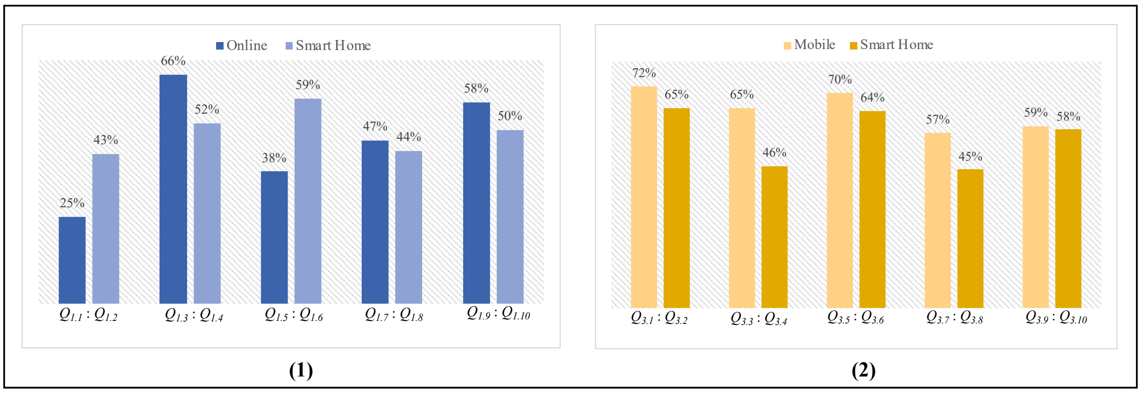


Figure 4.6.: (1): Q_1 and (2): Q_3 – Amount of participants rated *sensitive* and *highly sensitive* in the online and mobile compared to the smart home context .

4.1.4. Summary

In order to gain insights regarding RQ1 and RQ2, we addressed these questions, especially in user-centric experiment 1 (see Sec. 4.1.2) by asking participants about data streams, including data collection, processing and disclosure in smart home environments. Considering the survey results, participants outline that most (ca. 93%) of them are aware of data streams of devices. We see that frequent and average users seem more aware of data collection than the non-users from the derived user profiles. However, the results also outline that in many cases, the participants do not know the data types collected by the devices because only 58% indicated that they know the data types collected. Most participants also indicated that they want to have an overview of data collected by the devices about them and their surroundings. Further results also outline that most participants do not know third parties who have access to the collected data types. Only 24% of the participants mentioned that they know the third parties. Furthermore, evaluating the hypotheses in user-centric experiment 1 emphasizes that the participants want more transparency regarding the data collection and disclosure process in the smart home context. These results clearly outline that users need more transparency regarding collected and shared data types by devices. The lack of transparency could be caused by different reasons: (1) general information given by the provider without any specific transparency regarding collected and shared data types for the device functionality scope and (2) less user effort in finding these information due to the non-understandable presentation of that information and lack of clear information presentation.

Additionally, the user-centric experiment 2 also confirms that the sensitivity perception of users is data type-dependent in the smart home context. Those observations extend the results from the previous works, for instance [71–74, 76], which indicate that the sensitivity perception is context- and data type-dependent. Our results extend the results from previous works [71–74, 76] regarding the participant's sensitivity perception in the smart home

context. In addition, comparing the results in the smart home, online, and mobile context outlines that participants' sensitivity perception is high regarding collected biometrical data, like fingerprint and face ID, regardless of the context. Furthermore, the results show that the device does not significantly influence the sensitivity perception of the participants in the considered case. These observations conclude that participants' sensitivity perception, in the first place, strongly depends on the collected data type. Previous works outline that there are dependencies between participants' sensitivity perception, privacy perception and data sharing attitude because the users' sensitivity perception influences users' privacy perception [71, 77–80], which then in turn influences users' data sharing attitude. Therefore, it is essential that the UCCPs allow users to apply those controls while considering the sensitivity of the data types collected and shared by the used devices.
To sum up, most participants explicitly indicated they want to control data collection and disclosure in their smart home environments. The control willingness again shows that the participant's need for data stream controls is high because of missing transparency regarding data streams. The lack of transparency strengthens the fact that the participants are limited in gaining more awareness regarding data streams of devices.

4.2. User-Centric Data Stream Controls for Smart Home Devices

In this section, we present the derived UCCPs based on the results of the user-centric experiments from Sec. 4.1 to address the research gaps from Sec. 3.4. The presentation of the UCCPs in this section is twofold. First, we present the overview of the UCCPs and their categorization. Then, we explain the functionalities and setting options of each UCCP in the following subsections. The presented UCCPs are published in scientific papers [6, 7, 81]. We slightly adjusted the names of the UCCPs, corresponding variables and enhanced the UCCP settings compared to our publications, [6, 7, 81], in order to improve our derived UCCPs and the reading flow of this thesis.

4.2.1. Overview of User-Centric Data Stream Controls

In total, we derived six UCCPs, which allow data stream control from a user perspective in the smart home context. We categorized the UCCPs into three categories: (1) transparency, (2) implication and (3) access. In the category *transparency*, we summarized the UCCPs, allowing users to gain transparency and limit data collection. The second category *implication* contains UCCPs, which provide users details about sharing the collected data types. In the last category *access*, we summarized the UCCPs, which allow users to limit the data-sharing process while using devices in smart homes. Tab. 4.6 presents the overview of the six UCCPs with short descriptions.
Transparency: This category includes in total three UCCPs, and those are *Sensitivity*, *Mini-*

Category	UCCPs	Short Description
Transparency	UCCP: Sensitivity	Possibility to tag the collected data types of the device(s) as *high-, medium-* and *low-sensitive*
	UCCP: Minimization	Possibility to limit the collected data types by the device(s)
	UCCP: Granularity	Possibility to set in which granularity the data types are collected and saved for later data reviews and sharing
Implication	UCCP: Sharing Attitude	Possibility to weigh between arising benefits (social and personal) and risks (privacy and security) regarding the collected data types by the device(s)
Access	UCCP: Disclosure Limitations	Possibility to control the data sharing: Share or delete the collected data types by the device(s)
	UCCP: Access Limitations	Possibility to limit the sharing of the collected data types of device(s) by choosing the data consumers and usage purposes

Table 4.6.: Overview of the UCCPs for the data stream control in smart homes based on [6]

mization and *Granularity*. The UCCP *Sensitivity* allows users to tag the device and corresponding collected data as *high-, medium-* or *low-sensitive* depending on which data types are collected by the device. For this UCCP, it is relevant that the provider provides users with the following details regarding the collected data types by the corresponding device:

- Collected data types for main device functionalities
- Collected data types for optional device functionalities

An example based on a smart robot is presented in Tab. 4.7 regarding the collected data for main and optional functionalities. After supplying transparency regarding the collected data types by the devices according to the functionality scope, users can apply the UCCPs and decide which functionalities are essential for them. Allowing users to apply *Sensitivity* is important because, for example, based on the collected data types by a smart robot (see Tab. 4.7), users might tag the data types collected for optional functionalities as *high-sensitive* and the data types collected for the main functionality as *medium-* or *low-sensitive*, because for instance biometrical data as voice prints or human pictures in different status can be used by data consumers to manipulate, to harm or even to create digital profiles of the users in that specific home area depending on data consumers and usage purposes after data sharing. The users can consider the applied sensitivity in the UCCP *Sensitivity* when they later make data-sharing decisions regarding the collected data types by the corresponding device.

The UCCP *Minimization* in this category allows users to limit the collected data types by the devices. Based on the supplied transparency regarding the collected data types for main and

Functional Scope	Short Description of the Functionality	Collected Data Points
Main functionality	Automation of the house cleaning	Wifi name, wifi password, cleaning schedule, cleaning protocols, current status of the robot
Optional functionality	Efficient house cleaning, voice control based cleaning	Email, password, language, home map, voice print, email and password for Ecovacs app, human availability at home, video calling with persons at home, connecting with people nearby using the same device & name, address, country, robot name, video manager, pictures of home/people/furniture

Table 4.7.: Example based on a smart robot (based on [85])

optional functionalities, users can set which data types can be collected by the corresponding device according to their functional preferences. In order to apply the UCCP *Minimization*, the devices must allow users to limit all the collected data types by the device. Referring to the previous example with the smart robot from Tab 4.7, the users could allow only the collection of the data types relevant to carry out the main functionality of the smart robot and use only the corresponding services. In this case, the smart robot cannot collect the other data types regarding optional functionalities. If users want to share the data types later, they are only allowed to share the collected data types they allow to collect for the corresponding device. The last UCCP in this category is UCCP *Granularity*, which allows users to set the granularity of the collected data types by the smart devices. In order to apply *Granularity*, as with the other two UCCPs, users must get transparency regarding the collected data types by the device. After gaining transparency regarding the collected data types and limiting the data collection in UCCP *Minimization*, users can set on which granularity level the data types are collected and saved for the users' reviews and data disclosure. The granularity settings of the collected data differ according to the collected data type and are described in detail in Sec. 4.2.2. Based on the smart robot example collecting only data types for the main functionality (from Tab. 4.7), the users could set the corresponding granularity level for the collected data points, for example aggregating the cleaning schedule data on a daily, weekly, monthly level rather than presenting the detailed information regarding each cleaning session.

Implication: This category includes only one UCCP, UCCP *Sharing Attitude*. The UCCP *Sharing Attitude* requires that the users are supplied with information regarding associated risks and advantages in case of data sharing in an understandable way. The users should be able to gain transparency regarding the associated risks and advantages of data sharing. Based on this information, users can decide whether the associated advantages are more important than the associated risks after data sharing or vice versa. Based on these gained perceptions, users can make conscious decisions regarding disclosing the collected data types. In the case of the smart robot from the example (see Tab. 4.7), users could weigh that the associated advantages are higher than the associated risks in case of sharing the collected data types for the main functionality after setting the previous UCCPs.

Access: In this category, the UCCPs, UCCP *Disclosure Limitations* and UCCP *Access Limitations* are clustered. The UCCP *Disclosure Limitations* allows users to control the sharing of the collected data types by devices. Besides tagging, limiting the data collection and setting the data collection granularity, it is essential that the users can also limit the data types shared, considering the settings of the previous UCCPs. Allowing users to control the data sharing should include that the users are allowed to choose between sharing and not sharing the data types.

While UCCP *Disclosure Limitations* allows users to limit data sharing, UCCP *Access Limitations* allows users to control the data access for the shared data types of their devices. This UCCP allows users to specify the data consumers and corresponding usage purposes for the shared data types. In this way, *Access Limitations* allows users to limit the data access besides limiting the data sharing. Referring to the example with the smart robot from Tab. 4.7, users could set with *Disclosure Limitations* that the collected data types like cleaning schedule, cleaning protocols, and current status of the robot (from main functionalities) are only shared and that the other data types are not shared and deleted. Additionally, by applying *Access Limitations* in this example, users could set that the smart robot provider can only access and use the specific shared data types for improving the smart robot functionality in the users' daily routines and for general device improvements.

4.2.2. Setting Options of User-Centric Data Stream Controls

In the following, we describe the setting options and corresponding details of the derived UCCPs presented in Sec. 4.2.1. In the setting options of the UCCP *Sensitivity*, users must indicate their attitude regarding the sensitivity of the data types of the data streams. Users can tag the data types as *high-*, *medium-* or *low-sensitive*. In case users are unsure about the tagging or want to apply the default settings, then all the collected data types are tagged according to the system-defined sensitivity, which is explained in detail in Sec. 6.1 in Tab. 6.5. The assignment of sensitivity depends on the selected scope of the functionalities, as described in Sec. 4.2.1.

Complementary to *Sensitivity*, UCCP *Minimization* setting options ask users to set which data types should be collected. The data collection limitation depends on the selected functionality scope, as described in Sec. 4.2.1. Users can choose between main and optional functionalities and specify the data types that can be collected and not collected. The users must apply the setting options of the *Minimization* for all the collected data types of the chosen functionality scope. In this UCCP, the users receive more background information regarding the providers with privacy ratings of the providers (*pRat*). *pRat* are supplied to the users in order to increase the transparency of how trustworthy the providers process the collected data. The *pRat* are derived based on the mechanism from [62] including a 5-star-based rating system with the icon “i” next to the star-based rating. The icon “i” delivers users more explanation regarding the corresponding rating, for example, "...data of your smart home environment are directly saved in the providers’ cloud, and you cannot be sure who can get access to your data and for which purpose [62]" [7][2]. In case the user setting is missing, then the default settings are applied, where only the data types for the main functionality are collected.

After limiting the collected data types, in UCCP *Granularity*, the users can specify the granularity of the data types for data collection. The granularity level depends on the specific data type. Depending on which data types are collected, users must choose the suitable granularity level for the data types collected. The details regarding the different granularity levels are presented in the Tab. 4.8. The provider of the devices can enhance the granularity level if the existing granularity does not fit the collected data types. The default settings of *Granularity* include that the collected data types are aggregated and anonymized only monthly for numeric data and approximate data regarding personal information.

The setting options of the UCCP *Sharing Attitude* allow users to indicate their sense of privacy risks and advantages arising from sharing the collected data types. Users can choose between their own risk sensitivity and their own sense of advantages. The users must indicate their sense of privacy risks and advantages for all the collected data types after applying *Minimization*. The users are asked to carry out these user settings after reviewing the supplied information by the provider, like *Data Sharing Information Categories* (DSC) in Tab. 4.9 and an overview of personal information types in Tab. 4.10. The DSC presents users with the arising privacy risks and advantages from disclosing the collected data types. The overview of personal information types supplies users with the sensitivity of different information in various contexts, which are assigned to types of the DSC in order to allow users to assess the sensitivity of different data types and to understand the risks and advantages arising from disclosing the collected data types. The DSC in this UCCP are derived based on the related work in the context of privacy risks and advantages arising from data disclosure in the smart home context [24,25,37,62,70,71]. The overview of personal information types is derived based on the user-centric experiment 2 from Sec. 4.1.3 and based on existing works covering all the different collected data types [71,72]. The users also can add categories to DSC, and the default settings of *Sharing Attitude* are set so that the risk sensitivity is higher than the sense of advantages.

Furthermore, with the setting options of *Disclosure Limitations*, users can limit the data sharing regarding the collected data types after setting *Minimization*. The users can choose

[2] “The approach from Zimmermann et al. [62] is a 5-star-based rating system. This system is similar to a star-based product rating, for example, on Amazon. The 5-star ratings of each provider result from the given information by each provider and user experiences with the corresponding smart object." [7]

Granularity Level	Name of the Granularity Level	Example Options of the Granularity Level
n	Numeric data, like usage, routines, image quality, motion detection sensitivity	Aggregation to approximate data: daily, weekly, monthly, yearly or value ranges for a week, month, year
p	Personal information data	Aggregation to approximate data: instead of the address, the province and country, instead of the age, the age cluster of the user, instead of the name, the gender of the user, instead of the clear email address, the initials must be anonymized
l	Location data	Aggregation to approximate data: instead of the exact location, the k-anonymity mechanism is applied so that the approximate area is given and aggregated daily, weekly, and monthly, yearly.
va	Video and Audio data	Aggregation to approximate data: images and voices are anonymized and can be aggregated daily, weekly, monthly, and yearly, and to the number of persons talking or appearing in the audio and video recordings

Table 4.8.: Possible granularity levels regarding *UCCP Granularity*

between disclosing or deleting the collected data types. In case the users decide to use the default setting of the UCCP *Disclosure Limitations*, then the collected data types are not disclosed and are deleted. In case of data deletion, the users can also set the retention period of the collected data types. The retention period of the collected data types before deleting can be set by using two options: *specific period*, where the period includes options like after 3, 6, or 12 months or according to the storage capacity, where the storage refers to the capacity of the device or used cloud storage of the corresponding device. The default settings regarding the retention period are set to delete the data according to the storage capacity.

After setting the UCCP *Disclosure Limitations*, in the UCCP *Access Limitations*, the users can limit the access of the selected data types for sharing. Users can choose the third-party data consumers (*dCon*) accessing the shared data types, for instance, doctors, insurance companies, and government agencies. After limiting the data consumers, users can limit the usage purposes of the shared data types (*dPurp*) used by the corresponding data consumer, for example, personal health plans and statistical purposes. During the selection of the *dCon*, users are presented with a rating for each *dCon* based on the 5-star-rating approach, which

Abbreviation of Category	Category Name	Category Description
r_1: Associated privacy risks 1	Discrimination and Manipulation	Using to create special contract and discriminate the users, for example, manipulating the device owners with contracts.
r_2: Associated privacy risks 2	Burglaries and Misuse	Using the data to harm the users, for instance, breaking in after analysing data about home availability and smart door lock.
r_3: Associated privacy risks 3	Profiling	Using the data to track the users and manipulate the users and steal the user's identity.
r_4: Associated privacy risks 4	Carrier risks	Using the data to find out characteristics of the users, for example, analysing and disclosing such information can result in risks for future employers.
r_5: Associated privacy risks 5	Damaging	Using the data to damage the device owners, for instance, identity theft based on the disclosed sensor data.
r_6: Associated privacy risks 6	Personal Exposure	Using the data to publish things users are doing, for example, data disclosure could result in being exposed because they had done something they did not want their friends and family to know about, maybe also to carry out Propaganda, etc.
a_1: Associated social advantage 1	Personal Advantages	Using the data to provide user-specific contracts and users can earn money.
a_2: Associated social advantage 2	Social Advantages	Using for statistical aims, for example, using data for research works, market analysis.

Table 4.9.: Data sharing information categories DSC [7]

Information Type	**Directly**	**Linkable**	**DSC**					
			r_1	r_2	r_3	r_4	r_5	r_6
Body size	-	x	x	x	-	-	-	-
Voiceprint	-	x	x	-	-	-	-	-
Body and facial images	x	-	x	x	-	-	-	-
Biological characteristics / Biometrics	x	-	x	x	-	x	-	-
Recording of using health equipment	-	x	x	x	x	-	x	x

Table 4.10.: Overview of the personal information types and their associated risks from DSC [7]

is derived based on the approach from [62]. To get the rating, each *dCon* has to answer several questions, for instance, "...where the data types are saved, for which purpose the data types are used, and with which other companies/associations the data types are shared..." [7]. According to the answers, the *dCon* receive a rating based on the 5-star-rating approach. The default setting of this UCCP includes that the data types are not shared with any *dCon* for any *dPurp*.

4.2.3. Summary

To sum up, the results of the user-centric experiments from Sec. 4.1 outline that users are willing to have control over data streams when using devices. The participants specified that they want (1) to tag the sensitivity of the collected data types, (2) to control the data collection by applying minimization and granularity strategies, (3) to receive information about arising risks and advantages from data disclosure, (4) to control which collected data types are shared, with whom and for which purposes. These survey results confirm that participants want to have and apply the six derived UCCPs in this section. while using corresponding devices. To address the sensitivity-related findings from user-centric experiments, the derived UCCPs also allow users to indicate their data type-dependent sensitivity perception before applying strategies to limit the data types collected and shared based on arising risks and advantages, as listed below. The participants also confirmed in our user-centric experiments that they are motivated to use devices while applying the UCCPs mentioned above, which could, in return, increase user acceptance regarding device usage. User acceptance regarding devices could decrease more and more when those devices do not supply users with more transparency and control regarding the data streams. In the following, we present the summary of the derived UCCPs.

- UCCP *Sensitivity*: tag data types as *high-*, *medium-* or *low-sensitive*
- UCCP *Minimization*: specify data types as *collect* or *do not collect*
- UCCP *Granularity*: specify the granularity of the collected data types
- UCCP *Sharing Attitude*: indicate own risk sensitivity and sense of advantages
- UCCP *Disclosure Limitations*: specify data types as *disclose* or *delete*
- UCCP *Access Limitations*: specify the *data consumers* and *data usage purposes* for shared data types

5. Validation of User-Centric Data Streams with Existing Smart Home Devices

This Chapter presents the validation results of existing devices. We validated those devices to investigate whether they supply the derived UCCPs from Sec. 4.2. We considered the six main categories and the corresponding devices from Sec. 2.2. In the following sections, we present the methodology of the device validation (Sec. 5.1) and the details regarding the considered devices (Sec. 5.2). Subsequently, we outline the extent to which the derived UCCPs are supplied by the considered devices (Sec. 5.3).

5.1. Methodology of Smart Home Device Validation

In our validation, we focused on (1) gaining more insights regarding the device functionalities and (functionality-related) data streams. Furthermore, we also evaluated whether our derived UCCPs from Sec. 4.2 are applicable while using the corresponding devices. To gain those insights, we first considered the device descriptions and privacy policies of the providers to outline the data streams and supplied UCCPs by the corresponding device. The insights based on that information for each of the six devices are presented in Sec. 5.2. In the second step, we divided the activities from purchasing to setting up and using the device into six phases in order to outline the required data types in each phase transparently. All six phases are applicable for all the considered devices, except for the smart voice assistant Amazon Echo, where the phase *Setup Voice Assistants* is omitted because users do not have to set the voice assistant separately. Fig. 5.1 presents the six phases.

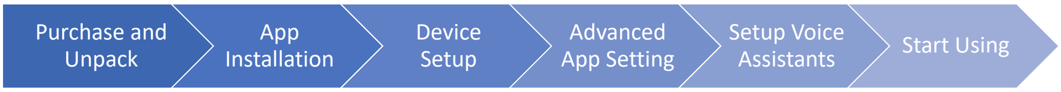


Figure 5.1.: Six phases of the activities from purchasing to setting up and using the devices

In each of the phases from Fig. 5.1, different data types are collected, processed and shared by the corresponding device. Those data types in each phase of corresponding devices are presented in Sec. 5.2 and in Appendix C. We considered about 76 individual and non-overlapping data types collected, processed and shared by the six considered devices from Sec. 2.2.

1. **Purchase and unpack:** User buys the device and unpack it at home

2. ***Application* (App) installation:** User installs and logs into the App of the corresponding device

3. **Device setup:** User switches on the device

4. **Advanced app** settings: User adjusts the required settings to start the device usage

5. **Setup voice assistants:** User adjusts the settings to control the device via voice assistants

6. **Start using**: User starts using the device, and its functionalities

5.2. Analysis of the Data Streams of Existing Smart Home Devices

In the following, we present the device descriptions, privacy policies of providers and the data streams of each device in each phase from Fig. 5.1. In order to outline the collected, processed and shared data types in each phase, we will present an example with the Reolink smart camera in this section since this device delivers most details regarding the data streams. To keep the reading flow, we present the data types of the data streams in each phase for the other five considered devices from Sec. 2.2 in Appendix C.

Security: According to market analysis from Bitkom and Statista [29,31,84], smart cameras are considered as the most popular devices regarding *security* in the smart home context. According to chip online test [86], the Reolink Argus 3 Pro is rated as the test winner for smart home cameras. In this category, we will focus on Reolink smart cameras to investigate the data types of the data streams for the functionality provision in detail (based on [87]). The main functionalities of smart cameras are that those devices allow the tracking of what is happening in the installed area of the camera. Examples are somebody entering the home or garage, leaving the house, having a live view of things happening in specific rooms, two-way talking with the people in the room, and lightning corners [87]. The Reolink smart cameras also capture the dynamics of long-term events, like the sun rising or flower blooming. The collected data types via the smart cameras can be stored locally by applying a micro SD card or remotely using the Reolink cloud [87]. All the detected motions are stored in the cloud and can be accessed via the Reolink App or web page, even with local storage. Reolink cloud ensures the secure data storage of the customers with the collaboration of *Amazon Web Services* (AWS), like the standard *Advanced Encryption Standard* (AES) algorithm for encrypted data transmission, *Rivest–Shamir–Adleman* (RSA) algorithm for secure key exchange and *Transport Layer Security* (TLS) standards, to keep users' information confidential [87]. Users can access the data on the Reolink cloud with their password and set which area during which time is monitored and which data types are collected [87]. Only the owners and people who get permission from the owners can access the cloud and can

Figure 5.2.: Reolink Argus 3 Pro smart camera [87]

access, delete and download the recordings of users' smart cameras. The recordings will be automatically deleted according to the cloud subscription terms and the user's storage time for the device. The Reolink smart cameras can also be controlled with Google Assistant. Fig. 5.2 presents the Reolink smart camera device. According to the privacy policies of Reolink, they promise the users to follow the data minimization principle in their data streams and allow users to control which data types are considered in data streams from which area. In their privacy policies [87], Reolink also outlines that they share the data types of the data streams according to a minimum necessary principle with third parties to perform essential functions [87]. Otherwise, the data types are not shared with third parties without users' consent. Despite these settings, the provider also says that not providing consent to collect and process required data types will affect the smart camera's functionality [87] but does not understandably and transparently inform users which data types are necessary for supplying main and optional functionalities. Tab. 5.1 presents the 31 data types of the Reolink smart camera's data streams resulting from our analysis. Reolink smart camera partially delivers information regarding 17 data types, which can be controlled by applying three UCCPs (*Minimization, Granularity* and *Access Limitations*) while using the smart camera.

Comfort and Lightning: According to the statistics from Bitkom and Statista [29, 31, 84], smart bulbs belong to the most important and most popular used device in the smart home context, which can be clustered into this category. According to chip online test [88], the PHILIPS Hue bulb White and Color Ambiance has been categorized as the best smart bulb. In this category, we will focus on this smart bulb and investigate the data types of the data streams for the functionality provision in detail (based on [89, 90]). According to PHILIPS, the main functionalities of the device are controlling the lights (up to ten) via Bluetooth remotely and setting lightening scenes according to user preferences, for instance, movie night, relaxing and cooking mode [89]. The smart bulb allows users to control its functionalities via voice control devices like Amazon Alexa and Google Assistant. In this context, PHILIPS recommends that users can install the PHILIPS Hue Bridge to expand its functionalities [89, 90]. PHILIPS Hue Bridge allows users to control up to 50 lights at home and automate the control while setting routines and timers [89, 90]. Examples are going

Phase	Data Type
Purchase & unpack	Name Address
App installation	Email Password Country
Device setup	Wifi name Wifi password Password for camera protection Name for camera IP address Camera recording area (adjusted)
Advanced App setting	Device information Battery usage data Image quality of recordings Date of the recording Motion detection sensitivity Audio & video recording schedule Audio & video recording duration Recording storage duration QR code for sharing recordings Long-term event recording Spotlight intensity & activation Cloud storage plan Motion recording
Setup voice assistants	Reolink Skill in Alexa Email (connection between Reolink camera and voice assistant, Alexa) Password (connection between Reolink camera and voice assistant, Alexa) Activated cameras for smart home control (in Reolink App) Transfered recordings with voice assistant (Display on Alexa via voice control)
Start using	Camera recordings Motion recording (audio & video - according to schedule & duration settings)

Table 5.1.: 31 Collected data types by the Reolink smart camera in the phases from Fig 5.1

Figure 5.3.: PHILIPS Hue smart bulb [89]

to sleep or wake-up routines and turning lights on when you come home. It also allows the installation of other PHILIPS smart devices, like motion sensors and switches [89, 90]. Fig. 5.3 presents the PHILIPS Hue smart bulb. In this case, the provider supplies in its privacy policies a set of examples regarding the data types of the data streams [91]. Based on PHILIPS's descriptions and our analysis, we derived 18 data types of the PHILIPS bulb's data streams (see Tab. C.1 in Appendix C). PHILIPS outlines, in general, that the collected data types are securely saved and processed in PHILIPS servers, servers of the used voice assistants and other third parties, like *Information and Communications Technology* (ICT), communication service, and payment provider suppliers [91]. Thereby, PHILIPS does not deliver any hints regarding applicable UCCPs while using the PHILIPS smart bulb. During our analysis, we identified two data types, which can be controlled by applying one UCCP (*Minimization*) while using the smart bulb.

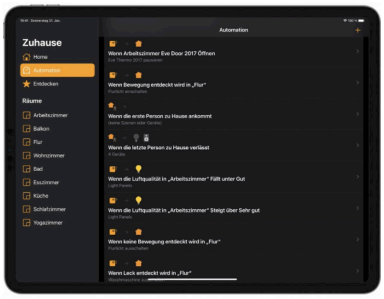
Energy Management: According to market analysis from Bitkom and Statista [29, 31, 84], smart thermostats are considered as one of the most important application areas in energy management in the smart home context. According to the chip online test [92], the eQ-3 Homematic IP smart heater thermostat – Evo is considered in the test comparison as the best smart heater thermostat. In this category, we will focus on this thermostat to investigate the data types of the data streams for the functionality provision in detail (based on [93]). The main functionalities of Homematic heater thermostats allow users to control the room temperature according to their preferences [93]. The thermostats can be controlled via App and voice assistants, like Amazon Alexa or Google Assistant. The smart heater thermostat also facilitates the dynamic balancing of the thermostat's temperature [93]. Additionally, Homematic heater thermostats also allow the adjustment of the room temperature up to 13 times per day and setting up to three adjustable personal heating preferences, called heating profiles [93]. These functionalities enable users to save energy consumption during heating. Fig. 5.4 presents the Homematic IP smart heater thermostat. However, the provider generally

Figure 5.4.: Homematic IP smart heater thermostat [93]

describes the privacy policies in their data protection regulations. To provide the smart heater thermostat functionalities, the data streams of the device include 17 different data types, which we derive based on our analysis and the provider does not transparently define (see Tab. C.2 in Appendix C). Homematic does not deliver specific information regarding supplied UCCPs while using the Homematic smart heater thermostat.

Household Appliances: According to the statistics from Bitkom and Statista [29, 31, 84], smart robot cleaners are considered as the most popular device in household appliances in the smart home context. According to chip online test [94], the Ecovacs Deebot X1 Omni is rated as the test winner. In this category, we will focus on the Ecovacs smart robot cleaner to investigate the collected data types for the functionality supplement in detail (based on [85]). The main functionality of the Ecovacs robot cleaner is the automatic cleaning of one's home. With the integrated cameras and autopilot functionality, the smart robot cleaner can avoid obstacles and recognize people [85]. The Ecovacs robot cleaner also includes a mapping technology, which allows users to create a map of their own house and save it to ensure efficient and fast navigation of the smart robot [85]. It also includes voice assistant technology, which allows users to control the start of the cleaning sessions of the smart robot and to follow users' voices during the cleaning sessions [85]. Fig. 5.5 presents the Ecovacs smart robot cleaner. Ecovacs smart robot provider generally presents privacy policies regarding the data streams. Based on our analysis, the data streams of the Ecovacs smart robot consider 19 data types to supply its functionalities (see Tab. C.3 in Appendix C). Ecovacs only delivers information regarding three data types, which can be controlled by applying two UCCPs (*Minimization* and *Disclosure Limitations*) while using the smart robot.

Figure 5.5.: Ecovacs Deebot X1 Omni smart robot [85]



Wenn Arbeitszimmer Eve Door 2017 Öffnen
Eve Thermo 2017 pausieren

Wenn Bewegung entdeckt wird in „Flur"
Flurlicht einschalten

Wenn die erste Person zu Hause ankommt
(keine Szenen oder Geräte)

Wenn die letzte Person zu Hause verlässt
4 Geräte

Wenn die Luftqualität in „Arbeitszimmer" Fällt unter Gut
Light Panels

Wenn die Luftqualität in „Arbeitszimmer" Steigt über Sehr gut
Light Panels

Wenn keine Bewegung entdeckt wird in „Flur"
Flurlicht ausschalten

Wenn Leck entdeckt wird in „Flur"
Waschmaschine ausschalten

Figure 5.6.: Apple HomeKit smart hub App [95, 96]

Networking and Control: According to the statistics from Chip [30], this category includes smart home hub systems, which allow control over a set of devices in a smart home environment. One of the most popular smart home control hubs besides Google Nest and Amazon Alexa is Apple Homekit App [95]. Users can connect their devices, controllable via Wi-Fi and Bluetooth, to the Home App of any iOS device to integrate the devices in the Apple Homekit App [95, 96]. Fig. 5.6 presents the Apple HomeKit App. According to privacy policies of Apple, the communication between the devices is security key-based and within the local network [96]. Once users have allowed the data sharing, the data types of the data streams are released from their local network or home area [96]. By our analysis, the data streams of the Apple HomeKit include 10 data types (see Tab C.4 in Appendix C). Apple only provides information regarding three data types, which can be controlled by applying one UCCP (*Disclosure Limitations*) while using the smart hub.

Entertainment: According to market analysis from Statista [31], smart voice assistants are considered as the most popular devices in the entertainment area in smart homes. According to current analyses [97], the smart voice assistant Amazon Echo is viewed as the market leader in this category. The main functionalities of Amazon Echo are controlling the

Figure 5.7.: Amazon Echo smart voice assistant [98]

interaction of the devices among each other via voice control [98, 99]. Fig. 5.7 presents the Amazon Echo voice assistant. According to Amazon privacy policy [98], the voice assistant allows users to control data type collection. Users can turn off the microphone and camera and enable visual data collection signs via a light indicator [98]. Additionally, Amazon Echo allows users to control the voice recording status and history regarding defined data types of the data streams of used devices [98]. Based on descriptions [99] and our analysis, we identified 13 data types of the Amazon Echo's data streams (see Tab. C.5 in Appendix C). Amazon Echo only delivers information regarding one data type, which can be controlled by applying two UCCPs (*Minimization* and *Disclosure Limitations*) while using the voice assistant.

5.3. Evaluation of User-Centric Data Stream Control Supplied by Existing Smart Home Devices and Summary

After analysing the data types of the data streams of the six above-mentioned devices (in Sec. 5.2), in this section, we outline the evaluation results regarding the supplied UCCPs. The following list presents the data types per device, which are controllable by supplied UCCPs from Tab. 4.6.

Reolink smart camera:

- *Minimization*
  - name, address, email, password, battery usage data, audio & video recording schedule and duration, recording storage duration, long-term event recording, email Reolink login, password Reolink login, current camera recording, motion recording

- *Granularity*
  - image quality of recording, motion detection sensitivity
- *Access Limitations*
  - QR code for sharing recordings

PHILIPS smart bulb:

- *Minimization*
  - name, address

Ecovacs smart robot:

- *Minimization* and *Disclosure Limitations*
  - video manager, pictures of home & people & furniture, voice print through life audio calls

Apple HomeKit smart hub:

- *Disclosure Limitations*
  - overview & name of devices, overview of the rooms, list of the devices in each room

Amazon Echo smart voice assistant:

- *Minimization* and *Disclosure Limitations*
  - current conversation & commands

We can see that the existing devices today do not fully supply the derived UCCPs from Sec. 4.2. Additionally, none of the considered devices transparently present which data types of the data streams are necessary for which functionality scope of the device. Our analysis shows that the Homematic smart heater thermostat does not supply any UCCPs. In general, we observe that the considered devices only partially supply the derived UCCPs from Sec. 4.2. Our analysis also outlines the following possibilities to control the data streams of the six considered devices (from Sec. 5.2) based on supplied UCCPs.

1. Only four out of the six considered devices allow users to minimise the collected data types *(Minimization)* partially: smart camera, robot, bulb, voice assistant
2. Besides *Minimization*, the smart camera also allows users to control the granularity *(Granularity)* and the access to the collected and shared data types *(Access Limitations)*
3. Three out of the six considered devices allow users to partially limit the disclosure of the collected data types *(Disclosure Limitations)*: smart robot, hub and voice assistant.

4. Three of the six considered devices allow users to apply at least the functionality of two UCCPs partially for the data types of the data streams:
    - Smart camera with *Minimization, Granularity & Access Limitations*
    - Smart robot with *Minimization & Disclosure Limitation*
    - Smart voice assistant with *Minimization & Disclosure Limitation*
5. Two of the six considered devices allow users to minimise the collected data types (*Minimization*) regarding the data types *name* and *address*: smart bulb and cameras

To sum up, the considered devices only supply users the derived UCCPs partially without delivering the entire transparency and control possibilities to the users regarding the data streams of the devices. Our analysis outlines that *Minimization* is the most applicable UCCP regarding the six devices. Additionally, the analysis also outlines that only the Reolink smart camera allows users to apply *Granularity* and *Access Limitations*, besides *Minimization*, and allows users to set the granularity of the collected data types and limit who can access collected data types.

6. Proposal with User-Centric Data Stream Controls for Smart Home Device Usage

This chapter presents our derived proposal for device classification and configuration based on derived UCCPs, including a progress report in order to address the research gaps regarding data-related privacy issues and data stream controls from a user perspective, as described in detail in Sec. 1. Our proposal is called **quality check** and consists of two steps. The two steps include a **quality check** (1) **for providers** and (2) **for users**. The quality check for providers evaluates the provision of the UCCP settings and classifies the corresponding device based on providers' inputs. We present in Sec. 6.1 the details regarding the evaluation logic and the quality check results for providers based on the considered devices from Sec. 5. Subsequently, in Sec. 6.2, we present the details regarding the quality check for users, which delivers users transparency regarding supplied UCCPs. It also allows users to set the supplied UCCP settings efficiently while receiving a progress report regarding applied settings during device usage.

6.1. Quality Check for Provider Including User-Centric Data Stream Controls

In this section, we first describe the process of the quality check for the provider in order to get the required inputs to calculate the quality check results. Second, we describe the calculation steps of the quality check for the provider

6.1.1. Process

The quality check process for the provider contains two main process steps. The main two process steps are (1) general input for transparency and (2) input for evaluating the supplied UCCPs.

Process Step 1: General Inputs for Transparency:

In this step, different inputs are considered to increase data collection transparency. The inputs must be supplied by the providers of the corresponding devices. The Fig. 6.1 presents the relevant inputs of the first process step. The relevant inputs are (1) smart device name, (2) mandatory data types for main functionality and (3) optional data types for optional functionality. Based on the given amount regarding mandatory and optional data types, the

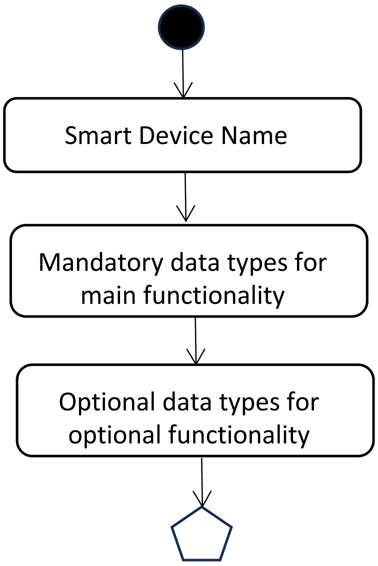

Smart Device Name
Mandatory data types for main functionality
Optional data types for optional functionality

Figure 6.1.: Substeps of the process step 1: quality check for provider

total data type amount of the data streams of the corresponding device is derived. The details of the relevant inputs are described in the following.

Smart Device Name: This input asks to enter the name of the device, providers want to evaluate. In order to give some examples, we refer to the devices, we considered in Sec. 5. In case the provider wants to evaluate the six devices from 5, then they have to specify the information as follows:

- Smart camera: Reolink Argus 3 Pro
- Smart bulb: Philips Hue Bulb White and Color
- Smart heater thermostat: eQ-3 Homematic IP Smart Heater Thermostat
- Smart robot: Ecovacs Deebot X1 Omni
- Smart hub: Apple Homekit
- Smart voice assistant: Amazon Echo

Mandatory Data Types for Main Functionality: This input requires the amount and the names of the mandatory data types regarding the data streams of the main functionality of the corresponding device. Tab. 6.1 presents exemplary mandatory data types regarding the Reolink smart camera. The main functionality of the Reolink smart camera is tracking a defined area. The other considered devices from Sec. 5 are presented in Tab. D.1 (in Appendix D).

Device	Amount of Mandatory Data Types	Mandatory Data Types
Smart camera: Reolink Argus 3 Pro	12 from 31	name, address, Wifi name, Wifi password, IP address, camera recording area (adjusted), device information, data of recording, QR code for sharing recordings, transferred recordings with voice assistant (Display on Alexa via voice control), camera recordings, motion recording (audio & video according to schedule & duration settings)

Table 6.1.: Example of the specific mandatory data types for the smart camera Reolink (derived based on [87])

Optional Data Types for Optional Functionality: In this input the amount and the names of the data types are captured, which the providers need to supply users optional functionalities during the usage of the corresponding device. Tab. 6.2 presents exemplary optional data types regarding the Reolink smart camera. The other considered devices from Sec. 5 are presented in Tab. D.2 (in Appendix D).

Based on the inputs regarding mandatory and optional data types, the total data type amount of the data streams are captured. Regarding the considered devices in Sec. 5, the following information is delivered based on the inputs regarding mandatory and optional data types:

- Smart camera: Reolink Argus 3 Pro: Amount: 31, for specific data types: See Tab. 5.1
- Smart bulb: Philips Hue Bulb White and Color: Amount: 18, for specific data types: see Tab. C.1
- Smart heater thermostat: eQ-3 Homematic IP Smart Heater Thermostat: Amount: 17, for specific data types: see Tab. C.2
- Smart robot: Ecovacs Deebot X1 Omni: Amount: 19, for specific data types: see Tab. C.3
- Smart hub: Apple Homekit: Amount: 10, for specific data types: see Tab. C.4
- Smart voice assistant: Amazon Echo: Amount: 13, for specific data types: see Tab. C.5

Process Step 2: Inputs for UCCP Functionality Evaluation

The second process step of the quality check for the provider needs inputs regarding the UCCPs. It must be indicated whether the UCCPs are supplied by the corresponding device. These inputs also must be supplied by the providers. These inputs are used for calculation

Device	Amount of Optional Data Types	Optional Data Types
Reolink Argus 3 Pro	19 from 31	email, password, country, password for camera protection, name for camera, battery usage data, image quality of recordings, motion detection sensitivity, audio & video recording schedule and duration, recording storage duration, long-term event recording, spotlight intensity & activation, cloud storage plan, motion recording, Reolink skill in Alexa, Email (connection between Reolink camera and voice assistant, Alexa), Password (connection between Reolink camera and voice assistant, Alexa), Activated cameras for smart home control (in Reolink App)

Table 6.2.: Example of the specific optional data types for the smart camera Reolink (derived based on [87])

steps to deliver the quality check results. Tab. 6.3 presents which input is relevant to evaluate which UCCP is supplied. In Fig. 6.2, we present an exemplary input query regarding the UCCP *Sensitivity*. First, it must be indicated whether the UCCP is supplied and in case it is supplied, then the amount and the names of the controllable data types are captured (see Fig. 6.2). This input query is executed for all six UCCPs, as presented in Tab. 6.3, in order to get insights regarding supplied UCCPs by the corresponding device.

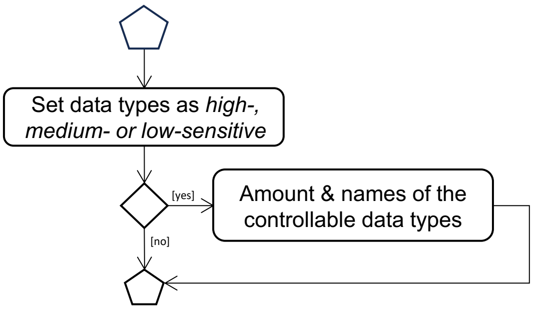

Set data types as *high-, medium- or low-sensitive*
[yes]
[no]
Amount & names of the controllable data types

Figure 6.2.: Input query of the process step 2 exemplary for *UCCP Sensitivity*: quality check for provider

Only for the supplied UCCPs the relevant inputs in the second step must be supplied by the providers. Referring again to the considered devices from Sec. 5, Tab. 6.4 presents

Input query results of the process step 2	Input for UCCP
Set data types as *high-, medium- or low-sensitive*	Sensitivity
Limit data types collected	Minimization
Set the granularity for collected data types	Granularity
Weigh between arising benefits and risks	Sharing Attitude
Limit the consumers of shared data types	Disclosure Limitations
Limit the usage purposes of shared data types	Access Limitations

Table 6.3.: Input queries for the evaluation of the supplied UCCPs

summarized results regarding the supplied UCCPs, as already presented in detail in Sec. 5.3. In Tab. 6.4, we present all the considered devices at this point, besides the example with the Reolink smart camera, in order to use these details in Sec. 6.1.2 during the description of the calculation steps of the quality check for providers.

Based on the inputs of the two process steps, the quality check calculation steps for providers are carried out to deliver the corresponding results. The calculation steps are presented in Sec. 6.1.2.

6.1.2. Calculation Steps with an Example

This section presents the calculation steps of the quality check for the provider. The calculation steps consist of nine steps named from A to I and are presented in Fig. 6.3. We first present relevant information for the calculation, and then we describe each calculation step in subsequent paragraphs.

Goal of the Calculation Steps: The aim of the calculation steps is to classify the provision of the UCCPs on a classification scale per device. The classification scale presents the classification categories for each UCCP per device. In order to calculate the classification category per UCCP, the calculation steps A to D must be carried out once per device and the steps E to I for each of the six UCCP per device. The result of step D (resulting from A to D) delivers us the achievable maximum value for the device and considered functional scope (main or optional) while applying UCCPs during device usage. In comparison, the result of step G (resulting from E to G) delivers us the possible value per UCCP for the considered functional scope, taking into account which data types can be controlled by which UCCP. The calculated ratio between these two values (results of steps D and G) allows us to

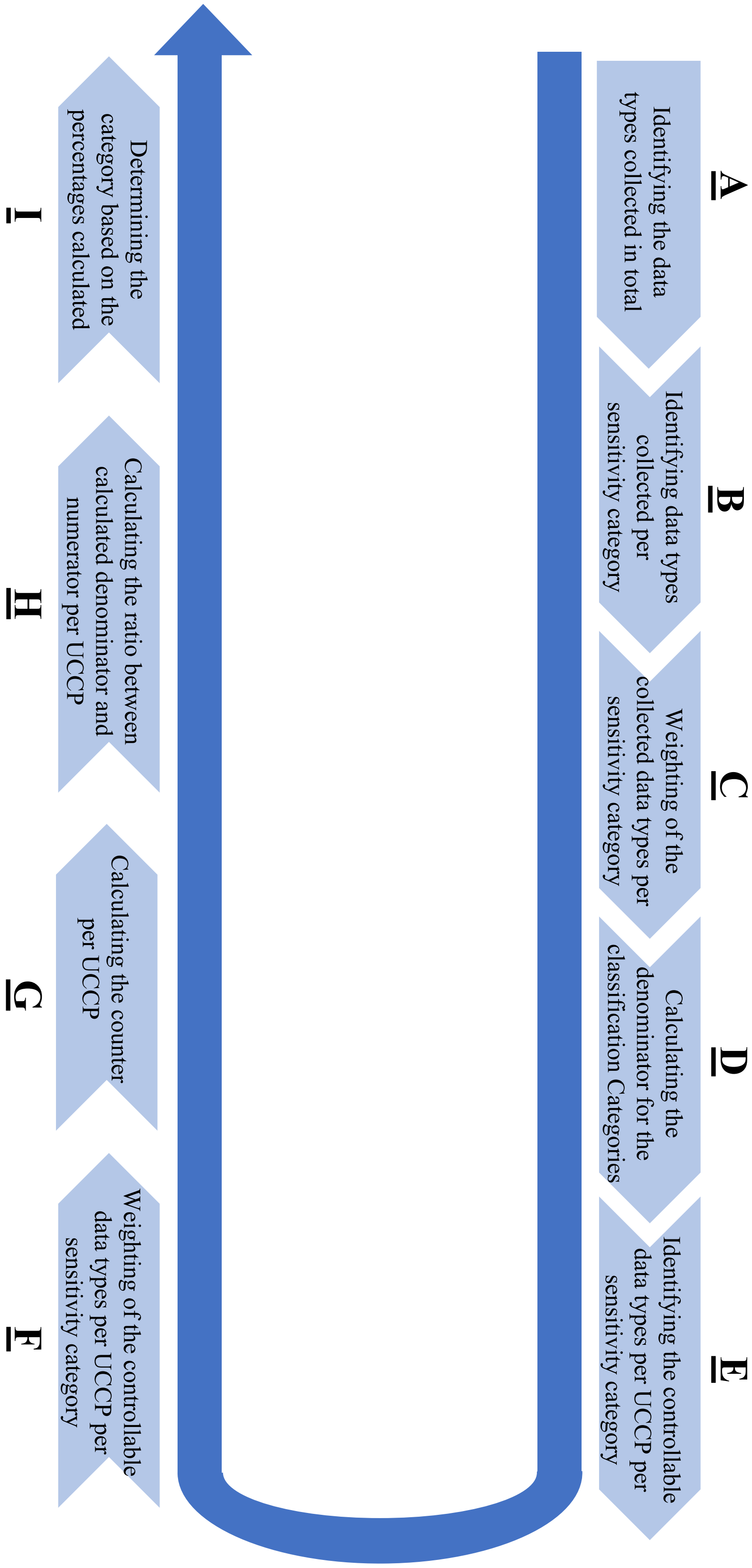

A
Identifying the data types collected in total
B
Identifying data types collected per sensitivity category
C
Weighting of the collected data types per sensitivity category
D
Calculating the denominator for the classification Categories
E
Identifying the controllable data types per UCCP per sensitivity category
F
Weighting of the controllable data types per UCCP per sensitivity category
G
Calculating the counter per UCCP
H
Calculating the ratio between calculated denominator and numerator per UCCP
I
Determining the category based on the percentages calculated

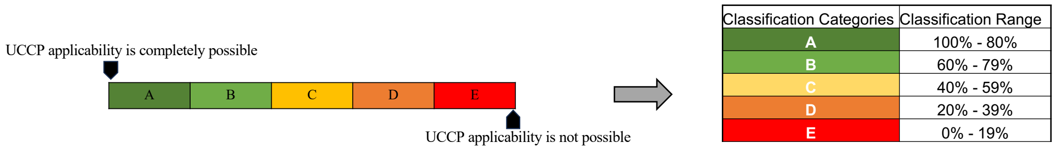
Figure 6.3.: Calculation process of the quality check for provider

Device	Sensitivity	Minimization	Granularity	Sharing Attitude	Disclosure Limitations	Access Limitations
Smart camera: Reolink Argus 3 Pro	0	14	2	0	0	1
Smart bulb: Philips Hue White and Colour	0	2	0	0	0	0
Smart heater thermostat: eQ-3 Homematic IP	0	0	0	0	0	0
Smart robot: Ecovacs Deebot X1 Omni	0	3	0	0	3	0
Smart hub: Apple Homekit	0	0	0	0	3	0
Smart voice assistant: Amazon Echo	0	1	0	0	1	0

Table 6.4.: Input query results of the process step 2 for supplied UCCPs for considered devices, as listed in Sec. 5.3

determine the classification category per UCCP per device in steps H and I.

In order to use the results of the quality check for end users and providers as well as to increase the recognition value, we derived the classification scale based on existing scales in the context of energy efficiency classes for household appliances, CO2 efficiency classes for motor vehicles and Nutri score of consumer food items (see [100–102]). Our classification scale consists of five classification categories from A to E, where A means UCCP is completely provided, and E means UCCP is not provided. The derived scale with five classification categories is presented in Fig 6.4. Each classification category includes a classification range, which we use to assign the results of the calculation process per UCCP on the derived scale. In Fig 6.4, the classification ranges are also presented. The results of the classification scale are presented on each device package.


UCCP applicability is completely possible
A
B
C
D
E
UCCP applicability is not possible

Classification Categories	Classification Range
A	100% - 80%
B	60% - 79%
C	40% - 59%
D	20% - 39%
E	0% - 19%

Figure 6.4.: Classification categories from A to E (derived based on [100–102])

Calculation Steps with an Example: After presenting the goal of the calculation steps, in this section, we present the details of the calculation steps from Fig. 6.3. The corresponding variables and calculation logic are presented in the following for all the calculation steps A to I. We will explain each step based on an example to increase the comprehensibility of each calculation step. In our applied example, we consider the possibility of *Minimization* applicability of the Reolink smart camera. As we outlined in the Tab. 5.1, the data streams of the Reolink smart camera include in total 31 different data types and allow users to apply the *Minimization* for 14 different data types while using main and optional functionality scope (see Tab. 6.4). We consider the controllable data types by applying *Minimization* while using the *main functionality* scope of the Reolink smart camera to explain the calculation steps A to I. The data streams of the *main functionality* of the Reolink smart camera include 12 data types (see Tab. 6.1).

Calculation Process Step A: The first step, A, consists of identifying the amount of the data types in total. The data types in total are defined as D. The D is the set of individual data types, which are defined as d_n. The following formula presents the mathematical expression of the step A:

$$D = (d_1, ..., d_n) \tag{6.1.1}$$

Based on the considered example of the Reolink smart camera, the following output is generated in step A based on the inputs from the provider.

$$D = (d_1, ..., d_{12}); |D_R| = 12 \tag{6.1.2}$$

In the following, we present a few examples regarding individual data types, d_n, from R.

- d_1 = "Name" (see Tab 6.1)
- d_2 = "Address" (see Tab 6.1)
- ...

Calculation Process Step B: The calculation step B includes identifying data types (d_n) per sensitivity category. The sensitivity is data type dependent, and it is relevant to consider in our proposal because users should be able to consider the sensitivity of the data types while setting the UCCPs of a device. The sensitivity categories for the data types are derived based on the results from Sec 4.1.3 and from previous works [71–74, 76, 103]. Tab. 6.5 presents an extract of the derived sensitivity categories considered in the calculation. The sensitivity categories consist of three categories. The three categories with the weighting factors are: (1) *high-sensitive* with 100 as weighting factor (f_{w_1}), (2) *medium-sensitive* with 30 as weighting factor (f_{w_2}) and (3) *low-sensitive* with 1 as weighting factor (f_{w_3}). In order to avoid manipulability of the quality check results, we decided to choose high values for the weighting factors with sufficient distance, compared to the values from existing works like [71, 72, 76], in which the highest sensitivity is measured with 10, medium with 5 and low sensitivity with 1. Additionally, this Weighting also allows us to consider the amount of controllable *high-sensitive* data types with a significant focus in the quality check results for the provider. The set of the data types per sensitivity category are defined as C_1, C_2 and C_3.

Data Type	Weighting Factor
Name	1
Address	100
Email	30
Password	100
Country	1
Wifi Name	100
Wifi Password	100
Password for Smart Device Protection	100
Name for the Device	1
IP-Address	100
Current smart Device recording	100
Device Information	1
Battery Usage Data	1
Image Quality of Recordings	1

Table 6.5.: Extract of the derived sensitivity categories for the calculation process based on [71, 72, 72–74, 76, 103] and results based on Sec 4.1.3

The following formulas present the mathematical expressions of the step B:

$$C_1 \subseteq D; \tag{6.1.3}$$

$$C_2 \subseteq D; \tag{6.1.4}$$

$$C_3 \subseteq D; \tag{6.1.5}$$

For all the abovementioned mathematical expressions, apply the following:

$$C_1 \cap C_2 = \emptyset; C_2 \cap C_3 = \emptyset; C_3 \cap C_1 = \emptyset \tag{6.1.6}$$

Referring to the considered example with the Reolink smart camera, the following result is generated in step B based on the input entered by the provider.

- $|C_1| = 9$ ("*high-sensitive*")

- $|C_2| = 0$ ("*medium-sensitive*")
- $|C_3| = 3$ ("*low-sensitive*")

Calculation Process Step C: In the calculation step C, the data types (D) are weighted with the weighting factors (f_{w_1}, f_{w_2}, f_{w_3}) of the three sensitivity categories ("*high-sensitive*", "*medium*", "*low-sensitive*"), as mentioned in Step B. The weighted data types per sensitivity category are defined as w_x. The following mathematical expressions present the weighting of the data types per sensitivity category:

$$w_1 = |C_1| \cdot f_{w_1} \tag{6.1.7}$$

$$w_2 = |C_2| \cdot f_{w_2} \tag{6.1.8}$$

$$w_3 = |C_3| \cdot f_{w_3} \tag{6.1.9}$$

Referring to the considered example, the Reolink smart camera, the following results are calculated in step C.

- $w_1 = 900\ (9 \cdot 100)$
- $w_2 = 0\ (0 \cdot 30)$
- $w_3 = 3\ (3 \cdot 1)$

Calculation Process Step D: Step D includes calculating the denominator for the classification categories, defined as w. The w consists of the sum from w_1, w_2 and w_3 from step C. The mathematical expression is presented in the following:

$$w = (w_1 + w_2 + w_3) \tag{6.1.10}$$

Based on the considered example of the Reolink smart camera (w_R), the following result is calculated in this step. w_R is the achievable maximum value for the Reolink smart camera and main functionality scope:

- $w_1 = 900$
- $w_2 = 0$
- $w_3 = 3$
- $w_R =$ **903**; (900 + 0 + 3)

Calculation Process Step E: In the calculation step E, the controllable data types per UCCP per sensitivity category are identified, which is defined as A^u_C. The identified controllable data types per UCCP u are classified into the three sensitivity categories C based on the

Tab. 6.5. The mathematical expressions of this step are presented in the following. The following steps must be carried out for each UCCP (from $u = 1$ to $u = 6$).

$$A_1^u \subseteq C_1 \tag{6.1.11}$$

$$A_2^u \subseteq C_2 \tag{6.1.12}$$

$$A_3^u \subseteq C_3 \tag{6.1.13}$$

Referring to the example with the Reolink smart camera and considering the applicability of *Minimization* ($u = 2$) for main functionalities, deliver the following results in step E.

- $|A_1^2| = 4$ (C_1: "*high-sensitive*")
- $|A_2^2| = 0$ (C_2: "*medium-sensitive*")
- $|A_3^2| = 1$ (C_3: "*low-sensitive*")

Calculation Process Step F: The calculation step F consists of the weighting of the controllable data types per UCCP per sensitivity category, which is defined as w_x^u. In the following, the mathematical expressions of this step are presented. The following steps are carried out for each UCCP (from $u = 1$ to $u = 6$).

$$w_1^u = |A_1^u| \cdot f_{w_1} \tag{6.1.14}$$

$$w_2^u = |A_2^u| \cdot f_{w_2} \tag{6.1.15}$$

$$w_3^u = |A_3^u| \cdot f_{w_3} \tag{6.1.16}$$

Step F delivers the following results based on the considered example with the Reolink smart camera, with the supplied UCCP, *Minimization* ($u = 2$).

- $w_1^2 = 400$ $(4 \cdot 100)$
- $w_2^2 = 0$ $(0 \cdot 30)$
- $w_3^2 = 1$ $(1 \cdot 1)$

Calculation Process Step G: Step G calculates the counter per UCCP to calculate the classification category for the considered UCCP. The calculated counter per UCCP is defined as w_u. The following mathematical expressions are carried out for each UCCP (from $u = 1$ to $u = 6$).

$$w_u = (w_1^u + w_2^u + w_3^u) \tag{6.1.17}$$

Referring to the example with the Reolink smart camera (w_{2R}) and the supplied UCCP, *Minimization* ($u = 2$), step G calculates the following results. w_{2R} is the possible value for the supplied *Minimization* and main functionality scope:

- $w_1^2 = 400$

- $w_{\frac{2}{2}} = 0$
- $w_{\frac{2}{3}} = 1$
- w_{2R} = **401**; = (400 + 0 + 1)

Calculation Process Step H: Step H includes calculating the ratio between the calculated denominator from step D and the numerator per UCCP from step G. The result of this step is defined as R^u. In the following, the mathematical expression is presented, and this must be carried out for each UCCP (from $u = 1$ to $u = 6$).

$$R^u = \frac{w_u}{w} \tag{6.1.18}$$

We received the following results in step H based on the considered example with the Reolink smart camera and provided *Minimization* for the main functionality scope (R^2_R).

- $w_{2R} = 401$
- $w_R = 903$

$$R^2_R = \frac{401}{903} = 44\% \tag{6.1.19}$$

Calculation Process Step I: In this step, we determine the classification category for each UCCP based on the results of step H from $u = 1$ to $u = 6$ for the main and optional functionality scope. We determine the corresponding classification range based on the results of R^u from step H and the classification ranges from Fig 6.4. Referring to the example based on the Reolink smart camera and the supplied *Minimization* for main functionality scope, the R^2_R is 44%, which is between 40% and 59%. In this case, the classification category for R^2_R is **C** according to the ranges from Fig 6.4. After applying the whole calculation process with steps A to I for all six UCCPs for the data types of the data streams of the Reolink smart camera, we receive the results for main and optional functionalities as presented in Fig 6.5. In case the provider is interested in improving the quality check results, then the quality check recommends allowing users to control at least all the *high-sensitive* data types by the corresponding device. The results in Fig. 6.5 are derived based on the results in Tab. 6.6 from step H per UCCP (u).

Calculation per UCCP in Steps H	Result for main functionalities	Results for optional functionalities
R_R^1	0%	0%
R_R^2	44%	66%
R_R^3	0%	0,5%
R_R^4	0%	0%
R_R^5	0%	0%
R_R^6	11%	0%

Table 6.6.: Quality check results of the Reolink smart camera from step H

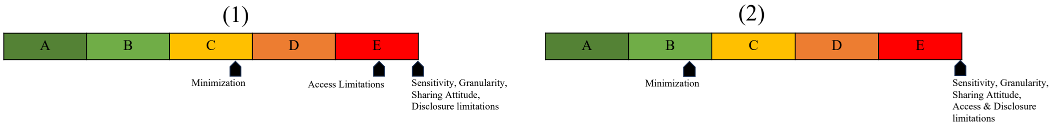


Figure 6.5.: Quality check results for provider for the Reolink smart camera, where scale (1) presents the results for main and scale (2) results for the optional functionality scope

Tab. 6.7 and 6.8 present the amount of the data types for supplied UCCPs by the considered devices while using the main and optional functionality scope. In Tab. G.2 and G.3, we only presented the devices that only provide UCCPs. The Fig. 6.6 and Tab. 6.9 present the quality check results for the provider for all the considered devices from Sec. 5 based on provider inputs and after applying the calculation steps A to I.

Device	UCCP	*high-sensitive*	*medium-sensitive*	*low-sensitive*
Smart bulb: Philips Hue White and Color	*Minimization*	1	0	1
Smart hub: Apple Homekit	*Disclosure Limitations*	1	0	0
Smart voice assistant: Amazon Echo	*Minimization*	1	0	0
	Disclosure Limitations	1	0	0

Table 6.7.: Controllable data types by the considered devices while using main functionality scope (derived based on the results from Sec. 5)

Device	UCCP	*high-sensitive*	*medium-sensitive*	*low-sensitive*
Smart robot: Ecovacs Deebot X1 Omni	*Minimization*	3	0	0
	Disclosure Limitations	3	0	0
Smart hub: Apple Homekit	*Disclosure Limitations*	1	1	0

Table 6.8.: Controllable data types by the considered devices while using optional functionality scope (derived based on the results from Sec. 5)

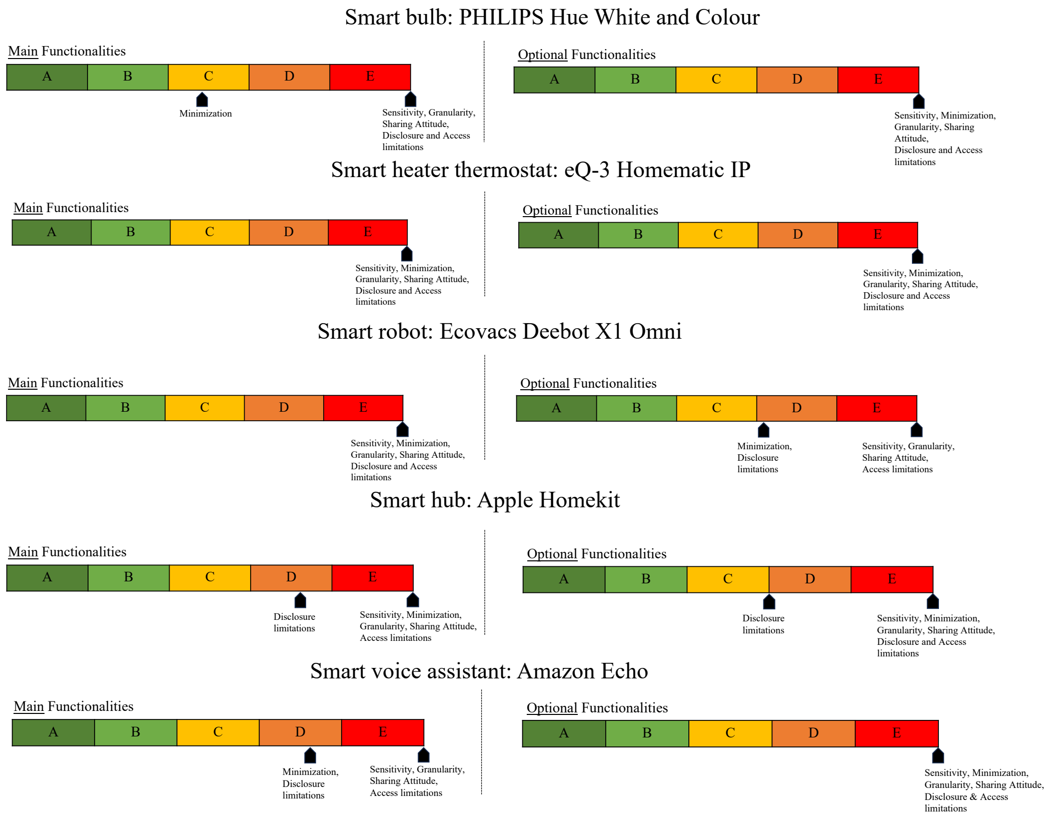


Figure 6.6.: Quality check results for provider for considered devices from Sec. 5

Device	Functionality Scope	*Sensitivity*	*Minimization*	*Granularity*	*Sharing Attitude*	*Disclosure Limitations*	*Access Limitations*
Smart bulb: Philips Hue White and Colour	main	0%	**50%**	0%	0%	0%	0%
	Optional	0%	0%	0%	0%	0%	0%
Smart heater thermostat: eQ-3 Homematic IP	main	0%	0%	0%	0%	0%	0%
	Optional	0%	0%	0%	0%	0%	0%
Smart robot: Ecovacs Deebot X1 Omni	main	0%	0%	0%	0%	0%	0%
	Optional	0%	**38%**	0%	0%	**38%**	0%
Smart hub: Apple Homekit	main	0%	0%	0%	0%	**23%**	0%
	Optional	0%	0%	0%	0%	**39%**	0%
Smart voice assistant: Amazon Echo	main	0%	**25%**	0%	0%	**25%**	0%
	Optional	0%	0%	0%	0%	0%	0%

Table 6.9.: Results of the quality check for the provider (in percentage) for considered devices from Sec. 5

To summarize, the comparison of the quality check results for the main and optional functionalities outlines the following aspects:

- Three of the devices allow users to apply UCCPs for more data types while using the optional than the main functionalities of the corresponding devices: Smart robot, camera and hub.

- Two of the devices allow users to apply UCCPs for more data types while using the main than the optional functionalities of the corresponding devices: Smart bulb and voice assistant

- One of the devices does not allow users to apply any UCCPs neither for the used main nor optional functionalities of the corresponding device: Smart heater thermostat

6.2. Quality Check for Users Including User-Centric Data Stream Controls

In this section, we describe the quality check solution for users to outline how device users can apply supplied UCCPs by providers. First, we present the process of the solution in Sec. 6.2.1, and second, we describe the solution based on the example Reolink smart camera in Sec. 6.2.2.

6.2.1. Process

In this section, we present the entire process of the quality check solution for users. This process allows users to set the supplied UCCPs by the provider for the corresponding device. During the setting up phase of the device, the users have to go through the supplied information and set different settings. Initially, the users must review the quality check results for the providers and the controllable data types before setting their preferred functionality scope. The users can choose between main and optional functionality. After choosing the functionality scope, users must decide whether they want to set the supplied UCCPs. In case users want to apply the settings of the supplied UCCPs, then the users are asked to go through the process steps presented in Fig. 6.7. Before setting the supplied UCCPs, users have to choose their preferred control layer. The control layers are derived based on the evaluation of the device functionalities in Sec. 5. In the smart home context, users can control their devices differently, depending on which devices are integrated into the corresponding smart home environment.

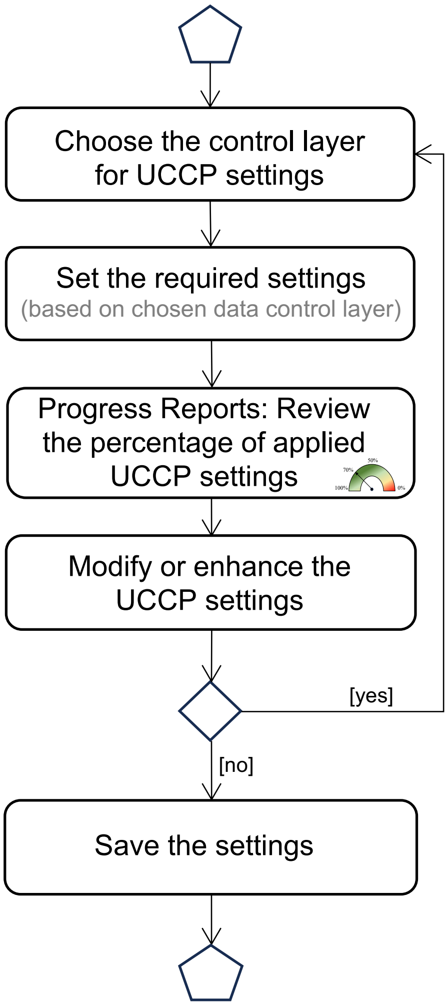

Choose the control layer for UCCP settings
Set the required settings
(based on chosen data control layer)
Progress Reports: Review the percentage of applied UCCP settings
50%
70%
100%
0%
Modify or enhance the UCCP settings
[yes]
[no]
Save the settings

Figure 6.7.: Process steps for setting the supplied UCCPs in the quality check for users

Our evaluation in Sec. 5 allows us to derive four data control layers. Users can choose their preferred data control layer while using devices. The four data control layers are (1) the smart hub layer, (2) the smart device layer, (3) the data type category layer and (4) the data type layer. Fig. 6.8 presents the derived four data control layers. The description of the four layers is presented in the following. Depending on users' control willingness and spent time setting such controls, users can choose their preferred data control layer in the quality check for users.

- **Smart Hub Layer:** Smart hubs are installed to interact with all the devices installed within a smart home. The UCCPs are set in the smart hub and thus into all the data types of the data streams of the installed devices in that specific smart home.

- **Smart Device Layer:** Data streams of each device include different data types. The UCCPs are set for a device and thus to all the data types of the specific device.

- **Data Type Category Layer:** Every data type is clustered into a data type category, and the data types within a data type category are controlled. The UCCPs are set for a data type category and, thus, to all data types in that category. We derived five data type categories while considering the data types collected by the six devices from Sec. 5. The derived five data type categories are (1) personal information, (2) personal preferences, (3) location data, (4) video & audio recordings, and (5) diagnostic, usage and performance data. Those five derived data type categories in this layer are presented with a few corresponding data type examples in Tab. 6.10. Fig. 6.9 presents the amount of the data types in each data type category for considered devices in Sec. 5.

- **Data Type Layer:** Each data type is controlled, and UCCPs are set for each data type.

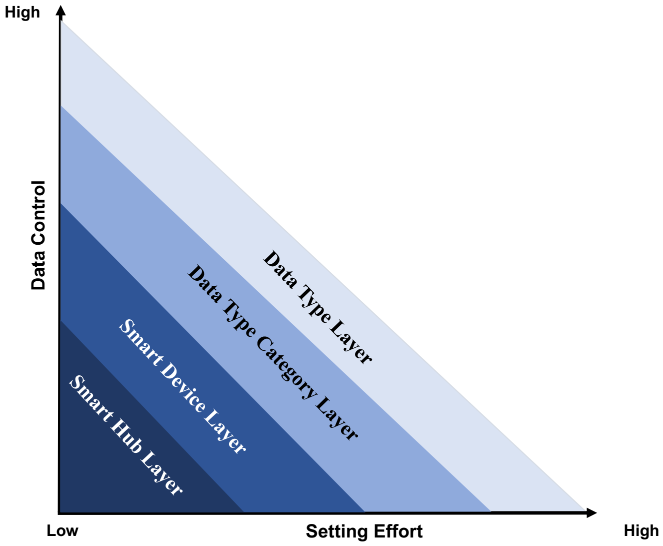


Figure 6.8.: Derived data control layers for users (based on the evaluation in Sec. 5)

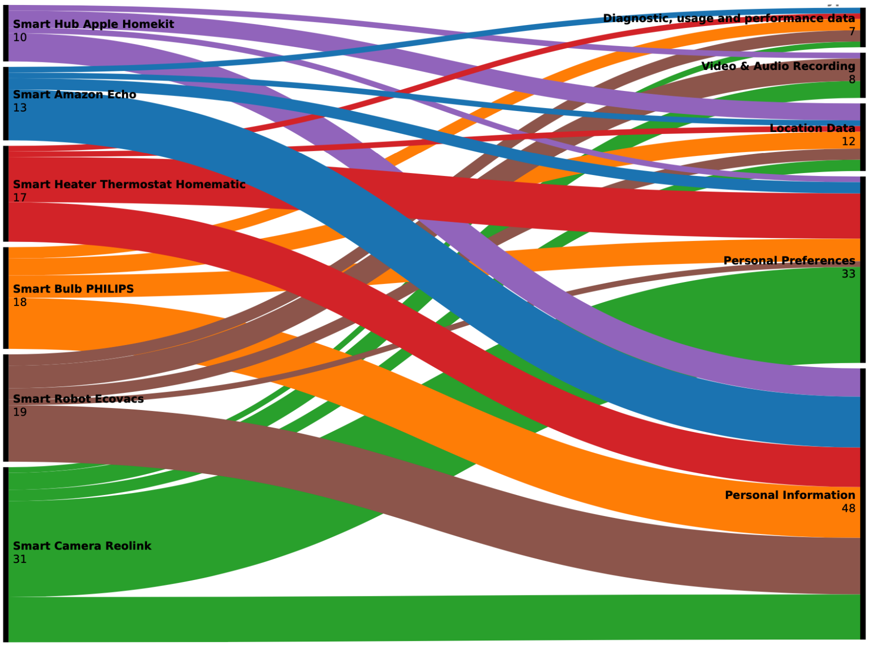


Figure 6.9.: Amount of the data types in each data types category of the six considered devices from Sec. 5

Data Type	Data Type Category
Name	Personal Information
Address	Personal Information
IP Address	Location Data
Device Information	Diagnostic, Usage and Performance Data
Audio & video recording schedule	Personal Preferences
Date of recording	Video & Audio Recording

Table 6.10.: Example extract for data types categories derived based on [71,72,76] and based on results in Sec. 5

After users choose the preferred data control layer, users can set the supplied UCCP settings. The amount of the setting options is adjusted depending on the chosen data control layer. While setting the UCCPs, users can also review the progress report presenting the percentage of the applied UCCP settings. The percentage of the applied UCCP settings is presented as outlined in Fig. 6.10 so that the users can modify or enhance the UCCP settings based on the progress resport. The percentage of the applied UCCP settings already reaches 65% to 70%

(in the green area in Fig. 6.10), in case users set the UCCP settings for all the controllable *high-sensitive* data types.

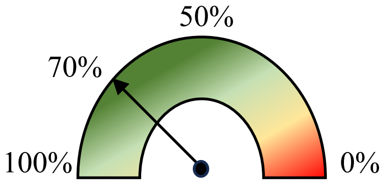


Figure 6.10.: Progress report for applied UCCP settings

After adjusting the settings, users can save those and continue with the setup process for device usage. If the users do not want to apply the settings of supplied UCCPs, then the users can also proceed with the setup process without going through the steps from Fig. 6.7. The setup process asks users to enter the requested data types and allows users to start using the corresponding device.

6.2.2. Process with an Example

This section will present the above-described process with the example of the Reolink smart camera. After reviewing the quality check results for the provider as presented in Fig. 6.5 (from Sec. 6.1), the users can choose whether they want to use the main, optional or entire functionality scope. The data types of the data streams regarding the chosen functionality scope are presented in Tab. 6.1 and Tab. 6.2 (in Sec. 6.1). If the user decides to use only the main functionality, then the data streams include only 12 out of 31 possible data types, and in the case of the entire functionality scope, 31 data types. Although the Reolink smart camera collects 31 data types, it only allows users to control 17 data types in total while applying three UCCPs (*Minimization*, *Granularity* and *Access Limitations*) partially. In our example, the user wants to apply the supplied UCCPs for all the controllable data types while using the main and optional device functionalities. The user first has to choose the data control layer. In the case of the Reolink smart camera, the data control layers include the following amount of setting options regarding the supplied UCCPs.

- Data Type Layer: 17 setting options per camera
  - 17 = 14 data types for *Minimization* + 2 data types for *Granularity* + 1 data types for *Access Limitations*
- Data Type Category Layer: 6 setting options per camera
  - 6 = 2 *Data Type Categories* multiplied with 3 supplied UCCPs
- Smart Device Layer: 3 setting options per camera

  - 3 = 1 device multiplied with 3 supplied UCCPs

- Smart Hub Layer: 3 setting options for all cameras in the smart home environment
  - 3 supplied UCCPs

After choosing the data control layer, users have to set the settings for the three supplied UCCPs, based on the setting options described in Sec. 4.2.2. After applying the UCCP settings, users can start using the Reolink smart camera. The requested data types by the camera are controlled according to users' UCCP settings during the setup phase, as described above.

6.3. Evaluation of the Proposal

In order to outline the added value of our proposed quality check solution, we compare our UCCP-based proposal with already existing approaches, [25, 37, 41, 42, 62]. The existing approaches from [25, 37, 41, 42, 62] provide relevant input for our proposal, as described in Sec. 3.2. While [37] presents a reference architecture including a trade-off decision model, [25] proposes an RBAC-based solution in the context of data sharing in smart homes. Furthermore, the works [41, 42] present a security and privacy-preserving solution as well as related negotiation mechanisms for the IoT context. Additionally, the work [62] presents a smart home configurator that informs users regarding privacy and security implications without informing users regarding specific data streams and possibilities to apply UCCPs for the corresponding device. Tab. 6.11 presents the results of this evaluation (based on [7, 81]). We considered eleven evaluation metrics, and those are derived from [6, 7, 13, 81, 104, 105]. In our evaluation, we used the following scale: ○ = no possibility; ◑ = partially possible; and ● = possible. In case the considered solution fully supplies the considered evaluation metric, then ● is used in Tab. 6.11. When the considered solution only partially supplies the considered evaluation metric, then ◑ is used in Tab. 6.11. The ○ is used in Tab. 6.11 when the considered solution does not supply the considered evaluation metric. To summarize our evaluation in Tab. 6.11, the qualitative evaluation of these approaches outlines that previous approaches do not allow users to control the entire data streams of the devices sufficiently. Hence, those approaches also do not supply users with options to apply derived UCCPs and receive progress reports regarding applied UCCP settings while using the devices. Furthermore, the previous approaches also do not supply any mechanisms for providers to evaluate and classify their devices regarding the UCCP provision for data streams of those devices.

6.4. Summary

To sum up, the participants of our user-centric experiments outlined that they want user-centric solutions to support them in gaining more transparency and applying UCCPs for devices. In order to address these requirements, we derived a UCCP-based quality check for providers and users. In the quality check solution for the provider, we considered the user need for more transparency regarding the following points: (1) what are the data types of the data streams, (2) for which device functionalities these data types are necessary and (3) which data types of the data streams are controllable by applying which UCCPs. The quality check results for providers are presented by using a scale derived based on existing scale systems in another user context, like energy and CO2 efficiency, to increase the familiarity of the results. Furthermore, the quality check results for the provider also deliver providers transparency regarding the controllable data types, and they can improve the results by allowing users to control more data types or by minimizing the amount of the data types of data streams.

Moreover, on the other hand, the quality check for users allows them to apply the supplied UCCPs in the user-preferred control layer. In this part, we considered the user needs regarding more control over data types of data streams while allowing users to apply those controls in an efficient and preferred way. The quality check for users presents in the device's setup process the controllable data types per UCCP to allow users to set the functionality scope, and UCCP settings. In the setup process, the solution also presents users with a progress report including the percentage of the applied UCCP settings to allow users to modify and enhance the applied UCCP settings. Furthermore, our proposal also allows the examination of the quality check results by a third party, like a neutral body like consumer protection, as in Germany Stiftung Warentest (details: [106]), in order to control whether the provider has correctly stated the inputs regarding data collection and usage during the device application.

Evaluation Metrics	Our Quality Check	Trade-off Decision Model [37]	RBAC-based Model [25]	Privacy Preservation Model [41]	Negotiation Model [42]	Smart Home Configurator [62]
Minimizing data types of data streams by users [7, 81, 105]	●	○	○	○	○	○
Aggregating data types of data streams by users [7, 81, 105] (*UCCP: Granularity*)	●	○	○	◐	○	○
Limiting data type processing according to specific device functionalities by users [105]	●	○	○	◐	○	○
Presenting the processed data type transparently to users	●	●	○	○	◐	◐
Limiting the data collection by users [6, 7, 13, 81] (*UCCP: Minimization*)	●	○	○	○	○	○
Limiting the data disclosure and access by users [6, 7, 13, 81] (*UCCP: Disclosure Limitations, Access limitations*)	●	◐	◐	◐	●	○
Setting the sensitivity perception regarding data streams by users [6] (*UCCP: Sensitivity*)	●	◐	○	○	◐	○
Presenting the impact of data disclosing transparently to users [7, 13, 81, 104], for instance, arising risks and advantages by disclosing (*UCCP: Sharing Attitude*)	●	◐	○	○	○	○
Evaluating devices regarding supplied UCCPs by providers	●	○	○	○	○	○
Presenting the supplied UCCPs transparently by providers	●	○	○	○	○	○
Presenting progress report regarding applied UCCP settings	●	○	○	○	○	○

Table 6.11.: Results of the qualitative analysis of the proposed quality check solution for providers and users of devices (based on [7, 81])

7. Discussion

In this chapter, we discuss the results of the Sec. 4 - 6. In the first step, we present in Sec. 7.1 the strengths and in Sec. 7.2 the limitations of our research work.

7.1. Implications of Considering User-Centric Data Stream Controls

This section outlines the strengths of our derived UCCPs and UCCP-based proposal. While in Sec. 7.1.1, we consider the implications from a user perspective, in Sec. 7.1.2, we consider the implications from a provider perspective.

7.1.1. Smart Home Device User

Our derived UCCPs and the proposed quality check proposal for users allow us to address the research gaps regarding GDPR requirements, like giving more transparency and control to users regarding data streams in the IoT, especially in the smart home context. We derived our main results regarding the UCCPs based on user-centric experiments. The results of the user-centric experiments outline that (1) although users are aware of the data streams, they have less transparency about the data types of the data streams, and (2) users have nearly no possibility to apply UCCPs while using devices. These results allow us to answer **RQ1: To what extent are users aware of the data streams of smart home devices?** as follows: users are generally aware of the data streams of devices, but the lack of transparency regarding the data streams while using devices leads to the conclusion that users are not fully aware of the data streams of the corresponding devices. Allowing participants to have more control over data streams while using the devices will improve their awareness regarding the data streams of the corresponding devices. Not supplying users with the required transparency and UCCPs for devices could lead to less user acceptance of those devices in the long term. Thus, our derived UCCPs allow users to apply minimization and aggregation strategies, capture their sensitivity perception and attitude for sharing, limit data disclosure and access. Furthermore, our UCCPs address user and GDPR requirements, allowing users to be involved regarding data stream control while using corresponding devices. Our user-centric experiments also outline that the users are motivated to use devices while having more transparency and possibilities for the application of UCCPs. Integrating the UCCPs in devices also influences the user acceptance of using those devices and associated technology advancement.

The derivation of UCCPs based on user-centric experiments allows us to answer the **RQ2: What are the UCCPs for smart home devices?** as follows: as mentioned above, the derived UCCPs allow users to control the data collection, processing and disclosure while using devices in the smart home context. Furthermore, the derived UCCPs also allow users to gain more transparency regarding arising risks and advantages from the disclosure of the collected data types and allow them to set their sensitivity perception.
Moreover, our quality check proposal allows users to apply the derived UCCPs after getting transparency regarding controllable data types based on the chosen functionality scope of the corresponding device. In daily life, most users tend not to spend time reading a lot of text and go through a lot of settings while using devices. In order to address this aspect, our quality check solution for users allows them to apply the UCCPs on their preferred control layer, depending on how much control the user wants to have and how much time s/he wants to spend setting the supplied UCCPs. Furthermore, it is possible to support the users while UCCP settings using machine learning algorithms. In this context, we have published in [81] our initial ideas on how the setting options of the UCCPs can be optimized and derived based on users' past behaviours and activities using machine learning algorithms. In our ideas in [81], we outlined how users can be supplied with data-sharing recommendations as input for UCCP settings. The recommendations are derived based on GDPR-related best practices, users' past activities and context-sensitive factors with the help of supervised and active machine learning methods, like *Support Vector Machines* (SVM) and decision tree. Further investigations must be carried out in order to outline how those machine learning algorithms can be trained based on data from different devices with different users over a long period of time to make such algorithms applicable in this context. In this way, we could increase the efficiency and the automation level of the quality check solution for users. Moreover, compared to other solutions, our solution for users also visualizes with a progress report whether they have exhausted the possible UCCP setting options and applied complete possible UCCPs when using the corresponding device. It is also conceivable that the smart home configurator from [62] can be enhanced with our proposed UCCP-based quality check solution in order to compensate the limitations of the configurator as mentioned in Sec. 3.2 and 3.4.

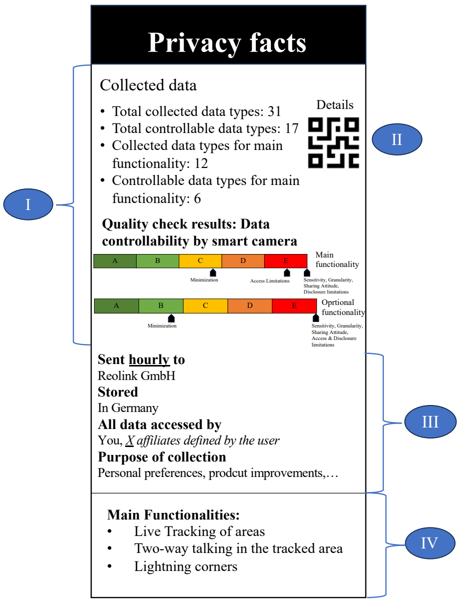
7.1.2. Smart Home Device Provider

The benefits of the derived UCCPs and its application within the quality check solution for providers are threefold: (1) giving more transparency to their users, (2) increasing user acceptance regarding their devices, and (3) deriving improvements for devices regarding data stream controls. Our validation of the existing devices in Sec. 5 outlined that the users have poor possibilities to gain transparency and apply UCCPs while using such devices. The validation results allow us to answer the **RQ3: To what extent do the existing smart home devices supply UCCPs?** as follows: our validation confirms that (1) those devices include various amounts of data types in their data streams and (2) that the application of the derived

UCCPs from Sec. 4.2 is hardly possible or sometimes not possible at all. Furthermore, the analysis also outlines that the providers do not supply the required transparency towards the users regarding which data types are collected, processed and shared for which functionality of each device. A few devices, for instance, smart heater thermostats and bulbs, do not or hardly supply any UCCPs, whereas the Reolink smart camera allows the partial application of the three UCCPs. The evaluation underlines that the providers generally describe the information regarding data streams and UCCPs. This lack of transparency regarding data streams and the limitations regarding UCCPs lead to a further decline in user acceptance regarding those devices. Therefore, supplying the UCCPs allows providers to increase transparency regarding the data streams and enable users to control the data types of those data streams.

In addition, our proposed quality check for providers allows them (1) to gain transparency, which UCCPs are possible, and simultaneously (2) to deliver users this transparency in the device purchase process. Our derived quality check proposal allows us to answer the **RQ4: How to consider UCCPs while using smart home devices?** as follows: applying the entire quality check proposal allows us to outline how the providers and users can use the derived UCCPs. The quality check allows providers to address the GDPR requirements in this context, like giving users more transparency and control regarding the data streams in the IoT and smart environment context (especially GDPR Art. 5, 9, 12, 13, 14, 15 and 19). Simultaneously, the application of the quality check solution will increase the user acceptance of devices in the long term. Moreover, we recommend presenting the results of the quality check for the provider on the corresponding device packages in order to allow users to gain transparency regarding data streams and their control before purchasing the corresponding device. For this purpose, it is possible to integrate the results of the quality check for providers in the *Privacy Facts Label* from [5].

We described the details regarding the *Privacy Facts Label* and its derivation in Sec. 2.3. Fig. 7.1 presents our proposed adjustments of the *Privacy Facts Label* based on the example Reolink smart camera. In the following, we explain the proposed adjustments. The first adjustments are presented as I: the scale of the quality check results for the provider and the amount of collected and controllable data types for main and entire device functionalities. This information allows users to view at first sight the amount of the collected and controllable data types on the device package and which UCCPs from Tab. 4.6 are supplied. This adjustment allows us to increase the transparency regarding the collection and control of data types towards the users already during the purchase process. The subsequent adjustment presented as II includes that the users receive details by scanning the presented QR code on the label regarding controllable data types per UCCP and collected data types clustered in main and optional functionality. Tab. E.1 in Appendix E presents an example extract regarding this information for the Reolink smart camera. Additionally, by scanning the QR code, the users can access the description of the UCCPs so that the users can see which UCCP is supplied for which collected data type. This second adjustment allows users to gain detailed transparency regarding specific controllable data types and required data types for



Privacy facts

Collected data

- Total collected data types: 31
- Total controllable data types: 17
- Collected data types for main functionality: 12
- Controllable data types for main functionality: 6

Details

Quality check results: Data controllability by smart camera

Sent <u>hourly</u> to
Reolink GmbH
Stored
In Germany
All data accessed by
You, *<u>X</u> affiliates defined by the user*
Purpose of collection
Personal preferences, prodcut improvements,…

Main Functionalities:

- Live Tracking of areas
- Two-way talking in the tracked area
- Lightning corners

I
II
III
IV
A
B
C
D
E
Main functionality
Minimization
Access Limitations
Sensitivity, Granularity, Sharing Attitude, Disclosure limitations
A
B
C
D
E
Oprtional functionality
Minimization
Sensitivity, Granularity, Sharing Attitude, Access & Disclosure limitations

Figure 7.1.: Adjusted *Privacy Facts Label* based on the quality check results for provider for Reolink smart camera based on [5]

the main and optional functionality scope already during the purchase process. The third adjustment presented as III includes the default settings regarding storage, data access and purpose of data collection. These default settings are presented as examples so that users can get a first impression regarding data storage, sharing and disclosure. In case users can limit the access and usage purposes of the shared data types as well as the granularity of the stored data types according to the derived UCCPs, then corresponding hints are presented under these adjustments, as we can see in Fig. 7.1 with *X affiliates defined by the user*. This presentation allows us to deliver user transparency regarding limiting the consumers accessing the shared data types on this label besides the details under the QR code. The last and fourth adjustment presents users with the main functionality of the corresponding device in order to increase users' transparency (see IV in Fig. 7.1).

Additionally, it is essential to integrate the derived UCCPs in the service layers of IoT architecture as well as in ISO standards, like ISO/IEC 30118-1, as described in Sec. 2.1. In this way, the providers can supply the derived UCCPs for users in a standardized way and also can address the GDPR requirements simultaneously.

To sum up, our derived UCCPs and proposed quality check solution for users and providers allow us to (1) address the open research gaps of the GDPR and (2) give users the possibility to control the data streams of the devices from a user perspective. Additionally, our solution allows providers to improve their device settings regarding UCCPs and increase the user acceptance of the corresponding devices. Besides the benefits, our work has a few limitations, which we outline in the following section.

7.2. Threats to Validity

In this section, we discuss the threats to validity aspects. We focus on the validity criteria from [107]. In Sec. 7.2.1, we consider the aspects of the construct validity, in Sec. 7.2.2, the aspects of the internal validity. Subsequently, in Sec. 7.2.3, we present the aspects of the external validity.

7.2.1. Construct Validity

According to [107], construct validity considers the extent to which a study is consistent with its intentions [108]. In the case of this thesis, the construct validity aspects are concerned with the extent to which our user-centric experiments and device validation conform to the derived UCCPs and the proposed quality check solution. To receive reliable results regarding the UCCPs based on user-centric experiments, we had to consider at least half of the world population, including all the countries. To mitigate this threat, we carried out user-centric experiments within this thesis, each with approx. 200 (in experiment 1) and 700 (in experiment 2) participants in order to receive representative results to derive the UCCPs. In addition, we designed our questionnaires from the user-centric experiments in English to reach a representative number of the participants. Also, we used different online platforms to reach frequent Internet users because those users are more likely to apply the technologies related to the Internet, like IoT, smart devices, and smart environments. Moreover, to achieve independent and uninfluenced results regarding our investigation intents, we structured our questionnaire-based experiment with several matrix and multiple-choice questions. The experiment questions were formulated neutrally so as not to influence the participants during the user-centric experiments. We also carried out pretests with volunteers to ensure neutrality and to avoid the manipulability of the experiment participants. Based on the participants' answers in the user-centric experiments, especially in the first user-centric experiment, we could derive the UCCPs presented in Sec. 4.2 and answer our RQs. Even though the answers of the user-centric experiments deliver the primary basis for our derived UCCPs, we have to consider that those answers represent participants' opinions and do not necessarily reflect their actual behaviour. Additionally, the proposed quality check in Sec. 6 must be validated with further user-, provider-, and device-centric experiments to increase the applicability of the UCCP-based proposal. To mitigate this threat, we evaluated the derived UCCPs based on six existing devices from the main service-oriented smart home categories, derived based on different comparative tests and statistics, like [29–31]. The six considered devices in our evaluation are the most popular and used devices in these six categories. Based on our evaluation of existing devices, we proposed a quality check solution for providers and users in our thesis. Additionally, in order to mitigate the threat of the applicability, we recommend integrating the results of the proposed quality check for providers in an existing and user-centric tested label for smart devices, *Privacy Facts Label*, from [5], as described in Sec. 7.1.2. Furthermore, the results of our proposed quality check for the provider can be examined at any time by a third party, for instance, neutral bodies like consumer protection, as in Germany Stiftung Warentest, in order to verify the accuracy of the results. In this way, the related manipulability aspects of the quality check results by providers are addressed. To improve the applicability and usability aspects, it is essential that a representative amount of providers implement the proposed quality check solution and introduce those devices into the smart home market. Based on these devices, user- and provider-centric applicability and usability tests in the real world can be carried out to derive improvements.

7.2.2. Internal Validity

Internal validity refers to the causal relations investigated by the study [107]. In our case, this aspect concerns the quantitative evaluation of our user-centric experiments and our evaluation of the six devices. In order to mitigate the threats in this context, we derived the causal relations after testing the derived hypotheses of the user-centric experiments. To test the derived hypotheses, we applied different statistical tests. The statistical tests are derived based on the fulfilled preconditions after discarding the invalid cases based on different criteria, for example, spent time and invalid answers. With this procedure, we could derive the causal relations regarding the UCCPs based on matching statistical tests to the collected data sample from the user-centric experiments. Furthermore, in our device evaluation, we followed the same process steps in order to analyse whether those devices allow the application of the derived UCCPs. In this way, we were able to outline the lack of the UCCP application in existing devices.

7.2.3. External Validity

Referring to [107], external validity includes the fact whether the general applicability of the derived results is given. This means that those results can also be applied to other contexts and not only in the considered context, like in any context where data types are collected, processed and shared to supply users with services or functionalities, for instance, websites and other online services. In our case, the external validity refers to whether the derived UCCPs and quality check proposal for providers and users can be applied to all devices and other smart environments. To receive reliable results regarding the applicability of the derived UCCPs, we had to consider all the individual data types of the data streams of all existing devices. To mitigate this threat, we considered six devices from the six main service-oriented smart home categories. Those six devices include different data types in their data streams. In total, we considered 76 individual and non-overlapping data types in our device evaluation. Based on the device evaluation, we were able to derive the quality check proposal for providers and users in order to outline how the derived UCCPs can be used to classify and configure existing devices. Considering the most popular smart home categories and devices while deriving the UCCPs and quality check proposal allow us to present a solution which can be applied in the smart home context in general. Furthermore, the quality check solution from Sec. 6 can also be applied to smart devices in other smart environments to increase user acceptance of corresponding devices. The derived quality check proposal also delivers the flexibility to consider the relevant UCCPs in the corresponding smart environment in case not all the UCCPs are necessary. This flexibility allows users and providers to apply the quality check solution in other contexts where data types are collected, processed and shared to supply users with different services in order to increase transparency and user acceptance regarding the supplied services.

8. Conclusion

In this section, we conclude the results of this thesis. In Sec. 8.1, we summarize the main results of this thesis, and in Sec. 8.2, we outline the further research options in the context of the topic of this thesis.

8.1. Summary

In this Thesis, we proposed six UCCPs and outlined how the existing devices can be classified and configured based on the derived UCCPs by presenting the quality check solution for providers and users. The proposed UCCPs allow the application of different data collection strategies, like minimization and aggregation, as well as data disclosure and access limitations, like limiting the data sharing and data consumers accessing the data types shared. Additionally, the derived UCCPs allow users to capture the sensitivity perception and sharing attitude of users for collected data types by the devices. The findings of the user-centric experiments outline that the users are willing to have UCCPs while using devices. The experiments with existing and most popular devices outline that those devices only poorly supply the UCCPs. The user-centric experiments also outline that the users are motivated to use the devices while applying UCCPs and having more transparency. This willingness, in turn, influences the user acceptance of devices.

Furthermore, the proposed quality check solution for providers allows providers to evaluate their devices to which extent they supply users the derived UCCPs while using the corresponding device. Based on the quality check results for the provider, they can improve their devices regarding the UCCPs and deliver those results to users to increase transparency regarding applicable UCCPs. We recommended that the results of the quality check for the provider should be presented on the device package in an already proposed label with a scale which is popular and familiar to users. This presentation lets users gain transparency regarding data streams and supplied UCCPs already in the device purchasing process. The quality check for users allows users to set the supplied UCCP settings based on their chosen functionality scope of the device and data control layer, depending on their willingness regarding data control and effort. Additionally, the quality check for users also provide a progress report regarding applied UCCP settings by the users. The approach can be implemented and applied by any provider and user without a deep understanding of our proposed solution. Moreover, our quality check solution also allows the examination of its results by a third party, for instance, a neutral body like consumer protection, Stiftung Warentest in Germany.

8.2. Outlook

The proposed UCCP-based quality check solution for providers and users outlines that it is possible to deliver users more data stream controls from a user perspective in the smart home context. Although our work present promising results, there are open research areas which can be addressed in the future. In the following, we outline the details regarding further research. The derived UCCPs and their consideration in the quality check proposal must be tested in real-world experiments to improve the applicability for users and providers of devices. Additionally, further investigations must be carried out to increase the usability of the quality check for users and providers. In this context, it is interesting to investigate to which extent machine learning algorithms can be applied to increase the efficiency and the automation level of the quality check solution for users. Moreover, further research should be carried out to investigate how the quality check solution can be considered in the IoT architecture-related ISO standards, like ISO/IEC 30118-1 from [26, 27] and considered in communication standards, like Matter protocols for devices. Enhancing IoT architectures with UCCPs as standards for devices will deliver possibilities to holistically address the open requirements of GDPR and increase user acceptance regarding device usage.

Bibliography

[1] D.-R. Berte, "Defining the iot," in *Proceedings of the international conference on business excellence*, vol. 12, no. 1, pp. 118–128, 2018.

[2] M. J. Bassman, C. Dahlke, and L. Russell, "Development of an interoperability tool for software engineering environments," in *Proceedings of the fifth Washington Ada symposium on Ada*, pp. 49–57, 1988.

[3] J. Dahmen, D. J. Cook, X. Wang, and W. Honglei, "Smart secure homes: a survey of smart home technologies that sense, assess, and respond to security threats," *Journal of reliable intelligent environments*, vol. 3, no. 2, pp. 83–98, 2017.

[4] Statista. (2023) Number of internet of things (iot) connected devices worldwide from 2019 to 2023, with forecasts from 2022 to 2030. [Online]. Available: https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide/

[5] A. Railean, "Improving iot device transparency by means of privacy labels," 2022.

[6] C. I. Wickramasinghe and D. Reinhardt, "A survey-based exploration of users' awareness and their willingness to protect their data with smart objects," in *Proceedings of the 14th IFIP Summer School on Privacy and Identity Management Data for Better Living - AI and Privacy*, vol. 576, pp. 427 – 446, 2020.

[7] C. Wickramasinghe and D. Reinhardt, "A user-centric privacy-preserving approach to control data collection, storage, and disclosure in own smart home environments," in *International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services (MobiQuitous)*, pp. 190–206, 2021.

[8] E. Zeng, S. Mare, and F. Roesner, "End User Security and Privacy Concerns with Smart Homes," *In Proceedings of the 13th USENIX Conference on Usable Privacy and Security (SOUPS '17),*, pp. 65–80, 2017.

[9] S. Dutta, S. S. L. Chukkapalli, M. Sulgekar, S. Krithivasan, P. K. Das, and A. Joshi, "Context Sensitive Access Control in Smart Home Environments," in *IEEE 6th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)*, pp. 35–41, 2020.

[10] A. Ukil, S. Bandyopadhyay, and A. Pal, “Iot-privacy: To be private or not to be private,” in *2014 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 123–124, 2014.

[11] W. Zhou, Y. Jia, A. Peng, Y. Zhang, and P. Liu, “The effect of iot new features on security and privacy: New threats, existing solutions, and challenges yet to be solved,” *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1606–1616, 2019.

[12] H. Nissenbaum, “Privacy as contextual integrity,” *Wash. L. Rev.*, vol. 79, pp. 119 – 158, 2004.

[13] 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),” *Official Journal of the European Union*, vol. L119/1, pp. 1 – 88, 2016.

[14] GDPR. GDPR art. 9 processing of special categories of personal data. [Online]. Available: https://gdpr-info.eu/art-9-gdpr/

[15] I. Consulting. "Art. 22 gdpr automated individual decision-making, including profiling". [Online]. Available: https://gdpr-info.eu/art-22-gdpr/

[16] A. D. Kounoudes and G. M. Kapitsaki, “A mapping of iot user-centric privacy preserving approaches to the gdpr,” *Internet of Things*, vol. 11, no. 100179, 2020.

[17] FIT Frauernhofer, “Besserer datenschutz in smart homes: 30 test-haushalte gesucht.” [Online]. Available: https://idw-online.de/de/news814774

[18] Bundesministerium für Wirtschaft und Klimaschutz (BMWK). Internet der dinge. [Online]. Available: https://www.bmwk.de/Redaktion/DE/Artikel/Digitale-Welt/internet-der-dinge.html

[19] I. G. S. Intiative, “Internet of things global standards initiative,” 2015. [Online]. Available: https://www.itu.int/en/ITU-T/gsi/iot/Pages/default.aspx

[20] M. D. Firoozjaei, R. Lu, and A. A. Ghorbani, “An evaluation framework for privacy-preserving solutions applicable for blockchain-based internet-of-things platforms,” *Security and Privacy*, vol. 3, no. 6, pp. 1 – 28, 2020.

[21] Statista. "smart home weltweit". [Online]. Available: https://de.statista.com/outlook/dmo/smart-home/weltweit

[22] J. Carretero and J. D. García, “The internet of things: Connecting the world,” *Personal and Ubiquitous Computing,*, vol. 18, no. 2, pp. 445–447, 2014.

[23] Y. B. Hamdan *et al.*, “Smart home environment future challenges and issues-a survey,” *Journal of Electronics*, vol. 3, no. 01, pp. 239–246, 2021.

[24] M. Barhamgi, M. Yang, C.-M. Yu, Y. Yu, A. Bandara, D. Benslimane, and B. Nuseibeh, “Poster: Enabling end-users to protect their privacy,” in *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security (ASIA CCS ’17)*, pp. 905–907, 2017.

[25] X. Huang, R. Fu, B. Chen, T. Zhang, and A. Roscoe, “User interactive internet of things privacy preserved access control,” *International Conference for Internet Technology and Secured Transactions,*, pp. 597–602, 2012.

[26] ISO, “Iso/iec 30118-1:2021 information technology - open connectivity foundation (ocf) specification -part 1: Core specification.” [Online]. Available: https://www.iso.org/standard/82127.html

[27] O. C. Foundation. Open Connectivity Foundation (OCF) Specification. [Online]. Available: https://openconnectivity.org/developer/specifications/

[28] S. K. N. Paulsen. (2021) "4 von 10 deutschen nutzen smart-home-anwendungen". [Online]. Available: https://www.bitkom.org/Presse/Presseinformation/4-von-10-Deutschen-nutzen-Smart-Home-Anwendungen

[29] K. M. N. Paulsen, “"43 prozent der deutschen nutzen smart-home-technologien".” [Online]. Available: https://www.bitkom.org/Presse/Presseinformation/Smart-Home-2022

[30] P. Krajewski. "die besten und beliebtesten smart home-anbieter". [Online]. Available: https://www.chip.de/artikel/CHIP-Umfrage-Die-gefragtesten-SmartHome-Produkte_182849918.html

[31] Statista. Besitz von smart-home-geräten in deutschland im jahr 2022. [Online]. Available: https://de.statista.com/prognosen/999788/deutschland-besitz-von-smart-home-geraeten

[32] J. Webster and R. T. Watson, “Analyzing the past to prepare for the future: Writing a literature review,” *MIS quarterly*, pp. xiii–xxiii, 2002.

[33] S. H. Hussain, S. Geetha, and M. A. Prabhakar, “Design and implementation of an adaptive model for sustainable home automation using internet of things (iot),” *International Journal of Advanced Engineering Technology*, vol. VII, no. 1, pp. 827–829, 2016.

[34] Y. J. Jia, Q. A. Chen, S. Wang, A. Rahmati, E. Fernandes, Z. M. Mao, A. Prakash, and S. J. Unviersity, “Contexlot: Towards providing contextual integrity to appified iot

platforms," *Network and Distributed System Security Symposium (NDSS),*, pp. 1–15, 2017.

[35] C. Perera, C. McCormick, A. K. Bandara, B. A. Price, and B. Nuseibeh, "Privacy-by-design framework for assessing internet of things applications and platforms," *In Proceedings of the 6th International Conference on the Internet of Things (IoT'16),*, pp. 83–92, 2016.

[36] T. Yu, V. Sekar, S. Seshan, Y. Agarwal, and C. Xu, "Handling a trillion (unfixable) flaws on a billion devices: Rethinking network security for the internet-of-things," in *Proceedings of the 14th ACM Workshop on Hot Topics in Networks*, no. 5, pp. 1–7, 2015.

[37] M. Barhamgi, M. Yang, C.-M. Yu, Y. Yu, A. K. Bandara, D. Benslimane, and B. Nuseibeh, "Enabling end-users to protect their privacy," in *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security,*, pp. 905–907, 2017.

[38] A. Chakravorty, T. Wlodarczyk, and C. Rong, "Privacy preserving data analytics for smart homes," in *Proceedings of 2013 IEEE Security and Privacy Workshops,*, pp. 23–27, 2013.

[39] X. Huang, P. Craig, H. Lin, and Z. Yan, "Seciot: A security framework for the internet of things," *Security and Communication Networks*, vol. 9, no. 16, pp. 3083–3094, 2016.

[40] A. Ouaddah, A. Abou Elkalam, and A. Ait Ouahman, "Fairaccess: A new blockchain-based access control framework for the internet of things," *Security and Communication Networks*, vol. 9, no. 18, pp. 5943–5964, 2016.

[41] M. Abu-Tair, S. Djahel, P. Perry, B. Scotney, U. Zia, J. M. Carracedo, and A. Sajjad, "Towards secure and privacy-preserving iot enabled smart home: Architecture and experimental study," *Sensors*, vol. 20, no. 21, pp. 1–14, 2020.

[42] K. Alanezi and S. Mishra, "Incorporating individual and group privacy preferences in the internet of things," *Journal of Ambient Intelligence and Humanized Computing,*, pp. 1–16, 2021.

[43] A. Alcaide, E. Palomar, J. Montero-Castillo, and A. Ribagorda, "Anonymous authentication for privacy-preserving iot target-driven applications," *Computers & Security*, vol. 37, pp. 111–123, 2013.

[44] H. Cao, S. Liu, Z. Guan, L. Wu, H. Deng, and X. Du, "An efficient privacy-preserving algorithm based on randomized response in iot-based smart grid," in *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People*

and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), pp. 881–886, 2018.

[45] J. Cao, B. Carminati, E. Ferrari, and K.-L. Tan, "Castle: Continuously anonymizing data streams," *IEEE Transactions on Dependable and Secure Computing*, vol. 8, no. 3, pp. 337–352, 2010.

[46] E. M. Chan, P. E. Lam, and J. C. Mitchell, "Understanding the challenges with medical data segmentation for privacy," in *Usenix Conference on Safety, Security, Privacy and Interoperability of Health Information Technologies*, pp. 1–10, 2013.

[47] L. Guo, M. Dong, K. Ota, Q. Li, T. Ye, J. Wu, and J. Li, "A secure mechanism for big data collection in large scale internet of vehicle," *IEEE Internet of Things Journal*, vol. 4, no. 2, pp. 601–610, 2017.

[48] X. Li, J. Niu, M. Z. A. Bhuiyan, F. Wu, M. Karuppiah, and S. Kumari, "A robust ecc-based provable secure authentication protocol with privacy preserving for industrial internet of things," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 8, pp. 3599–3609, 2017.

[49] J. Su, D. Cao, B. Zhao, X. Wang, and I. You, "epass: an expressive attribute-based signature scheme with privacy and an unforgeability guarantee for the internet of things," *Future Generation Computer Systems*, vol. 33, pp. 11–18, 2014.

[50] X. Wang, J. Zhang, E. M. Schooler, and M. Ion, "Performance evaluation of attribute-based encryption: Toward data privacy in the iot," *2014 IEEE International Conference on Communications (ICC)*, pp. 725–730, 2014.

[51] J.-C. Yang and B.-X. Fang, "Security model and key technologies for the internet of things," *The Journal of China Universities of Posts and Telecommunications*, vol. 18, pp. 109–112, 2011.

[52] L. Yang, A. Humayed, and F. Li, "A multi-cloud based privacy-preserving data publishing scheme for the internet of things," in *Proceedings of the 32nd Annual Conference on Computer Security Applications*, pp. 30–39, 2016.

[53] W. Yang, N. Li, Y. Qi, W. Qardaji, S. McLaughlin, and P. McDaniel, "Minimizing private data disclosures in the smart grid," in *Proceedings of the 2012 ACM conference on Computer and communications security*, pp. 415–427, 2012.

[54] M. Keshavarz and M. Anwar, "The automatic detection of sensitive data in smart homes," in *HCI for Cybersecurity, Privacy and Trust: First International Conference, HCI-CPT 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings 21*. Springer, pp. 404–416, 2019.

[55] A. Subahi and G. Theodorakopoulos, "Detecting iot user behavior and sensitive information in encrypted iot-app traffic," *Sensors*, vol. 19, no. 21, p. 4777, 2019.

[56] N. Sadeh, J. Hong, L. Cranor, I. Fette, P. Kelley, M. Prabaker, and J. Rao, "Understanding and capturing people's privacy policies in a mobile social networking application," *Personal and ubiquitous computing*, vol. 13, no. 6, pp. 401–412, 2009.

[57] L. Fang and K. LeFevre, "Privacy wizards for social networking sites," in *Proceedings of the 19th international conference on World wide web*, pp. 351–360, 2010.

[58] B. Knijnenburg and H. Jin, "The persuasive effect of privacy recommendations for location sharing services," *SSRN Electronic Journal*, vol. 2399725, 2013.

[59] J. Xie, B. P. Knijnenburg, and H. Jin, "Location sharing privacy preference: Analysis and personalized recommendation," in *Proceedings of the 19th international conference on Intelligent User Interfaces*, pp. 189–198, 2014.

[60] G. Pallapa, S. K. Das, M. Di Francesco, and T. Aura, "Adaptive and context-aware privacy preservation exploiting user interactions in smart environments," *Pervasive and Mobile Computing*, vol. 12, pp. 232–243, 2014.

[61] M. S. N. Khan, S. Marchel, S. Buchegger, and N. Asokan, "Chowniot: Enhancingi iot by automated handling of ownershio change," *Privacy and Identity Management. Fairness, Accountability, and Transparency in the Age of Big Data (13th IFIP)*, pp. 1–15, 2018.

[62] V. Zimmermann, E. Dickhaut, P. Gerber, and J. Vogt, "Vision: Shining light on smart homes - supporting informed decision-making of end users," in *Proceedings of 2019 IEEE European Symposium on Security and Privacy Workshops (EuroS PW)*, pp. 149–153, 2019.

[63] S. Zheng, N. Apthorpe, M. Chetty, and N. Feamster, "User perceptions of smart home iot privacy," in *Proceedings of the ACM on Human-Computer Interaction*, vol. 2, no. CSCW. ACM, pp. 200, 2018.

[64] K. Martin and H. Nissenbaum, "Measuring privacy: an empirical test using context to expose confounding variables," *Columbia Science and Technology Law Review*, vol. 18, p. 176, 2016.

[65] N. Apthorpe, Y. Shvartzshnaider, A. Mathur, D. Reisman, and N. Feamster, "Discovering smart home internet of things privacy norms using contextual integrity," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 2, Article 59, pp. 1–23, 2018.

[66] H. Lee and A. Kobsa, "Understanding user privacy in internet of things environments," in *2016 IEEE 3rd World Forum on Internet of Things (WF-IoT).* IEEE, pp. 407–412, 2016.

[67] P. E. Naeini, S. Bhagavatula, H. Habib, M. Degeling, L. Bauer, L. F. Cranor, and N. Sadeh, "Privacy expectations and preferences in an iot world," in *Thirteenth Symposium on Usable Privacy and Security (SOUPS)*, pp. 399–412, 2017.

[68] E. McReynolds, S. Hubbard, T. Lau, A. Saraf, M. Cakmak, and F. Roesner, "Toys that listen: A study of parents, children, and internet-connected toys," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 5197–5207, 2017.

[69] E. S. Udoh and A. Alkharashi, "Privacy risk awareness and the behavior of smartwatch users: A case study of indiana university students," in *2016 Future Technologies Conference (FTC)*, pp. 926–931, 2016.

[70] V. Zimmermann, M. Bennighof, M. Edel, O. Hoffmann, J. Jung, and M. Wick, "'home, smart home' -exploring end users' mental models of smart homes," *Mensch und Computer 2018-Workshopband,*, pp. 401 – 417, 2018.

[71] G. Milne, G. Pettinico, F. Hajjat, and E. Markos, "Information sensitivity typology: Mapping the degree and type of risk consumers perceive in personal data sharing," *Journal of Consumer Affairs*, vol. 51, no. 1, pp. 133–161, 2016.

[72] J. Rumbold and B. Pierscionek, "What are data? a categorization of the data sensitivity spectrum," *Big Data Research*, vol. 12, pp. 49 – 59, 2018.

[73] M. N. Al-Ameen, T. Tamanna, S. Nandy, M. M. Ahsan, P. Chandra, and S. I. Ahmed, "We don't give a second thought before providing our information: Understanding users' perceptions of information collection by apps in urban bangladesh," in *Proceedings of the 3rd ACM SIGCAS Conference on Computing and Sustainable Societies*, pp. 32–43, 2020.

[74] K. Fietkiewicz and A. Ilhan, "Fitness tracking technologies: Data privacy doesn't matter? the (un)concerns of users, former users, and non-users," in *Proceedings of the 53rd Hawaii International Conference on System Sciences*, pp. 1– 10, 2020.

[75] O. Kulyk, B. Reinheimer, L. Aldag, P. Mayer, N. Gerber, and M. Volkamer, "Security and privacy awareness in smart environments–a cross-country investigation," in *International Conference on Financial Cryptography and Data Security*, pp. 84–101, 2020.

[76] E.-M. Schomakers, C. Lidynia, D. Müllmann, and M. Ziefle, “Internet users’ perceptions of information sensitivity - insights from germany,” *International Journal of Information Management*, vol. 46, pp. 142–150, 2019.

[77] A. Balapour, H. R. Nikkhah, and R. Sabherwal, “Mobile application security: Role of perceived privacy as the predictor of security perceptions,” *International Journal of Information Management*, vol. 52, no. 102063, 2020.

[78] N. Guhr, O. Werth, P. P. H. Blacha, and M. H. Breitner, “Privacy concerns in the smart home context,” *SN Applied Sciences*, vol. 2, no. 2, pp. pp. 1–12, 2020.

[79] M. Jozani, E. Ayaburi, M. Ko, and K.-K. R. Choo, “Privacy concerns and benefits of engagement with social media-enabled apps: A privacy calculus perspective,” *Computers in Human Behavior*, vol. 107, pp. pp. 106–260, 2020.

[80] K. B. Sheehan and M. G. Hoy, “Dimensions of privacy concern among online consumers,” *Journal of Public Policy & Marketing*, vol. 19, no. 1, pp. pp. 62–73, 2000.

[81] C. I. Wickramasinghe, “Best-practice-based framework for user-centric privacy-preserving solutions in smart home environments,” in *International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services (MobiQuitous)*, pp. 101–120, 2022.

[82] "zvei: Jeder fünfte deutsche nutzt smart-home-anwendungen". [Online]. Available: https://www.baulinks.de/webplugin/2018/0470.php4

[83] M. Brandt, “So smart sind deutschlands haushalte,” 2021. [Online]. Available: https://de.statista.com/infografik/3105/anzahl-der-smart-home-haushalte-in-deutschland/

[84] S. K. N. Paulsen. "4 von 10 deutschen nutzen smart-home-anwendungen". [Online]. Available: https://www.bitkom.org/Presse/Presseinformation/4-von-10-Deutschen-nutzen-Smart-Home-Anwendungen

[85] Ecovacs. "deebot omni robot". [Online]. Available: https://www.ecovacs.com/de

[86] S. Finkel, “Überwachungskamera: Testsieger mit solar bei stiftung warentest.” [Online]. Available: https://www.chip.de/news/Solar-Ueberwachungskamera-Testsieger-bei-Stiftung-Warentest-jetzt-reduziert_184318262.html

[87] Reolink. "reolink argus 3 pro smart camera". [Online]. Available: https://reolink.com/product/argus-3-pro/

[88] M. L. D. Nusser, “Smarte lampen test: 46 modelle im vergleich.” [Online]. Available: https://www.chip.de/artikel/Smarte-Lampen-Test-2022-46-Modelle-im-Vergleich_150456426.html

[89] Philips-hue.com, “A60 - smarte lampe e27 - 800.” [Online]. Available: https://www.philips-hue.com/de-de/p/hue-white---color-ambiance-doppelpack-e27/8719514328365#specifications

[90] Lampenwelt.de, “Philips hue whitecolor ambiance led e27 9w 1100lm.” [Online]. Available: https://www.lampenwelt.de/philips-hue-white-color-ambiance-led-e27-9w-1100lm.html#product-collateral

[91] P. Hue, “Philips hue datenschutzhinweise.” [Online]. Available: https://www.philips-hue.com/de-de/support/legal/privacy-policy

[92] C. Testcenter, “Die besten smarten heizkörperthermostate.” [Online]. Available: https://www.chip.de/bestenlisten/Bestenliste-Smarte-Heizkoerperthermostate--index/index/id/1450/

[93] Homematic, “Heizkörperthermostat - evo.” [Online]. Available: https://homematic-ip.com/de/produkt/heizkoerperthermostat-evo

[94] R. Voss and S. Finkel. "saugroboter-test: Die besten staubsaugerroboter". [Online]. Available: https://www.chip.de/artikel/Staubsaugerroboter-Test-Die-besten-Sauger-im-Vergleich_113901610.html

[95] B. Kolbow-Lehradt, “Homekit im test: Mit apple 2021 das smarthome steuern, kommandieren, automatisieren.” [Online]. Available: https://t3n.de/news/apple-homekit-im-test-1185589/

[96] A. Inc., “Homekit-datensicherheit.” [Online]. Available: https://support.apple.com/de-de/guide/security/sec49613249e/web

[97] M. Wendel. "die 5 beliebtesten sprachassistenten im überblick". [Online]. Available: https://www.homeandsmart.de/smart-home-sprachassistenten

[98] A. Amazon. "alexa, entwickelt, um deine daten zu schützen". [Online]. Available: https://www.amazon.de/Datenschutzportal-fÃijr-Alexa/b?ie=UTF8&node=17084415031

[99] J. Cohen, “Amazon’s alexa collects more of your data than any other smart assistant.” [Online]. Available: https://uk.pcmag.com/smart-home/136455/amazons-alexa-collects-more-of-your-data-than-any-other-smart-assistant

[100] BMWK: Bundesministerium für Wirtschaft und Klimaschutz, “Energieverbrauchskennzeichnung von neuen pkw.” [Online]. Available: https://www.bmwk.de/Redaktion/DE/Artikel/Energie/energieverbrauchskennzeichnung-von-pkw.html

[101] Verbraucherzentrale, “Energie-labels: eine übersicht,” 04 2023. [Online]. Available: https://www.verbraucherzentrale.de/wissen/umwelt-haushalt/nachhaltigkeit/energielabels-eine-uebersicht-5751

[102] Bundesministerium für Wirtschaft und Klimaschutz (BMWK). Nutri-score. [Online]. Available: https://www.bmel.de/DE/themen/ernaehrung/lebensmittel-kennzeichnung/freiwillige-angaben-und-label/nutri-score/nutri-score_node.html

[103] C. Park, Y. Kim, and M. Jeong, “Influencing factors on risk perception of iot-based home energy management services,” *Telematics and Informatics*, vol. 35, no. 8, pp. 2355–2365, 2018.

[104] J. Lin, W. Yu, N. Zhang, X. Yang, H. Zhang, and W. Zhao, “A survey on internet of things: Architecture, enabling technologies, security and privacy, and applications,” *IEEE Internet of Things Journal*, vol. 4, no. 5, pp. pp. 1125–1142, 2017.

[105] M. C. Oetzel and S. Spiekermann, “A systematic methodology for privacy impact assessments: A design science approach,” *European Journal of Information Systems*, vol. 23, no. 2, pp. pp. 126–150, 2014.

[106] S. Warentest. [Online]. Available: https://www.test.de/thema/

[107] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *"Experimentation in software engineering"*. Springer Science & Business Media, 2012.

[108] J. Erbel, “Scientific workflow execution using a dynamic runtime model,” 2022.

A. Appendix A: User-Centric Experiments 1 - Survey Questions [6]

Q_1: Have you heard about the Internet of Things (IoT)?
Possible answers: *Yes / No / I prefer not to answer this question*

Q_2: If yes, in which context have you heard about IoT?
Possible answers: *smart home / Industry 4.0 / smart factory / smart city / smart things / Others: free text box for participants / I prefer not to answer this question*

Q_3: Indicate the smart objects or applications (apps) that you know or use in your everyday life? (multiple choice possible)
Possible answers: *smart fridge / controlling home technology apps / smart grid apps / smart bulbs / smart alarm clock / smart toothbrush / smart washing machine / smart voice control objects, such as Amazon echo / augmented/virtual reality glasses / smart scale / smart health devices / smart door/window locks / smartphone / Others: free text box for participants / I do not utilize smart objects / I prefer not to answer this question*

Q_4: If you already use smart objects, how do you get access to the collected data from your smart objects, through an app or web interface? (multiple choice possible and please click on respective smart object to choose the option between app and web interface) Possible answers: *smart fridge / controlling home technology apps / smart grid apps / smart bulbs / smart alarm clock / smart toothbrush / smart washing machine / smart voice control objects, such as Amazon echo / augmented/virtual reality glasses / smart scale / smart health devices / smart door/window locks / smartphone / I prefer not to answer this question*

Q_5: How frequently do you use a device connected to the Internet, such as smart scale, fridge, wearables, watch, etc.? (smartphone, computer, smart TV does not count as smart devices in this question).
Possible answers: *more than 20 times per day / less than 20 times per day / once per day / very rare / other options: free text box for participants / I do not use any smart objects / I prefer not to answer this question*

Q_6: When using smart objects, I believe that those objects collect information about myself and my environment.

5-level Likert scale: *1 strongly disagree / 2 disagree / 3 partly agree / 4 agree / 5 strongly agree / Don't know / Prefer not to answer*

Q_7: I believe that I know the information collected by smart objects. 5-level Likert scale: *1 strongly disagree / 2 disagree / 3 partly agree / 4 agree / 5 strongly agree / Don't know / Prefer not to answer*

Q_8: Please indicate which kind of information you believe the smart objects collect about you. (multiple choice possible)
Possible answers: *location data / browsing data / data about your health, such as weight, movements, food purchase / personal data, e.g. bank details, relationships from voice control, finger prints from door locks etc. / other information: free text box for participants / I prefer not to answer this question*

Q_9: I believe that I know the parties who have access to collected data and receive the collected data from my smart objects. (Parties can be: hospital, doctors, insurance companies, institutes using data for statistics, etc.)
Possible answers: *I know / If you know, please mention in short key points the parties / I do not know / I prefer not to answer this question*

Q_{10}: I take special measures (such as switching off some services etc.) to protect my privacy when using smart objects.
Possible answers: *Yes / If yes, please indicate the measures you usually take and the conditions / No / I prefer not to answer this question*

Q_{11}: In a few years, I believe that it will be difficult to live without using smart objects.
5-level Likert scale: *1 strongly disagree / 2 disagree / 3 partly agree / 4 agree / 5 strongly agree / Don't know / Prefer not to answer*

Q_{12}: What do you think about the advantages you have by using smart home objects? (Participants had the possibility to rate on each statement by using the following Likert Scale.)
5-level Likert scale: *1 strongly disagree / 2 disagree / 3 partly agree / 4 agree / 5 strongly agree / Don't know / Prefer not to answer*

- facilitating to fulfill the everyday and routine tasks.
- high comfort and convenience
- low error rates (humans make mistakes more easily and frequently than machines)

- automatic adjustments of settings regarding my current lifestyle
- the smart objects record interesting information about myself and my surroundings.
- smart objects outline the optimization potentials regarding my everyday work or my health plan etc.
- specific things are automatically done by smart objects and releasing you from these tasks so you can spend time for more important things.
- no advantages
- Others: free text box for participants

Q_{13}: Please enter the answer for the following question to make sure, that you are not a robot: 150 + (2x2) =

Q_{14}: What do you know about privacy issues in the context of Internet of Things, specifically, to what extent do you understand potential privacy risks? (such as third parties get access to your data / to your house or to your bank account etc.).
Possible answers: *Please enter your answer here / I prefer not to answer this question*

Q_{15}: I believe that smart objects can endanger my privacy.
5-level Likert scale: *1 strongly disagree / 2 disagree / 3 partly agree / 4 agree / 5 strongly agree / Don't know / Prefer not to answer*

Q_{16}: Please enter your answer regarding the following statements. (Participants had the possibility to rate on each statement by using the following Likert Scale.)
5-level Likert scale: *1 strongly disagree / 2 disagree / 3 partly agree / 4 agree / 5 strongly agree / Don't know / Prefer not to answer*

- $Q_{16.1}$: I would like to have more information about the data collected by smart objects about me in a smart home environment.
- $Q_{16.2}$: I would like to have an overview of all the information collected by my smart objects.
- $Q_{16.3}$: I would like to be able to control which information is collected about myself in a smart home environment.
- $Q_{16.4}$: I would like to know in real-time about the data collected in my smart home environment.

- $Q_{16.5}$: I would like to have a summary of the collected data over a given period, e.g. daily, weekly, monthly, etc.
- $Q_{16.6}$: I would like to have more information about the associated risks to my privacy by sharing the collected data.
- $Q_{16.7}$: I would like to have more information about the associated personal and social advantages by sharing the collected data from my smart home environment.
- $Q_{16.8}$: I would like to be able to control the data shared by my smart objects.
- $Q_{16.9}$: I would like to be able to determine which information is used for which purpose.
- $Q_{16.10}$: I would like to be able to determine who is able to access my data.
- $Q_{16.11}$: I would be willing to spend time to audit the data collected about myself in a smart home environment.
- $Q_{16.12}$: I would prefer having an automated system taking privacy decisions for me after learning about my privacy risk awareness.
- $Q_{16.13}$: I would like to have clear policies with the provider regarding the collected data from my own smart home environment.

Q_{17}: Indicate the factors that motivate you to use smart home objects while having the control about the data collected.
Possible answers: *Feeling myself secured and protected / It is not possible for third party data consumers to draw a digital biography from my daily routine / Having control about the usage of collected data about me / Others: free text box for participants / I prefer not to answer this question*

Q_{18}: How old are you?

Q_{19}: What is your gender?
Possible answers: *Male / female / I prefer not to answer this question*

Q_{20}: What is your nationality?

Q_{21}: What is your annual income range? (Euro values or equivalent in local currency)?
Possible answers: *Less than 25.000 Euro / 25.000 Euro - 40.000 Euro / 40.000 Euro - 75.000 Euro / 75.000 Euro - 100.000 Euro / More than 100.000 Euro / I prefer not to answer this question*

Q_{22}: Please enter any comment, you would like to make regarding this survey. Please do not write any personal information in this field so that your answers remain anonymous.

B. Appendix B: User-Centric Experiments 2 - Survey Questions

Q_1: The following data types are collected in online services to which you have signed up and in your smart home environment that you control. Please indicate your sensitivity perception using the scale provided regardless of the usage purposes and whether the data are collected and processed locally or in a cloud. 5-level Likert scale: *1 Non Sensitive / 2 Less Sensitive / 3 Medium Sensitive / 4 Sensitive / 5 Highly Sensitive / I don't know / I don't know / Prefer not to answer*

Numeration	Statement
$Q_{1.1}$	The online service of the energy supplier collects energy consumption data of your heaters.
$Q_{1.2}$	The smart heaters collect energy consumption data of your heaters.
$Q_{1.3}$	The online purchasing service collects data on your purchasing habits.
$Q_{1.4}$	The smart fridge collects data on your purchasing habits.
$Q_{1.5}$	The online medical service collects data on your weight.
$Q_{1.6}$	The smart scale collects data on your weight.
$Q_{1.7}$	The online medical service collects data how often you do sports.
$Q_{1.8}$	The smart watch collects data how often you do sports.
$Q_{1.9}$	The online banking service collects your face-ID.
$Q_{1.10}$	The smart door lock collects your face-ID.

Table B.1.: Statements from the survey questions: Q_1

Q_{2a}: The following data types are collected in your smart home environment that you control. Please indicate your perception of linkability of the presented data types regardless of the usage purposes and whether the data are collected and processed locally or in a cloud. The rated statements are presented in Tab.B.2. Possible answers: *1 Directly linkable to you / 2 Indirectly linkable to you / 3 Directly linkable to you / Prefer not to answer*

Q_{2b}: The following data types are collected in your smart home environment that you control. Please indicate your sensitivity perception using the scale provided regardless of the usage purposes and whether the data are collected and processed locally or in a cloud. The rated statements are presented in Tab.B.2. 5-level Likert scale: *1 Non Sensitive / 2 Less Sensitive / 3 Medium Sensitive / 4 Sensitive / 5 Highly Sensitive / I don't know / I don't know / Prefer not to answer*

Numeration	Statement
$Q_{2a.1}$ and $Q_{2b.1}$	Biometrical data, such as fingerprint or face IDs from your smart door lock.
$Q_{2a.2}$ and $Q_{2b.2}$	Voiceprint from your smart speakers.
$Q_{2a.3}$ and $Q_{2b.3}$	Physical measurements, such as weight and height from your smart scale.
$Q_{2a.4}$ and $Q_{2b.4}$	Private details such as home address from your smart watch.
$Q_{2a.5}$ and $Q_{2b.5}$	Purchasing habits from your smart fridge.
$Q_{2a.6}$ and $Q_{2b.6}$	Personal preferences, such as sexual preferences, political affiliation, and religion from your smart speaker.
$Q_{2a.7}$ and $Q_{2b.7}$	Diagnoses and medical history from your smart wearable.
$Q_{2a.8}$ and $Q_{2b.8}$	Information about your family and environments from your smart speaker.
$Q_{2a.9}$ and $Q_{2b.9}$	Lifestyle information, such as energy consumption data, sleeping habits, doing sports from your smart heaters and watch.
$Q_{2a.10}$ and $Q_{2b.10}$	Information about your availability at home from your smart smoke detectors.

Table B.2.: Statements from the survey questions: Q_{2a} and Q_{2b}

Q_3: The following data types are collected in mobile applications to which you have signed up and in your smart home environment that you control. Please indicate your sensitivity perception using the scale provided regardless of the usage purposes and whether the data are collected and processed locally or in a cloud. 5-level Likert scale: *1 Non Sensitive / 2 Less Sensitive / 3 Medium Sensitive / 4 Sensitive / 5 Highly Sensitive / I don't know / I don't know / Prefer not to answer*

Numeration	Statement
$Q_{3.1}$	The mobile banking app collects your fingerprint.
$Q_{3.2}$	The smart door lock collects your fingerprint.
$Q_{3.3}$	The mobile home security app collects information about your availability at home.
$Q_{3.4}$	The smart smoke detectors collect information about your availability at home.
$Q_{3.5}$	The mobile social media apps collect information about your private life, such as relationship status, age, name, address, details about family members. etc..
$Q_{3.6}$	The smart speaker collects information about your private life, such as relationship status, age, name, address, details about family members. etc..
$Q_{3.7}$	The mobile location app collects data on current location at home.
$Q_{3.8}$	The smart bulbs collect data on current location at home.
$Q_{3.9}$	The mobile sleeping app collects data on your sleeping habits.
$Q_{3.10}$	The smart bed sensors collect data on your sleeping habits.

Table B.3.: Statements from the survey questions: Q_3

Q_4: The following data types are collected in your smart home environment that you control. Please indicate your sensitivity perception using the scale provided regardless of whether the data are collected and processed locally or in a cloud. 5-level Likert scale: *1 Non Sensitive / 2 Less Sensitive / 3 Medium Sensitive / 4 Sensitive / 5 Highly Sensitive / I don't know / I don't know / Prefer not to answer*

Q_5: Please enter the answer for the following question to make sure, that you are not a robot 12 *1 – 0 Possible answers: *0 / 11 / 12*

Q_6: How old are you? Possible answers: *18 - 30 years / 31 - 45 years / >45 years*

Q_7: What is your gender? Possible answers: *Male / Female / Diverse*

Q_8: Please indicate, in which country have you resided the most during the past 15 years? Possible answers: *the list of the countries in the world was presented.*

Q_9: In your household, do you have any smart home objects? (e.g. smart door locks, bulbs, cameras, fridges, voice control, scale, etc.) Possible answers: *No, I don't want to get any smart home objects / No, I am thinking of getting smart home objects / Yes, I own one of the smart home objects / Yes, I own couple of smart home objects / I prefer not to answer*

Q_{10}: Which of the smart home objects do you already own or want to get? Possible answers: *I don't own / I plan to buy / I own / I prefer not to answer*

- Smart door locks
- Smart window sensors
- Smart fridge
- Smart bulbs
- Smart scale
- Smart voice control, such as amazon echo
- Smart cameras
- Smart heaters
- Smart washing machine
- Smart water tap sensor
- Smart health care wearables, such as blood pressure monitor
- Smart smoke detector
- Smart bed sensors
- Smart coffee machine
- Smart oven
- Smart cooker

Q_{11}: Please enter any comment you would like to make regarding this survey. Please do not write any personal information in this field so that your answers remain anonymous.

Numeration	Statement
$Q_{4.1}$	The data on your energy consumption from smart heaters are shared to create general statistics on energy consumption.
$Q_{4.2}$	The data on your energy consumption from smart heaters are shared to make individual energy offers to you.
$Q_{4.3}$	The data on your purchasing habits from smart fridge are shared to create general statistics on sales volumes.
$Q_{4.4}$	The data on your purchasing habits from smart fridge are shared to provide you with discount offers.
$Q_{4.5}$	The data on your physical measurements from smart scale are shared to create general statistics on physical measurements of the society.
$Q_{4.6}$	The data on your physical measurements from smart scale are shared to provide you with individualized lifestyle tips.
$Q_{4.7}$	The data on your availability at home from smart smoke detectors are shared to create general statistics about the society.
$Q_{4.8}$	The data on your availability at home from smart smoke detectors are shared to provide you with different security offers for your home.
$Q_{4.9}$	The data on your private life, such as age, relationship, name, from smart speaker are shared to create general statistics about the social distribution.
$Q_{4.10}$	The data on your private life, such as age, relationship, name, from smart speaker are shared to provide you with seasonal and individualized offers.
$Q_{4.11}$	The biometrical data (face-ID, fingerprint) from smart door lock are shared to run general statistics on whether similarities exist within the society.
$Q_{4.12}$	The biometrical data (face-ID, fingerprint) from smart door lock are shared to identify you at public places, such as airport, train station, etc..
$Q_{4.13}$	The data on your lifestyle, such as sporty movements, sleeping habits, from smart bulbs, bed sensors and watch are shared to create general statistics on how healthy people live.
$Q_{4.14}$	The data on your lifestyle, such as sporty movements, sleeping habits from smart bulbs, bed sensors and watch are shared to provide you with individual healthcare insurance offers.

Table B.4.: Statements from the survey questions: Q_4

C. Appendix C: Data Types of the Data Streams of the Evaluated Smart Home Devices

In the following, the data types of the data streams of the five considered devices from service-oriented smart home categories are presented.

Phase	Data Type
Purchase & unpack	Name Address
App installation	Email (Hue App) Password (Hue App) Country
Device setup	Name for Bulb Room of the Bulb Name for Bluetooth Account Email for Bluetooth Account
Advanced App setting	Phone Location Data Device Information Data about your sleeping rhythm Lightning according to the music Lightning scenes: sunset, concentration, reading, cinema mode, etc.
Setup voice assistants	Email (connection between Philips Hue and voice assistant) Password (connection between Philips Hue and voice assistant) Activated bulbs for smart home control (in Amazon Echo App)
Start using	Current status of the lights: on/off

Table C.1.: 18 collected data types by the PHILIPS smart bulb in the phases from Fig 5.1

Phase	Data Type
Purchase & unpack	Name Address
App installation	Email Password Device Information Room name of the thermostat Name of the thermostat Display alignment
Device setup	Temperature setting of the room (thermostat installed room) Wifi name Wifi password
Advanced App setting	Setting the preferred temperatures for specific days and times (Preferred heating profiles) Users' heating preferences
Setup voice assistants	Email (connection between Homematic thermostat and voice assistant) Password (connection between Homematic thermostat and voice assistant) Activated thermostats in corresponding rooms for smart home control (in Alexa App)
Start using	Current temperature of the thermostat in the corresponding room

Table C.2.: 17 collected data types by the Homematic smart heater thermostat in the phases from Fig 5.1

Phase	Data Type
Purchase & unpack	Name Address
App installation	Email Password Country Language
Device setup	Wifi name Wifi password Robot name
Advanced App setting	Home Map (room structure/area names) Cleaning schedule Cleaning protocol (date, time, areas, etc.) Video manager (detection of items, carpets at home) Pictures of home/people/furniture (by video manager) Voiceprint through live audio call functionality (with video manager) People and Furniture and other obstacles on the home map
Setup voice assistants	Email (connection between Ecovacs robot and voice assistant) Password (connection between Ecovacs robot and voice assistant)
Start using	Current status of the robot: cleaning/stationed

Table C.3.: 19 collected data types by the Ecovacs smart robot in the phases from Fig 5.1

Phase	Data Type
Purchase & unpack	Name Address
App installation	Apple ID Apple ID Password Overview & Name of smart devices at home Overview of the rooms (where the devices are placed)
Device setup	List of the devices in each room
Advanced App setting	Control behaviour of individual smart devices at home
Setup voice assistants	Voiceprint through voice-based device control
Start using	Current status of all the smart devices at home

Table C.4.: 10 collected data types by the Apple HomeKit smart hub in the phases from Fig 5.1

Phase	Data Type
Purchase & unpack	Name Address
App installation	Email (Amazon Account) Password (Amazon Account) Language Room names of the Amazon Echo
Device setup	Wifi name Wifi password Voice print for the control
Advanced App setting	Bluetooth speaker information (through connection with Amazon Echo)
Start using	Current conversation & commands Users' contacts Users' preferences

Table C.5.: 13 collected data types by the Amazon Echo smart voice assistant in the phases from Fig 5.1

D. Appendix D: Details regarding Inputs for Quality Check for Provider

In the following, the details of the quality check for providers regarding the five considered devices from service-oriented smart home categories are presented.

Device	Amount of Mandatory Data Types	Mandatory Data Types
Smart bulb: PHILIPS Hue White and Color	4 from 18	name, address, device information, current status of the lights: on/off
Smart heater thermostat: eQ-3 Homematic IP	7 from 17	name, address, device information, Temperature setting of the room (thermostat installed room), Wifi name and password, current temperature of the thermostat in the corresponding room
Smart robot: Ecovacs Deebot X1 Omni	5 from 19	name, address, Wifi name and password, current status of the robot: cleaning/stationed
Smart hub: Apple Homekit	6 from 10	name, address, Apple ID, Apple ID password, list of the devices in each room, current status of all the devices at home
Smart voice assistant: Amazon Echo	5 from 13	name, address, Wifi name and password, current conversation & commands

Table D.1.: Example of the specific mandatory data types for the five considered devices from Sec. 5 (derived based on [85, 89, 90, 93, 95, 96, 98, 99]).

Device	Amount of Optional Data Types	Optional Data Types
Smart bulb: Philips Hue White and Color	14 from 18	email and password (Hue App), country, name for bulb, room of the bulb, name and email for Bluetooth account, phone location data, data about your sleeping rhythm, lightning according to the music, lightning scenes: sunset, concentration, reading, cinema mode, etc., email and password (connection between Philips Hue and voice assistant), activated bulbs for smart home control (in Amazon Echo App)
Smart heater thermostat: eQ-3 Homematic IP	10 from 17	email, password, room name of the thermostat, name of the thermostat, display alignment, setting the preferred temperatures for specific days and times (preferred heating profiles), users' heating preferences, email and password (connection between Homematic thermostat and voice assistant), activated thermostats in corresponding rooms for smart home control (in Alexa App)
Smart robot: Ecovacs Deebot X1 Omni	14 from 19	email, password, country, language, robot name, home map (room structure/area names), cleaning schedule, cleaning protocol (date, time areas, etc.), video manager (detection of items, carpets at home), picture of home-/people/furniture (by video manager), voice print through live audio call functionality (with video manager), people and furniture and other obstacles on the home map, email and password (connection between Ecovacs robot and voice assistant)
Smart hub: Apple Homekit	4 from 10	overview & name of devices at home, overview of the rooms (where the devices are placed), control behaviour of individual devices at home, voiceprint through voice-based device control
Smart voice assistant: Amazon Echo	8 from 13	email and password (Amazon account), language, room names of the Amazon Echo, voice print for the control, Bluetooth speaker information (through connection with Amazon Echo), users' contacts and preferences

Table D.2.: Example of the specific optional data types for the five considered devices from Sec. 5 (derived based on [85, 89, 90, 93, 95, 96, 98, 99]).

E. Appendix E: Details regarding *Privacy Facts Label* Adjustments

In the following, the details behind the QR code on the adjusted *Privacy Facts Label* for the Reolink smart camera are presented.

Data Type	UCCP Applicability	Functionality Scope
C1: Name	*Minimization*	Main
C2: Address	*Minimization*	Main
C3: Email	*Minimization*	Optional
C4: Password	*Minimization*	Optional
C5: Country	-	Optional
C6: Wifi Name	-	Main
C7: Wifi Password	-	Main

Table E.1.: Example extract for the Reolink smart camera

F. Appendix F: Results of the Groupwise-analyses regarding Q_{2b}

In the following, we present the details regarding the results of the groupwise-analyses based on the Kruskal-Wallis H tests between user categories (Q_9), age groups (Q_8), considered countries (Q_6) and statements from Q_{2b}.

Statements	Significant Difference	Between non- & planning user	Between non- & less familiar user	Between non- & familiar user
$Q_{2b.1}$	0.0006834	-	✓	-
$Q_{2b.2}$	4.88e-05	-	✓	✓
$Q_{2b.3}$	6.709e-08	-	✓	✓
$Q_{2b.4}$	1.641e-06	✓	✓	✓
$Q_{2b.5}$	8.136e-07	✓	✓	✓
$Q_{2b.6}$	0.0005041	✓	-	✓
$Q_{2b.7}$	0.0003934	-	✓	✓
$Q_{2b.8}$	9.887e-06	-	✓	✓
$Q_{2b.9}$	1.868e-07	✓	✓	✓
$Q_{2b.10}$	6.312e-07	✓	✓	✓

Table F.1.: Results of the groupwise-analyses based on the Kruskal-Wallis H tests between user categories (Q_9) and the statements from Q_{2b} ("✓" = significant difference and "-" = no significant difference)

Statements	Significant Difference	Between 18 - 30 & <45 years	Between 30 - 45 years & <45 years
$Q_{2b.1}$	0.007261	✓	-
$Q_{2b.2}$	0.005775	✓	✓
$Q_{2b.3}$	0.00893	-	✓
$Q_{2b.4}$	0.04019	-	-
$Q_{2b.5}$	0.001771	✓	✓
$Q_{2b.6}$	0.03803	-	✓
$Q_{2b.7}$	-	-	-
$Q_{2b.8}$	0.005279	✓	✓
$Q_{2b.9}$	0.0002945	✓	✓
$Q_{2b.10}$	0.0004146	✓	✓

Table F.2.: Results of the groupwise-analyses based on the Kruskal-Wallis H tests between age groups (Q_8) and the statements from Q_{2b} ("✓" = significant difference and "-" = no significant difference)

Statements	Significant Difference	Between Germany & UK	Between UK & Switzerland	Between UK & Austria	Between Germany & Austria
$Q_{2b.1}$	-	-	-	-	-
$Q_{2b.2}$	-	-	-	-	-
$Q_{2b.3}$	0.01727	-	✓	-	-
$Q_{2b.4}$	0.002981	✓	-	✓	-
$Q_{2b.5}$	2.774e-06	✓	✓	✓	-
$Q_{2b.6}$	-	-	-	-	-
$Q_{2b.7}$	-	-	-	-	-
$Q_{2b.8}$	-	-	-	-	-
$Q_{2b.9}$	0.0001501	✓	✓	-	-
$Q_{2b.10}$	-	-	-	-	-

Table F.3.: Results of the groupwise-analyses based on the Kruskal-Wallis H tests between countries, Germany, UK, Switzerland and Austria (Q_6) and the statements from Q_{2b} ("✓" = significant difference and "-" = no significant difference)

G. Appendix G: Results of the Groupwise-analyses regarding Q_1 and Q_3

In the following, we present the details regarding the Results of the groupwise-analyses based on the Kruskal-Wallis H tests between user categories (Q_9), age groups (Q_8), considered countries (Q_6) and statements from Q_1 and Q_3.

Statements	Significant Difference	Between non- & planning user	Between non- & less familiar user	Between non- & familiar user
$Q_{1.1}$	0.0003232	✓	✓	✓
$Q_{1.2}$	3.008e-06	✓	✓	✓
$Q_{1.3}$	0.0003026	✓	-	✓
$Q_{1.4}$	0.004632	✓	-	-
$Q_{1.5}$	0.01268	✓	-	-
$Q_{1.6}$	0.003911	-	-	✓
$Q_{1.7}$	0.0001475	✓	-	✓
$Q_{1.8}$	9.398e-05	-	-	✓
$Q_{1.9}$	0.001033	✓	✓	-
$Q_{1.10}$	0.0004532	✓	-	✓

Table G.1.: Results of the groupwise-analyses based on the Kruskal-Wallis H tests between user categories (Q_9) and the statements from Q_1 ("✓" = significant difference and "-" = no significant difference)

Statements	Significant Difference	Between non- & planning user	Between non- & less familiar user	Between non- & familiar user
$Q_{3.1}$	-	-	-	-
$Q_{3.2}$	0.0001828	-	✓	✓
$Q_{3.3}$	2.948e-05	✓	✓	✓
$Q_{3.4}$	1.338e-05	✓	✓	✓
$Q_{3.5}$	0.001072	-	✓	✓
$Q_{3.6}$	2.834e-05	-	✓	✓
$Q_{3.7}$	7.698e-07	✓	✓	✓
$Q_{3.8}$	1.459e-08	-	✓	✓
$Q_{3.9}$	2.389e-06	✓	✓	✓
$Q_{3.10}$	2.82e-05	-	✓	✓

Table G.2.: Results of the groupwise-analyses based on the Kruskal-Wallis H tests between user categories (Q_9) and the statements from Q_3 ("✓" = significant difference and "-" = no significant difference)

Statements	Significant Difference	Between 18 - 30 & <45 years	Between 30 45 years & <45 years
$Q_{1.1}$	0.01775	✓	✓
$Q_{1.2}$	0.004434	✓	✓
$Q_{1.3}$	0.002386	✓	✓
$Q_{1.4}$	0.0005877	✓	✓
$Q_{1.5}$	0.02177	✓	-
$Q_{1.6}$	0.02728	✓	-
$Q_{1.7}$	0.009542	✓	-
$Q_{1.8}$	0.007757	✓	-
$Q_{1.9}$	0.004849	✓	-
$Q_{1.10}$	0.01025	✓	-

Table G.3.: Results of the groupwise-analyses based on the Kruskal-Wallis H tests between age groups (Q_8) and the statements from Q_1 ("✓" = significant difference and "-" = no significant difference)

Statements	Significant Difference	Between 18 - 30 & <45 years	Between 30 45 years & <45 years
$Q_{3.1}$	0.0001609	✓	✓
$Q_{3.2}$	0.005326	✓	-
$Q_{3.3}$	6.739e-05	✓	✓
$Q_{3.4}$	0.0009209	✓	✓
$Q_{3.5}$	0.02865	-	-
$Q_{3.6}$	0.04429	-	-
$Q_{3.7}$	0.2192	-	-
$Q_{3.8}$	0.001007	✓	✓
$Q_{3.9}$	5.319e-06	✓	✓
$Q_{3.10}$	0.002795	✓	-

Table G.4.: Results of the groupwise-analyses based on the Kruskal-Wallis H tests between age groups (Q_8) and the statements from Q_3 ("✓" = significant difference and "-" = no significant difference)

Statements	Significant Difference	Between Germany & UK	Between UK & Switzerland	Between UK & Austria	Between Germany & Austria
$Q_{1.1}$	-	-	-	-	-
$Q_{1.2}$	0.0001528	✓	✓	-	-
$Q_{1.3}$	1.226e-06	✓	✓	✓	-
$Q_{1.4}$	0.04165	-	-	-	-
$Q_{1.5}$	0.02264	✓	-	-	-
$Q_{1.6}$	0.04328	-	-	-	-
$Q_{1.7}$	-	-	-	-	-
$Q_{1.8}$	-	-	-	-	-
$Q_{1.9}$	5.478e-06	✓	✓	✓	-
$Q_{1.10}$	0.0006761	✓	-	✓	-

Table G.5.: Results of the groupwise-analyses based on the Kruskal-Wallis H tests between countries, Germany, UK, Switzerland and Austria (Q_6) and the statements from Q_1 ("✓" = significant difference and "-" = no significant difference)

Statements	Significant Difference	Between Germany & UK	Between UK & Switzerland	Between UK & Austria	Between Germany & Austria
$Q_{3.1}$	-	-	-	-	-
$Q_{3.2}$	-	-	-	-	-
$Q_{3.3}$	-	-	-	-	-
$Q_{3.4}$	-	-	-	-	-
$Q_{3.5}$	0.0226	-	-	-	✓
$Q_{3.6}$	-	-	-	-	-
$Q_{3.7}$	-	-	-	-	-
$Q_{3.8}$	-	-	-	-	-
$Q_{3.9}$	3.072e-05	✓	✓	✓	-
$Q_{3.10}$	0.0001868	✓	✓	✓	-

Table G.6.: Results of the groupwise-analyses based on the Kruskal-Wallis H tests between countries, Germany, UK, Switzerland and Austria (Q_6) and the statements from Q_3 ("✓" = significant difference and "-" = no significant difference)