TALLINN UNIVERSITY OF TECHNOLOGY
School of Business and Governance

Marina Kurokawa

Risk-based data classification of household energy consumption data from a data protection perspective

Master's thesis

Technology Governance and Digital Transformation Program

Supervisor: Dr. Alexandros Pazaitis
Co-supervisor: Dr. Chris Giotitsas

Tallinn 2023

I hereby declare that I have compiled the thesis independently
and all works, important standpoints and data by other authors
have been properly referenced and the same paper
has not been previously presented for grading.

The document length is 12,812 words from the introduction to the end of the conclusion.

Marina Kurokawa ……………………………
(date)

TABLE OF CONTENTS

ABSTRACT

Privacy is one of the fundamental human rights, and the concept of data protection was developed to protect personal privacy in a digitized society. With the development of Big Data and algorithms, data are not only direct sources of information but also the basis for interfaces integrating other types of information, including sensitive information. Research to develop methods for inference from electricity consumption data is progressing, and the ability to infer more kinds of information and the accuracy of inferences about such information can be expected to increase in the future as research progresses and is used in conjunction with other data sources. In addition, it has been pointed out that household energy consumption data can be useful for improving private and public services in several ways. On the other hand, energy consumption data sharing through data-sharing platforms is not active due to various problems, and sharing data based on the consent of data subjects without considering the characteristics of the data may lead to privacy violations.

With this in mind, how we protect personal identification for ethical and efficient use of household electricity consumption data is critical from a privacy perspective. However, research on household electricity consumption data from a privacy invasion perspective is scarce, except for research on technical system components. And even the GDPR, the EU's comprehensive data protection law, does not explicitly refer to electricity consumption data. This study aims to help understand household electricity consumption data from a privacy perspective and help to fill current research gaps in this area. In this paper, we have explored the recent arguments and challenges for understanding household electricity consumption data from a privacy risk perspective and proposed and applied a new framework to household electricity data.

Keywords: Data protection, privacy, GDPR, energy data, smart meter data

1. INTRODUCTION

1.1. Background

While the use of natural person data offers great societal benefits, it can also threaten individual privacy, making the balance between these two aspects one of the most important discussions related to Big Data (Garfinkel, 2015; Polonetsky & Tene, 2013). Combining data from different sources greatly increases the potential for data use (Boland et al., 2017), but sharing data with other parties can also pose the risk of data misuse (Nguyen et al., 2017). In terms of using Big Data from external sources in addition to their own data, the financial and medical industries are more advanced. These industries have already begun to use Big Data for experimental and real-world use cases. For example, disease prediction and prevention (M. Chen et al., 2017) and AI-assisted early cancer detection (Hunter et al., 2022) are good examples in medicine and healthcare. While using Big Data can help improve services and solve societal problems, we also need to consider privacy. In particular, personal information, referred to as "personal data" or "personally identifiable information," must be strictly protected based on the right to privacy (Voss, 2016). This need is also reflected in European Union (EU) legislation. In 2018, the EU introduced a new data protection regulation, the General Data Protection Regulation (GDPR), which strictly regulates personal data collection, use, and management.

In addition to the healthcare and financial industries, the energy industry is also expected to use energy-related data soon. The proliferation of smart meters and related technologies enables the collection of near real-time energy consumption data. The application of these data is expected to contribute not only to energy supply and demand optimization (Humayun et al., 2022) or energy optimization in individual households (Misra et al., 2020) but also to public benefits, such as tracking residents who have not evacuated in the event of a disaster (Oshima et al., 2022), police investigation (Murrill et al., 2012) and electricity fraud detection (Badr et al., 2023). In addition to these applications, other uses can be expected if companies can use energy data as part of Big Data with other data from various sources as well as data use in the financial or healthcare industries.

Although the potential for using energy data is promising, research on the use of energy data lags behind that of finance and healthcare. While research on risk-based handling and assessment methods considering the risk of data types is active in health-related data (Al-Fedaghi & Al-Azmi, 2012; Dash et al., 2019; Malgieri & Comandé, 2017), research from a similar perspective is scarce in the area of energy consumption data use. In terms of regulation, for example, the GDPR explicitly defines "data concerning health" but does not explicitly mention energy data. However, household energy consumption data are considered personal information and relate to individuals' private lives (Cuijpers & Koops, 2013). It is possible to provide information about their private lives, such as their routine activities and lifestyles (Muneeb et al., 2019) and some demographic information (Grünewald & Diakonova, 2019). As a way of sharing smart meter data with maintaining privacy, data-sharing platforms have been introduced in some countries, such as the United Kingdom (Webborn et al., 2019). However, they are little used due to high participation requirements and expensive transaction fees (Z. Chen et al., 2023; Webborn et al., 2019).

The background to the lack of active research and legislative consideration is that (1) the switch to smart meters has started relatively late in Europe (Kochański et al., 2020; Potdar et al., 2018), (2) the adoption rates of home energy management systems (HEMS) or smart energy homes are also low, (3) consumers perceive energy consumption data as relatively insensitive data than other data (Teng et al., 2022). However, these conditions have changed in recent years. The penetration rate of smart meters is 56 percent of electricity customers in the EU27+3 in 2022 (Berg Insight, 2022), following pioneers such as Sweden (completed by 2009) and Finland (completed by the end of 2013) (Tounquet & Alaton, 2019).

In addition, the recent energy crisis in Europe and societal demands in terms of sustainability have stimulated European interest in saving energy (Osborne Clarke, 2022). This situation would increase the type and amount of energy data available and the demand for it. This circumstance will also increase the use of residential energy data, but this has not yet been sufficiently discussed. In order to tackle skyrocketing energy costs, the European Commission launched the Digitization Action Plan in October 2022 (the European Commission, 2022b). This plan aims to drive the digitization of energy infrastructure in Europe, and it also mentions the need to protect the privacy of energy data and the possibility of sharing data, including with third parties (the European Commission, 2022). To make use of the data, the European Commission has organized several workshops on use cases of energy data sharing (the European

Commission, 2022b). However, most of the reported cases are related to renewable energy, nuclear power and other politically challenging issues, local energy, and energy optimization of industries and large buildings, and there is a lack of discussion on personal energy consumption (Gouriet et al., 2022; the European Commission, 2022a).

Based on these considerations, to contribute to the future use of household electricity data, we will explore possible privacy risks of household electricity data, and how we can understand household electricity consumption data in the context of European regulations.

1.2. Research objectives and structure of the research

This research aims to understand household electricity consumption data from a privacy perspective and to develop a new classification framework for household electricity consumption data to contribute to the future across sector data use. To achieve this goal, we start with a literature review of relevant topics to provide an overview of the topics and the current discussions around them. We then review the concept of personal data and the currently proposed approaches to distinguish personal data from other types of data (i.e., not energy data, such as health data) and propose a new framework from a privacy perspective. We then organize the types of household electricity consumption data and explore how we can categorize them using the proposed framework. Therefore, two main outcomes are expected from this research: 1) the development of a new framework for classifying household electricity consumption data from a privacy perspective, and 2) the application of the framework to household flow electricity data to contribute to the understanding of household electricity consumption data and future data use. The discussion concludes with considerations of the results and recommendations for using household electricity consumption data.

1.3. Scope of the research

In this paper, we focus on household electricity consumption data[1], so industrial sector energy data and generation and transmission data are not included in the discussion. On the other hand, household electricity consumption data[1] are the target of analysis, so smart meter data[1], HEMS

[1] The definitions of these terms are explained in section 1.5.

data[1], and smart home energy consumption data[1] are also research subjects. The discussion of privacy in the context of energy data also includes technical aspects such as security and communication. However, this paper focuses on the risk of privacy invasion of data corresponding to personally identifiable information (PII)[1], and other aspects are out of scope. Therefore, this paper does not address technical methods of privacy protection. The subject of our study is household electricity consumption data (i.e., timestamped electricity consumption per unit of time). And datasets consisting of combinations of household electricity consumption data and other data types are not the study's subject. Since the research targets data to use in the EU, we focus on regulations in the EU, and regulations in other areas and country-specific regulations in individual EU countries are not considered.

1.4. Research Questions

Therefore, the main research question of this paper is:

- What types of household electricity consumption data can be used within and across sectors while maintaining privacy?

And sub-questions for supporting the main questions are:

- What are the privacy risk factors in the current discussion on household electricity consumption data from a data protection perspective?

- How can household electricity consumption data be classified in terms of data protection and risk?
- What types of household electricity consumption data can be personal data?

1.5. Definitions of terms

Here, we define the terms we use in this paper to avoid misunderstanding and maintain consistency of terms.

Personal information: There is no globally agreed definition of the term "personal information" (Schwartz & Solove, 2014). In this paper, we define the term following Garfinkel's (2015) definition as “any information relating to an individual” (Appendix. 1).

Personally identifiable information (PII) and ***Personal data***: Personally identifiable information (PII) or personal data is personal information that identifies or identifiable a natural person. These two terms are often used as synonyms, but "PII" is a more general term and is also used in international standards such as the International Organization for Standardization (ISO) and some countries such as the United States (Boyne, 2018; Garfinkel, 2015). And "personal data" is the term used in EU regulations such as the GDPR. Whether the two terms indicate similar meanings depends on the definition of PII (Garfinkel, 2015). Because PII is used as a unique noun with its own definitions as well as a general term, the term PII may refer to a different scope of data depending on where it is used (Goddard, 2017; Schwartz & Solove, 2014). Therefore, when we use the term "PII" in this paper, it is a general term that only means "information which directly refers to a specific natural person or from which it may be possible to infer a specific natural person" and is not linked to definitions in specific laws. And the term "personal data" is used when we talk about EU regulations, and it follows the EU definition (explained later in Section 3.2.2). In this paper, we use these terms as we have defined them here, taking into account the definitions and meaning of the term in previous research, regardless of the words used in the original papers (Unless a special annotation is attached). Therefore, please note that the term may not be the same as the term used in an original document.

Pseudonymization, ***anonymization, de-identification***, and ***re-identification***: the two terms: ”pseudonymization” and “anonymization” refer to data processing techniques in the context of data protection (Finck & Pallas, 2020). In this paper, we define the terms with reference to the definitions of Garfinkel (2015) and the definitions in GDPR as follows:

- ***anonymization (verb: anonymize)***: the process of completely removing any link between the data and the data subject, where the link cannot be recovered once removed.
- ***pseudonymization (verb: pseudonymize)***: the process of removing any link between the data and the data subject, whereby the data cannot be linked to the data subject as a result of the process. However, through the use of additional data, the data may identify or make identifiable the data subject.
- ***de-identification (verb: de-identify)***: the process of removing any links between the data and the data subject. The term refers to both anonymization and pseudonymization.

- ***re-identification (verb: re-identify)***: the process of restoring identifiability to de-identified data. It involves re-linking data and data subjects and attempting to identify individuals from de-identified data.

Electricity generator, ***distributor***, and ***retailer***: "Electric utilities" have these three functions. Traditionally, a single electricity company has held all three functions, but today each function is spun off into a separate company, with different companies performing different functions (Knezović et al., 2015).

- ***Electricity generator***: an electricity generator is a company that produces electricity. It has a power plant and sells and delivers the electricity it generates. This type of company is beyond the scope of this paper and will not be discussed in detail.

- ***Electricity distributor***: An ***electricity distributor*** or ***distribution system operator (DSO)*** is a company that manages the distribution of electricity and delivers electricity from generators to end users in a given geographical area (European Distribution System Operators for Smart Grids (EDSO), 2023). In this paper, we use the term distribution system operator (DSO), which is commonly used in the European Union, as we focus on the European system.

- ***Electricity retailer***: An electricity retailer is a company that provides electricity services to end consumers. They enter into electricity supply contracts with end users, issue bills, accept payments, provide customer support, etc. (Tushar et al., 2021).

Household electricity consumption data: We define the term as data related to electricity consumption in a household. In this paper, this data includes 1) smart/analog meters installed by utility companies and 2) data from a home energy management system (HEMS).

Smart meter: There is also no internationally agreed definition of the term "smart meter" (Koponen et al., 2008), but in general, the term "smart meter" is used to refer to a meter for measuring and transmitting energy consumption data (electricity, gas, water) and energy supply-related data in near real-time (Alahakoon & Yu, 2013). Technically, a "smart meter" is also a component of a HEMS or smart device that performs the functions mentioned above. However, to distinguish it from other data, in this paper, we refer to electricity meters with the functions explained above, which are installed in households by utility companies and are primarily used for reading and billing.

Home energy management system (HEMS): Home energy management system or residential energy management system is a system that allows consumers to measure, monitor, record, control, and help optimize energy use in their homes (Adeli & Hedman, 2020). HEMS can be offered as a standalone service or as part of smart home functionality (Adeli & Hedman, 2020), and this paper will address both cases.

HVAC (heating, ventilation, and air conditioning): A general term for air conditioning systems and equipment that control the temperature, humidity, and air quality of an indoor space, including heating and cooling equipment.

Near real-time: the term "near real-time" refers to a slight time delay from real-time, which refers to the network transmission or data processing system (Alliance for Telecommunications Industry Solutions, 2001; Collins Dictionaries, 2023). For example, when sensors detect events, there is a small time delay between the "occurrence," the "detection," and the "notification of the occurrence. " The data is processed in near real-time but has this small time delay, which is referred to as "near real-time data."

Load disaggregation and ***Non-Intrusive Load Monitoring (NILM)***: Load disaggregation is the decomposition of data from dwelling unit smart meters into the type of equipment used in the household by applying techniques such as machine learning and statistics (Angelis et al., 2022). NILM is the technique used for load disaggregation. Current research shows that applying NILM to smart meter data at intervals of 15 minutes or less can determine the type of appliances used in the home with high accuracy (Teng et al., 2022).

2. RESEARCH METHODS

Two main research methods are used in this study: Literature review and desk-based research. As mentioned in the research objectives section, this research aims to achieve two main outcomes: a framework for classifying household electricity consumption data from a privacy perspective and a classification of household electricity consumption data based on this framework. This paper consists of three main parts: 1) reviewing the literature on related topics, 2) building a data classification framework, and 3) categorizing the types of household electricity consumption data and applying the framework to them. The methods and data sources used in each part are explained below.

1) Reviewing the literature on related topics

First, we conduct a literature review on the relevant topics. The purpose of this part is to provide baseline knowledge and organize the discussion points and current research gaps. The literature reviews we conduct in this section are on the following topics:

- Privacy and data protection
- Personally identifiable information (PII)
- Data protection and related regulations in the EU
- Systems related to household electricity consumption data (electricity metering and HEMS)
- Privacy discussion around household electricity consumption data

2) Building a data classification framework

We conduct a literature review on methods for classifying or distinguishing data from a privacy perspective. We also conduct desk-based research on the legal requirements in the EU area and review the literature on it. Then, we review the methods proposed in the previous research and develop a new framework for classifying household electricity consumption data from the data protection perspective.

3) Categorizing the types of household electricity consumption data and applying the framework to them

To build the list of types of household electricity data, we first work on collecting information about the types of household electricity data. In this section, we mainly collect primary data. To collect information on smart meters, in addition to a literature review, we obtained current information on smart meters for households in each EU country through desk research. First, we collected information from the current version of the report on the deployment and progress of smart meters in the EU published by the European Commission in 2019 (Tounquet & Alaton, 2019). To update the information, we also review each EU member state's smart meter data requirements and determine if the requirements have changed. The list of national regulators in the EU (Appendix. 2) was taken from a report (Tounquet & Alaton, 2019) examining the deployment and progress of smart meters in the EU. Since more than half of the EU countries did not meet the original timeline for switching from analog meters or first-generation smart meters (i.e., smart meters that do not meet the European Commission's minimum requirements) to smart meters (Vitiello et al., 2022), we also collected information from the report previously published by the European Commission (Tounquet & Alaton, 2019).

To collect the data type of HEMS, we used a dataset published by Pritoni et al. (2018) as primary data. This dataset is collected data from a survey of purchasable residential management devices, and all data collected in the survey are available in the tabular form of an Excel file. Note the data set was originally established with the purpose of sharing raw data from 308 residential energy management devices/systems available in the United States (Pritoni et al., 2018). However, we found that the dataset also covers the major service providers in European markets, and there is no difference between markets regarding the data type. Therefore, we believe the data can be used as input for our research. In addition to this survey data, we also refer to studies published by the researchers who worked on this survey (Ford et al., 2017; Pritoni et al., 2018). After completing the above procedures, we apply the framework to the list of household electricity consumption data types that we created.

Having completed the above procedures, we apply the framework to the list of household electricity consumption data types. While we explain the organization of the data types in terms of individual identifiability using the framework, we also address which household energy consumption data require special consideration and how this might change in the future.

3. LITERATURE REVIEW

In this section, we provide readers with basic knowledge and current discussions on related topics by reviewing related research.

3.1 Privacy and Data Protection

3.1.1 Background of privacy and data protection

The right to privacy is one of the fundamental human rights and includes the right to respect the private life, private communications, and family life of the individual (Robertson, 1973). And it is defined in Article 12 of the Universal Declaration of Human Rights (The Universal Declaration of Human Rights, 1948). The idea of data protection derives from the right to privacy but is distinguished as two separate rights (De Hert, 2012). The right to data protection is the right to protect personally identifiable information from misuse and abuse (Bennett, 1992).

Solove (2000) looks back at the history of data-keeping in the United States and points out that privacy and data protection did not become a serious concern until modern times. According to the research, prior to the 19th century, personal information was kept in a small local community where everyone knew everyone else and kept information based on human memory, and it was mostly spread as gossip (Solove, 2000). Thus, the amount of data and its dissemination were limited, and the information retention period was relatively short. This situation changed in the 1960s and 1970s when public and private entities first began to digitize their records (Bygrave, 2010; Solove, 2008). Records in digital format and computer technologies made it easy to store enormous amounts of data, retain them for relatively long periods, and combine individual sources of information.

Bygrave (2010) analyzes the ideal concepts of privacy and data protection and their rationales of them by comparing their conceptualization in different countries. Bygrave (2010) shows that the conceptualization and emphasized aspects of privacy and data protection vary by country and region and depend on a complex context of societal factors such as culture, philosophies, emotional needs, as well as technological factors. Bennett & Raab (2017) note that the focus of privacy and data protection discussions and related concerns have varied over time, even in the same country. They also note that regulations and policies have changed over time to keep pace with each critical concern (Bennett & Raab, 2017). In addition, Poullet (2021) points out that current data protection laws, including the GDPR, were written to address current critical issues, given the level of technology at the time of drafting. Hence, data protection laws reflect the discussion in the country (Poullet, 2021).

3.1.2 Data protection in the EU

In the EU, the right to privacy is defined in Article 8 of the European Convention on Human Rights (European Convention on Human Rights, 1950). And data protection itself is recognized in constitutional instruments as a fundamental human right alongside the right to privacy in the EU (Bygrave, 2010; Charter of Fundamental Rights of the European Union, 2000).

As mentioned earlier (Section 3.1.1), the understanding of privacy and how personal data are attempted to be protected varies from country to country or region to region, and it is influenced by the different contexts of each country. Bygrave (2010) points out that the bureaucratic nature of data protection discussions and the comprehensive regulations on data protection are the characteristics of the privacy discussion in Europe and that they originate from the agony of past totalitarian oppression.

3.1.3 Regulation for data protection: GDPR

This tendency is also reflected in legislation, and the attempt to comprehensively regulate personal information is a feature of the European data protection discussion (Boyne, 2018; Bygrave, 2010; Park, 2019). The GDPR stems from data protection requirements in the EU (Poullet, 2021) and is a comprehensive data protection framework that applies to personal data (Erickson, 2018). The GDPR is directly applicable in all EU member states and European Economic Area countries without the need for a local law to come into force and therefore has the function of a basis of data protection regulation (Schellinger et al., 2022).

GDPR consists of 99 articles (legal requirements) and 173 recitals (supporting documents that provide additional information for understanding) (Wills, 2019). As the GDPR is a comprehensive regulation, it defines various points of data protection. Still, given the purpose of this research, we will focus on 1) the types of data to be protected, 2) data processing, 3) the legal basis for processing personal data, and 4) de-identification (anonymization and pseudonymization).

1) Types of data to be protected

The GDPR takes a binary approach and distinguishes between personal data and non-personal data (Finck & Pallas, 2020). And the GDPR defines that data protection is only applied to personal data, and non-personal data is not within the scope of the regulation (Finck & Pallas, 2020). In other words, the data that can be considered personal data need to be treated in accordance with the GDPR, while non-personal data are not subject to the requirements of the data protection regulation. Therefore, whether the data can be considered personal data or not is the first and one of the most critical assessment points in the GDPR and most data protection regulations in today's society (Finck & Pallas, 2020; Onik et al., 2018; Schwartz & Solove, 2014). To make the law work smoothly, the GDPR includes specific examples of personal data in addition to the definition of personal data in the articles (Article 4 of GDPR, 2016). However, it is not practical to specify all specific types of personal data in the legislation and revise it with state of the art (Voss, 2016) because, as we discussed earlier, privacy concepts and perceptions change depending on context and time (Bygrave, 2010), and technological developments also change whether data concern "identifiable" or "non-identifiable" individuals (Schwartz & Solove, 2011). As a result, most data protection laws, including the GDPR, now use the concept of personal data for data protection rather than listing specific types of data.

2) Data processing

The GDPR defines the term "processing" in Article 4 (2) and covers a wide range of procedures related to data operations. In the GDPR, the term "data processing" refers to any procedure carried out on data (Gazi, 2020), and some examples of processing are given, such as collecting, recording, storing, disclosing, disseminating, etc. (Article 4 (2) of the GDPR, 2016). Therefore, any organization or individual that processes personal data is considered a data controller or data processor and is subject to the GDPR (Gazi, 2020; Gil González & de Hert, 2019).

3) Legal basis for processing personal data

When data collectors want to process personal data, the GDPR requires a legal basis for the lawful processing of personal data (Article 5 (1) (a) of the GDPR, 2016). It defines six types of the legal basis for this:

1) Consent of the data subject
2) Necessity for the performance of a contract
3) Compliance with a legal obligation
4) Necessity for the protection of vital interests
5) For public interest
6) For the legitimate interest

If the case is not according to the nature of one of the individuals, it is prohibited that the processing personal data (Gil González & de Hert, 2019). As can be seen from these reasons, most of the legal grounds are about the necessity of the data processing or a contribution to the public interest, except the ground for obtaining consent from the data subjects. Therefore, in practice, the majority of data processing is based on consent (Suripeddi & Purandare, 2021).

4) De-identification (anonymization and pseudonymization)

The GDPR does not explicitly define the term "anonymization" explicitly, but the term "pseudonymization" is defined to explain de-identification (Article 4 (5) of the GDPR, 2016). The GDPR defines "pseudonymization" as follows:

> “the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person” (Article 4 (5) of the GDPR, 2016)

According to the GDPR, pseudonymized data is originally personal data but no longer contains personal data because the part that identifies a natural person has been removed (Stalla-Bourdillon & Knight, 2016). The term "pseudonymization" is used in the GDPR with almost the same meaning as the term "de-identification," which we defined in Section 1.5. As the definition in the GDPR shows, pseudonymized data also includes the case where individual data subjects can be distinguished by a pseudonymization key after personal data have been removed (Kotschy, 2016). In this case, the pseudonymization key is stored separately under strict control (Kotschy, 2016). This way of pseudonymization is commonly used, especially in research applications of health data where privacy protection is needed during processing, but the value of the data would be compromised if it were completely separated from the individual (Shuaib et al., 2021).

Furthermore, Recital 26 of the data should be treated as personal data if the de-identified data can be re-identified with little effort (Recital 26 of the GDPR, 2016). Therefore, there is no explicit definition of "anonymized" data (i.e., data where any link between the data and the data subject is completely removed and the link cannot be re-established) under the GDPR, as it is difficult to determine whether the link cannot be re-established or not (Stalla-Bourdillon & Knight, 2016). It is also stated that whether or not data can be re-identified is influenced by technological developments, implementation costs, motivation to do so, etc., and that these factors must be taken into account (Recital 26 of the GDPR, 2016).

3.1.4 EU data protection regulations and documents other than GDPR

The European Data Protection Board (EDPB), established on the basis of Article 68 of the GDPR (Etteldorf, 2019), also issues general guidance to understand the GDPR better. However, Finck & Pallas (2020) point out that there are contradictions between the recitals of the GDPR and the EDPB guidelines, which leads to confusion. Recital 26 of the GDPR (About treatment of de-identified data, detailed in Section 4.2) and the guidelines published by the Article 29 Working Party (the predecessor of the current EDPB) take different approaches to assess whether the de-identified data are personal data or not, and they recommend different methodologies (Finck & Pallas, 2020; Stalla-Bourdillon & Knight, 2016). This inconsistency is because this guidance was issued before the GDPR, and the GDPR recommends a different approach, but the EDPB has not yet updated the guidance for handling de-identified data (Peloquin et al., 2020).

There is another directive related to data protection for household energy in the EU, Directive (EU) 2019/944, which imposes some additional requirements on energy-related data, such as electricity consumption data must be provided transparently to customers (Lavrijssen et al., 2022). In contrast, with respect to personal data, it states that privacy and data protection measures for energy consumption data and related customer data should comply with EU data protection legislation and does not specify additional requirements (Directive (EU) 2019/944, 2019). Therefore, in this research, we treat the approach of Recital 26 as the currently recommended approach for distinguishing personal data in the EU.

3.2 Personal Identifiable Information (PII)

3.2.1 Idea and role of PII

As explained in the definitions of terms (Section 1.5), PII is data that uniquely identifies a natural person and requires special treatment in most developed countries (Gieser, 2015). The concept of personally identifiable information is linked to privacy and data protection, and data that can indicate an individual's private life are considered PII, regardless of whether the information is important or trivial (Al-Fedaghi & Al-Azmi, 2012).

In the current legal approach to data protection and privacy, the concept of personally identifiable information is at the center of the discussion (Schwartz & Solove, 2014). This is because most current data protection laws, including the GDPR, use the concept of PII (regardless of the term used) (Finck & Pallas, 2020; Schwartz & Solove, 2014), and its definition defines the scope and boundaries of whether data should be strictly protected by regulation (Schwartz & Solove, 2011).

3.2.2 Definition in GDPR: "personal data"

The GDPR defines "personal data" corresponding to PII in Article 4 (1) as follows:

> "'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person" (Article 4 (1) of the GDPR, 2016)

As we can see, it defines personal data as "information relating to an identified or identifiable natural person" (Article 4 (1) of the GDPR, 2016). And it also gives examples of personal data used with the phrase "in particular," identifiers such as names or data that may relate to a sensitive identity. Bleier et al. (2020) note that the GDPR renovated the definition of personal data by considering data relating to an indirectly identifiable natural person as personal data in the same way as directly identifiable personal data (Bleier et al., 2020). As can be seen from the definition, personal data is data concerning natural persons, whereas data concerning legal persons or dead persons are not personal data (Finck & Pallas, 2020).

In addition to this definition, the GDPR defines data with a high privacy risk in Article 9 as "special categories of data" whose processing requires special care and a high level of security.

Highly sensitive data are enumerated as “special categories of data” in Article 9 (1) of the GDPR, such as ethnic origin, political opinions, genetic or health-related data, and so on (Article 9 (1) of the GDPR, 2016).

3.2.3 Discussions and challenges in distinguishing between PII and non-PII

As noted earlier, not all types of PII are listed as examples, so data collectors must assess whether the data meet the criteria for PII. The discussion of determining PII is not just about a single type of data but also about a data set that consists of two or more data types. If the dataset contains PII, the dataset is also considered PII and must be treated in the same manner (Otjacques et al., 2007). Therefore, there are three main points of distinction between PII and non-PII.

1) Whether the single data type is an identifier or can be an identifier of a natural person

2) Whether the data set contains data types that identify or can potentially identify a natural person

3) Whether the data set can identify or potentially identify a natural person

The first argument concerns whether the data itself qualifies as personal data. This point is mainly discussed for data that can be identifiers, like names and identification numbers, but are not explicitly mentioned in the law (Onik et al., 2018). The second point concerns whether the dataset contains PII, so it relates to the first point. Mostly, it is about de-identified data and whether the degree of de-identification is sufficient to preserve anonymity (Finck & Pallas, 2020). The third point is about the possibility of inferring an individual from the dataset, regardless of whether the dataset contains PII or not. For example, one study found that 87% of individuals in the United States can be uniquely identified by combining only three pieces of information: Zip code, gender, and date of birth (Sweeney, 2000). In this case, the combination of these three data can be considered PII even though it is not explicitly defined in the regulation.

However, there is not yet a generally accepted method for distinguishing whether or not data is PII (Garfinkel, 2015; Schwartz & Solove, 2014). As the concept of data protection is influenced by various social factors and the level of technological development, it is not easy to establish static criteria; therefore, the distinction between personal and non-personal data is one of the main challenges and ongoing discussions in the field of data protection, and regulators and researchers are trying to develop and propose methods (Detailed in Section 4).

3.3 Energy-related data source in a household

3.3.1 Data from smart/analog meters installed by utility companies

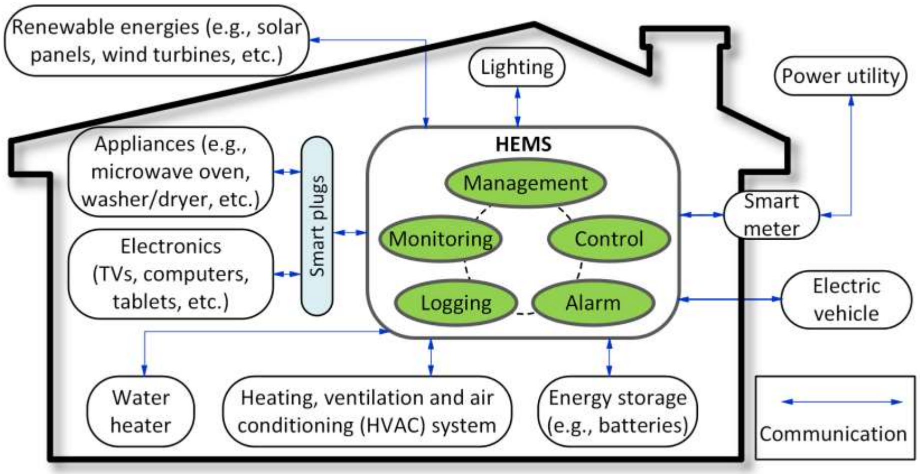
In order to pay for the electricity consumed in the household, users must submit their consumption data to the electricity retailers with whom they have a contract (King & Jessen, 2014). Since meters are installed to calculate energy bills, users can not opt out of this data collection (Ramokapane et al., 2022). Smart meters measure energy consumption and transmit this information to the electricity retailer without the need for an electricity company staff to be on site. The interval of measurement and transmission of information depends on the utility or the regulations of the country where the meter is operated (Tounquet & Alaton, 2019). It is typically between 15 minutes and one hour and is at most near real-time (Tounquet & Alaton, 2019). The electricity consumption data collected by the smart meter is the time-stamped electricity consumption (kWh) per defined interval (Pereira et al., 2022). Unlike HEMS, smart meters primarily measure and transmit household electricity consumption data but do not have functions such as device control. Because smart meters transmit more detailed data than analog meters, there is a view that opt-out options for smart meters should be offered (King & Jessen, 2014). In the Netherlands, for example, the right to opt out of smart meters (either switching off the radio frequencies of smart meters or using analog meters) has been enshrined in law (Geels et al., 2021).

Who collects the data on electricity consumption depends on the country and varies even within the EU area (Tounquet & Alaton, 2019). However, the electricity distribution system operators (DSO) or electricity retailers are responsible for or have a right to access the data in most countries (Tounquet & Alaton, 2019). The EU regulation requires utilities to ensure that the end customer has access to their smart meter data (Directive (EU) 2019/944, 2019). Thus, even if a DSO is a data collector, the electricity retailer is subject to the regulation as a data processor because it must process smart meter data for the purpose of sharing usage with users and billing (Tounquet & Alaton, 2019). Electricity retailers must collect and store their customers' personal information, including identifiers such as name, address, telephone numbers, etc., for their administrative and billing operations. The structure of data storage and linkage depends on the individual company, but generally, customer data and electricity consumption data are linked via ID (e.g., customer ID, contract number, smart meter ID, etc.) (Drkušić, 2017) (Appendix 3)

3.3.2 HEMS

As defined in Section 1.5, in this paper, we use the term HEMS to refer to both a standalone energy management system and a subset of smart home features related to energy management. HEMS, as the name suggests, is a system designed for home energy management, and when the term HEMS is used, it often includes the system itself and the services that run on HEMS. The idea of HEMS has a history of more than a hundred years (Adeli & Hedman, 2020), but the development of the current form of HEMS began in the 1970s (Lissa et al., 2021). From today's perspective, HEMS is a system that consists of both software and hardware and is designed to manage and control multiple functions (Kumar et al., 2019). HEMS has five key functions; monitoring, controlling, logging, managing, and alarming energy consumption and related activities (Alden et al., 2019). When the term is used in the context of smart grids, HEMS sometimes also refers to the functions of power generation (e.g., power generation by photovoltaic panels in the home) and power storage in addition to the main functions (Zafar et al., 2020). However, this research aims to understand household electricity consumption data, so we consider the five functions: "monitor," "control," "log," "alarm," and "manage" energy consumption in a home as HEMS core functions. In the current HEMS, these functions are realized and optimized based on machine learning; therefore, data and data processing play a crucial role (Mahapatra & Nayyar, 2022).

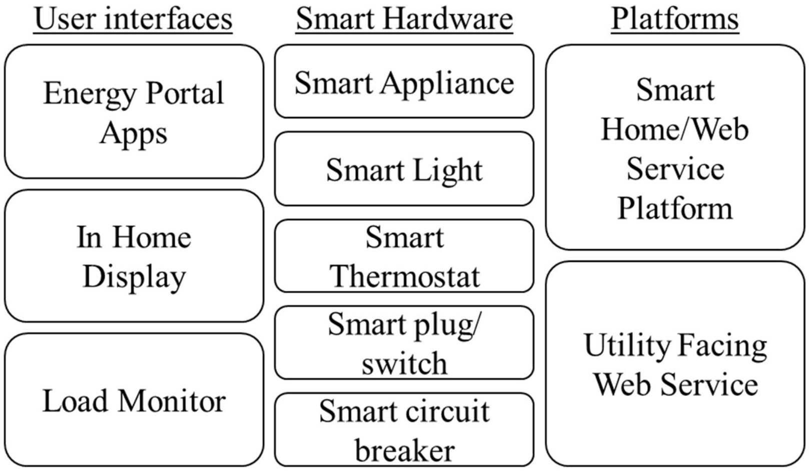
Figure 1. Schematic of a typical example of an architecture of HEMS


Renewable energies (e.g., solar panels, wind turbines, etc.)
Lighting
Power utility
Appliances (e.g., microwave oven, washer/dryer, etc.)
Smart plugs
HEMS
Management
Monitoring
Control
Logging
Alarm
Smart meter
Electronics (TVs, computers, tablets, etc.)
Electric vehicle
Water heater
Heating, ventilation and air conditioning (HVAC) system
Energy storage (e.g., batteries)
Communication

Source: Alden et al. (2020)

The types of devices and functions present in the HEMS system depend on the type of devices installed in the home and the services they use, but the major components of HEMS, which are potential data collection points, are shown in Figure 2.

Figure 2. Classification of possible components of HEMS



Source: Ford et al. (2017) and Alden et al. (2020) - edited by author

Note that we only show components that can serve as data collection points, so other components, such as network or security components, are not included in Figure 2. Here we give a brief overview of smart hardware (potential energy data collection points inside a house).

1) Smart appliance

Smart appliances are home appliances with software and IoT technologies that help users monitor, remotely control, or efficiently manage them by communicating with systems (LaMarche et al., 2021). Appliances with functions that meet the definition of smart appliances are mainly large appliances such as washing machines, refrigerators, dryers, and dishwashers (Ford et al., 2017). These devices collect data on the operation, status of products, energy consumption, etc., depending on the specific functions of each product (Pritoni et al., 2018).

2) Smart light

Smart lights contain sensors, microprocessors, and switches with remote control functions or relays in conventional lights, which can provide users with remote or automatic control. According to a study on commercialized smart devices, no smart lighting products measure energy consumption data, and these products tend to focus on the control function (Ford et al., 2017).

3) Smart thermostat

A smart thermostat is a product that monitors and controls energy consumption and enables energy savings by optimizing cooling and heating. The main components of smart thermostats are sensors, optimization algorithms, physical control devices for cooling and heating, and network communication functions (Ford et al., 2017). Smart thermostats contain sensors or/and are connected to sensors, and these sensors measure the data required for their algorithms, such as temperature, humidity, degree of brightness, occupancy, etc.(Pritoni et al., 2018). Users can also remotely control cooling and heating systems via smartphone apps. Some energy management systems use smart thermostats to predict occupants' future activities for the purpose of energy optimization (Ford et al., 2017).

4) Smart plug/smart switch

Smart plugs and smart switches are products that help non-smart devices work like smart devices (Alden et al., 2019). They are placed between an electrical outlet and a device. These products allow users to remotely control and monitor the status and energy consumption of the device (Ford et al., 2017).

4) Smart circuit breaker

Smart circuit breakers have almost the same function as smart plugs and smart switches, but the control and monitoring unit is per circuit breaker. They allow users to monitor power consumption based on branch circuits or some devices in the circuit and remote control based on circuit breakers (Alden et al., 2019).

Who processes the data in the HEMS service depends on the service providers and the service that the customer actually uses. For example, in the case of the traditional HEMS service, the manufacturer of the energy management system is also the service provider and collects the data on energy consumption in the home (Machorro-Cano et al., 2020; Nacer et al., 2017). In the case of a service provided by HEMS as part of the smart home, the smart home service provider is usually a data processor (Nacer et al., 2017). And the feature providers on the platform may also be data processors (Ford et al., 2017; Nacer et al., 2017). For example, Amazon Alexa, Google Home/Nest, and Apple HomeKit are popular smart home platforms that process data, including energy usage data (Kim et al., 2020). Smart appliance manufacturers, such as manufacturers of smart refrigerators, are typical examples of smart home function providers. They may be a data processor of household electricity consumption data, but whether they collect or perform other processing depends on the device and service (Ford et al., 2017). Smart home platformers and some smart home service providers, in particular, collect and use a variety of data in addition to energy consumption data (Kim et al., 2020; Zhang et al., 2018) (Appendix 4). Some HEMS and smart home apps allow users to voluntarily enter additional information, such as the number of occupants and the use and area of individual rooms, or collect other personal information, such as voice data in the case of voice control and the content of conversations (Ammari et al., 2019). As mentioned earlier, data collected from individual households is always personal data, which includes personal data such as the user's home address and account information (Machorro-Cano et al., 2020; Nacer et al., 2017). Again, the data processed by the actors on the platform will differ depending on the services and products chosen by the user, but other types of data may be stored together with non-energy consumption data or linked by IDs or other means (Drkušić, 2019; Nacer et al., 2017) (Appendix 5).

3.4 Privacy discussion around household electricity consumption data and preceding research

3.4.1 Privacy concerns and discussions about household electricity data

We present an overview of typical concerns and discussions around household electricity consumption data from the literature. First, there is a discussion about whether the measurement of electricity consumption itself raises the possibility of privacy abuses (Bugden & Stedman, 2019; Jakobi et al., 2019; Sovacool et al., 2017). This privacy discussion is especially argued for installing smart meters by electricity companies since electricity is a basic infrastructure in developed countries and is unavoidable in daily life (Singh et al., 2015). There are resistance movements from consumer groups against installing smart meters for households (Sovacool et al., 2017), and their main claims are privacy and health impact concerns (AlAbdulkarim et al., 2012; Chamaret et al., 2020). In Europe, some countries, such as Sweden, have completed the transition to smart meters (Tounquet & Alaton, 2019), while there is resistance in some countries that is delaying the deployment of smart meters (Chamaret et al., 2020; Cuijpers & Koops, 2013).

Second, there are discussions and concerns related to technical security issues in measuring and storing electricity consumption data. For example, some issues have been raised about communication security that allows outsiders to read data in the current system (Khattak et al., 2019), vulnerabilities in ICT products and services (Nwankwo et al., 2022), and vulnerabilities in data repositories that store measurement data (Asghar et al., 2017), but these technical issues are beyond the scope of this paper and will not be discussed in detail.

Third, the discussions on the use of household electricity data. McKenna et al. (2012) categorize the discussions in the literature on household electricity consumption data in terms of data use and summarize that there are five types of them. The five discussions of the use of household electricity consumption data are 1) unlawful use, 2) unexpected commercial use, 3) law enforcement use, 4) use by other parties for legal purposes, and 5) use by family members or roommates (McKenna et al., 2012).

3.4.2 Effect of Bigdata and Machine Learning

Electricity consumption data itself is simply data indicating the amount of electricity consumed in a given time period (e.g., kWh/h) with a time stamp. However, it may be possible to derive more personal information by looking at data continuously and applying research findings and methods. Teng et al. (2022) point out that the range of information that can be derived from electricity consumption data is expanding through the application of machine learning and Big Data techniques. Many empirical studies have shown that information about personal behavior and characteristics can be derived from energy data with some degree of accuracy. For example, information that can be derived from household electricity consumption data includes household characteristics such as marital status, number of residents, social class of a main income earner (Beckel et al., 2013; Grünewald & Diakonova, 2019), assets, and appliance ownership (AlAbdulkarim et al., 2012; Welikala et al., 2019), routine lifestyle habits, such as when they sleep, cook, go out, watch TV (D. Chen et al., 2015; Muneeb et al., 2019; Stankovic et al., 2016), and even preferences, such as the type of TV program watched (Greveler et al., 2012), and having a cup of coffee and a toast for breakfast (Molina-Markham et al., 2010).

In addition to the type of personal information that can be derived from electricity consumption data, cases with high accuracy of guessing models have also been reported. Teng et al. (2022) summarize the results of NILM experimental studies on smart meter data and show that load disaggregation with high accuracy is possible using smart meter data with an interval of 15 minutes or less (Appendix. 6). It has already deployed automated chatbots in pilot projects that provide energy-saving advice and feedback, such as recommendations to replace inefficient appliances, based on results analyzed by NILM in several EU countries (Khazaei et al., 2019).

Chalmers et al. (2015) has succeeded in developing an algorithm for detecting abnormal behavior with a high accuracy of 99.17% in a demonstration experiment using 30-minute power consumption data of major home appliances and machine learning. This technology is expected to have practical applications in the medical and nursing fields, such as remote monitoring of dangerous behaviors in patients with Alzheimer's disease and other conditions (Chalmers et al., 2015).

What information can be inferred with what level of accuracy depends on the resolution of the electricity consumption data (i.e., what unit of time, what unit of electricity consumption (whole house, breaker, appliance unit, etc.)) (Véliz & Grunewald, 2018), but it has been shown that some information can be determined with some accuracy even with low-resolution data, such as per hour or per day (Teng et al., 2022). It is also noted that the system can be further improved in studies using labeled data that collect and combine behavioral data from individuals in the target area (Teng et al., 2022).

Chen et al. (2023) point out that even after the data applied de-identification measures, there is a possibility that inferences about individuals and their private lives can be drawn from the energy consumption data itself. To protect individual privacy, they suggest generating synthetic data with the same data structures and statistical characteristics as the original data and using them instead of the de-identified original data (Z. Chen et al., 2023).

3.4.3. Legal understanding of household electricity consumption data

As mentioned in the introduction, energy consumption data are not explicitly referred to as personal data in the GDPR. However, as we have already discussed, energy consumption data can be used for inferring other personal information and personal life, so they can be considered personal data according to the definition in Article 4 of the GDPR (Blanke, 2020; Martinez et al., 2020; Wachter & Mittelstadt, 2019). It is not a rule that is legally binding, but the expert group of the European Commission's Smart Grid Task Force recommends that energy consumption data and some other types of energy-related data be treated as personal data (Smart Grid Task Force, 2018). As this Recommendation relates to data protection in smart grids (Smart Grid Task Force, 2018), it assumes smart meters that function as part of that function (i.e., next-generation smart meters, whose rollout in the EU has been delayed) and that are installed on a contractual basis. In addition to this recommendation, for example, in Spain, an EU member state, the Supreme Court ruling on personal data also mentioned the possibility that energy consumption data could be personal data (One Trust, 2019).

4. ANALYTICAL FRAMEWORK

Above we have presented basic knowledge and background on the subject and introduced the current discussion on it. In this section, taking these factors into account, we discuss the frameworks for distinguishing PII from non-PII and propose and develop an analytical framework for classifying household electricity consumption data. First, we review the literature that proposes methods for distinguishing PII from non-PII based on a risk-based approach. We provide an overview of the proposed frameworks and point out shortcomings in assuming that they apply to household electricity consumption data. Second, we present the main frameworks for developing a new framework. Finally, we develop a new framework and explain our proposed framework.

4.1 Methods for distinguishing PII and non-PII

Following the risk-based approach, methods to distinguish between PII and non-PII have been proposed and discussed, particularly in the research area of health data use. For example, Al-Fedagshi et al. (2012) propose a sensitivity-based gradational classification of PII through health data research using context-free analysis. They propose to decompose the components of the dataset, calculate the degree of sensitivity by component, and suggest effective risk-based data management and the potential for data utilization by extracting non-PII portions (Al-Fedaghi & Al-Azmi, 2012).

Onik et al. (2018) define PII as identifiers themselves or data with identifiers[2] and define data that do not contain an identifier but may be able to identify individuals as "potential PPI (PPII)" and present examples of PII and PPII (Appendix. 7). They note that some types of data are not always capable of identifying an individual on their own, but can become identifying data when combined with other information (Onik et al., 2018). They suggest that this data (i.e., PPII) should be covered by privacy protections just as PII is (Onik et al., 2018). However, as mentioned earlier, judgments based on static lists to determine which data are PII or PPII are likely to be difficult to implement in practice because the number of data types is enormous and increases from time to time.

Garfinkel (2015) introduces a framework for classifying data according to the degree of de-identification, called the "Data identifiability spectrum" (Detailed in Section 4.3). In this framework, all data are classified into five classes with respect to their identifiability with an individual natural person (Detailed in Section 3.2). This framework is based on a risk-based approach, and Garfinkel also notes that the level of risk changes with technological development and points out that regular policy and the result of classification review are important (Garfinkel, 2015).

Malgieri & Comandé (2017) propose a method for distinguishing sensitive personal data based on the inherent characteristics of the data and the ability to infer other types of personal data. They explore health-related personal data that require additional careful treatment under the GDPR. Malgieri & Comandé (2017) point out that while health-related data require rigorous and secure treatment, data that are not health data but have some correlation with health data and can be used to make inferences about an individual's health are not regulated under the current regulation (Malgieri & Comandé, 2017). They caution against the risk of inferring health data from non-

[2] They rely on the National Technical Information Service (NTIS) definition of PII, noting that PII there includes only "directly" identifiable information, such as data with names or other identifiers.

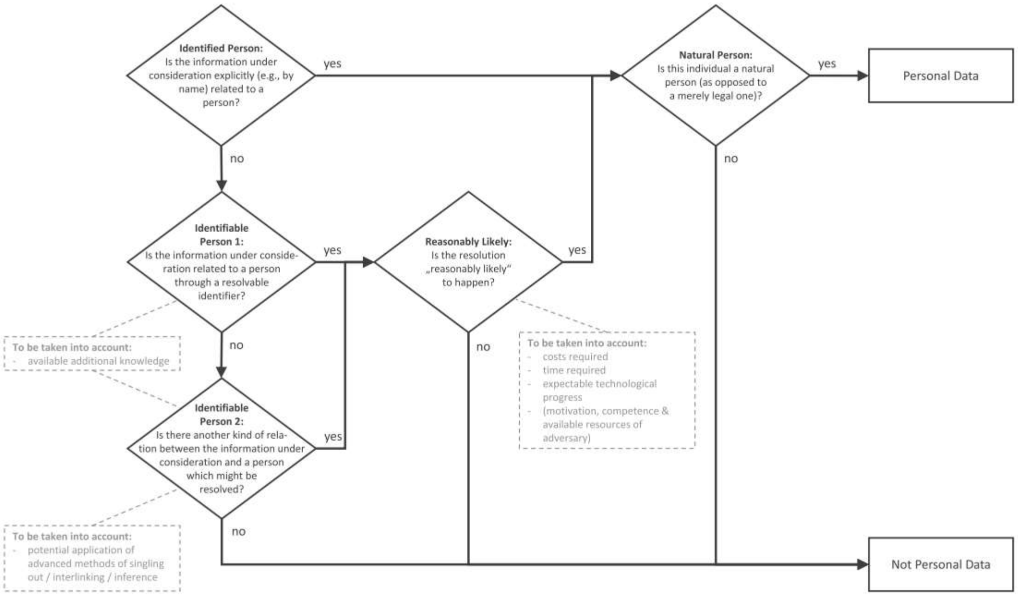
health personal data or non-personal data and suggest that these data be recognized as "quasi-health data" and treated as the health data that can be inferred from these data (Malgieri & Comandé, 2017). Based on a risk-based approach, they suggest that assessing the "computational distance" of sensitive personal data, i.e., the feasibility of deriving specific personal data (in this case, health data), has to be substantively determined. They also note that the feasibility of derivation depends on technological development and will change and that combining different types of data would increase the feasibility of deriving other types of data (Malgieri & Comandé, 2017).

4.2. Procedure for distinguishing personal data from non-personal data under the GDPR

Under the GDPR, it is defined that pseudonymized data must be treated as personal data, while anonymized data is recognized as non-personal data (Kuner et al., 2021). The GDPR adopted the risk-based approach for distinguishing personal data (Helminger & Rechberger, 2022), and Recital 26 explains how organizations can classify de-identified data as pseudonymized data (i.e., personal data) or anonymized data (i.e., non-personal data) (GDPR, Recital 26, 2016).

Recital 26 explains that objective factors such as cost, time, and the development of technologies should be taken into account for considering the possibility of identifying a natural person (Mourby et al., 2018). Finck & Pallas (2020) presents the flowchart for distinguishing personal data according to Recital 26 of the GDPR (Figure.3). Recital 26 calls for considering whether or not the de-identification method is "reasonably likely." However, there are no specific criteria for "reasonableness"(Finck & Pallas, 2020; Kotschy, 2016; Stalla-Bourdillon & Knight, 2016), and Finck & Pallas (2020) point out that this room for interpretation can lead to confusion and inconsistent implementations. However, it is a realistic approach in the field of data protection, where technological advances are remarkable.

Figure 3. Flowchart of distinguishing “personal data” and “not personal data” under GDPR Recital 26

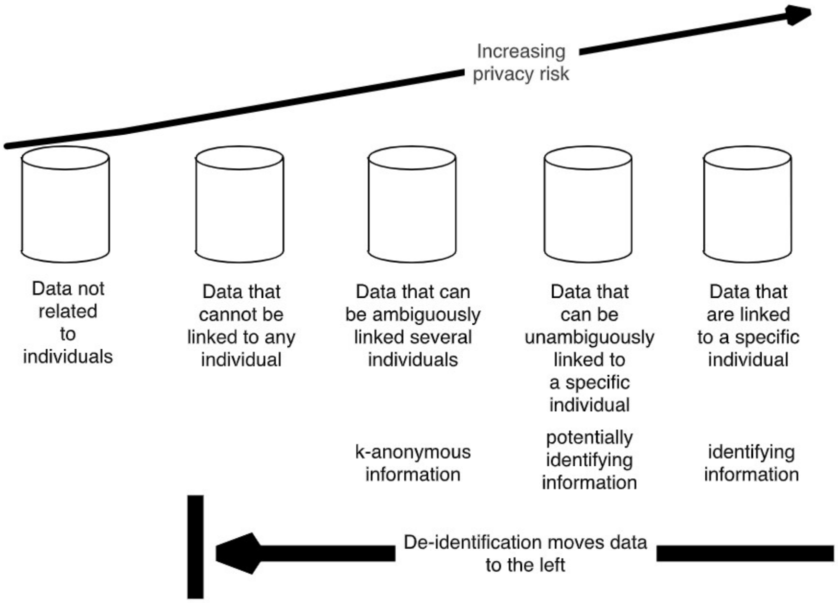

Identified Person:
Is the information under consideration explicitly (e.g., by name) related to a person?
yes
no
Natural Person:
Is this individual a natural person (as opposed to a merely legal one)?
yes
Personal Data
no
Identifiable Person 1:
Is the information under consideration related to a person through a resolvable identifier?
yes
no
Reasonably Likely:
Is the resolution „reasonably likely" to happen?
yes
no
To be taken into account:
- available additional knowledge
To be taken into account:
- costs required
- time required
- expectable technological progress
- (motivation, competence & available resources of adversary)
Identifiable Person 2:
Is there another kind of relation between the information under consideration and a person which might be resolved?
yes
no
To be taken into account:
- potential application of advanced methods of singling out / interlinking / inference
Not Personal Data

Source: Finck & Pallas (2020)

4.3. Data identifiability spectrum

As we introduced in Section 4.1, Garfinkel (2015) provides a framework called the "Data identifiability spectrum" for classifying data based on a risk-based approach. In this framework, all data is classified into five classes based on the degree of de-identification and linked to the level of privacy risk. Garfinkel (2015) classifies data that is inherently unrelated to personal information as data with the lowest privacy risk and data that is explicitly linked to an identifier as data with the highest privacy risk. And in between, there are three categories: 1) data that relate to an individual but where it is not possible to identify the natural person, 2) data that can be narrowed down to multiple candidates, and 3) data that are de-identified but where it is almost possible to identify an individual (Garfinkel, 2015).

Figure 4. Data identifiability spectrum


Increasing privacy risk
Data not related to individuals
Data that cannot be linked to any individual
Data that can be ambiguously linked several individuals
Data that can be unambiguously linked to a specific individual
Data that are linked to a specific individual
k-anonymous information
potentially identifying information
identifying information
De-identification moves data to the left

Source: Garfinkel (2015)

While this framework allows for the classification of individual data in terms of privacy risk, it does not cover the perspective of "reasonableness" (see Section 4.2), which is required by regulations such as GDPR and proposed in some relevant studies (Finck & Pallas, 2020; Kuner et al., 2021; Schwartz & Solove, 2011; General Data Protection Regulation (GDPR), 2016). Therefore, in this paper, we propose a classification based on this framework, with a modified framework that adds specific rationality perspectives mentioned in other studies and regulations.

4.4. Perspective on the possibility of inference

Although the approach presented in Recital 26 anticipates future technological developments and adopts a flexible approach, some argue that the GDPR does not provide sufficient legal protection for Big Data or algorithmic inference of personal data (Wachter & Mittelstadt, 2019). Unlike the California Consumer Privacy Act, which provides protection for personal data at high risk of being inferred by algorithms, the GDPR does not explicitly address such inferences (Blanke, 2020; Wachter & Mittelstadt, 2019). The element of “inference” was considered in the A29WP but not explicitly mentioned in the GDPR (Blanke, 2020; Finck & Pallas, 2020). However, the wording in Recital 26 indicates that the 'singling out' method is only one of the possible methods (Kuner et al., 2021) and does not reject it.

As discussed above, various types of personal information can be inferred from household electricity consumption data. Note that several studies have reported that particularly sensitive personal data under Article 9(1) in GDPR (e.g., ethnicity and health-related data) can also be inferred with some degree of accuracy from household electricity consumption data (Section 4.1). Therefore, we believe that we cannot ignore this risk when discussing household electricity consumption data from a privacy perspective.

4.5. Evaluation framework for household electricity consumption data based on privacy risk

Based on the above, we propose a new framework for assessing household energy consumption data from a privacy perspective. This framework follows the risk-based approach adopted by the GDPR and takes into account some key factors, such as technological development, the technological level of de-identification, and the feasibility of inferring information. The vertical axis of the framework shows the level of de-identification by current technological standards. The horizontal axis indicates the feasibility of re-identification considering cost, time, and labor. In this framework, the top left has the highest privacy risk, and the bottom right has the lowest risk of identifying a particular natural person.

Figure 5. Evaluation framework for household electricity consumption databased on privacy risk

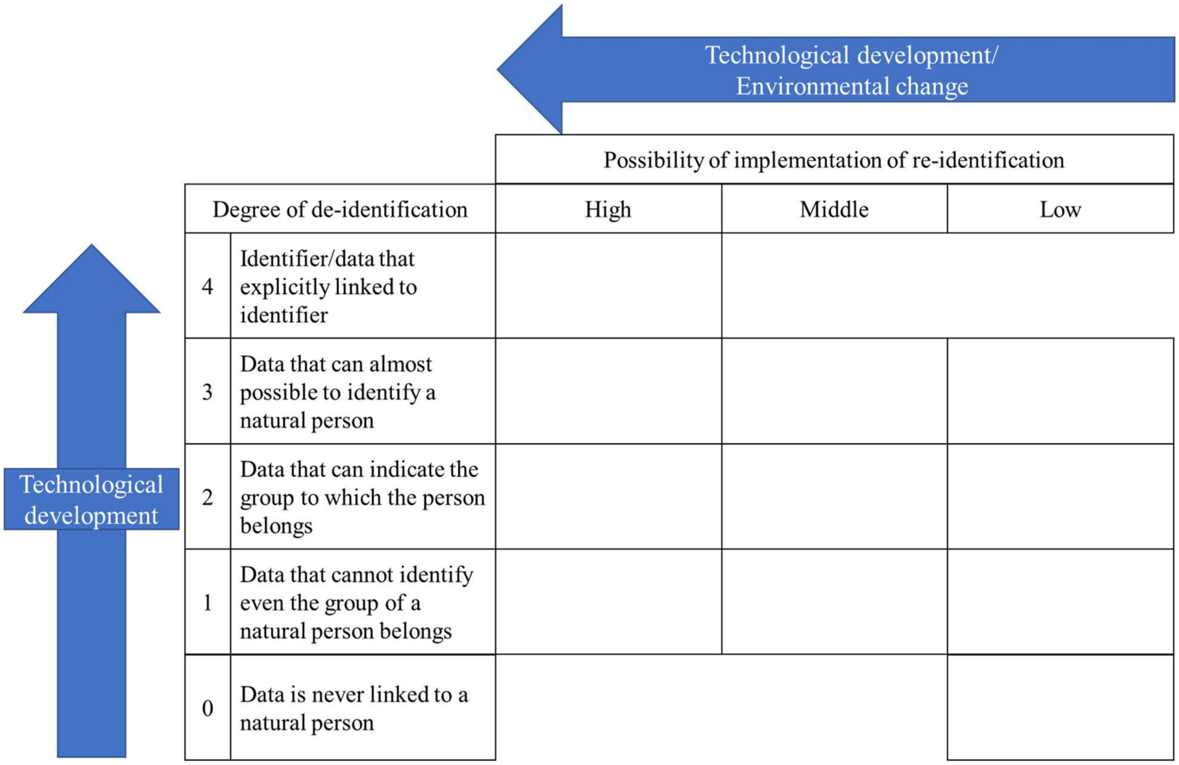
		Possibility of implementation of re-identification		
Degree of de-identification		High	Middle	Low
4	Identifier/data that explicitly linked to identifier			
3	Data that can almost possible to identify a natural person			
2	Data that can indicate the group to which the person belongs			
1	Data that cannot identify even the group of a natural person belongs			
0	Data is never linked to a natural person			

Source: Author

As we discussed above, the degree of de-identification of individual data (i.e., how difficult it is to re-identify the data) changes with future technological development. In this framework, the vertical axis depends on the method used by the data collector and the technological level of de-identification. Thus, if there is a technological development, such as the invention of new re-identification methods, the level of de-identification may decrease, and the plotted data type will move up the map. On the horizontal axis, technological developments and environmental changes, such as implementation cost reductions or the proliferation of high-speed processing systems, may increase the possibility of re-identification, and the plotted data type will move to the left.

Therefore, in order to manage and use data while maintaining privacy, it is important to have a good understanding of the current state of de-identification and re-identification technologies and to keep up with current information.

Figure 6. Image of a shift in privacy risk


Technological development/
Environmental change
Possibility of implementation of re-identification
Degree of de-identification
High
Middle
Low
4
Identifier/data that explicitly linked to identifier
3
Data that can almost possible to identify a natural person
2
Data that can indicate the group to which the person belongs
1
Data that cannot identify even the group of a natural person belongs
0
Data is never linked to a natural person
Technological development

Source: Author

After reviewing the studies on inference possibilities and the current discussion on household electricity data, we propose the qualitative criteria for the evaluating level of de-identification and feasibility of re-identification as follows:

Table 1. Degree of de-identification

	Degree of de-identification	Explanation	Examples
4	Identifier/data that explicitly linked to an identifier	Identifier itself or data set that contains an identifier.	Full name, national identification number, passport number, and data that contain these information
3	Data that can almost possible to identify a natural person	Personal information that is not an identifier or does not contain an identifier, but it is almost possible to identify a specific natural person.	Combination of zip code, date of birth, and last three digits of the mobile number

2	Data that can indicate the group to which the person belongs	Personal information can be used for narrowing down candidates for data subjects to the group to which they belong. Data with k-anonymity.	Company name and department where the individual works, a combination of age, gender, and occupation in a particular city
1	Data that cannot identify even the group of a natural person belongs	Data is personal information, but that can not be used for even narrowing down candidates for data subjects to the group level to which they belong.	Digits of postal code, today's dinner menu,
0	Data is never linked to a natural person	Non-personal information and data were never linked with any natural person from the begging.	Temperature data collected by sensors installed by the government in a public park

Source: Author

Table 2. Possibility of implementation of re-identification

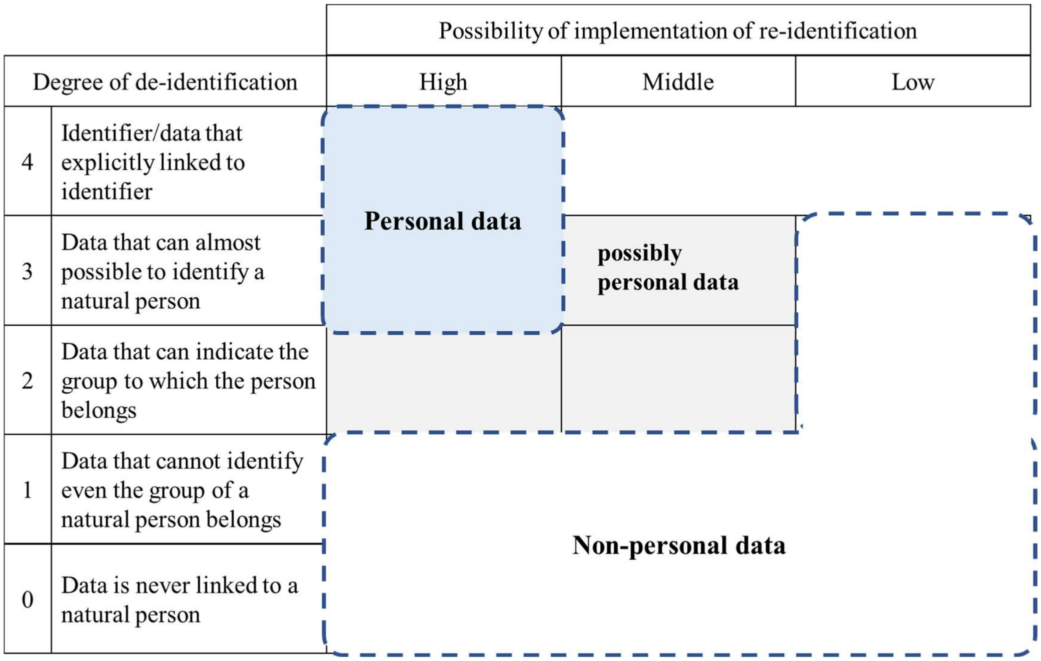
Possibility of re-identification	Explanation
High	- It is possible to identify a specific natural person without applying any measures. - It is possible to associate de-identified data with an identifier with low costs and efforts, such as the pseudonymized data, but the de-identification data list is available. - There is/are method(s) for inferring a natural person or group(s) of natural persons with reasonable accuracy. And the method(s) are easily applicable without much modification, e.g., using open-source predictive models with few regional differences.
Middle	- It is possible to link de-identified data to an identifier, but it is costly or takes a lot of time, or there is little benefit to implementing it, such as well-protected pseudonymized data. - There is/are established/proposed method(s) to re-identify a natural person, but it is costly or time-consuming to implement, or there is little benefit.

	- There is/are established/proposed methods to re-identify a natural person or group(s) of natural persons with reasonable accuracy, but implementation is costly or time-consuming or has little benefit. For example, additional data collection or research is required to modify the established model for a particular domain.
Low	- There is no established/proposed method to link the data and an identifier or other re-identifying method(s) to identify a natural person from the data (i.e., anonymized data). - There is no established/proposed method to inferring a natural person or group(s) of natural persons with reasonable accuracy.

Source: Author

In light of Recital 26 of the GDPR, the areas of personal data and non-personal data are shown in Figure 7. The grey areas are "pseudonymized" or "anonymized" data (in the definition of the GDPR) and cannot be judged to be clearly personal data under current law; however, depending on the terms of the data, there is a possibility that the data may be judged as personal data. In addition, data in this grey area may clearly become personal data in the future due to recent technological developments, etc., so special attention should be paid to this area.

Figure 7. Relation with the GDPR Recital 26



		Possibility of implementation of re-identification		
Degree of de-identification		High	Middle	Low
4	Identifier/data that explicitly linked to identifier	Personal data		
3	Data that can almost possible to identify a natural person		possibly personal data	
2	Data that can indicate the group to which the person belongs			
1	Data that cannot identify even the group of a natural person belongs		Non-personal data	
0	Data is never linked to a natural person			

Source: Author

5. SUMMARY OF RESULTS: CLASSIFICATION OF HOUSEHOLD ELECTRICITY CONSUMPTION DATA

In this section, we summarize the results for the types of household electricity consumption data, as well as the results and considerations for applying our new framework to these data. First, we present a summary of the types of household electricity consumption data that are commonly available in the EU (the methodology used was described in detail in Section 2). Then, we apply the framework proposed in Section 4.5 to these data. Finally, we explain the results of applying the framework and the use of household electricity consumption data.

5.1. Household electricity consumption data

The types of household electricity consumption data were obtained using the methodology explained in Section 2. From the research on the situation of smart electricity meters in the 27 EU countries, we obtained information on different time intervals of data. From the research on smart devices used in HEMS, we obtained information on 115 smart device products that collect electricity consumption data in households (Appendix. 8). As explained earlier, electricity consumption data are time-stamped data indicating the amount of electricity consumed in a given time interval. Therefore, each type of household electricity data differs only in the following three respects:

1) Data collection points (i.e., which device is used to collect the data)
2) Data interval (i.e., the unit of time interval, e.g., every 15 minutes, every hour)
3) Unit of data measurement (e.g., per appliance, per household)

Hence, we summarize the data type result according to these three aspects. The identified data types are listed in Table 3. All data listed in Table 3 are time-stamped electricity consumption data, except for data from analog meters. The unit of volume for measuring electricity consumption data is kWh for all data types.

Table 3. The types of electricity consumption data in households by data collection point

Data collection point	Data interval	Volume unit	Unit of data measurement
Smart meter	Near real-time	kWh	Household
	10 minutes	kWh	Household
	15 minutes	kWh	Household
	30 minutes	kWh	Household
	1 hour	kWh	Household
Analog meter	Monthly	kWh	Household
Smart appliance	Near real-time	kWh	Appliance
Smart thermostat	Near real-time	kWh	Household
	Near real-time	kWh	Room
Smart plug/switch	Near real-time	kWh	Appliance
Smart circuit breaker	Near real-time	kWh	Circuit

Source: Author - categorized data from Tounquet & Alaton (2019), data set published in Pritoni et al. (2018), Pritoni et al. (2018), Ford et al. (2017), and Alden et al. (2019)

Generally, a unit of data measurement is per household in the case of a meter installed by an electricity company. More precisely, the data units of smart meters and analog meters installed by utilities are per contract. However, residential electricity contracts are usually per household, so " per contract" is almost synonymous with "per household." For ease of understanding, we refer to the unit as the household.

As discussed in Section 3.3, the types of data collected by smart devices on HEMS and their data intervals depend on the products and services they use. For example, some products only have the actual control function and have no other four HEMS functions (detailed in Section 3.3.2). Indeed, in this case, these products do not collect consumption data. Therefore, this table shows the possible household electricity consumption data types derived from research, but this does not mean that all smart devices collect such electricity consumption data.

The data collected by smart appliances are per appliance unit, which means the electricity consumption data of smart appliances show the electricity consumption by using a specific appliance. Smart appliances currently on the market that collect energy data are primarily those that consume a high percentage of the electricity used by appliances in the home, such as smart washing machines, smart refrigerators, smart dryers, and smart dishwashers (Ford et al., 2017). Most products of this type have all or most of the five functions of HEMS (control, monitoring, management, alarming, and logging) on a near real-time basis, so the time interval of data is also near real-time (Alden et al., 2019; Ford et al., 2017).

The main role of the smart thermostat is to optimize HVAC systems, so it collects electricity consumption data in near real-time. The measuring unit of electricity consumption data is a whole house or a single room, as it is connected to the HVAC unit. For smart plugs and switches and smart circuit breakers, most products aim to control the operation and power consumption and collect electricity consumption data in near real-time. Note, even though smart plugs are technically capable of collecting data on a plug-by-plug basis (i.e., instead of connecting the plug directly to a device, multiple devices are connected to the plug via a power strip), the unit of measure is the appliance unit because the product is basically designed to be managed on an appliance basis. Smart lighting systems are not included in this list because there are currently no commercial products that collect energy consumption data. The reason for this is that lighting can determine its power consumption with a high degree of accuracy from its status (on/off and brightness settings), so measuring actual power consumption data is not very useful for the additional cost (Ford et al., 2017).

In addition, as discussed in Section 3.3, for the various smart devices that make up HEMS, other types of data are collected from sensors and other sources in addition to electricity consumption data. The types of data collected vary from smart device to smart device and from energy management system to energy management system, but the types of data that can be collected simultaneously in this study are summarized as examples in Appendix 8. Also, the composition of the database (location of storage, type of data stored, type of data stored together or whether data is combined from different sources, duration of data storage, etc.) varies depending on the devices and services installed in each household, whether or not the data collected is stored, and whether or not it is considered integrated HEMS.

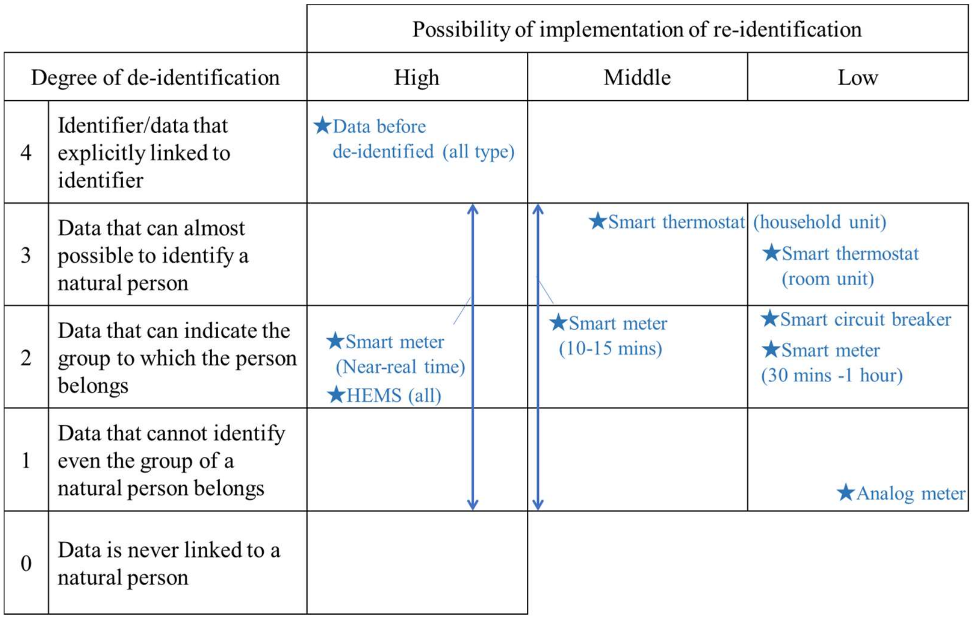
5.2. Mapping of household electricity consumption data from a privacy perspective

Figure 8 shows the results of applying the framework proposed in Section 4.5 to the household electricity consumption data types presented in 5.1, considering the inherent characteristics and current research discussed in Section 3.4. As explained in the introduction, this study is concerned with standalone household electricity consumption data, so the entire dataset as HEMS or smart home service with its myriad of possible combinations is not included in Figure 8. Since smart/analog meters and smart devices of HEMS are installed in a natural person's home and electricity consumption results from a natural person's activity, no household electricity data is classified as non-personal information. Therefore, the question arises whether it can be personal data or personal information but not personal data.

First, household electricity consumption data are usually associated with an identifier or a unique key that can be easily linked to the identifier (details in Section 3.3). Therefore, before any de-identification method is applied, these data are clearly PII and also personal data under the GDPR.

Next, we consider the electricity consumption of each household, which itself has the ability to infer a particular natural person. Although the data obtained from smart electricity meters have a longer data interval of 15 minutes compared to the real-time data from HEMS, there is much research on derivation methods. Therefore, it is possible to derive multiple combinations of personal information from the data of a single smart meter, allowing inferences to be made about specific groups of people. For example, NILM and other algorithmic load decomposition, profiling from geographic information, and methods for inferring different types of information, such as individual lifestyles are being studied. Whether it is possible to just narrow down to a group with the same characteristics or to a specific individual level depends on the size of the sample of data subjects. That is, the number of extracted data samples was not large enough, k-anonymity may not be guaranteed, and it is technically possible to identify the specific individual.

Figure 8. Mapping of household electricity consumption data from each data collecting point



		Possibility of implementation of re-identification		
Degree of de-identification		High	Middle	Low
4	Identifier/data that explicitly linked to identifier	★Data before de-identified (all type)		
3	Data that can almost possible to identify a natural person		★Smart thermostat (household unit)	★Smart thermostat (room unit)
2	Data that can indicate the group to which the person belongs	★Smart meter (Near-real time) ★HEMS (all)	★Smart meter (10-15 mins)	★Smart circuit breaker ★Smart meter (30 mins -1 hour)
1	Data that cannot identify even the group of a natural person belongs			★Analog meter
0	Data is never linked to a natural person			

Source: Author

5.3. Considerations and recommendations

As part of the use of Big Data, research is being conducted on analyzing and profiling household energy consumption data. Depending on the nature and conditions of the data, even with the current state of the art, some inferences can be made with some degree of accuracy, and further progress is expected.

Even when only electricity consumption data from household units are available, it is technically possible to some degree to narrow down the data subject through a combination of demographic and lifestyle information derived from the energy consumption data. Still, caution is needed with information that could identify an individual. In particular, when the population size is small, such as when only residents of a specific region are extracted as a dataset, the possibility that individuals can be identified from energy consumption data cannot be dismissed, so we suggest that this should be treated with additional caution.

Suppose energy consumption data is data that is separated from the identifier of the data listed in 5.1, i.e., energy consumption data alone with a timestamp tied to the data collection device number. In this case, only a limited amount of data can be derived from the device-related data alone, and it is expected that the data itself is unlikely to be personal data. Therefore, factoring out and using only these low-risk data may be a realistic approach that preserves usage costs and privacy. Smart meters, however, under their legal requirements, may not be non-personal data if the sentence in Recital 26 is interpreted to mean that if personal data are included in the original database, the extracted data are also treated as personal data unless the original data are destroyed (Stalla-Bourdillon & Knight, 2016).

In this Figure 8, only the case of energy consumption data alone is presented. On the other hand, as shown in Section 3.3.2, some smart devices also collect data other than electricity consumption data. For example, some smart thermostats incorporate temperature and motion sensor data, HVAC operating information, and energy price information into their algorithms. Therefore, depending on the configuration of the system, including the database, the combination of multiple data of different types may lead to the possibility of identifying individuals. A comprehensive assessment of whether or not the data falls under the category of personal data is required, taking into account not only energy consumption data alone but also data collected or stored at the same time. Therefore, it is necessary to consider the privacy risk of the combined data, including storage and data collection points.

6. CONCLUSION

The right to privacy is one of the fundamental human rights, and the concept of data protection was developed to protect personal privacy in a digitized society. With the development of Big Data and algorithms, data not only provide information directly, but also allow inferences to be made about various other information, including sensitive information. Research to develop methods for making inferences from electricity consumption data is progressing, and the ability to infer more and more types of information and the accuracy of inferences about such information can be expected to improve in the future as research progresses and is used in conjunction with other data sources.

With this in mind, how we protect personal identification for ethical and efficient use of household electricity consumption data is critical from a privacy perspective. However, research on household electricity consumption data from the perspective of privacy invasion is scarce, with the exception of research on technical system components. And even the GDPR, the EU's comprehensive data protection law, does not clearly refer to electricity consumption data.

In this paper, we explored the current arguments and challenges for using household electricity consumption data from a privacy risk perspective and proposed and applied a new framework. The aim of this study was to understand household electricity consumption data from a privacy risk perspective and propose a framework to help with data use and data management based on a risk-based approach. There has not been much discussion in the area of research on electricity consumption data and its regulation, so we wanted to help fill the current research gaps in this area.

Our proposed framework follows the risk-based approach of the GDPR and considers key factors that impact privacy invasion risks. Namely, we define the criteria for classifying data based on their privacy risk using the degree of de-identification from a technological perspective and the feasibility of deriving information. We propose to allocate the electricity consumption data of each household based on these criteria to understand the degree of privacy risk of each data and the feasibility of using them. Given the nature of household electricity consumption data and the purpose of privacy protection, we also propose to update the mapping periodically to keep pace with technological developments.

As we have argued above, the discussion of the use of electricity consumption data has not been adequately discussed or legislated, despite its potential. We believe that this risk-based framework helps to understand data and support to avoid privacy-invasive uses or, conversely, to avoid excessive restrictions on uses that do not actually pose a high privacy risk.

LIST OF REFERENCES

Adeli, E., & Hedman, G. (2020). *Home Energy Management Systems: A Research Study on the European and Nordic Market*.

AlAbdulkarim, L., Lukszo, Z., & Fens, T. (2012). Acceptance of privacy-sensitive technologies: smart metering case in The Netherlands. *Third International Engineering Systems Symposium CESUN*.

Alahakoon, D., & Yu, X. (2013). Advanced analytics for harnessing the power of smart meter big data. *2013 IEEE International Workshop on Inteligent Energy Systems (IWIES)*, 40–45.

Alden, R. E., Han, P., & Ionel, D. M. (2019). Smart plug and circuit breaker technologies for residential buildings in the us. *2019 8th International Conference on Renewable Energy Research and Applications (ICRERA)*, 1018–1021.

Al-Fedaghi, S., & Al-Azmi, A. A. R. (2012). Experimentation with Personal Identifiable Information. *Intelligent Information Management*, *04*(04), 123–133. https://doi.org/10.4236/iim.2012.44019

Alliance for Telecommunications Industry Solutions. (2001). *American National Standard T1.523-2001*. Telecom Glossary 2000. http://www.atis.org/glossary/

Ammari, T., Kaye, J., Tsai, J. Y., & Bentley, F. (2019). Music, Search, and IoT: How People (Really) Use Voice Assistants. *ACM Trans. Comput. Hum. Interact.*, *26*(3), 11–17.

Angelis, G.-F., Timplalexis, C., Krinidis, S., Ioannidis, D., & Tzovaras, D. (2022). NILM applications: Literature review of learning approaches, recent developments and challenges. *Energy and Buildings*, *261*, 111951. https://doi.org/10.1016/j.enbuild.2022.111951

Asghar, M. R., Dán, G., Miorandi, D., & Chlamtac, I. (2017). Smart meter data privacy: A survey. *IEEE Communications Surveys & Tutorials*, *19*(4), 2820–2835.

Badr, M. M., Ibrahem, M. I., Kholidy, H. A., Fouda, M. M., & Ismail, M. (2023). Review of the Data-Driven Methods for Electricity Fraud Detection in Smart Metering Systems. *Energies*, *16*(6), 2852.

Beckel, C., Sadamori, L., & Santini, S. (2013). Automatic socio-economic classification of households using electricity consumption data. *Proceedings of the Fourth International Conference on Future Energy Systems*, 75–86. https://doi.org/10.1145/2487166.2487175

Bennett, C. J. (1992). *Regulating privacy: Data protection and public policy in Europe and the United States*. Cornell University Press.

Bennett, C. J., & Raab, C. D. (2017). *The governance of privacy: Policy instruments in global perspective*. Routledge.

Berg Insight. (2022, November 23). *Smart electricity meter penetration rate in Europe reached 56 percent at the end of 2022*. https://www.berginsight.com/smart-electricity-meter-penetration-rate-in-europe-reached-56-percent-at-the-end-of-2022

Blanke, J. M. (2020). Protection for 'Inferences Drawn': A Comparison Between the General Data Protection Regulation and the California Consumer Privacy Act. *Global Privacy Law Review*, *1*(2).

Bleier, A., Goldfarb, A., & Tucker, C. (2020). Consumer privacy and the future of data-based innovation and marketing. *International Journal of Research in Marketing*, *37*(3), 466–480. https://doi.org/10.1016/j.ijresmar.2020.03.006

Boland, M. R., Karczewski, K. J., & Tatonetti, N. P. (2017). Ten Simple Rules to Enable Multi-site Collaborations through Data Sharing. *PLOS Computational Biology*, *13*(1), e1005278. https://doi.org/10.1371/journal.pcbi.1005278

Boyne, S. M. (2018). Data Protection in the United States. *The American Journal of Comparative Law*, *66*(suppl_1), 299–343. https://doi.org/10.1093/ajcl/avy016

Bugden, D., & Stedman, R. (2019). A synthetic view of acceptance and engagement with smart meters in the United States. *Energy Research & Social Science*, *47*, 137–145.

Bygrave, L. A. (2010). Privacy and data protection in an international perspective. *Scandinavian Studies in Law*, *56*(8), 165–200.

Chalmers, C., Hurst, W., Mackay, M., & Fergus, P. (2015). Smart meter profiling for health applications. *2015 International Joint Conference on Neural Networks (IJCNN)*, 1–7.

Chamaret, C., Steyer, V., & Mayer, J. C. (2020). "Hands off my meter!" when municipalities resist smart meters: Linking arguments and degrees of resistance. *Energy Policy*, *144*, 111556. https://doi.org/10.1016/j.enpol.2020.111556

Chen, D., Kalra, S., Irwin, D., Shenoy, P., & Albrecht, J. (2015). Preventing occupancy detection from smart meters. *IEEE Transactions on Smart Grid*, *6*(5), 2426–2434.

Chen, Z., Li, J., Cheng, L., & Liu, X. (2023). Federated-WDCGAN: A federated smart meter data sharing framework for privacy preservation. *Applied Energy*, *334*, 120711.

Collins Dictionaries. (2023). *Definition of "near real-time."* Collins English Dictionary. https://www.collinsdictionary.com/dictionary/english/near-real-time

Cuijpers, C., & Koops, B.-J. (2013). Smart metering and privacy in Europe: Lessons from the Dutch case. *European Data Protection: Coming of Age*, 269–293.

Dash, S., Shakyawar, S. K., Sharma, M., & Kaushik, S. (2019). Big data in healthcare: management, analysis and future prospects. *Journal of Big Data*, *6*(1), 54. https://doi.org/10.1186/s40537-019-0217-0

De Hert, P. (2012). A Human Rights Perspective on Privacy and Data Protection Impact Assessments. In *Privacy Impact Assessment* (pp. 33–76). Springer Netherlands. https://doi.org/10.1007/978-94-007-2543-0_2

Drkušić, E. (2017, May 10). *A Data Model for an Electric Power Production System*. Vertabelo. https://vertabelo.com/blog/a-data-model-for-an-electric-power-production-system/

Drkušić, E. (2019, January 24). *The Smart Home Data Model*. Vertabelo. https://vertabelo.com/blog/the-smart-home-data-model/

Erickson, A. (2018). Comparative Analysis of the EU's GDPR and Brazil's LGPD: Enforcement Challenges with the LGPD. *Brook. J. Int'l L.*, *44*, 859.

Etteldorf, C. (2019). EDPB on the Interplay between the ePrivacy Directive and the GDPR. *Eur. Data Prot. L. Rev.*, *5*, 224.

European Distribution System Operators for Smart Grids (EDSO). (2023, April 30). *What is a DSO?* https://www.edsoforsmartgrids.eu/about-dsos/what-is-a-dso

Directive (EU) 2019/944, Pub. L. No. Document 32019L0944 (2019). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019L0944

Charter of Fundamental Rights of the European Union, (2000). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:12012P/TXT&from=EN

Finck, M., & Pallas, F. (2020). They who must not be identified—distinguishing personal from non-personal data under the GDPR. *International Data Privacy Law*, *10*(1), 11–36. https://doi.org/10.1093/idpl/ipz026

Ford, R., Pritoni, M., Sanguinetti, A., & Karlin, B. (2017). Categories and functionality of smart home technology for energy management. *Building and Environment*, *123*, 543–554. https://doi.org/10.1016/j.buildenv.2017.07.020

Garfinkel, S. (2015). *De-identification of Personal Information: (pp. 1-46)*. US Department of Commerce, National Institute of Standards and Technology.

Gazi, T. (2020). Data to the rescue: how humanitarian aid NGOs should collect information based on the GDPR. *Journal of International Humanitarian Action*, *5*(1), 1–7.

Geels, F. W., Sareen, S., Hook, A., & Sovacool, B. K. (2021). Navigating implementation dilemmas in technology-forcing policies: A comparative analysis of accelerated smart meter diffusion in the Netherlands, UK, Norway, and Portugal (2000-2019). *Research Policy*, *50*(7), 104272.

Gieser, D. (2015). *What is Personally Identifiable Information?* [Law School Student Scholarship. 686]. Seton Hall University.

Gil González, E., & de Hert, P. (2019). Understanding the legal provisions that allow processing and profiling of personal data—an analysis of GDPR provisions and principles. *ERA Forum*, *19*(4), 597–621. https://doi.org/10.1007/s12027-018-0546-z

Goddard, M. (2017). The EU General Data Protection Regulation (GDPR): European Regulation that has a Global Impact. *International Journal of Market Research*, *59*(6), 703–705. https://doi.org/10.2501/IJMR-2017-050

Gouriet, M., Barancourt, H., Boust, M., Calvez, P., Laskowski, M., Taillandier, A.-S., Tilman, L., Uslar, M., & Warweg, O. (2022). The Energy Data Space: The Path to a European Approach for Energy. In *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage* (pp. 535–575). Springer International Publishing Cham.

Greveler, U., Glösekötterz, P., Justusy, B., & Loehr, D. (2012). Multimedia content identification through smart meter power usage profiles. *Proceedings of the International Conference on Information and Knowledge Engineering (IKE)*, 1.

Grünewald, P., & Diakonova, M. (2019). The specific contributions of activities to household electricity demand. *Energy and Buildings*, *204*, 109498. https://doi.org/10.1016/j.enbuild.2019.109498

Helminger, L., & Rechberger, C. (2022). Multi-party computation in the GDPR. *Privacy Symposium 2022: Data Protection Law International Convergence and Compliance with Innovative Technologies (DPLICIT)*, 21–39.

Humayun, M., Alsaqer, M. S., & Jhanjhi, N. (2022). Energy Optimization for Smart Cities Using IoT. *Applied Artificial Intelligence*, *36*(1). https://doi.org/10.1080/08839514.2022.2037255

Jakobi, T., Patil, S., Randall, D., Stevens, G., & Wulf, V. (2019). It is about what they could do with the data: A user perspective on privacy in smart metering. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *26*(1), 1–44.

Khattak, A. M., Khanji, S. I., & Khan, W. A. (2019). Smart meter security: Vulnerabilities, threat impacts, and countermeasures. *Proceedings of the 13th International Conference on Ubiquitous Information Management and Communication (IMCOM) 2019 13*, 554–562.

Khazaei, M., Stankovic, L., & Stankovic, V. (2019). *Trends and challenges in smart metering analytics*. University of Strathclyde.

Kim, S., Park, M., Lee, S., & Kim, J. (2020). Smart home forensics—data analysis of IoT devices. *Electronics*, *9*(8), 1215.

King, N. J., & Jessen, P. W. (2014). For privacy's sake: Consumer "opt outs" for smart meters. *Computer Law & Security Review*, *30*(5), 530–539.

Knezović, K., Marinelli, M., Codani, P., & Perez, Y. (2015). Distribution grid services and flexibility provision by electric vehicles: A review of options. *2015 50th International Universities Power Engineering Conference (UPEC)*, 1–6.

Kochański, M., Korczak, K., & Skoczkowski, T. (2020). Technology Innovation System Analysis of Electricity Smart Metering in the European Union. *Energies*, *13*(4), 916. https://doi.org/10.3390/en13040916

Koponen, P., Saco, L. D., Orchard, N., Vorisek, T., Parsons, J., Rochas, C., Morch, A. Z., Lopes, V., & Togeby, M. (2008). Definition of smart metering and applications and identification of benefits. *Deliverable D3 of the European Smart Metering Alliance ESMA (Available at Www. Esma-Home. Eu, Members Area)*, *42*.

Kotschy, W. (2016). *The new General Data Protection Regulation-Is there sufficient pay-off for taking the trouble to anonymize or pseudonymize data*.

Kumar, P., Lin, Y., Bai, G., Paverd, A., Dong, J. S., & Martin, A. (2019). Smart grid metering networks: A survey on security, privacy and open research issues. *IEEE Communications Surveys & Tutorials*, *21*(3), 2886–2927.

Kuner, C., Bygrave, L. A., Docksey, C., Drechsler, L., & Tosoni, L. (2021). The EU General Data Protection Regulation: A Commentary/Update of Selected Articles. *Update of Selected Articles (May 4, 2021)*.

LaMarche, J., Cheney, K., Christian, S., & Roth, K. (2021). *Home energy management: products & trends (pp. 165-175)*. eScholarship, University of California.

Lavrijssen, S., Espinosa Apráez, B., & ten Caten, T. (2022). The Legal Complexities of Processing and Protecting Personal Data in the Electricity Sector. *Energies*, *15*(3), 1088. https://doi.org/10.3390/en15031088

Lissa, P., Deane, C., Schukat, M., Seri, F., Keane, M., & Barrett, E. (2021). Deep reinforcement learning for home energy management system control. *Energy and AI*, *3*, 100043.

Machorro-Cano, I., Alor-Hernández, G., Paredes-Valverde, M. A., Rodríguez-Mazahua, L., Sánchez-Cervantes, J. L., & Olmedo-Aguirre, J. O. (2020). HEMS-IoT: A Big Data and Machine Learning-Based Smart Home System for Energy Saving. *Energies*, *13*(5), 1097. https://doi.org/10.3390/en13051097

Mahapatra, B., & Nayyar, A. (2022). Home energy management system (HEMS): Concept, architecture, infrastructure, challenges and energy management schemes. *Energy Systems*, *13*(3), 643–669.

Malgieri, G., & Comandé, G. (2017). Sensitive-by-distance: quasi-health data in the algorithmic era. *Information & Communications Technology Law*, *26*(3), 229–249.

Martinez, J., Ruiz, A., Puelles, J., Arechalde, I., & Miadzvetskaya, Y. (2020). Smart grid challenges through the lens of the european general data protection regulation. *Advances in Information Systems Development: Information Systems Beyond 2020 28*, 113–130.

McKenna, E., Richardson, I., & Thomson, M. (2012). Smart meter data: Balancing consumer privacy concerns with legitimate applications. *Energy Policy*, *41*, 807–814.

Misra, M., Singh, P., & Alhelou, H. H. (2020). Energy Optimization for Smart Housing Systems. *Journal of Informatics Electrical and Electronics Engineering (JIEEE)*, *1*(1), 1–6. https://doi.org/10.54060/JIEEE/001.01.005

Molina-Markham, A., Shenoy, P., Fu, K., Cecchet, E., & Irwin, D. (2010). Private memoirs of a smart meter. *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, 61–66.

Mourby, M., Mackey, E., Elliot, M., Gowans, H., Wallace, S. E., Bell, J., Smith, H., Aidinlis, S., & Kaye, J. (2018). Are 'pseudonymised' data always personal data? Implications of the GDPR for administrative data research in the UK. *Computer Law & Security Review*, *34*(2), 222–233. https://doi.org/10.1016/j.clsr.2018.01.002

Muneeb, U. H., Rehmani, M. H., Kotagiri, R., Zhang, J., & Chen, J. (2019). Differential privacy for renewable energy resources based smart metering. *Journal of Parallel and Distributed Computing*, *131*, 69–80. https://doi.org/10.1016/j.jpdc.2019.04.012

Murrill, B. J., Liu, E. C., & Thompson, R. M. (2012). *Smart meter data: Privacy and cybersecurity*. Congressional Research Service, Library of Congress.

Nacer, A., Marhic, B., & Delahoche, L. (2017). Smart Home, Smart HEMS, Smart heating: An overview of the latest products and trends. *2017 6th International Conference on Systems and Control (ICSC)*, 90–95.

Nguyen, V. M., Brooks, J. L., Young, N., Lennox, R. J., Haddaway, N., Whoriskey, F. G., Harcourt, R., & Cooke, S. J. (2017). To share or not to share in the emerging era of big data: perspectives from fish telemetry researchers on data sharing. *Canadian Journal of Fisheries and Aquatic Sciences*, *74*(8), 1260–1274.

Nwankwo, I., Stauch, M., Radoglou-Grammatikis, P., Sarigiannidis, P., Lazaridis, G., Drosou, A., & Tzovaras, D. (2022). Data Protection and Cybersecurity Certification Activities and Schemes in the Energy Sector. *Electronics*, *11*(6), 965. https://doi.org/10.3390/electronics11060965

One Trust. (2019, August 19). *Spain: Supreme Court issues decision on energy consumption data and personal data definition*. Data Guidance. https://www.dataguidance.com/news/spain-supreme-court-issues-decision-energy-consumption

Onik, M. M. H., Al-Zaben, N., Yang, J., Lee, N.-Y., & Kim, C.-S. (2018). Risk Identification of Personally Identifiable Information from Collective Mobile App Data. *2018 International Conference on Computing, Electronics & Communications Engineering (ICCECE)*, 71–76. https://doi.org/10.1109/iCCECOME.2018.8659213

Osborne Clarke. (2022, October 6). *Energy saving requirements are emerging across Europe*. https://www.osborneclarke.com/insights/energy-saving-requirements-are-emerging-across-europe

Oshima, H., Ishizone, T., Nakamura, K., & Higuchi, T. (2022). Occupancy Detection for General Households by Bidirectional LSTM with Attention. *IECON 2022 – 48th Annual Conference of the IEEE Industrial Electronics Society*, 1–7. https://doi.org/10.1109/IECON49645.2022.9968594

Otjacques, B., Hitzelberger, P., & Feltz, F. (2007). Interoperability of E-Government Information Systems: Issues of Identification and Data Sharing. *Journal of Management Information Systems*, *23*(4), 29–51. https://doi.org/10.2753/MIS0742-1222230403

Park, G. (2019). The changing wind of data privacy law: A comparative study of the European Union's General Data Protection Regulation and the 2018 California Consumer Privacy Act. *UC Irvine L. Rev.*, *10*, 1455.

Peloquin, D., DiMaio, M., Bierer, B., & Barnes, M. (2020). Disruptive and avoidable: GDPR challenges to secondary research uses of data. *European Journal of Human Genetics*, *28*(6), 697–705. https://doi.org/10.1038/s41431-020-0596-x

Pereira, L., Costa, D., & Ribeiro, M. (2022). A residential labeled dataset for smart meter data analytics. *Scientific Data*, *9*(1), 134.

Polonetsky, J., & Tene, O. (2013). Privacy and big data: making ends meet. *Stan. L. Rev. Online*, *66*, 25.

Potdar, V., Chandan, A., Batool, S., & Patel, N. (2018). Big energy data management for smart grids—Issues, challenges and recent developments. *Smart Cities: Development and Governance Frameworks*, 177–205.

Poullet, Y. (2021). Data protection or privacy? In *Deep diving into data protection: 1979-2019: celebrating 40 years of research on privacy data protection at the CRIDS* (pp. 463–468). Larcier.

Pritoni, M., Ford, R., Karlin, B., & Sanguinetti, A. (2018). Home energy management (HEM) database: A list with coded attributes of 308 devices commercially available in the US. *Data in Brief*, *16*, 71–74.

Ramokapane, K. M., Bird, C., Rashid, A., & Chitchyan, R. (2022). Privacy design strategies for home energy management systems (hems). *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–15.

Robertson, A. H. (1973). *Privacy and Human Rights: Reports and Communications Presented at the Third International Colloquy about the European Convention on Human Rights*. Manchester University Press.

Schellinger, B., Völter, F., Urbach, N., & Sedlmeir, J. (2022). Yes, I do: Marrying blockchain applications with GDPR. *E-Government*, *19*, 22.

Schwartz, P. M., & Solove, D. J. (2011). The PII problem: Privacy and a new concept of personally identifiable information. *NYUL Rev.*, *86*, 1814.

Schwartz, P. M., & Solove, D. J. (2014). Reconciling personal information in the United States and European Union. *Calif. L. Rev.*, *102*, 877.

Shuaib, M., Alam, S., Alam, M. S., & Nasir, M. S. (2021). Compliance with HIPAA and GDPR in blockchain-based electronic health record. *Materials Today: Proceedings*.

Singh, R., Wang, X., Mendoza, J. C., & Ackom, E. K. (2015). Electricity (in) accessibility to the urban poor in developing countries. *Wiley Interdisciplinary Reviews: Energy and Environment*, *4*(4), 339–353.

Smart Grid Task Force. (2018). *Expert Group 2: Regulatory Recommendations for Privacy, Data Protection and Cyber-Security in the Smart Grid Environment* (Vol. 2).

Solove, D. J. (2000). Privacy and power: Computer databases and metaphors for information privacy. *Stan. L. Rev.*, *53*, 1393.

Solove, D. J. (2008). *Understanding privacy* (Public law and legal theory working paper No. 420). Harvard University Press, May.

Sovacool, B. K., Burke, M., Baker, L., Kotikalapudi, C. K., & Wlokas, H. (2017). New frontiers and conceptual frameworks for energy justice. *Energy Policy*, *105*, 677–691. https://doi.org/10.1016/j.enpol.2017.03.005

Stalla-Bourdillon, S., & Knight, A. (2016). Anonymous data v. personal data-false debate: an EU perspective on anonymization, pseudonymization and personal data. *Wis. Int'l LJ*, *34*, 284.

Stankovic, L., Stankovic, V., Liao, J., & Wilson, C. (2016). Measuring the energy intensity of domestic activities from smart meter data. *Applied Energy*, *183*, 1565–1580. https://doi.org/10.1016/j.apenergy.2016.09.087

Suripeddi, M. K. S., & Purandare, P. (2021). Blockchain and GDPR – A Study on Compatibility Issues of the Distributed Ledger Technology with GDPR Data Processing. *Journal of Physics: Conference Series*, *1964*(4), 042005. https://doi.org/10.1088/1742-6596/1964/4/042005

Sweeney, L. (2000). Simple demographics often identify people uniquely. *Health (San Francisco)*, *671*(2000), 1–34.

Teng, F., Chhachhi, S., Ge, P., Graham, J., & Gunduz, D. (2022). Balancing privacy and access to smart meter data: an Energy Futures Lab briefing paper. In *An Energy Futures Lab Briefing Paper*.

European Convention on Human Rights, (1950). https://www.echr.coe.int/documents/convention_eng.pdf

the European Commission. (2022a). *Workshops: Digitalisation of the energy system*. https://commission.europa.eu/events/workshops-digitalisation-energy-system-2022-02-16_en

the European Commission. (2022b, February 21). *NEWS ARTICLE: Digitalisation of energy: best practices for data sharing*.

the European Commission. (2022, October 18). *Digitalising the energy system - EU action plan* (COM(2022) 552 final). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions.

General Data Protection Regulation (GDPR), Pub. L. No. (EU) 2016/679 (2016). https://eur-lex.europa.eu/eli/reg/2016/679/oj

Article 4 (1) of the GDPR, Pub. L. No. (EU) 2016/679, General Data Protection Regulation (2016). https://eur-lex.europa.eu/eli/reg/2016/679/oj

Article 4 (2) of the GDPR, Pub. L. No. (EU) 2016/679, General Data Protection Regulation (2016). https://eur-lex.europa.eu/eli/reg/2016/679/oj

Article 4 (5) of the GDPR, Pub. L. No. (EU) 2016/679, General Data Protection Regulation (2016). https://eur-lex.europa.eu/eli/reg/2016/679/oj

Article 4 of GDPR, Pub. L. No. (EU) 2016/679, General Data Protection Regulation (2016). https://eur-lex.europa.eu/eli/reg/2016/679/oj

Article 5 (1) (a) of the GDPR, Pub. L. No. (EU) 2016/679, General Data Protection Regulation (2016). https://eur-lex.europa.eu/eli/reg/2016/679/oj

Article 9 (1) of the GDPR, Pub. L. No. (EU) 2016/679, General Data Protection Regulation (2016). https://eur-lex.europa.eu/eli/reg/2016/679/oj

Recital 26 of the GDPR, Pub. L. No. (EU) 2016/679, General Data Protection Regulation (2016). https://eur-lex.europa.eu/eli/reg/2016/679/oj

The Universal Declaration of Human Rights, Pub. L. No. General Assembly resolution 217 A (1948). https://www.un.org/en/about-us/universal-declaration-of-human-rights#:~:text=Article%2012,against%20such%20interference%20or%20attacks.

Tounquet, F., & Alaton, C. (2019). *Benchmarking smart metering deployment in the EU-28 [Final Report]*.

Tushar, W., Yuen, C., Saha, T., Chattopadhyay, D., Nizami, S., Hanif, S., Alam, J. E., & Poor, H. V. (2021). Roles of retailers in the peer-to-peer electricity market: A single retailer perspective. *Iscience*, *24*(11), 103278.

Véliz, C., & Grunewald, P. (2018). Protecting data privacy is key to a smart energy future. *Nature Energy*, *3*(9), 702–704. https://doi.org/10.1038/s41560-018-0203-3

Vitiello, S., Andreadou, N., Ardelean, M., & Fulli, G. (2022). Smart Metering Roll-Out in Europe: Where Do We Stand? Cost Benefit Analyses in the Clean Energy Package and Research Trends in the Green Deal. *Energies*, *15*(7), 2340. https://doi.org/10.3390/en15072340

Voss, W. G. (2016). General Data Protection Regulation, Privacy Shield, and the Right to Delisting. *The Business Lawyer*, *72*(1), 221–234. https://www.jstor.org/stable/26419118

Wachter, S., & Mittelstadt, B. (2019). A right to reasonable inferences: re-thinking data protection law in the age of big data and AI. *Colum. Bus. L. Rev.*, 494.

Webborn, E., Elam, S., McKenna, E., & Oreszczyn, T. (2019). Utilising smart meter data for research and innovation in the UK. *ECEEE Summer Study*, *2019*, 1387–1396.

Welikala, S., Thelasingha, N., Akram, M., Ekanayake, P. B., Godaliyadda, R. I., & Ekanayake, J. B. (2019). Implementation of a robust real-time non-intrusive load monitoring solution. *Applied Energy*, *238*, 1519–1529. https://doi.org/10.1016/j.apenergy.2019.01.167

Wills, L. (2019, May 1). *A Very Brief Introduction to the GDPR Recitals*. American Bar Association. https://www.americanbar.org/groups/litigation/committees/minority-trial-lawyer/practice/2019/a-very-brief-introduction-to-the-gdpr-recitals/#:~:text=As%20mentioned%2C%20the%20GDPR%20consists,context%20to%20supplement%20the%20articles.

Zafar, U., Bayhan, S., & Sanfilippo, A. (2020). Home energy management system concepts, configurations, and technologies for the smart grid. *IEEE Access*, *8*, 119271–119286.

Zhang, N., Mi, X., Feng, X., Wang, X., Tian, Y., & Qian, F. (2018). Understanding and mitigating the security risks of voice-controlled third-party skills on amazon alexa and google home. *ArXiv Preprint ArXiv:1805.01525*.

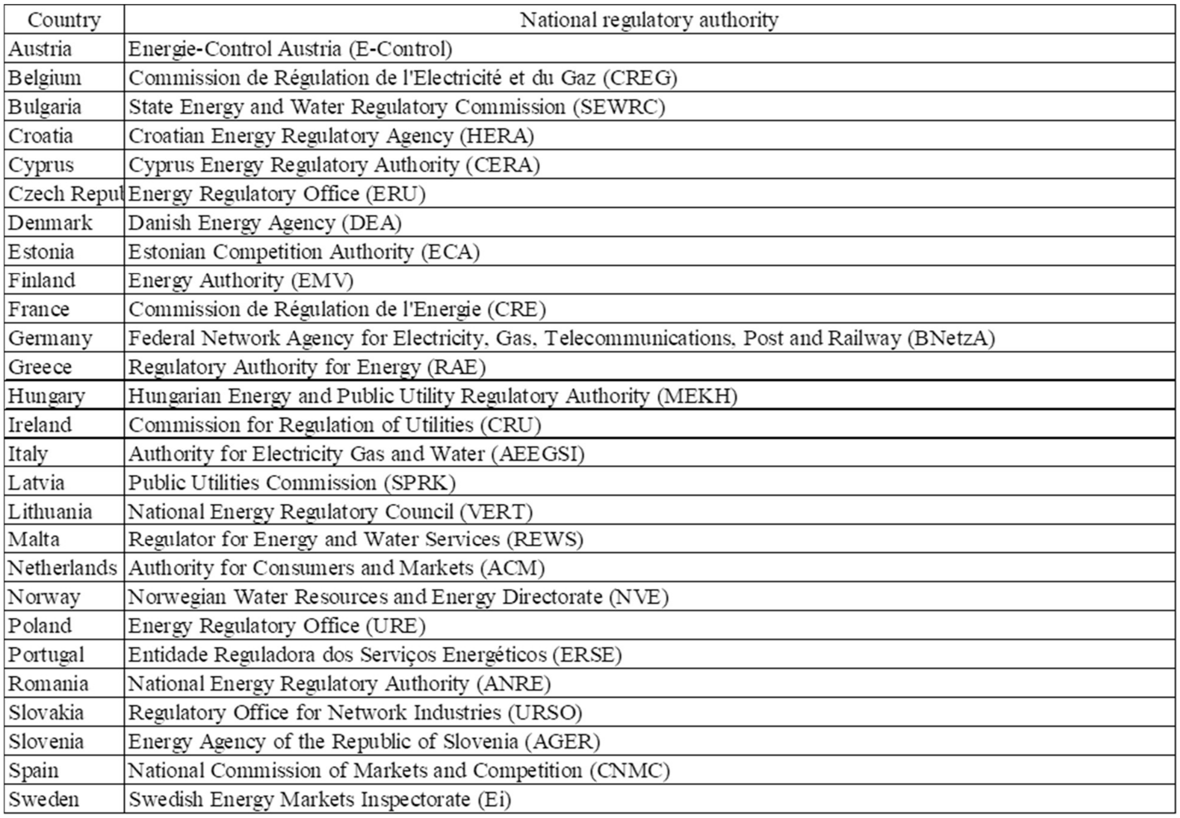
APPENDICES

Appendix 1. Definition of terms related to personal information in this paper

Classification			Definition
Information	Personal information	PII	Information attributable to a specific individual
		Non-PII personal information	Information relating to an individual but not attributable to a specific individual
	Non-personal information		Information that not relating to any individual

Source: Author

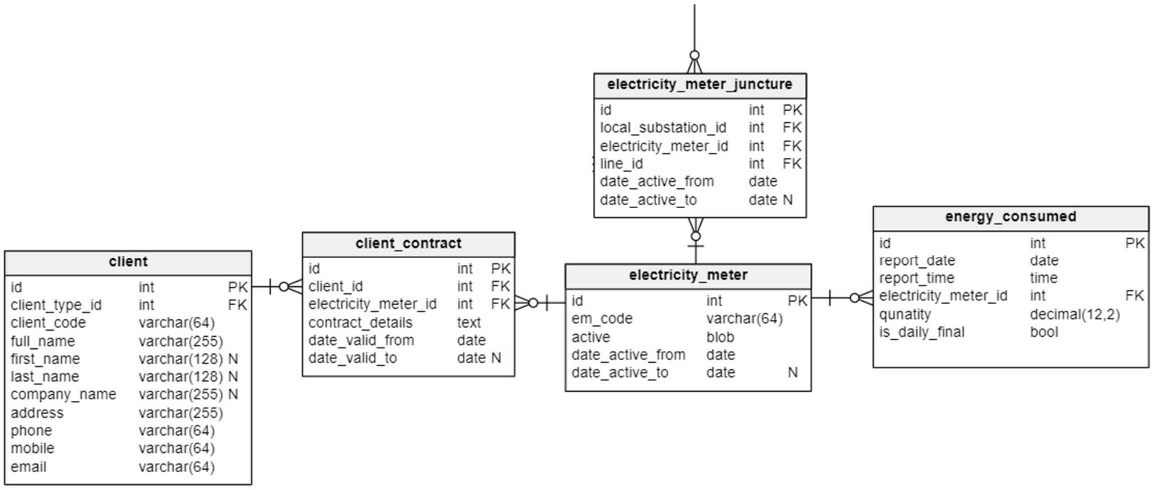
Appendix 2. List of EU countries and its national regulatory authorities



Country	National regulatory authority
Austria	Energie-Control Austria (E-Control)
Belgium	Commission de Régulation de l'Electricité et du Gaz (CREG)
Bulgaria	State Energy and Water Regulatory Commission (SEWRC)
Croatia	Croatian Energy Regulatory Agency (HERA)
Cyprus	Cyprus Energy Regulatory Authority (CERA)
Czech Repul	Energy Regulatory Office (ERU)
Denmark	Danish Energy Agency (DEA)
Estonia	Estonian Competition Authority (ECA)
Finland	Energy Authority (EMV)
France	Commission de Régulation de l'Energie (CRE)
Germany	Federal Network Agency for Electricity, Gas, Telecommunications, Post and Railway (BNetzA)
Greece	Regulatory Authority for Energy (RAE)
Hungary	Hungarian Energy and Public Utility Regulatory Authority (MEKH)
Ireland	Commission for Regulation of Utilities (CRU)
Italy	Authority for Electricity Gas and Water (AEEGSI)
Latvia	Public Utilities Commission (SPRK)
Lithuania	National Energy Regulatory Council (VERT)
Malta	Regulator for Energy and Water Services (REWS)
Netherlands	Authority for Consumers and Markets (ACM)
Norway	Norwegian Water Resources and Energy Directorate (NVE)
Poland	Energy Regulatory Office (URE)
Portugal	Entidade Reguladora dos Serviços Energéticos (ERSE)
Romania	National Energy Regulatory Authority (ANRE)
Slovakia	Regulatory Office for Network Industries (URSO)
Slovenia	Energy Agency of the Republic of Slovenia (AGER)
Spain	National Commission of Markets and Competition (CNMC)
Sweden	Swedish Energy Markets Inspectorate (Ei)

Source: Author made – from Tounquet & Alaton (2019)

Appendix 3. Example of data table structure in electricity retailers (smart meter data and customer-related part)



electricity_meter_juncture		
id	int	PK
local_substation_id	int	FK
electricity_meter_id	int	FK
line_id	int	FK
date_active_from	date	
date_active_to	date	N

client		
id	int	PK
client_type_id	int	FK
client_code	varchar(64)	
full_name	varchar(255)	
first_name	varchar(128)	N
last_name	varchar(128)	N
company_name	varchar(255)	N
address	varchar(255)	
phone	varchar(64)	
mobile	varchar(64)	
email	varchar(64)	

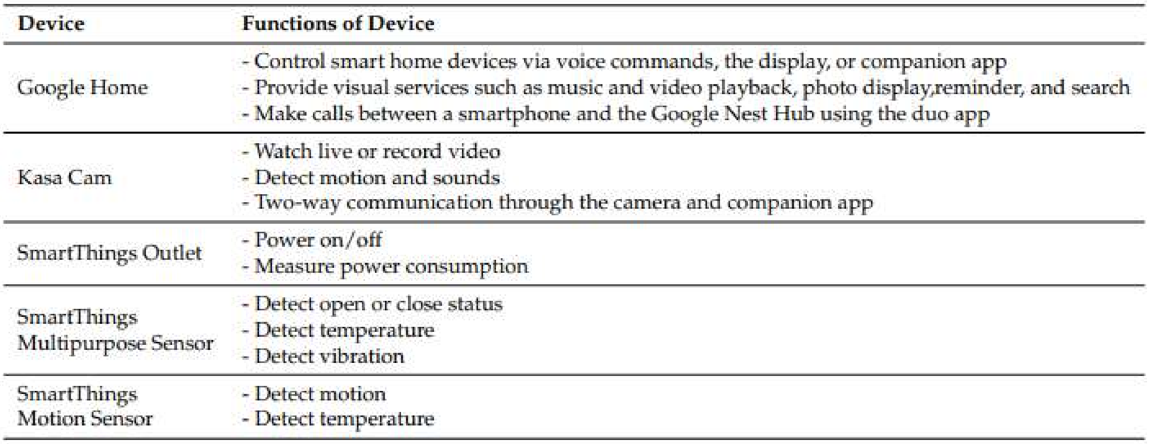
client_contract		
id	int	PK
client_id	int	FK
electricity_meter_id	int	FK
contract_details	text	
date_valid_from	date	
date_valid_to	date	N

electricity_meter		
id	int	PK
em_code	varchar(64)	
active	blob	
date_active_from	date	
date_active_to	date	N

energy_consumed		
id	int	PK
report_date	date	
report_time	time	
electricity_meter_id	int	FK
qunatity	decimal(12,2)	
is_daily_final	bool	

Source: Drkušić (2017) – Author edited

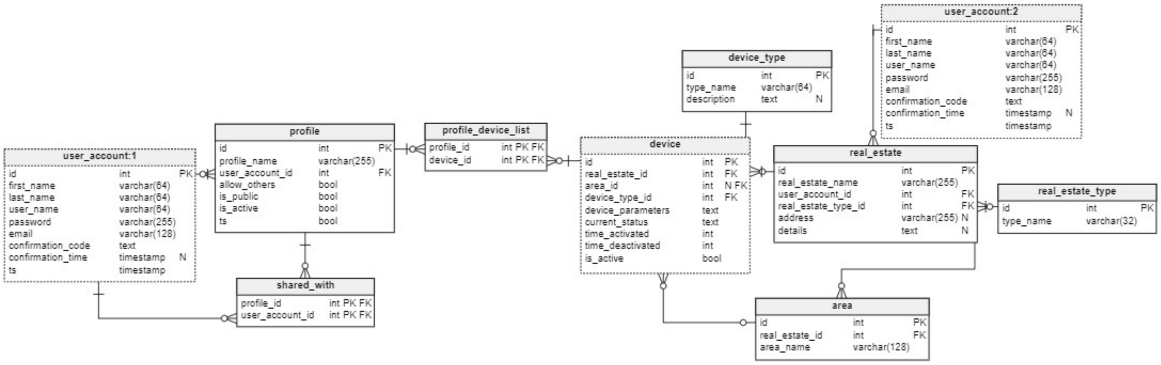
Appendix 4. Examples of data types collected via smart home device



Device	Functions of Device
Google Home	- Control smart home devices via voice commands, the display, or companion app - Provide visual services such as music and video playback, photo display,reminder, and search - Make calls between a smartphone and the Google Nest Hub using the duo app
Kasa Cam	- Watch live or record video - Detect motion and sounds - Two-way communication through the camera and companion app
SmartThings Outlet	- Power on/off - Measure power consumption
SmartThings Multipurpose Sensor	- Detect open or close status - Detect temperature - Detect vibration
SmartThings Motion Sensor	- Detect motion - Detect temperature

Source: Kim et al. (2020)

Appendix 5. Example of data table structure from a smart device



user_account:1		
id	int	PK
first_name	varchar(64)	
last_name	varchar(64)	
user_name	varchar(64)	
password	varchar(255)	
email	varchar(128)	
confirmation_code	text	
confirmation_time	timestamp	N
ts	timestamp	

profile		
id	int	PK
profile_name	varchar(255)	
user_account_id	int	FK
allow_others	bool	
is_public	bool	
is_active	bool	
ts	bool	

shared_with	
profile_id	int PK FK
user_account_id	int PK FK

profile_device_list	
profile_id	int PK FK
device_id	int PK FK

device_type		
id	int	PK
type_name	varchar(64)	
description	text	N

device		
id	int	PK
real_estate_id	int	FK
area_id	int	N FK
device_type_id	int	FK
device_parameters	text	
current_status	text	
time_activated	int	
time_deactivated	int	
is_active	bool	

user_account:2		
id	int	PK
first_name	varchar(64)	
last_name	varchar(64)	
user_name	varchar(64)	
password	varchar(255)	
email	varchar(128)	
confirmation_code	text	
confirmation_time	timestamp	N
ts	timestamp	

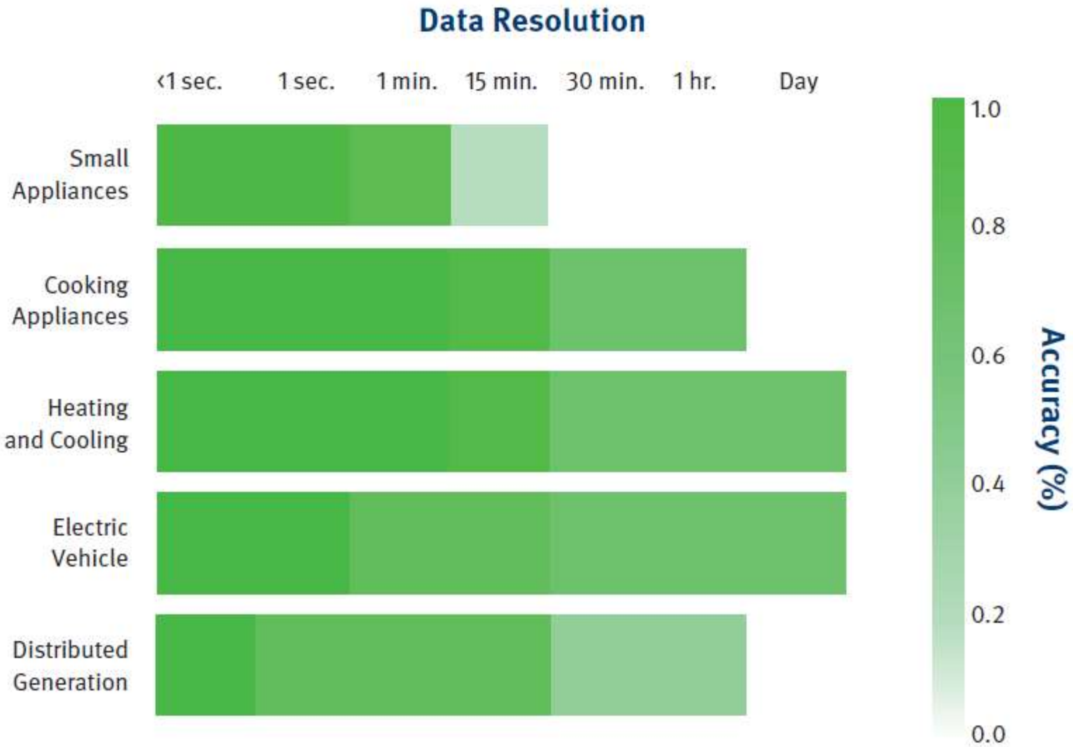
real_estate		
id	int	PK
real_estate_name	varchar(255)	
user_account_id	int	FK
real_estate_type_id	int	FK
address	varchar(255)	N
details	text	N

real_estate_type		
id	int	PK
type_name	varchar(32)	

area		
id	int	PK
real_estate_id	int	FK
area_name	varchar(128)	

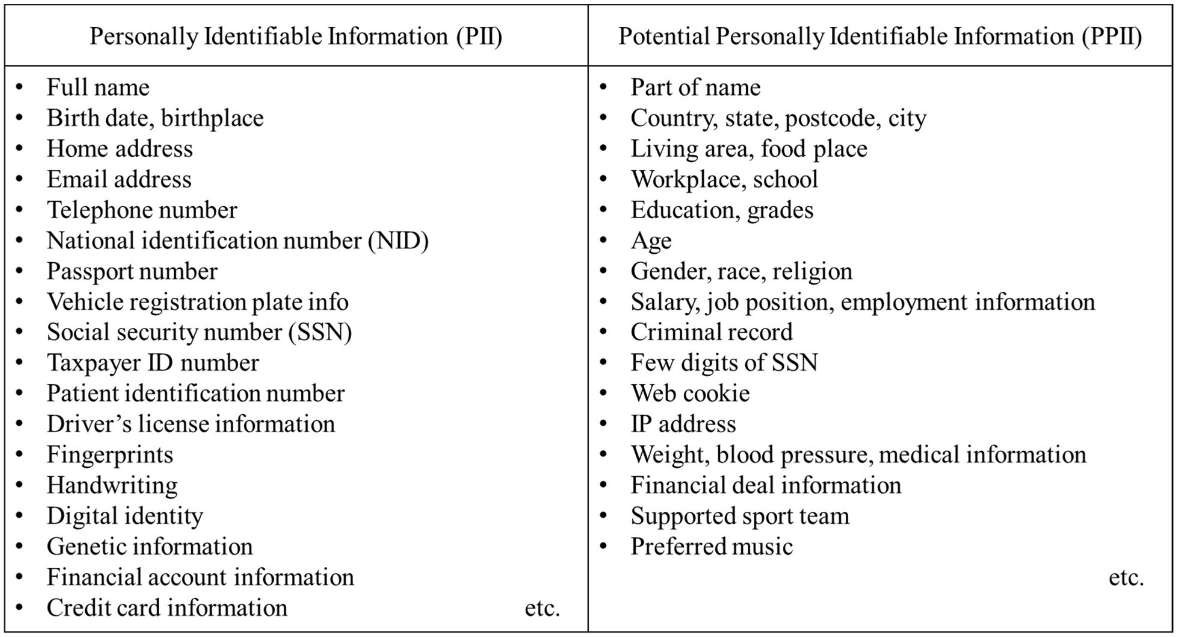
Source: Drkušić (2019) – Author edited

Appendix 6. Accuracy of precision of each interval of smart meter data by NILM



Source: Teng et al. (2022)

Appendix 7. Examples of Personally identifiable information (PII) and Potential personally identifiable information (PPII) by Onik et al. (2018)



Personally Identifiable Information (PII)	Potential Personally Identifiable Information (PPII)
• Full name • Birth date, birthplace • Home address • Email address • Telephone number • National identification number (NID) • Passport number • Vehicle registration plate info • Social security number (SSN) • Taxpayer ID number • Patient identification number • Driver's license information • Fingerprints • Handwriting • Digital identity • Genetic information • Financial account information • Credit card information etc.	• Part of name • Country, state, postcode, city • Living area, food place • Workplace, school • Education, grades • Age • Gender, race, religion • Salary, job position, employment information • Criminal record • Few digits of SSN • Web cookie • IP address • Weight, blood pressure, medical information • Financial deal information • Supported sport team • Preferred music etc.

Source: Onik et al. (2018) – updated by Author

Appendix 8. Smart devices that collect electricity consumption data and other collected data (excluding electricity consumption data and control history data)

Model	Developer	Product category	Data types (except electricity consumption data)						
			Temperature	Humidity	Motion	Light	Occupancy	Location	Other
GC-TBZ48 Z-Wave Programmable Thermostat	2Gig	Smart Thermostat	✓						
Smart Dimmer 6	Aeon Labs / Aeotec	Smart Plug							
Smart Switch 6	Aeon Labs / Aeotec	Smart Plug							
Z-Wave Smart Strip	Aeon Labs / Aeotec	Smart Plug							
smart thermostat	Alarm.com	Smart Thermostat	✓				✓		*1
Eversense	Allure Energy	Smart Thermostat	✓				✓		
Neo Pro	Ankuoo	Smart Plug							
WeMo Insight Switch	Belkin	Smart Plug							
ComfortChoice Touch	Carrier	Smart Thermostat	✓						
ComfortChoice Edge	Carrier	Smart Thermostat	✓						
ComfortChoice Legacy	Carrier	Smart Thermostat	✓						
COR	Carrier	Smart Thermostat	✓	✓					
Pearl Thermostat	Centralite	Smart Thermostat	✓	✓					
3-Series Lamp Module	Centralite	Smart Plug							
3-Series Appliance Module	Centralite	Smart Plug							
Azela Appliance Module	Centralite	Smart Plug							
Control4 Wireless Thermostat by Aprilaire	Control4	Smart Thermostat	✓	✓			✓		
Programmable Communicating Thermostats	Cooper Industries - EATON	Smart Thermostat	✓						
Load Control Switches	Cooper Industries - EATON	Smart Switch							
Smart Plug DSP-W110	D-Link Systems	Smart Plug							
Smart Plug DSP-W215	D-Link Systems	Smart Plug							
ecobee3	ecobee	Smart Thermostat	✓	✓	✓		✓		
Smart Si Thermostat	ecobee	Smart Thermostat	✓	✓					
Edimax SP-2101	Edimax	Smart Plug							
Eve Energy	Elgato	Smart Plug							
Emberplug AV+	Embertec	Smart Plug							
Emberstrip 8AV+	Embertec	Smart Plug							
Emberstrip PC+	Embertec	Smart Plug							
Emberstrip 8PC+	Embertec	Smart Plug							
Emberstrip 8AV+ Bluetooth Sensor	Embertec	Smart Plug							
Sensi Wi-Fi Thermostat	Emerson	Smart Thermostat	✓						
Smart Energy Thermostat (EE542-1Z)	Emerson	Smart Thermostat	✓						
Pioneer Smart Thermostat	Energate	Smart Thermostat	✓						
Energate HolHom (foundation thermostat and DR gateway)	Energate	Smart Thermostat	✓						
HōlHōm Smart Plug	Energate	Smart Plug							
Wired Load Control Switch	Energate	Smart Switch							
Glass Series T-100-H Thermostat	Evolve	Smart Thermostat	✓						
Glass Series T-1500 Thermostat	Evolve	Smart Thermostat	✓						
Glass Series TB-200 Thermostat/Temperature Sensor	Evolve	Smart Thermostat	✓						
T-100-H Thermostat	Evolve	Smart Thermostat	✓						
T-1500 Thermostat	Evolve	Smart Thermostat	✓						
TB-200 Thermostat/Temperature Sensor	Evolve	Smart Thermostat	✓						
A1730 Thermostat	Fidure	Smart Thermostat	✓						
First Alert Onelink Thermostat	First Alert	Smart Thermostat	✓						
Plug-In Smart Switch (ZigBee)	GE	Smart Plug							
Plug-In Smart Dimmer (ZigBee)	GE	Smart Plug							
In-Wall Smart Switch (ZigBee)	GE	Smart Switch							
In-Wall Smart Dimmer (ZigBee)	GE	Smart Switch							
Lyric	Honeywell	Smart Thermostat	✓	✓			✓		
Thermostat (YTH8320ZW1007/U)	Honeywell	Smart Thermostat	✓						
Wi-Fi Smart Thermostat RTH9580WF	Honeywell	Smart Thermostat	✓	✓					
Wi-Fi 7-Day Programmable Thermostat RTH6580WF	Honeywell	Smart Thermostat	✓						
Wi-Fi 7-Day Programmable Touchscreen Thermostat RTH8580WF	Honeywell	Smart Thermostat	✓						
Wi-Fi Smart Thermostat with Voice Control RTH9590WF	Honeywell	Smart Thermostat	✓	✓					
VisionPRO Wi-Fi 7-Day Programmable Thermostat	Honeywell	Smart Thermostat	✓	✓					
Smart Thermostats	Insteon	Smart Thermostat	✓	✓					
Lumina RF Programmable Thermostat	Leviton	Smart Thermostat	✓	✓					
ThinQ Super-Capacity French Door Refrigerator	LG	Smart Appliance							*2 *3
Smart ThinQ Washer	LG	Smart Appliance							
Smart ThinQ Dryer	LG	Smart Appliance							
Smart ThinQ Oven	LG	Smart Appliance							
Ls 60i Smart Thermostat	Lockstate CONNECT	Smart Thermostat	✓						
Ls 90i Smart Thermostat	Lockstate CONNECT	Smart Thermostat	✓	✓					
Iris Smart Thermostat	Lowes	Smart Thermostat	✓						
Iris Smart Plug	Lowes	Smart Plug							
GEO 7-Day Wi-Fi Programmable Thermostat in White	LUX	Smart Thermostat	✓						
Nest Learning Thermostat	Nest	Smart Thermostat	✓	✓	✓	✓	✓	✓	
Lightpad Dimmer	Plum	Smart Switch							

Appendix 8 continued

Model	Developer	Product category	Data types (except electricity consumption data)						
			Temperature	Humidity	Motion	Light	Occupancy	Location	Other
Thermostat CT 50 + Wi-Fi Module	Radio Thermostat of America	Smart Thermostat	✓						
Thermostat CT 32 + ZigBee module	Radio Thermostat of America	Smart Thermostat	✓						
Thermostat CT 32 + Z-Wave module	Radio Thermostat of America	Smart Thermostat	✓						
Thermostat CT 80 + WiFi module	Radio Thermostat of America	Smart Thermostat	✓	✓					
Thermostat CT 80 + ZigBee module	Radio Thermostat of America	Smart Thermostat	✓	✓					
Thermostat CT 80 + Z-Wave module	Radio Thermostat of America	Smart Thermostat	✓	✓					
Thermostat CT 100	Radio Thermostat of America	Smart Thermostat	✓						
Thermostat CT 101	Radio Thermostat of America	Smart Thermostat	✓						
Thermostat CT 110	Radio Thermostat of America	Smart Thermostat	✓						
TZ 45 Thermostat	RCS Technology	Smart Thermostat	✓						
TZB 45 Thermostat	RCS Technology	Smart Thermostat	✓						
TE45 RDS Thermostat	RCS Technology	Smart Thermostat	✓						
TW 45 Thermostat	RCS Technology	Smart Thermostat	✓						
Wiser Air Smart Thermostat	Schneider Electric	Smart Thermostat	✓	✓					
Smappee Comfort Plug	Smappee	Smart Plug							
HA Dual-Relay Controller ZBLC15 (4033A)	Smartenit	Smart Switch							
HA Metering Dual-Load 30A Controller ZBMLC30 (4040B)	Smartenit	Smart Switch							
HA Metering Single Load Controller ZBMLC15 (4034A)	Smartenit	Smart Switch							
HA Metering Smart Plug ZBMPlug15 (5010Q)	Smartenit	Smart Plug							
SmartPower Outlet	SmartThings	Smart Switch							
PICOwatt	Tenrehte	Smart Plug							
ThinkEco Modlet Starter Kit	ThinkEco	Smart Plug							
WiFi smartAC Kit	ThinkEco	Smart Plug							
SmartAC Kit (For Gateway)	ThinkEco	Smart Plug							
Modlet BN WiFi	ThinkEco	Smart Plug							
ComfortLink™ II XL950	TRANE	Smart Thermostat	✓						
ComfortLink™ II XL850	TRANE	Smart Thermostat	✓						
XL824	TRANE	Smart Thermostat	✓						
XL624	TRANE	Smart Thermostat	✓						
Bright 700-10 - Thermostat	Tri Cascade	Smart Thermostat	✓						
i-bright7	Tri Cascade	Smart Plug							
Bright 20-10 Outlet - Plug	Tri Cascade	Smart Switch							
Bright 60-10 Switch - Light	Tri Cascade	Smart Switch							
ColorTouch	VENSTAR	Smart Thermostat	✓	✓					
Voyager	VENSTAR	Smart Thermostat	✓						
UFO Power Center	Visible Energy	Smart Plug							
Smart Dryer	Whirlpool	Smart Appliance							*2 *3
Smart Washer	Whirlpool	Smart Appliance							*2
Smart Dishwasher	Whirlpool	Smart Appliance							*2
Smart Refrigerator	Whirlpool	Smart Appliance							*2 *3
load monitoring smart plug	Wittech WiTenergy	Smart Plug							
York® Affinity™ Residential Communicating Control	York	Smart Thermostat	✓						
Thermostat	Zen	Smart Thermostat	✓						
Zuli SmartPlug	Zuli	Smart Plug							
MCP39005	Microchip Technology	Smart Circuit breaker							
MCP39006	Microchip Technology	Smart Circuit breaker							
ADE9000	Analog Device	Smart Circuit breaker							

*1 Camera data
*2 Temperature inside the appliance
*3 Humidity inside the appliance

Source: Author - from data from Pritoni et al. (2018) and Alden et al. (2019)

Appendix 9. Non-exclusive license

A non-exclusive license for reproduction and publication of a graduation thesis[3]

I ___________________________________ (*author's name*)

1. Grant Tallinn University of Technology free licence (non-exclusive licence) for my thesis

____

____

____,

(*title of the graduation thesis*)

supervised by____,

(*supervisor's name*)

1.1 to be reproduced for the purposes of preservation and electronic publication of the graduation thesis, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright;

1.2 to be published via the web of Tallinn University of Technology, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright.

2. I am aware that the author also retains the rights specified in clause 1 of the non-exclusive licence.

[3] *The non-exclusive licence is not valid during the validity of access restriction indicated in the student's application for restriction on access to the graduation thesis that has been signed by the school's dean, except in case of the university's right to reproduce the thesis for preservation purposes only. If a graduation thesis is based on the joint creative activity of two or more persons and the co-author(s) has/have not granted, by the set deadline, the student defending his/her graduation thesis consent to reproduce and publish the graduation thesis in compliance with clauses 1.1 and 1.2 of the non-exclusive licence, the non-exclusive license shall not be valid for the period*

3. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights, the rights arising from the Personal Data Protection Act or rights arising from other legislation.

_______________ (date)