information

Review

Data Ownership: A Survey

Jad Asswad [1,*] and Jorge Marx Gómez [2]

[1] Data Integration & Processing Group, R&D Devision Energy, OFFIS Institute for Information Technology, 26121 Oldenburg, Lower Saxony, Germany
[2] Department of Computer Science, Division of of Business Informatics/VLBA, Carl von Ossietzky University of Oldenburg, 26129 Oldenburg, Lower Saxony, Germany; jorge.marx.gomez@uol.de
* Correspondence: jad.asswad@offis.de; Tel.: +49-441-9722-186

Abstract: The importance of data is increasing along its inflation in our world today. In the big data era, data is becoming a main source for innovation, knowledge and insight, as well as a competitive and financial advantage in the race of information procurement. This interest in acquiring and exploiting data, in addition to the existing concerns regarding the privacy and security of information, raises the question of who should own the data and how the ownership of data can be preserved. This paper discusses and analyses the concept of data ownership and provides an overview on the subject from different point of views. It surveys also the state-of-the-art of data ownership in health, transportation, industry, energy and smart cities sectors and outlines lessons learned with an extended definition of data ownership that may pave the way for future research and work in this area.

Keywords: data ownership; data sovereignty; big data; privacy; security; internet of things; survey; state-of-the-art

Citation: Asswad, J.; Marx Gómez, J. Data Ownership: A Survey. *Information* **2021**, *12*, 465. https://doi.org/10.3390/info12110465

Academic Editor: Barbara Pes

Received: 31 August 2021
Accepted: 20 October 2021
Published: 10 November 2021

1. Introduction

We are living nowadays in the era of big data, where the amount of data is increasing steadily on a minute basis. According to an article published by Forbes in 2018, there are 2.5 quintillion bytes of data generated daily, and this number is growing rapidly [1]. The Industrial Development Corporation (IDC) estimated back in 2014 that by 2020 we will have 44 trillion gigabytes of data, 10 times more than we had in 2013, which means there will be around 5200 gigabytes of data per person [2]. Data is surrounding us in all aspects of our lives and is being generated from numerous sources like social media, information and communication technology (ICT) devices, as well as from sensors, machines and other internet of things (IoT) devices. IDC estimates that in 2025 there will be 79.4 zettabytes of data generated from 41.6 billion connected IoT devices from all around the world [3]. Along with this growth of data, the interest of acquiring and exploiting the data is increasing, raising questions about the privacy and security of data and even the question of who owns the data and how the ownership of data can be preserved.

The term "Data Ownership" has been mentioned in the literature since 1981 [4–7]; nevertheless, there is still no single standardized definition of the term that covers all perspectives on the issue. From a data and information management point of view, data owners are defined as The ones responsible and accountable for the quality of data. On the other hand, data ownership is defined often as the possession of complete control over the data and its rights, including the right to grant rights over the data to others. But during the years and coherent with the disrupting technologies at a time, the ownership of data has been seen and interpreted differently. From the rise of databases and the exploitation of networks, to the adoption of information systems and the spread of data through the internet, to the explosion of data through social media, to the vision of smart cities and the revolution of the Internet of Things, the issue of data ownership had many shapes and numerous implications. Even by examining the topic, the diverse fields of research had

different perspectives on the issue: some focused on the issue from an ethical and moral perspective [8–12], and some focused on the legal and juridical implications [13–16], while others tackled the problem from a technical point of view, seeking a feasible solution to the problem [17–19].

Due to the lack of a clear definition of data ownership, the term has been associated and even confused with data governance, data management, and aspects related to these domains. As to be seen in this study, a wide range of data ownership literature address one or multiple aspects of data governance and data management, like data sharing, data security, data privacy, data integrity, data sovereignty, data stewardship, data transparency, data traceability and others. The presented work does not argue about what to be considered a part of data ownership research, but rather highlights the wide spectrum of aspects that could be perceived about data ownership. Mainly, in light of the fact that the absence of a unified definition of the term is letting the door open for different interpretations and embodiments of data ownership.

The subject of data ownership is gaining in the recent few years an increasing attention. From the one hand, it affects besides the privacy and the security of data, the sharing of the raising amount of data, which presents huge opportunities across and beyond the boundaries of any organizational body. On the other hand, the emergence of technologies, that facilitate the securing and exploitation of data, like the toolsets of big data and data science and the mechanisms of blockchain, contributes also to the recent interest in the subject. The availability of instruments that preserve and guarantee sovereignty and ownership over the data promotes innovation that includes data owners, consumers, prosumers and all kind of stakeholders that might have interests over the data.

In addition to the importance that data ownership presents to any business or organization, ensuring the ownership of users over their data builds the trust between them and businesses and enable them to be part of the process, the product or the service, in which their data is part of. On the other hand, it empowers users to have full control over their data as the real owner of the data.

This work aims to give an overview over the subject of data ownership and tries to reflect knowledge from the different observations concerning this matter. The study tries to cover the problem of data ownership from different perspectives. At the beginning, a general background of data ownership and its technicality will be given. Subsequently, the issue of data ownership will be reviewed from a juridical point of view and especially in European/German regulations.

The main part of the work addresses the state-of-the-art concerning data ownership. For this purpose, five main sectors that deal intensively with the issue will be reviewed, namely: health, transportation, industry, energy, and smart cities. The study concludes by summarizing the insights derived from the conducted state-of-the-art overview and by discussing the comprehension of the subject of data ownership and its definition, as well as suggesting a possible course of action for handling the issue.

2. What Is Data Ownership?

2.1. Overview

Despite the absence of a unified view on the subject, data ownership is widely considered as the possession of complete control over the data and its rights including, but not limited to access, creation, generation, modification, analysis, use, sell, or deletion of the data, in addition to the right to grant rights over the data to others. However, this view simplifies the matter and oversees the entanglements of the issue that has no straightforward answer to its complications and cannot be summarized in one definition.

One of the major complications is the perception of data ownership and the willingness of data holders to share their data. Nowadays with the help of the constantly evolving technologies, companies and organizations can generate insights and values from raw data, and when the data has value, the issue of data ownership emerges [20]. Data ownership represents a major concern that influences the beneficial use and share of information. In

organizational context, information sharing behavior is crucial and strategic. However, such behavior implies the willingness of concerned parties to share information. From a philosophical point of view, one of the major factors affecting the sharing behavior is the perceptions of ownership by the stakeholders, which are more likely to affect individuals' behavior more than the legally forced ownership [21]. In order to argue about the perception of data ownership, the basic concept of property ownership needs to be understood. According to Hart [21], ownership is the rights to use, control the use, and remain in control of whatever is owned, without the interference from others. This concept of ownership might be applicable to data in many cases, yet data can differ distinctly from the traditional sense of property and therefore it requires further deliberations. Dissimilar to typical physical property, data is often characterized by intangibility, non-exclusivity, usage tolerance (do not deteriorate through use), and age susceptibility (do often deteriorate with age). These unique characteristics of data are in contrast to typical items of properties entangle the defining of data ownership and complicate the answer of the question: whose ownership perceptions is relevant [21].

From a marketing point of view, Gabisch and Milne [22] address the perception of data ownership by asking the question: how can the expectations of the consumers regarding privacy and data ownership be affected by compensating them for their personal information? Despite the fact that the expectation of the consumers might vary according to the sensitivity of the shared information with the marketers, the study found that receiving compensation reduces the expectations of privacy and security, especially when the compensation has monetary nature [22].

The issue of data ownership is rather relevant for organization for various reasons, including the aim to create an information sharing behavior within the organization, facilitating managerial decision-making based on data ownership guidelines, and establishing technological advancement in the areas of information systems and data management systems. Nevertheless, with the emergence of social media and IoT devices, the question of data ownership has extended beyond organizational context and has reached individual level, raising concerns about the security and the privacy of personal-related data and introducing a new conflict concerning data ownership.

Emerging topics such as Human Data Interaction (HDI) are addressing these concerns by putting the people at the center of data driven applications and supporting them in the management of their personal data [23,24]. HDI addresses also the ownership of IoT data raising the question of who owns the IoT data and who should access the data that is produced by the people [23]. Despite the wide realization of the problem in the IoT community, there is still no unified structure for addressing the ownership of IoT data. Mashhadi et al. [23] for example proposed in their work three different models for HDI in IoT domain with the aim to find a suitable data ownership model based on the sensitivity of the collected data and the consent that is granted to the users to interact with their data.

On the same subject, Mineraud et al. [25] evaluated in a gap analysis the IoT landscape concerning various issues including ownership, privacy and security of IoT data with the aim to improve and optimize the limitation of the current IoT ecosystems. The analysis highlighted the fact that for most IoT platforms data ownership is a major concern and it is rarely fully guaranteed. In most platforms, data is received as raw data and it is stored without encryption and without adequate security measures. According to the authors, full data ownership is only granted by the solutions where the user's data is stored locally. Based on that, the authors suggested that IoT platforms should provide mechanisms that allow the user to have full ownership and competent visibility over the data. In which, the owner will have the control over the raw data and be able to determine who or which platforms or resources can access the data [25].

Furthermore, the age of IoT introduced a new conflict as data generated by IoT devices concerns numerous stakeholders who aim to claim ownership over the data. In the example of networked cars, the data generated by a car summarizes information that interests the car owner, the car user, the manufacturer of the car and the manufacturer of the sensor, the

navigation service provider, the insurance company, the internet provider, in addition to the authorities [26]. Farkas highlights in his study this conflict of interests, raising the question of ownership, including the control over the collection, transfer and access of IoT data. In addition, he analyzed the law governing ownership and explored possible necessity to adapt new regulations in this regard [26]. According to his research, the adaptation of new law for governing IoT data entails negative and positive sides at the same time. On the one hand, a new law would facilitate the emergence of new business models where data would be better protected and it would regulate the market leading to more efficiency dealing with data. On the other hand, such regulation might limit the sharing and the free access of information, and probably would lead to monopolies that would demolish innovations [26]. However, the legal aspect of data ownership is another complication that will be further investigated later in the course of this study.

Saarikko et al. [27] approach also in their paper the conflict of ownership allocation, emphasizing on the idea that there is no single approach suitable for all cases and that we have to accept the fact that for now, data ownership will be tackled differently in different contracts and in different businesses and countries. They state in their study that in some cases it is seen that the person whose actions generated the IoT data has the right of ownership, where in other cases the ownership is considered to be the right of the party who collected and analyzed the data, which is in many cases the manufacturer of the device that generated the data. In addition to the argument that the information resulted from analyzing the data should be owned by the supplier [27].

The issue of ownership allocation remains unfortunately ambiguous as the question of who should own the data remains without a good answer. Should the one whose information contained in the data, or the one that created the data by collecting the information of the user own it? As both of them and maybe others can be defined as the owner of the data. Bertino [28] argues in her paper about the benefit of replacing the concept of data ownership with the concept of data stakeholders. In this case, a data item can be affiliated to multiple stakeholders that can access and use the same data source. However, Bertino states the fact that in some cases the different interested persons might be unaware of the existence of the other stakeholders, which can lead to conflicts at various levels [28].

Arguing also against the basic idea of data ownership in research and specifically in ecology research, Hampton et al. [29] encourage the shift from data ownership thinking towards data stewardship thinking, in order to increase the transparency and the sharing capabilities of ecology research data and to change the mindset in the field of ecology towards open science. Ecologists and researchers in general treat their data as properties, limiting the possibility to share data and increasing the possibility of repeating the same mistakes by other researchers conducting the same research. Through the shift to open science, Hampton et al. are advocating the researchers to see themselves as collectors and the ones that share their findings and their data with the scientific community rather than the owners of the data [29].

The complications of data ownership extend further to include the integration of data ownership concepts in IT landscapes. Organizations and companies aspire to establish one master data that hosts data, processes and information systems under the umbrella of one master data management (MDM). Besides the technological challenges of achieving such framework, data-related challenges remain the main ones and are mostly poorly tackled by organizations. A clear data ownership definition is the first step and one of the main preconditions to implement and establish one master data management. Well established data ownership concepts ensure the quality of the data and affect the attitude of employees toward the preservation of the integrity and the quality of data [19].

However, the integration of data ownership in information systems or database management systems is not straightforward and it entails a deep understanding of the components and the development of the system and of the way that the system will be controlled accordingly. Data ownership is mainly involved in the control dimension of a database architecture. Within the control dimension it is often confused between

"ownership" and "use" privileges. The usage privileges are concerned about the right to access, modify, create and manipulate data. Where ownership is about the right to control these privileges that are given to others. In this sense, even if the user is granted the same rights as the owner, the authority to grant, alter or withdraw these privileges is what distinguishes the owner. The previous distinction differentiates between the usage and the ownership privileges, but it does not determine exclusively the owner of the data. In some cases, a higher or maybe a centralized authority can interfere or revoke the ability of the owner to grant privileges, which means that the owner is not in all cases the exclusive owner of the data, which leaves the door open for further interpretations of the answer of who really owns the data [17]. According to Van Alstyne et al. different ownership results are derived from different incentive requirements for the data and the ownership of this data. In their work they introduce a theoretical framework that presents seven normative principles to help approaching the issue of data ownership and guide organizations addressing these incentives [17].

Fundamentally, the technical implementation of data ownership concepts is as heterogeneous as the issue itself. Since the 80's, the idea of enabling two parties to jointly compute a function over their inputs while preserving the privacy of their inputs was discussed. Secure computation and privacy-preserving computation inspires a verity of protocols and frameworks, that was and still relevant in securing and preserving the privacy of contributions in the communication and the jointly computation between two or more parties [30–34]. The need to preserve the privacy in data processing extends to include data mining research. Privacy-preserving data mining aims to develop data mining models without accessing or revealing the private information of the data sets. The basic idea is to use algorithms and techniques to exclude or modify personal, sensitive or private information in the computation process in a way that preserve the privacy of the data [35–37].

With the rapid growth of the amount of data, concepts like data market, data marketplace and data trading platform which enable the trading and the exchange of data between parties have emerged. One of the major challenges of these concepts is to provide an efficient data trading while preserving the security and the privacy of the trading [38]. Nonetheless, promising studies are proposing solutions that introduce secure data trading and exchange platforms, that increase on the one hand the utility of data and preserve on the other hand the privacy of the trading data between stakeholders [39,40].

Other evolving technologies as well as new technologies like blockchain are also opening the door for new possibilities to tackle emerging challenges. Despite the novelty of the blockchain technology, its characteristics provide possible solution for data problems, including the issue of data ownership. Karafiloski and Mishev address in their literature review the importance and the possibilities of blockchain, highlighting the fact that the decentralized nature of blockchain solutions can be a way to approach issues like privacy, security, data ownership and the control over the data without the interference or the need of a third party [41].

2.2. Data Ownership from a Juridical Point of View

The legal aspect of data ownership is complicated and not easy to discuss. As seen earlier in this study, the handling of data ownership differs in relation to the case, the country, the business and even the context in which a data ownership concept is needed. In addition, as of today there is still roughly no clear regulation that concerns one to one the problem of data ownership. Nevertheless, this part of the paper tries to briefly discuss data ownership from a juridical point of view and gives an overview of the current state of regulations and upcoming changes in this concern in Europe and Germany.

The German Federal Ministry of Transport and Digital Infrastructure carried out in 2017 a detailed study about ownership regulation for mobility data from technical, economic and legal perspectives [16]. The study summarizes comprehensively the current state of the law in Germany concerning data ownership in general and not only for mobility

data. The study discusses three level of data allocation to an authorized person. The first level is the data ownership itself and it concerns the complete assignment of data along full authority over the data to an owner. The second level is the encumbrance of data ownership and it deals with restrictions in favor of a third party or the public who can claim right over the data against the exclusive right of the owner. The third and last level is the assignment of power of disposal and right of use, and it concerns the granting of the right to use data by the data owner to a third party, taking into account the limitation through restrictions over the data.

In order to investigate if the current law supports any of these three levels and especially the basic assignment of data to a legal subject, the study gives first an overview of the German constitutional protection of data. As the fundamental rights in the German constitution are defensive in their nature, they provide protection to legal positions from intrusions, rather than addressing explicitly the actual assignment of ownership. The German basic law (GG) [42] tackles however the issue of guaranteeing ownership, where it provides protection if the data are ownership-capable in the constitutional sense. Which means it can protect the physical data storage medium on which the data in question can be stored. The basic law protects also trade and business secrets, the persons affected by the process of data collection and it guarantees the confidentiality and integrity of information technology systems and the secrecy of telecommunications [16].

The study continues beyond constitutional law to discuss the assignment of data ownership from the point of view of simple laws like data protection law, copyright law, criminal law, fairness law and general civil law.

Concerning data protection, the European General Data Protection Regulation (GDPR) [43] was adopted in 2016 and became legally binding since 25 May 2018 across Europe. Unfortunately, the GDPR does not address data ownership specifically, but it represents a crucial guideline for basic privacy and data protection. It brings as well to the table the Charter of Fundamental Rights of the European Union (CFR) [44] and the European Convention on Human Rights (ECHR) [45], which guarantees the general right to respect for private and family life, housing and communication and the explicit right of protection of personal-related data. GDPR in addition to the German federal data protection act (BDSG) [46] and the German telemedia act (TMG) [47] embody a variety of basic data privacy and protection principles. They address the requirements for data processing and the handling of personal-related data. In addition, they protect the use of collected data and adopt the principle of transparency concerning the nature and the scope of data processing. The data protection laws embrace also the necessity principle of "Data Minimization" by forcing the anonymization or the deletion of data, once it is no longer required or needed in personal-related form anymore. Furthermore, they support the principles of "privacy by design" in the selection and design of data processing systems including the use of anonymization and pseudonymization. On personal level, they protect widely the rights of the concerned persons, to information, correction, deletion and blocking, in addition to the right of data portability and the right to be forgotten. Data portability grants data subjects the right to freely extract their data from one controller and transfer it to another. Despite the importance of data portability in granting the data owner the control over where to have his or her own personal data, several interpretation of data portability in the GDPR are possible, which might lead to confusions that require further clarification [48].

Ultimately, data protection laws and principles provide a sort of protection for individuals, however, it is limited on multiple aspects. On the one hand, they are only relevant for personal-related data. On the other hand, they are concerned mainly with the right of defense due to its protective nature and it does not grant the power of disposal over the data to the individuals. As a result, data protection law is not directly concerned with the ownership of data and it cannot be explicitly used for data assignment and ownership allocation.

The German copyright law (UrhG) [49], however, provides its protection regardless of whether it is about personal-related data or not. It acknowledges personal intellectual

creations and protects the proprietary and intellectual interests of the creators. According to UrhG, the originator is the creator of the work and the one who has the copyright, even if he is an employee or a ghost writer, and not the employer or the contractors like it is the case for example in the USA. In fact, the employer needs to acquire a derived right of use in order to be able to use the work. The copyright law includes also related rights (neighboring rights) that protect work that are not "creative" but serve as services and achievements of a different kind. These rights are originated with the funder or the investor and not with the person who performs the work, like for example, the protection of scientific expenditure, performing artist, sound carrier manufacturer, broadcasting companies, database producer and press publisher.

Similar to the copyright law, the German criminal law (StGB) [50] does not differentiate between personal or non-personal related data. It protects however the entitled party against illegal use or interference on the used data. The criminal law also protects also the secrecy of the entitled party concerning the information contained in the data and punishes unauthorized access to data. Despite the protection provided by the criminal law, it is difficult to assess the limitation of the crime if the data is not assigned to an authorized person who might have an owner-like power of disposal. Unfortunately, such allocation of ownership is not regulated by the criminal law, which assumes the preexistence of such allocation. According to jurisdiction "scriptural act", the power of data disposal belongs to the one who stored the data. But on the other hand, Welp [51] views the "scriptor" as the person who generates data and causes the storage or the transmission of data, by entering data, storing it or triggering the input of external data. Therefore, according to this point of view, the power of disposal is justified when the data is generated and saved and is not determined based on the ownership of the data carrier or the personal-related content of the data. As a result, the criminal law does not grant a comprehensive ownership right and it does not contain regulation about positive rights of use, but rather it protects the entitled person against limited list of crimes from a third party.

The German Unfair Competition Act (UWG) [52] is responsible for the protection of business and trade secrets. It secures the confidentiality of secret business-related information by protecting the secrets of companies from betrayal by its own employees, from industrial espionage, secrets keeping, and the enticement or solicitation of betrayal. However, the protections under this law concern only business and trade secrets data and it is irrelevant for personal-related data.

The German civil law (BGB) [53] looks at the issue of data allocation and assignment from different angles. It deals with the ownership of objects, and that does not include data, as it is not considered an object. According to BGB, objects exist only in physical forms: solid, liquid or gaseous. That means that this law protects only the physical carrier of data and not data itself. A similar problem can be seen in the tort law of BGB, which deals with the violation of proprietary rights of ownership, and therefore it can only protect data partially due to the missing physicality of data. The tort law protects however against the manipulation of the data carrier and hence it protects indirectly the integrity of the embodied data. That means that any deletion or modification of the data stored in the data carrier constitutes a damage to property. According to BGB, the use advantages of an item are considered a type of the utilization of the item. Based on this, the owner of the item is assigned the first right of use, and he can in his turn transfer the right of use to a third party. Unfortunately, data cannot be considered a use advantage of an item, as data is fundamentally different from the typical use advantages. Usually, the use advantages are linked to a subject, are volatile, and resulted from dealing with the subject, whereas data exists permanently, can be reproduced and transferred, and is subject to various other regulations. As a result, the German civil law does not contribute directly to the assignment of ownership to legitimate owner. It provides integrity protection of ownership of the data carrier through the tort law and it protects against the violation of other regulations.

Summarizing the current state of the law in Germany, the study states that different laws generally protect data. Nevertheless, these laws differ in their protections and might

contradict each other under certain circumstances and depending on their scope and their point of view of the assignment of the power of disposal. The study shows also that the current state of law in Germany does not cover data ownership in principle but rather it provides incomplete divergent protection rights [16].

The regulations of data ownership in Europe faces many difficulties and various literature highlighted the uncertainty in dealing with ownership problems in general and the ownership of personal data as a crucial and ethical case [9,14,15]. Van Asbroeck et al. [14] mention in their overview of the EU legal framework regarding data ownership, the lack of clear regulation in regards of data ownership in Europe where personal data cannot be owned but rather protected from access or use by anyone other than the data subject to whom the data refer. They stress also the complexity of granting rights over the data as different stakeholders claiming ownership over the data as they played a part at some point in the value chain of the data, by for example creating, analyzing, altering or simply using the data. As sort of a solution, van Asbroack et al. suggest the introduction of non-exclusive ownership in data sets in form of traceability obligation that grants the legal protection against abuse or manipulation of the data [14].

Janeček [15] confirms in his study the complexity of the issue and the complications that might arise while regulating the ownership of personal data. He states in his research that despite the effort for protecting personal data (like the GDPR) and the recent debates about regulating the right of data ownership at the European Level, there is still uncertainty and no clear regulations that deal with the aspects of data ownership. There is even a confusion of what is considered personal data and what is considered non-personal data, especially with the fact that analyzing non-personal data can generate personal data. Art. 4 (1) of the GDPR refers to personal data as the information that relates to an identified or identifiable person "data subject". However, According to Janeček, the GDPR refers in other occasions to personal data as personal information, which leads to confusion regarding the difference between data and information and the distinction between the ownership of personal data versus the ownership of personal information. He highlights that in legal debates; the deference between information and data is illustrated as the distinction between the form of information (the syntactic level of information) and the meaning of information (the semantic level of information). From the legal point of view, semantic information cannot be protected, as that would restrict the free access to information. On the other hand, syntactic information in form of sequenced data can be protected. However, data can be processed differently, generating a variety of results that some of them might contain personal information. In this case, the same data can be considered both personal and non-personal depending on its context. Nevertheless, some data are originally personal data, if it contains personal information intrinsically (for example: the human DNA). This data is excluded from the ownership debates for legal and ethical consideration. And that leaves the door open for debates regarding the data that is personal but only extrinsically (like for example the length of a DNA) [15].

Janeček discusses also in his research the two legal approaches of answering the question of "Why the law should allow someone to own something?"; the top-down approach (the positivist ownership) and the bottom-up approach (the natural law approach to ownership). The top-down approach suggests that the law creates the ownership and that ownership would not exists without it. Where the bottom-up approach argues that ownership exists regardless of the legal system and that, the responsibility of the law is to protect ownership. Nevertheless, both approaches cover the four elements supporting ownership, which are control, protection, valuation, and allocation of the owned subject. The element of control discusses the full control over the data, by the meaning of access, store, sharing, sell, alter, process and even deletion of the data. The element of protection deals with the legal protection of the interest over the data against infringements and interference. The element of valuation highlights the importance of transparent valuation of data as a worthy subject to be owned. The element of allocation deals with the question of to whom should the ownership rights over the data be allocated. Janeček explores

both approaches in the context of ownership of personal data in IoT and highlights the advantages and disadvantages of each of them in relation to the four elements of ownership. In his study, he addresses the limitations of both approaches and suggests a revised bottom-up approach that fits the problem of ownership of personal data emphasizing the necessity to differentiate between information and data and between intrinsically and extrinsically personal data in order to apply ownership rights regulations on personal data. As for data allocation, the answer is not straightforward and should be derived from privacy considerations, which leads to the conclusion that the data subject is not always the one who should own the data [15].

As to be seen in this overview of the juridical and legal aspect of data ownership, the status quo of regulations and laws does not cover the entire complications of the issue. Despite the immense efforts in the last years, especially in Europe, towards the protection of the security and privacy of data and towards granting individuals more rights over their personal data, including the right to be forgotten, the questions of data ownership remain without clear answers. Nevertheless, the search for viable solutions and concepts can be noticed in multiple domains and it is explored in the following chapter.

3. Data Ownership State-of-the-Art

In order to gain a wider insight and better understanding of data ownership, this part of the paper explores previous studies, projects, concepts and even technical implementations tackling the subject of data ownership directly or indirectly. In this regard, the state-of-the-art from different sectors is presented and examples from five major sectors that are relevant to this research like health, transportation, industry, energy, and smart cities are highlighted.

3.1. Data Ownership in Health Sector

Health data is a sensitive issue in health sector as it deals mainly with data that is personal in its nature, in addition to other sensitive data. The need of data protection and regulations addressing the privacy of health data have been discussed extensively in the literature and different points are worth mentioning in this concern. On one hand, the necessity to protect personal health data (PHD) and to preserve the patients right to own their health data. On the other hand, the importance of facilitating the access to health data and data sharing for research and prognosis purposes and the support of data-driven medicine.

In a systematic review, Van Panhuis et al. [54] address the issue of the use and share of public health data and identify barriers that prevent data sharing in the health sector. The paper summarizes twenty potential barriers and categorizes them in six different categories: technical, motivational, economic, political, legal and ethical. Focusing on the possession and the privacy of health data, the legal category addresses two main barriers derived from 18 reviewed papers. First, the ownership and copyright barrier, that poses the question "who owns public health data?" the public or the agencies that collect the data, and argues about the use of copyrights to restrict rather than spread access to health data. The second barrier deals with the protection of privacy, where several restrictive policies are applied on health data. Such policies are forced due to privacy concerns, as it is difficult in many cases to distinguish between personal-related data and anonymized data. That led to the use of aggregated data (without any personal identifiers), which can be insufficient for certain applications [54]. In a similar study, Bietz et al. [55] identify data ownership as one of six challenges to use PHD for research purposes. According to their study, despite the willingness of individuals to anonymously share their PHD for the sake of research, barriers concerning data ownership may hinder the possibility of data sharing [55]. One of the suggested solution to facilitate the sharing of medical data is a framework that includes techniques of binning and digital watermarking to overcome the problems of data privacy and copyright protection [56].

The issue of data ownership became a main concern, mainly with the emergence of big data in health sector. In spite of the immense potentials of big data applications in health care and especially in concern of disease management from diagnosis to prevention to personalized treatment [57–59], the rise of big data promotes also the questions of data privacy, security, stewardship and governance [57]. In order to benefit from the huge amount of health data and ensure secure access to electrical medical records (EMR) without violating privacy and ownership rights, data ownership and security policies need to be established [59]. A variety of solutions tackle the issue of data ownership and propose implementations that aim to exploit the potentials of the vast amount of health data and to facilitate data sharing between the different stakeholders taking into consideration privacy, transparency, security, and sovereignty concerns.

Care.data is an initiative introduced by the Health and Social Care Information Centre in the UK in 2013. The Program collects individuals' data from the National Health Service (NHS) in the UK with the aim to link, store, and share the data for research purposes, as well as for commercial benefits in businesses [60]. The data was extracted from general practitioners' practices and the individuals who are registered in these practices got informed that their data would be uploaded to the program unless they object it. The initiative failed and was abandoned in 2016 due to confidentiality issues mainly, where the initiative suffered from the lack of patient awareness of the program and failed to demonstrate the advantages of data sharing [60–62].

In another pursuit to liberate PHD and meet their extended potentials in research and clinical practice, a concept of a data management system that grants individuals the right to own their data was proposed [63]. The idea is to create a platform that integrates individual's data from different sources: doctor records like tests, scans, and doctor visits, in addition to patient-generated data from wearables and gadgets. The data within the platform need to acquire seven characteristics: accessible anywhere and always available to the originator; controlled by the originator; unique and verifiable as belonging to a person; privacy-enabled; secure; independent of any third party and answers data provenance/lineage questions, that is, when, where and from whom the data came. The platform is represented as a medical data ownership engine, that consists of a data-sharing flywheel, which connects two feedback loops: the external wisdom of the individual's body and the wisdom of the population participants as a big medicine resource. The aforementioned individuals' data are fed into the flywheel to form a medical knowledge resource, that provides in its turn feedback to individuals concerning treatments and prevention measures [63]. The technical implementation of the platform was proposed by a nonprofit, social benefit organization 'UnPatient'. The intend was to use the mechanisms of blockchain and its related technologies to solve the data provenance problem and to provide a patient identification mechanism that differentiates between personal data and health data. Unfortunately, there is no follow up papers or studies about the data ownership engine or its implementation and the website unpatient.org is not reachable. However, other studies have explored the potentials of blockchain-based implementations in solving the problems of data ownership and sharing in health sector.

In an innovative study, Liang et al. [64] propose a user-centric and blockchain-based health data sharing and collaboration solution. The solution is a mobile user controlled system built on the hyperledger blockchain fabric [65], which is a permissioned blockchain with a modular architecture that allows the implementation of different use cases and plug-gable models. Fabric is a distributed ledger for smart contracts (chaincode) that supports consensus protocols, security, and privacy preservation [65,66]. The system consists of six components and actors: user, wearable devices, healthcare provider, insurance company, cloud database and the blockchain. The PHD is collected through the user from the wear-able devices and then uploaded to the cloud database through a mobile application. The healthcare providers and insurance companies can request data from the user, which can in his turn, as the owner of the data, grant, deny or revoke data access at any time. All

system's activities from data collection, upload, request, and access to data sharing are recorded on the blockchain to ensure integrity and trustworthiness [64].

Another highlight blockchain-based implementation in health sector is MedRec [67]. MedRec is a decentralized record management system that deals with patients' EMRs and grants them control and access over their data. The system facilitates data sharing and manages data ownership by preserving confidentiality, accountability and authentication across the platform. The modular design of the platform on an Ethereum blockchain supports the use of smart contracts that control data access and data retrieval and log patient-provider relationships. Providers of EMRs can add data about a patient, which in his turn can verify the record before accepting or rejecting it. The Patient has full authority on his own data and can authorize access on the data or sharing of the data between providers. The different stakeholders can be part of the blockchain network as miners, in which they can access anonymized data in return for securing the integrity of the blockchain via a proof of work [67,68].

The last example is from the republic of Estonia, one of the most successful countries in the provision of public digital services, which grant the citizen the right to own their data and provide security, confidentiality, integrity and availability over the data. The Estonian eHealth system is one important highlight of the e-society of Estonia. The system is a secure digital health record system that lets patients access their health data through the national ID and their signature. The data is accessible to specialists by default, but patients can revoke the access any time and for any case [69].

In addition to the aforementioned examples and the literature and studies mentioned in this chapter, other literature [70–78] discuss aspects of data ownership in health sector, but are not explained in detail since the presented examples above are up to date and cover the most aspects relevant to the scope of this paper.

3.2. Data Ownership in Transportation Sector

Intelligent Transportation Systems (ITS) are generating an increasing amount of data that can be exploited in various platforms to facilitate traffic management, pollution control, safe and accurate navigation and other beneficial applications. Nevertheless, the rising amount of data raises also concerns regarding security and privacy [79–82].

One of these applications is the Floating Car Data system (FCD). FCD collects time of travel, position, direction and velocity information from mobile phones in the vehicle and forms an essential source of traffic information. Many individuals oppose FCD projects due to privacy concerns regarding permanent traceability, in addition to trackable liability in case of breaking traffic rules (e.g., speed limit violations) [83].

Vehicular Ad Hoc Networks (VANETs) are another example of applications that deal with privacy issues. VANET is a wireless communication network based on the principles of Mobile Ad Hoc Networks (MANETs). Through VANET, data generated from a vehicle is distributed to neighbor vehicles. This data contains safety related information like accidents, road condition and car behavior such as emergency braking. Among the gathered data, the car can also collect personal-related information, which can lead to privacy concerns by the users of such networks [84]. Vehicle-to-Vehicle (V2V or Car2Car) communication within VANET and Car-to-Infrastructure communication can be encompassed under the umbrella of a more advanced and promising technology, namely Car2X. The Car2X technology enables the car to communicate not only with other cars but also with its surrounding through road-side access points enabling more reliability and accuracy of the shared information [85,86]. Based on the Car2X technology, an initiative for the exchange of safety-relevant traffic data was started. The initiative is a cooperation between original equipment manufacturer (OEM) leaders in the automotive industry (Daimler, BMW, Ford and Volvo) and navigation specialists (TomTom and HERE) in order to experiment in a pilot project the communication of information about hazardous situations in real-time via Car2X technology. The project is supported by six European countries: Germany, Spain, Finland, Luxembourg, the Netherlands and Sweden and aims to develop a platform that

facilitates the communication of anonymized safety-relevant traffic data. The solution motivates the different providers to bring their own data, where those who share their data get access to data from other providers [87].

A common and perceptible concern in the literature is the regulation of privacy protection as well as data sharing in ITSs. Lederman et al. [79] imply in their paper that it is not possible to address all ITS technologies and applications in a single privacy protection policy. Alternatively, they propose a concept for understanding privacy concerns. The idea is that in regards of privacy protections and data sharing, any ITS technology or application needs to address four dimensions: data collection, potential criminal implications on the collected data, Personally Identifiable Information (PII) within the data; and the possibility of secondary usage of the collected data. According to their work, this dimensions can be assured through two types of solutions: privacy-by-design solutions, which integrates privacy protection measures within data collection systems, and privacy-by-policy solutions, which provide guidelines for data collection and usage [79].

On the same subject, Hoh et al. [81] propose an architecture that fulfills the requirements of data privacy and data integrity in ITS. The architecture deals with the privacy issue by separating communication and authentication functions that rely on personal-related information like pseudonyms and identities from data analysis operations that require access to information about position. The proposed architecture proves that the presented data suppression techniques can reduce the privacy risks by preventing data mining algorithms from reconstructing private data from anonymous database samples [81].

Alongside the evolution and the improvement of ITSs, new technologies and innovative opportunities are emerging, opening the door to new possibilities for improvements and advancements. Examples of such opportunities are the concept of crowdsourcing and the cloud technology. Based on a cloud infrastructure and a crowdsourcing-based architecture, a user-driven Cloud Transportation System (CTS) was proposed [88]. The user is the center of the architecture and the source of information, the system collects user information and data to build traffic models and provides feedback predictions and route optimizations based on the collected data. The Cloud with its distributed computing mechanisms serves as a processing platform and provides a storage for the growing amount of traffic data [88]. The integration of a proper security and privacy preservation measures in the system was pointed out but not implemented in the CTS prototype presented in the paper.

With the advancement of car sensors, the car is becoming an Internet of Thing (IoT) platform that harvests the data from various innovative sensors. Joy and Gerla [80] present in their paper the concept of Internet of Vehicles (IoV), which is an intelligent vehicle grid, capable of making its own destination decisions using V2I and V2V communications between peers. They also introduce the "Haystack" privacy concept, which strengthens the privacy as more data owners participate. The mechanism provide accuracy as well as privacy, as it allows data owners to privatize their data locally and independently before submitting it in a cryptographic private write to a cloud service, which in its turn aggregates the responses from different data owners. The mechanism randomises the answers, so the aggregator can calculate and extract the true value, but at the same time can not know the origin of the data, as if the data owner is hiding in a "Haystack" [80].

IoV is one of many examples that is taking advantage of the availability of huge amount of data being generated from the vehicles. ITS is shifting from technology driven system into a data-driven intelligent transportation system (D^2ITS) that exploits efficiently the collected data from various resources. D^2ITSs are privacy-aware people-centric systems that allow users to access and interactively utilize data through data-driven functionalities and services [89]. Nevertheless, with the rapid growth of transportation data, it is becoming more difficult to deal with the growing amount of data. However, with the emerge of big data and data analysis, a wide range of opportunities is opened to develop platforms, architectures and concepts [90–92]. These opportunities enhance data collection,

storage and streaming and provide advanced data analytics including real-time, historical, predictive, video and image analytics [93].

Blockchain is another disruptive technology that shifts the legacy ITS towards a decentralized, secure and privacy-preserving ITS. The blockchain network can provide a reliable and trustworthy way to exchange data between peers (vehicles) and assure at the same time the privacy of the data source. Multiple literature focus on the implementations of blockchain in the transportation sector. One of these examples was proposed by Hirtan and Dobre [94]. In their paper, they propose a blockchain-based model for privacy-preservation in ITS. The model groups the clients into clusters according to the location to facilitate the data processing and to prevent delays and overheads. Through an offline blockchain, the clients can cooperate and exchange data securely based on privacy policies defined by the users themselves like enabling or disabling speed sharing or location sharing. Users can also exclude specific areas from data sharing [94].

In a preliminary study on blockchain-based ITS (B^2ITS), Yuan and Wang [95] introduce an ITS-oriented blochchain model that aims to standardize and characterize a typical architecture of a blockchain-based system in transportation sector. The model consists of seven layers that list and describe the specifications of such architectures.

In an attempt to realize a blockchain-based ITS system, the blockchain technology was used by Singh and Kim to propose a data sharing framework for Intelligent Vehicles (IV) [96,97]. In their work, they introduce an Intelligent Vehicle-Trust Point (IV-TP) mechanism. IV-TP is a blockchain-based mechanism for secure and reliable communication between IVs. In addition to the blockchain technology, IV-TP is based also on VANET for managing communication and on Vehicular Cloud Computing (VCC) which provides the users of VANETs low cost cloud-based computational services [96,97].

In another embodiment, the advantages of the blockchain was exploited to develop a blockchain-based key management platform for Vehicle Communication Systems (VCSs) [98]. The platform consists of two parts. The first part is a distributed blockchain network for simplifying key transfer among security managers, which capture departure information, encapsulate block to transport keys and execute rekeying to vehicles within the same security domain. The second part is a dynamic transaction collection period scheme that is responsible for reducing the key transfer time during vehicle handover [98].

3.3. Data Ownership in Industry

The sovereignty over data is becoming a main concern in industrial context. In recent years, the data is shifting from being the result or the enabler of a process, a product or a service to be the product or the service itself and the key for business innovation. With such shift, the value of data for businesses increased and so has the need to protect the data. In 2014, the Fraunhofer Society in Germany started an initiative under the name Industrial Data Space (IDS) with data sovereignty of data ownership at its core. The initiative was later renamed to International Data Spaces, also known as IDS, while retaining the same objectives and serving the same purposes but with an extended European and international ambitions.

The aim of the initiative is to create a virtual data space that supports enterprises from different industries and allows them to exchange and share data in a secure, controlled, and scalable way through standardized interfaces [99]. In addition, IDS serves as a link between the Internet of Things (also known as Industrie 4.0 in Germany) and the Smart Service World (data-driven and digitized services) through a Smart Data Management, that represents a modern form of a distributed and decentralized data management approach [100].

The main keystones of the IDS initiative are its key elements, a role concept, and a reference architecture model.

In order to achieve its goals and to create a trusted data network, IDS embraces a set of key elements, namely: data sovereignty; data governance; secure data exchange; decentral approach (distributed architecture); network of platforms and services; economies of scale

and networking effects; open approach (neutral and user-driven); and trust. These key elements pave the way for new applications or components to be part of the IDS [100].

A central component of IDS is the role concept that describes the role of the different participant stakeholders within the framework of the business ecosystem and facilitates data exchange between them. The concept presents five roles connected within the IDS to each other: the Data Provider, the Data User, the Broker, the AppStore Operator, and the Certification Authority. The data provider owns the data sources, from which it can offer certain data for participants in the IDS while maintaining its sovereignty over the data. The data provider is also responsible for its own data, in which it needs to describe the data sources and makes it available with a suitable data model and terms and conditions of data use. The data user obtains data from data providers via the broker. The Broker plays the role of an intermediary between the data user and the data provider in addition to its role as a source registry for data sources. It provides functions to data providers to publish data sources, to data users to find data sources, and to both of them to make agreements on the use and the provision of the data. In addition, the broker supervises the data exchange process from the search of data source, to data exchange transactions, to rollback support in case of faulty or incomplete processes. It can also perform data quality checks or even data analysis for large data volumes. The role of the AppStore operator is the provision of functions, that facilitate the development of data services (software) within the business ecosystem. In addition, it provides the data services via the AppStore. The certification authority checks whether the defined requirements in the IDS, which the participants have collectively defined as standards and norms, are met. This includes activities such as accompanying the application process until the certificate is handed over, accepting test reports from the testing laboratory or issuing rejection notices. In order to be able to carry out these activities efficiently, the certification body works closely together with testing bodies and accreditation bodies [100].

To crystalize the concept of the IDS and its elements, a reference architecture was developed to highlight the generalization of concepts, functionalities, and processes of the IDS [99,101].

According to the white paper of the IDS reference architecture model, the model consists of four architectures: Business Architecture, Data and Service Architecture, Software Architecture, and Security Architecture [100]. The business architecture contains all the concepts that are economically necessary for the success of the IDS. The Data and Service Architecture describes the functions that need to be implemented in IDS's applications. The security architecture of the IDS cover a variety of security aspects that can be combined in different forms to support different level of security. The software architecture describes how the data and service architecture must be implemented within the scope of the IDS [100].

The papers of the different versions of the reference architecture model of IDS describe the model further and provide a detailed view of the various components of the model and its structure [99,101]. The reference architecture model is structured in five layers and three cross-sectional perspectives (Figure 1).

The Business Layer defines the different roles of the stakeholders of the IDS and determines the roles' possible activities and connections. The Functional Layer regulates the functional requirements of the IDS and its features. The Process Layer manages the processes of the interactions between the components of the IDS. The Information Layer makes use of linked data principles for describing both the static and the dynamic aspects of the IDS's constituents through a conceptual model. The System Layer describes the components of the logical software and takes into consideration the integration, configuration, deployment, and extensibility of the components. In addition, the model is empowered with three cross-sectional perspectives that are directly related to the five layers of the reference architecture model. The perspectives are: Security, Certification, and Governance [99,101]. The security perspective is responsible for maintaining the trust among the different stakeholders. It secures the identification of the participants, commu-

nication, data exchange, as well as the use of data after the exchange. The certification perspective represents the core of data sovereignty in the IDS reference architecture model and is responsible for the certification of stakeholders and components. The governance perspective orchestrates the requirements to achieve a trusted collaboration environment in a business ecosystem and is responsible for the identification of roles, functions, and processes within the IDS. In addition, the governance perspective addresses directly the issue of data ownership in general and the support of the data owner in specific [99,101].

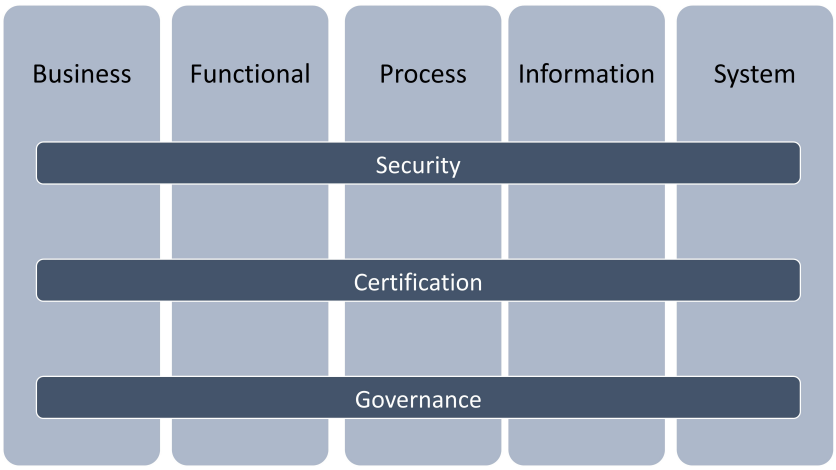

Figure 1. Own Depiction of the General Structure of IDS Reference Architecture Model [101].

The International Data Space presents with its reference architecture a roadmap for the preservation of ownership and sovereignty of data across any given business. Nevertheless, it also opens the door for greater exploitation opportunities outside the industrial context.

In an even larger scale, the GAIA-X project was born in 2019 with the aim to establish a data infrastructure and ecosystem for Europe with European values and standards. Similar to IDS, the heart of the GAIA-X initiative revolves around data sovereignty as the driver to develop a trustworthy data infrastructure that supports innovation and interoperability in the digital and data-driven economy. The Project is a collaboration between businesses, politics and science, with participant from about 300 organizations across Europe (as of June 2020). International Data Spaces Association is also part of GAIA-X, and its reference architecture serves as a foundation for designing data sovereignty based architectures [102].

The core of GAIA-X is the federation services that build the base of an open ecosystem and form the foundation for interoperability and interconnection across the project. They cover requirements for the technical operations of GAIA-X and they include, among others, services for certification and compliance, identity and trust, sovereign data exchange, standardization, in addition to a federated catalogue that encompasses directory to all providers and services, and other specifications. Moreover, these set of services and other possible services that might be added to this set will be developed according to the principles of security by design as well as privacy by design [102,103].

From an architectural point of view, GAIA-X is structured into two parts, the data ecosystem and the infrastructure ecosystem. Both ecosystems form with the federation services the three level of the GAIA-X architecture. The data ecosystem is concerned mainly with data as it embraces data spaces according to EU data strategy and enables the development of advanced smart services that support data-driven innovations across industry. On the other hand, the infrastructure ecosystem provides the needed infrastructure as well as services for dealing with the data [104].

3.4. Data Ownership in Energy Sector

The electrical grid is evolving steadily and new technologies are emerging and shifting the grid to become a data-dependent smart grid. The smart meter presents as a key element of the smart grid many advantages for controlling the grid and for optimizing the energy

consumption as well as the energy transmission, distribution and generation. In spite of its advantages, smart metering raises concerns about the privacy of the consumer [105]. In addition, there is uncertainty and no consent regarding data ownership in smart metering infrastructure [106,107].

The problem of privacy in smart grids is discussed in various literature, and various solutions to overcome the privacy problem were proposed. In a survey conducted in 2017, different privacy-preserving solutions for smart meter data were reviewed. The study shows that in spite of the proposed solutions and the state-of-the-art technologies on smart meter data privacy, there is still some open issues and limitations that threaten the privacy of data [108]. Apart from privacy solutions and technologies, the necessity of privacy laws that determine which data the customer owns and which data must be protected by utilities is another thing to be considered [109]. Such privacy concerns can hinder the innovation process and may prevent users and the community in general to support and be active stakeholders in the introduction of smart grids and the smart meter infrastructure [110].

One of the proposed solutions to establish a privacy-preserving smart grid is the SmartPrivacy concept [111]. SmartPrivacy is a conceptual model presented by Dr. Ann Cavaukian and aims to establish a smart grid that is not only resistant to attacks and natural disasters but also privacy preserving by decreasing data leakage and breaches of personal data. In addition, it gives consumers control over the consumption and over their personal information and ensures their participation as a key element of the smart grid. The concept emphasizes the principle of privacy by design in building a smart grid that integrates the advantages of a typical smart grid with the vision of SmartPrivacy. A smart grid with SmartPrivacy embodies a variety of characteristics. For example, the smart grid should be intelligent by being capable of collecting the minimum needed personal information without affecting the range and quality of the services, and it should collaborate and communicate transparently with consumers about the collection, use and disclosure of their personal data. The grid should also be an efficient smart grid that grows and expands without compromising the privacy and the security of personal information and disposes the no longer needed information. In addition, it is crucial for the grid to support the communication between consumers and utility to enable consumers to customize their preferences regarding their personal data and to give their consents prior to personal information disclosure to a third party. Furthermore, the grid should also be opportunistic, quality-focused, resilient to data leakage and information breaches, and certainly green [111].

In another solution for privacy preservation in smart home and smart grid context, a platform for privacy and security for analytics of smart home sensor data without compromising data utility was proposed. The approach does not transform the stored data, but rather it replaces personal related information via cryptographic techniques with hashed values before storing them. The platform maintains a separate identifier dictionary storage, with hashed and actual identifier values for re-identification of identifiers and processing results. The presented architecture ensures the privacy of the shared results and secures the data throughout its life cycle [112].

While the previous solution focused on encryption as a possible approach for solving the privacy problem, another solution suggested the anonymization as an alternative to tackle the problem [113]. The method anonymizes smart meter data to avoid the association of the readings with a particular smart meter or customer. The mechanism uses escrow-based anonymization on high-frequency metering data that contains personal information and is transmitted to the utility every few minutes. This frequent data is required for operational reasons and does not need to refer to a specific consumer or smart meter, but needs to refer to the location in the network. On the other hand, low frequency metering data is not anonymized as for billing and management purposes it needs to refer to a specific person or smart meter. Nevertheless, it is transmitted on weekly or monthly basis to preserve privacy [113]. Although the presented solution does not solve the privacy problem entirely, but it adds an additional layer of privacy to the smart grid.

Molina-Markham et al. [114] tackled the issue from another angle, where they propose a solution that allows the extraction of usage patterns and billing relevant information from a smart meter using off-the-shelf statistical methods without invading the privacy of households and without detailed information about the consumer's activities or the appliances signatures.

Shifting towards the decentralization of energy and smart grid systems, blockchain mechanisms present an opportunity to address both the ownership and the privacy of energy data. in this context, Aitzhan and Svetinovic [115] propose a decentralized energy trading system (PriWatt) that is build based on blockchain, multi-signatures, and anonymous encrypted messaging streams to maintain privacy and security. The system enables anonymous communication between peers and secures the transaction on the network. Using smart contracts, the system allows users to trade energy ownership and negotiate trading prices [115]. Similar to PriWatt, a proof-of-concept of a Local Energy Market (LEM) was designed and implemented on a private blockchain. The LEM allows the prosumers and the consumers to communicate and trade with each other without interference of central intermediary. The evaluation of LEM illustrates the benefit of a blockchain-based smart grid on different levels and shows the potential of decentralized smart grid in reducing the energy costs while providing a secure and reliable platform for both consumers and prosumers [116].

3.5. Data Ownership in Smart Cities

As shown so far in this capital, data is surrounding us in different aspects of our lives. Therewith, data ownership and some related issues like privacy, security and sovereignty of data is becoming a bigger concern in vital sectors, from health, mobility, energy, industrial, to other major sectors that are part of our lives and our cities. Smart cities provide like any city all these services; yet, they are built on data. Data is the center of any smart city, and the Smart Cities Council confirms this view by stating that: "A smart city gathers data from smart devices and sensors embedded in its roadways, power grids, buildings and other assets. It shares that data via a smart communications system that is typically a combination of wired and wireless. It then uses smart software to create valuable information and digitally enhanced services" [117]. With data as the core of a smart city, the sovereignty and ownership of data is one of the main concerns for managing a successful smart city that preserves its citizens the right to own, control, share and demolish their own data. During the Open Data Forum in 2014, data ownership was considered a major barrier standing against the innovation of information products that play a crucial role in the development of smart cities [118].

The significance of data ownership in the context of smart cities is illustrated in different literature. In the work of Edelenbos et al. [119], they present a research agenda for data-sensitive governance of smart cities with four main clusters of research. One of these clusters is the legal and social aspects of smart cities, which focuses mainly on the issue of data ownership and privacy in the context of smart city. Khan et al. [120] state in their work that ownership and privacy implications over the generated data from the connected environments within smart cities can influence also the adoption rate for services, especially if it is about sensitive or personal related data.

On another level, Morozov and Bria [121] look at the issue from another point of view, where they consider data ownership an opportunity rather than a burden. They articulate in their research about rethinking the smart city concept and establishing innovation and cooperation based smart city models that rely on democratic data ownership regimes. Such regimes remove the power from dominating companies and give citizens and cities the right to own the data produced within the cities and the power to be able to use data in public services and to exploit it in beneficial policies. By shifting towards a democratic data ownership regimes, cities will be able to make data available to startups, SMEs, and innovators across vertically [121]. However, in order to achieve that, it is significant to understand data that can be found in smart cities and to identify the current owner of this

data. Bischof et al. [122] discuss in their study the need for common semantic and data models for data generated within the smart cities in order to encourage and facilitate the development of new services. In pursuance of establishing a semantic description model, the study presents a list of data that can be retrieved from smart cities. the list is derived from the EU FP7 CityPulse project [123]. Besides the description of the data, the list defines the owner (publisher) of each data category in a smart city. It is to be noticed from the list, that the variety of data that can be collected in a smart city belongs to different owners and not exclusively to the city and its citizens. This emphasizes the necessity to consider the descriptions for data privacy and security as well as data ownership concepts in the design of models for smart cities [122]. A similar study in the context of the FIESTA-IoT project [124] present and analyze the different semantics and data models of smart cities with the aim to gain value by linking them in order to enable cooperation between the cities. This study present also a similar list containing smart cities data and its descriptions derived from both projects CityPulse and FIESTA-IoT. The various services of smart cities require different considerations of data ownership and this requirements need to be considered by designing common semantic descriptions based on the different protocols and standards of semantics and data models presented in the paper [125].

As mentioned earlier, data is the building blocks of smart cities. Nevertheless, this data includes a variety of personal related information that raises privacy and security concerns among different applications of smart cities. Zhang et al. [126] address in their paper the issue of security and privacy in smart city applications. In their overview, they shed the light on several challenges and present state-of-the-art solutions for some applications. Despite the promising solution presented in the paper, the authors emphasize on the importance of the collaboration between municipalities, regulation departments, industry, academia, and business companies. In addition, the authors advise on delegating a trusted third party for monitoring and auditing, and also underline the importance of securing data privacy, availability and management concurrently [126].

In a study addressing the privacy concerns in smart cities, van Zoonen [127] argues about two dimensions that represent the privacy concerns of the people. The data dimension describes the perception of data from being personal and sensitive in its nature to impersonal. Where the purpose dimension describes the purpose of the collected data, varying from being for services or for surveillance purposes [127].

Based on this two dimensions, van Zoonen conceptualizes a two dimensional framework that hypothesizes suitable applications and technologies in smart cities according to the degree of privacy concerns of the people, from hardly any concern to controversial. The privacy framework shown in (Figure 2) is represented through both the purpose of data dimension and the kind of data dimension. The intersection of both dimensions forms the four possible quadrants to place the application or the technology. Based on prior research on the privacy concerns of the people and based on the understanding of the landscape of the smart city, a technology or an application can be placed in one of the quadrants. According to the platform, a smart city application can range from raising hardly concern by handling impersonal data and serving service purpose (quadrant IV), to being controversial by handling personal data and serving surveillance purpose (quadrant II) [127]. The availability of such platform sheds the light on areas where data ownership preservation in the sense of privacy protection and the right to control and share the data might be significant.

Kakarontzas et al. [128] argue that having a suitable software architecture with adequate business processes is crucial for the development of smart city applications and projects. Supporting this idea, they conducted a survey to collect the necessary requirements for smart cities. Based on the gathered requirements, they proposed a conceptual enterprise architectural framework that serves as the foundation for launching smart city projects and applications. The conducted survey contains a category of questions dedicated to data sources. The results of the questions that are related to data access shows that the access level among city's administrative staff is 39% role-based access and 22% project

dependent access, whereas 60% of access by the city's stakeholders is project dependent. The access level of end users is divided between 33% individual data access and 28% project dependent. Concerning the question of data ownership, the results reveal that 50% of city's data belongs to the municipality, 39% are manipulated by the project organization and 28% are owned by various stakeholders. It is also important to mention that the survey reveals that in many cases, the ownership of the data is joined between different organizations or stakeholders. Regarding security, it is interesting to see that 72% of the surveyed smart cities do not use encryption and only 22% used encryption based on the application and its data. Based on the results of the survey, seven generic functional and quality requirements for smart cities where identified as follow: interoperability, usability, authentication and authorization, availability, recoverability, maintainability, and confidentiality. The identified requirements are used then to identify architectural patterns, which in their turn form the conceptual application architecture form. Authentication, authorization and confidentiality are managed through the mechanisms of the web server and the application itself, while the anonymity of information as well as other data ownership services are case specific and to be determined by the conceptualization of a specific smart city project. The presented framework is only a concept that need to be tested and evaluated. However, it should serve as the starting point to cover most architectural drivers of smart city's applications. The framework can be customized to be able to meet the different functional and non-functional requirements of a specific smart city [128].

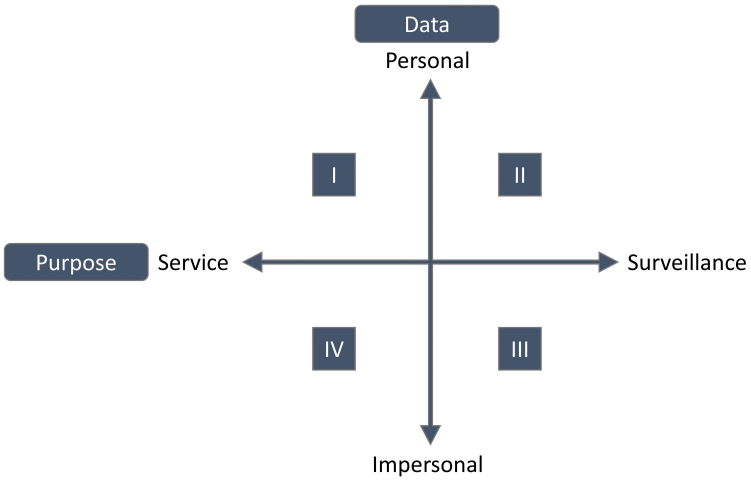

Figure 2. Privacy Framework according to van Zoonen [127].

In the work of Khan et al. [120], they present a framework that addresses specifically the security and privacy concerns in smart cities. The framework was developed based on an onion model that illustrates the different smart city stakeholders and specifies the relevant roles for the development of the various stages and components of the system. The 'Smart Secure Service Provisioning' (SSServProv) framework deals with trusted acquisition of data as well as secure and privacy-aware provisioning of services in smart cities. The platform ensures the privacy and security of the citizens' private and sensitive data and allows services only from legitimate service providers. It includes token based authentication, secure communication protocol and authorization mechanisms. In addition, the framework enriches data security and privacy with the data confidentiality and the data anonymization components. The data confidentiality component provides private and confidential data only to trusted service providers and users. It facilitates also the processing and persistence of data in untrusted domains by its cryptographic mechanisms. On the other hand, data anonymization secures private and personal related data as it decouples data from its owner and reduces the possibilities of privacy infringement as data cannot be traced to its owner [120].

In addition to data ownership related concerns, the rising amount of data and the emergence of open data in the context of smart cities present both opportunities and challenges that municipalities of smart cities need to address. In a state-of-the-art study, Molinari et al. [129] provide an overview of big data and open data in the city of Trento, Italy and propose a possible development in these matter. The paper identifies four enablers that facilitate the exploitation of big data and open data benefits in smart cities: cultural, organizational, governance, and technological enablers. By the cultural enablers, the authors emphasize on the idea of open data as an opportunity for change. This implies that for a city to be "smart", it has to surpass the traditional culture of being the sole owner of public data, and it should enable its citizens to access, discuss and comment these data. On the other hand, the authors express the need for rules for data circulation by governance enablers, where there should be rules of data ownership and privacy across public sectors to regulate citizen participation and data sharing [129].

The applications of big data and especially big data analysis provide huge opportunities for the development and the advancement of smart cities, but they also come with challenges that need to be addressed. One of the challenges is the ability to share data between different entities and stakeholders of a smart city without violating the citizens' rights of privacy. Another issue is the data ownership where most smart city entities claim the ownership of data despite the fact that some of this data deals with personal related and private information, like health, financial, location, and other personal information [130]. Pramanik et al. [131] describe in their paper a paradigm shift on three crucial dimensions in smart cities' context: health, data, and city itself. The new paradigm describes the shift towards smart cities that enable a variety of smart services including smart health, which in their turn are empowered by big data analysis and applications. Security and privacy present one of the major challenges facing this shift, especially when dealing with sensitive and personal information like health data. The work incarnates the new paradigm in a conceptual framework of a big data enabled smart healthcare system (BSHSF). The proposed framework can overcome a variety of challenges including security and privacy challenges through anonymization and cryptographic models. Other issues remain to be addressed in smart cities' big data smart systems like the right of possession and control over data [131].

In another work of Soto et al. [132], the issue of data ownership among other challenges are considered the requirements for the development of the ALMANAC smart city platform. The presented platform addresses the issue of interoperability among heterogeneous systems and services and embraces the concept of federation in smart cities through its decentralized nature. This will allow systems and applications within the existing landscape to cooperate and work with each other while keeping the possibilities to operate independently.

Another interpretation of decentralization in smart cities can be seen in the work of Ramachandran et al. [133]. The authors introduce in their work a decentralized marketplace for smart cities that facilitates data sharing without compromising the privacy of the data owners. The platform was designed and deployed using blockchain and other distributed ledger technologies. Using smart contract, potential buyers can acquire data products posted by data owners [133].

The exploitation of the blockchain technology and smart services in smart cities context has various advantages and uses and it can be seen in different literature. Following is as a short overview of some of these examples. Biswas and Muthukkumarasamy [134] present in their work a security framework, that ensures the privacy and security in the communication within a smart city. On the other hand, Sharma et al. [135] propose in their study a distributed blockchain based vehicular network architecture in smart city (Block-VN). The developed architecture is reliable and secure and ensures the privacy of the data through private key encryption and anonymization. Similar to Block-VN, Michelin et al. [136] introduce in their work SpeedyChain, a blockchain-based framework for data sharing among smart vehicles in smart cities while maintaining privacy, security

and integrity in a decentralized manner. The framework provide private and secure communication models for vehicles as well as roadside infrastructures, allowing them to trust the source of the received data. The framework ensures also privacy by adding time expiration limit for the usage of a vehicle's key. SpeedyChain allows fast addition of data to the blocks by using a Blockchain design that decouples the data stored in the transaction from the block header [136]. Using also the blockchain, Sun et al. [137] propose a conceptual framework for sharing services. The framework presents three dimensions: human, technology and organization that build the foundation of sharing services that embrace transparency, privacy and trust.

As to be seen in the various examples, the benefit of the exploitation of blockchain is undeniable. Rivera et al. [138] confirm this view in their review of the existing research of digital identity on blockchain-based implementation in smart cities' environments. The paper emphasizes the crucial role of blockchain in smart cities for the preservation of privacy and digital identity as well as a handy tool for authentication and security.

3.6. Summary

The different publications and literature concerning the subject of data ownership across different vital sectors shed the light on multiple issues that need to addressed and mentioned in the context of data ownership studies. To help capture the big picture of the conducted state-of-the-art, the reviewed papers were summarized in Table 1.

Table 1. Data Ownership State-of-the-Art Summary.

Sector	Paper	Year	DO	DS	BC	BD	SE	PR	Type of Research
Health	[54]	2014	✓	✓				✓	Review
	[55]	2015	✓	✓			✓	✓	Study
	[56]	2005	✓	✓				✓	Framework
	[57]	2015	✓			✓	✓	✓	Overview
	[58]	2014	✓	✓		✓			Overview
	[59]	2013	✓	✓		✓		✓	Study
	[60]	2016	✓	✓		✓	✓	✓	Platform (C)
	[63]	2015	✓	✓			✓	✓	Platform (C)
	[64]	2017	✓	✓	✓		✓	✓	Platform (S)
	[67,68]	2016	✓	✓	✓		✓	✓	Platform (S)
	[69]	2017	✓	✓	✓		✓	✓	Platform (S)
Transportation	[90]	2016				✓			Overview
	[91]	2019				✓	✓	✓	Platform
	[92]	2018				✓	✓	✓	Architecture
	[93]	2019				✓	✓	✓	Review
	[95]	2016			✓	✓	✓	✓	Study
	[96,97]	2017		✓	✓		✓		Framework
	[98]	2017			✓		✓	✓	Framework
	[94]	2018	✓		✓		✓	✓	Architecture (S)
	[79]	2016	✓	✓			✓	✓	Survey
	[86]	2008					✓	✓	Overview
	[89]	2011						✓	Survey
	[81]	2006					✓	✓	Architecture
	[82]	2002		✓			✓	✓	Framework
	[83]	2008	✓				✓	✓	Approach (C)
	[80]	2017	✓				✓	✓	Study
	[84]	2005					✓	✓	Overview
	[88]	2012					✓	✓	System (ITS)
	[85]	2010					✓	✓	Framework

Table 1. *Cont.*

Sector	Paper	Year	DO	DS	BC	BD	SE	PR	Type of Research
Industry	[99,101]	2018	✓	✓			✓	✓	Initiative (RA)
	[100]	2016	✓	✓			✓	✓	Initiative (C)
	[102,103]	2020	✓	✓			✓	✓	Initiative (C)
	[104]	2020	✓	✓			✓	✓	Initiative (RA)
Energy	[108]	2017					✓	✓	Survey
	[110]	2013	✓					✓	Study
	[107]	2011	✓	✓			✓	✓	Study
	[111]	2010		✓			✓	✓	Concept (CM)
	[109]	2013					✓	✓	Study
	[113]	2010					✓	✓	Method (C)
	[114]	2010					✓	✓	Architecture (C)
	[105]	2017						✓	Study
	[112]	2013	✓	✓			✓	✓	Framework
	[106]	2016	✓	✓				✓	Overview
	[115]	2018					✓	✓	System (POC)
	[116]	2018					✓		Approach (C)
Smart City	[128]	2014	✓				✓		Architecture (C)
	[119]	2018	✓			✓		✓	Study
	[118]	2014	✓				✓	✓	Study
	[127]	2016				✓		✓	Framework
	[121]	2018	✓	✓	✓	✓		✓	Study
	[126]	2017	✓				✓	✓	Overview
	[122]	2014	✓				✓	✓	Overview
	[125]	2018	✓						Overview
	[120]	2017	✓	✓			✓	✓	Framework
	[129]	2014	✓			✓	✓	✓	Study
	[130]	2015		✓		✓	✓	✓	Review
	[131]	2017		✓		✓	✓	✓	Framework (CM)
	[132]	2015	✓	✓					Platform (A)
	[133]	2018	✓	✓	✓		✓	✓	Platform (DM)
	[134]	2016			✓		✓	✓	Framework (CM)
	[137]	2016		✓	✓				Framework (CM)
	[135]	2017		✓	✓		✓	✓	Model (A)
	[138]	2017			✓		✓	✓	Review
	[136]	2018		✓	✓		✓	✓	Framework

Legends: Concept (C), System (S), Intelligent Transportation System (ITS), Reference Architecture (RA), Proof of Concept (POC), Model (M), Conceptual Model (CM), Architecture (A), and Data Marketplace (DM).

The table categorizes the reviewed papers according to the sectors considered in this state-of-the-art study. Furthermore, in order to gain an overview of the findings, for each reviewed paper the table lists the type of conducted research, and the topics discussed in the paper. For the sake of this study, only a couple of the relevant topics were considered, namely: Data Ownership (DO), Data Sharing (DS), Blockchain (BC), Big Data (BD), Security (SE), and Privacy (PR).

As to be seen in the table, the privacy of data, regardless of its type or size, is a main concern across the different sectors. On the other hand, it is to be noticed that the ownership of data is not always the main issue and is not an equal concern in the different sectors. Due to the nature of health sector dealing mostly with sensitive and personal data, the issue of data ownership is present in its publications for long time. While in the other sectors considered in this review, the issue of data ownership is not present as much and it is only getting some attention in the last few years. One way to explain this observation might be that health data and since its existence is personal data that tells a lot about the person's health and life and not only the person's way of life. On the other hand, with the

evolvement of technology and the emergence of social media, sensors and personalized technologies, the matter of data ownership came into question across different aspects and beyond the health sector. Examples of this observation can be stumbled upon throughout the reviewed publications in smart cities context, where data ownership is becoming more visible to researchers in recent years. A smart city is the best example of the modern life and its various aspects, where personalized technologies, smart services as well as various sensors are capturing the essence of a person's life while gathering or generating a big amount of personal related data.

For the sake of writing the previous state-of-the-art, a thorough literature review were conducted. The reviewed papers were derived from google scholar, web of science, and ORBIS the library of the University of Oldenburg. In addition to the search of publications, the web of science portal was used to obtain valuable insights about the subject of data ownership and its research. Through the portal, it was possible to analyze the results of the searched topics. According to the web of science, 560 is the number of publications addressing data ownership as a main topic, starting with the first publication in 1981 until the time of conducting this study mid 2020. The results of the conducted review in this study match the analyzed results from the web of science. The results shows the domination of health related research among the top 25 research areas in the topic of data ownership with more than 39% of the publications, including publication in the areas of health care sciences services, medical informatics, legal medicine, medical ethics, public environmental occupational health, general internal medicine, research experimental medicine and others. It also confirms the fact that the interest in data ownership as a research subject increased noticeably in the recent years where more than 80% of the publications where written between 2010 and 2020 (Figure 3).

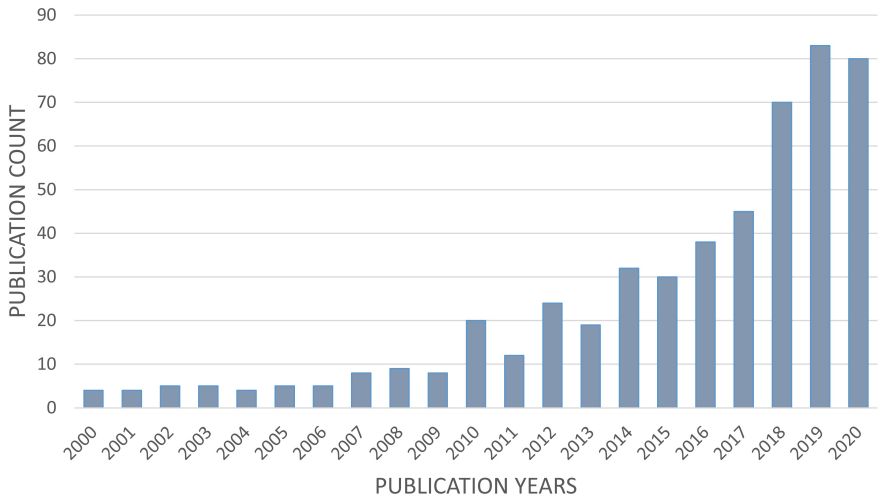

Figure 3. Data ownership publications by year (source: Web of Science).

The reviewed papers in this state-of-the-art address the issue of data ownership from different perspectives and vary in the approach of discussing or solving the issue. Some of these publications focus on surveying and studying the subject and its different aspects, while other authors direct their research towards reviewing the progress of the matter and giving overview of the existing solutions in this concern. Nevertheless, some researchers fixate on developing a viable solution to the problem. The proposed ideas vary from a basic concept for an approach, a method, a conceptual model, an architecture or a platform, to actual presentable implementations of concepts in the form of a platform, a system, a framework, or an architecture.

The presence of new technologies is also noticeable across the different publications. The rising amount of data is without a doubt one of the reasons that provoke the issue of data ownership; however, as mentioned in various publications, the different mechanisms

that came along the emergence of big data count as main enablers to solve the problem. On the other hand, the blockchain technology and its handy toolset of smart ledger, smart contracts and other services and mechanisms are becoming more and more relevant in solving the issues of privacy, security, sovereignty and ownership of the transferred and shared data.

To summarize this chapter, it is important to highlight the inclusion and exclusion criteria of this state-of-the-art. The conducted review included a wide range of literature that address the issue of data ownership directly or indirectly. It also included different examples from 5 major sectors. The review shed the light on the confusion and the complications of the issue, and highlighted the recent interest in dealing with the ownership of data across different aspects our lives. On the other hand, the review excluded literature that only mention data ownership as a side aspect or focus on one related aspect as the main topic of the work, like for example, the development or the implementation of a security mechanism. Excluded were also the literature that do not present novel ideas or are very similar to an included work.

The presented review gave an overview over the state-of-the-art of data ownership from the perspective of five sectors. Nevertheless, there is other aspects and examples of data ownership that can be explored and were not included in this study.

4. Discussion & Lessons Learned

In spite of the novelty of data ownership as a trending issue, its relevance to the modern data-driven world has become undeniable. Data has been described repeatedly as the "new gold" or the "new oil", and as a matter of fact, this view is not far from reality. Data has become the heart of new advancements in all aspects of our lives. It is the driver for competitive advantages and innovations and it is the source for new insights as it fuels the engines of data science, machine learning and artificial intelligence. On top of that, our personal data says a lot about us including our health, behavior, consumption, mobility, and our way of life in general, and this information can be used to our benefit and to optimize our life. However, to be the new gold or the new oil is not concluded by its advantages, as it comes with a complication; everyone wants to own it!

The ownership of data is not straightforward like the ownership of subjects or properties, and it entails various complications as to be seen in this paper. Going through the literature concerning data ownership, it is noticeable that there is an obscurity regarding the subject. On the one hand, there is as of today no unified definition of the term, but rather individual definitions from the point of view of the authors. On the other hand, the main question concerning data ownership: "Who owns the data?" remains without a plain answer, as there is still no regulations that cover the different variation of the subject. In addition, even when the owner is identified or agreed upon, it is not clear, what are the aspects that need to be covered in order to preserve data ownership of the owner. Does guaranteeing the security and the privacy of the data suffice, or does the preservation of data ownership extend to cover aspects like data sharing, data integrity, data sovereignty, data stewardship, data transparency, data provenance, data lineage, data linkage, and others?

From the point of view of this study, this question does not have a forthright answer and the answer is always case dependent. Any data aspect that fulfills the objectives and the requirements of a case is relevant to the development of a data ownership concept for this specific case. In addition, the sensibility and the type of the handled data, whether it is personal or non-personal data should also be taken into consideration. On top of these data aspects, the development of a data ownership concept should have regards to the ethical and legal aspects of the case in general and according to the country and the organizational body where the concept is to be implemented.

Similar to the development of the concept, the technical implementation of the concept is not restricted to one or a set of solutions that are always applicable to any case. Nevertheless, technologies like blockchain and smart services, federated architectures,

distributed architectures, and data spaces and lakes provide promising possible solutions for the realization of environments that support the preservation of data ownership. In addition, in order to choose suitable technologies for the implementation, data ownership requirements of the case including legal and ethical requirements should be taken into consideration. In some cases, a simple access and role management might be sufficient. Where in others, the requirements might extend to include the need for control management, version management, provision management, anonymization, encryption, traceability and other functionalities. Nonetheless, it is also advantageous by the development of data ownership concepts to consider approaches like privacy by design and secure by design.

In the light of the conducted overview and the lessons learned from the different studies addressing the issue, from the standpoint of this study data ownership is "the possession of data and the control of data including the ability to grant rights over data to others, taking into account the type of data, the legal and ethical aspects, the responsibility for data, and the needed requirements to preserve and achieve this possession" (Figure 4).

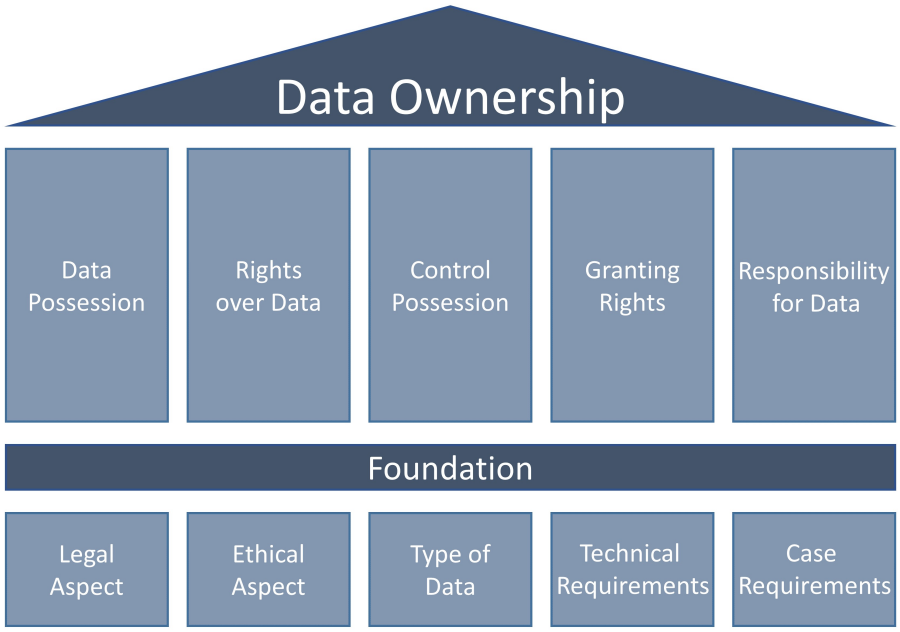

Figure 4. Data Ownership Conceptual Model.

The presented conceptual model in Figure 4 encapsulates the findings of this work, as well as the presented extended definition of data ownership from the point of view of this study. In our opinion, in order to address data ownership in any context, a set of aspects and characteristics need to be covered and embraced. Under the umbrella of a viable data ownership, the possession of data and the control over the data should be preserved. It also should encompass the ownership of the different rights over the data and the possibility to grant some of these rights to a third party. In addition, owning the data means also taking the responsibility over the possessed data.

The mentioned characteristics defines the meaning of possessing the ownership over the data, but from our point of view, in order to design and conceptualize a viable data ownership, the concept should be based on a strong foundation. This foundation determines to what extend and in what respect the characteristics of the concept should be contemplated. First of all, the foundation includes the understanding of the legal circumstances bounding the case. Including but not limited to the law of the country, organizational regulations, binding agreements, manufacturer's restrictions and others. In addition, the foundation takes into account the ethical aspect of the considered case, especially when dealing with medical, personal or sensitive data, which leads us to the next aspect. Understanding the type of data that we are dealing with is a crucial part of building a strong foundation. The

type of data might affect the characteristics of the concept or the complexity of the concept to be able to address this type of data. Furthermore, the foundation of a data ownership concept should be based on a good understanding of the case. For example, on what scale should the concept be implemented: within a platform, a system, an organization, a city, a country? Or to what extend should the concept cover in this case in order to preserve the ownership of the data? Finally, the foundation should also include a good understanding of the technical requirements of the case. For example, knowing if the concept will be implemented on top of a legacy or existing system or will it be built from scratch.

Five major sectors were reviewed in this work, but the necessity to address data ownership extends to all aspects of our lives, wherever data is being generated. Our daily routines are being monitored with data of sensors, cameras and location trackers. We are generating data with our consumption, whether it is domestic consumption like electricity, water or gas, or consumption in shopping centers, stores or activity centers. We are generating data in our daily use of ICT-devices for work or leisure purposes like for social media. The importance of the ownership of data might vary from one aspect to another, but it remains a relevant and crucial issue that deserves to be discussed and be taken into account. This relevance of data ownership is noticeable in the emerging trends and technologies, which address more and more the issue of data ownership and the privacy and security of data, like the concept of self-sovereign identity (SSI) [139], verifiable credentials (VCs) [140,141] and distributed data mesh architecture [142]. In addition to the numerous emerging IT-startups that are developing innovative solutions addressing data ownership, like EcoSteer [143], GoKnown [144], Mine [145], LeapXpert [146], Genomics Personalized Health [147] and others.

5. Conclusions

This paper set out a study of data ownership, highlighting its necessity and advantages and discussing the various complications surrounding the subject. The first part of the paper provided an overview of the different aspects of data ownership. Apart from the personal ethical dimension of data ownership, we have seen that it extends to reach organizational and beyond organizational contexts. In addition, we have seen the challenges that came along the emergence of IoT including the amount of data that is being generated and the numerous stakeholders that are interested in acquiring the data.

The paper then examined the current laws and regulations regarding data ownership on European and German level. We have seen that in spite of the immense efforts in the last years, the status quo of laws could not answer the question of who should own the data and are not yet sufficient to protect and preserve the ownership of data. In the second part of this paper, we surveyed the state-of-the-art on the conceptualisation and the implementation of data ownership concepts. We have focused on five major domains: health, transportation, industry, energy, and smart cities. The survey shows the reflection on the subject from different point of views and reveals the variations in approaching or solving the issue of data ownership.

To conclude, this study gave an extended definition of data ownership and argued that there is no single right answer to the questions of who should own the data and how to preserve the ownership of the data. It argued also that in order to address data ownership, different aspects need to be taken into consideration. Data ownership is always dependent on the case and its requirements including legal, ethical, organizational, and technical requirements.

Author Contributions: Conceptualization, J.A.; investigation, J.A.; writing—original draft preparation, J.A.; writing—review and editing, J.A. and J.M.G.; visualization, J.A.; supervision, J.M.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Bundesministerium für Wirtschaft und Energie (BMWi) grant number 03EE3016D.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: This work is part of the project "WiSA big data—Wind farm virtual Site Assistant for O&M decision support—advanced methods for big data analysis".

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

IDC	Industrial Development Corporation
ICT	Information and Communications Technology
IoT	Internet of Things
HDI	Human Data Interaction
MDM	Master Data Management
GG	German Basic Law
GDPR	European General Data Protection Regulation
CFR	Charter of Fundamental Rights of European Union
ECHR	European Convention of Human Rights
BDSG	German Federal Data Protection Act
TMG	German Telemedia Act
UrhG	German Copyright Law
StGB	German Criminal Law
UWG	German Unfair Compitition Law
BGB	German Civil Law
PHD	Personal Health Data
EMR	Electrical Medical Records
NHS	National Health Service (England)
ITS	Intelligent Transportation System
FCD	Floating Car Data System
VANET	Vehicular Ad Hoc Network
MANET	Mobile Ad Hoc Network
V2V	Vehicle-to-Vehicle
OEM	Original Equipment Manufacturer
PII	Personally Identifiable Information
CTS	Cloud Transportation System
IoV	Internet of Vehicles
D^2ITS	Data-Driven Intelligent Transportation System
B^2ITS	Blockchain-Based Intelligent Transportation System
IV	Intelligent Vehicle
IV-TP	Intelligent Vehicle-Trust Point
VCC	Vehicular Cloud Computing
VCS	Vehicle Communication System
IDS	Industrial/International Data Spaces
LEM	Local Energy Market
SSServProv	Smart Secure Service Provisioning
BSHSF	Big Data enabled Smart Healthcare System Framework
Block-VN	Blockchain-Based Vehicular Network
SSI	Self-Sovereign Identity
VCs	Verifiable Credentials

References and Notes

1. Marr, B. How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read. *Forbes*, 21 May 2018.
2. IDC. *Executive Summary: Data Growth, Business Opportunities, and the IT Imperatives | The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things*; IDC: Needham, MA, USA, 2014.
3. IDC. *The Growth in Connected IoT Devices Is Expected to Generate 79.4ZB of Data in 2025, According to a New IDC Forecast*; IDC: Needham, MA, USA, 2019.
4. Martin, J. *An End-User's Guide to Data Base*; Prentice-Hall: Hoboken, NJ, USA, 1981.
5. Cook, C. The Data Ownership Controversy. *Inf. Rec. Manag.* **1982**, *16*, 52–54.
6. Brathwaite, K.S. Resolution of Conflicts in Data Ownership and Sharing in a Corporate Environment. *ACM SIGMIS Database Database Adv. Inf. Syst.* **1983**, *15*, 37–42. [CrossRef]
7. Weldon, J.L. Who Owns the Data? *Inf. Syst. Manag.* **1986**, *3*, 54–57. [CrossRef]
8. Parry, O.; Mauthner, N.S. Whose Data Are They Anyway? Practical, Legal and Ethical Issues in Archiving Qualitative Research Data. *Sociology* **2004**, *38*, 139–152. [CrossRef]
9. Hand, D.J. Aspects of Data Ethics in a Changing World: Where Are We Now? *Big Data* **2018**, *6*, 176–190. [CrossRef] [PubMed]
10. Horner, J.; Minifie, F.D. Research Ethics II: Mentoring, Collaboration, Peer Review, and Data Management and Ownership. *J. Speech Lang. Hear. Res.* **2011**, *54*. [CrossRef]
11. Lehtiniemi, T.; Ruckenstein, M. The Social Imaginaries of Data Activism. *Big Data Soc.* **2019**, *6*, 2053951718821146. [CrossRef]
12. De Bruin, B.; Floridi, L. The Ethics of Cloud Computing. *Sci. Eng. Ethics* **2017**, *23*, 21–39. [CrossRef]
13. Harison, E. Who Owns Enterprise Information? Data Ownership Rights in Europe and the US. *Inf. Manag.* **2010**, *47*, 102–108. [CrossRef]
14. Van Asbroeck, B.; Debussche, J.; César, J. *Building the European Data Economy: Data Ownership*; White Paper; Bird & Bird: London, UK, 2017.
15. Janeček, V. Ownership of Personal Data in the Internet of Things. *Comput. Law Secur. Rev.* **2018**, *34*, 1039–1052. [CrossRef]
16. Denker, P.; Graudenz, D.; Schiff, L.; Schulz, S.E.; Hoffmann, C.; Jöns, J.; Jotzo, F.; Goeble, T.; Hornung, G.; Friederici, F.; et al. "Eigentumsordnung" für Mobilitätsdaten? 2017.p. 174.
17. Van Alstyne, M.; Brynjolfsson, E.; Madnick, S. Why Not One Big Database? Principles for Data Ownership. *Decis. Support Syst.* **1995**, *15*, 267–284. [CrossRef]
18. Vilminko-Heikkinen, R.; Pekkola, S. Changes in Roles, Responsibilities and Ownership in Organizing Master Data Management. *Int. J. Inf. Manag.* **2019**, *47*, 76–87. [CrossRef]
19. Silvola, R.; Jaaskelainen, O.; Kropsu-Vehkapera, H.; Haapasalo, H. Managing One Master Data—Challenges and Preconditions. *Ind. Manag. Data Syst.* **2011**, *111*, 146–162. [CrossRef]
20. Al-Khouri, A.M. Data Ownership: Who Owns "My Data". *Int. J. Manag. Inf. Technol.* **2012**, *2*, 1–8. [CrossRef]
21. Hart, D. Ownership as an Issue in Data and Information Sharing: A Philosophically Based View. *Australas. J. Inf. Syst.* **2002**, *10*, 7. [CrossRef]
22. Gabisch, J.A.; Milne, G.R. The Impact of Compensation on Information Ownership and Privacy Control. *J. Consum. Mark.* **2014**, *31*, 13–26. [CrossRef]
23. Mashhadi, A.; Kawsar, F.; Acer, U.G. Human Data Interaction in IoT: The Ownership Aspect. In Proceedings of the 2014 IEEE World Forum on Internet of Things (WF-IoT), Seoul, Korea, 6–8 March 2014; pp. 159–162.
24. Crabtree, A.; Mortier, R. Human Data Interaction: Historical Lessons from Social Studies and CSCW. In *ECSCW 2015: Proceedings of the 14th European Conference on Computer Supported Cooperative Work, Oslo, Norway, 19–23 September 2015*; Springer International Publishing: Cham, Switzerland, 2015; pp. 3–21. [CrossRef]
25. Mineraud, J.; Mazhelis, O.; Su, X.; Tarkoma, S. A Gap Analysis of Internet-of-Things Platforms. *Comput. Commun.* **2016**, *89–90*, 5–16. [CrossRef]
26. Farkas, T.J. Data Created by the Internet of Things: The New Gold without Ownership. *Rev. Prop. Inmater.* **2017**, *23*, 5. [CrossRef]
27. Saarikko, T.; Westergren, U.H.; Blomquist, T. The Internet of Things: Are You Ready for What's Coming? *Bus. Horizons* **2017**, *60*, 667–676. [CrossRef]
28. Bertino, E. Big Data—Security and Privacy. In Proceedings of the 2015 IEEE International Congress on Big Data, Santa Clara, CA, USA, 29 October–1 November 2015; IEEE: New York City, NY, USA, 2015; pp. 757–761. [CrossRef]
29. Hampton, S.E.; Anderson, S.S.; Bagby, S.C.; Gries, C.; Han, X.; Hart, E.M.; Jones, M.B.; Lenhardt, W.C.; MacDonald, A.; Michener, W.K.; et al. The Tao of Open Science for Ecology. *Ecosphere* **2015**, *6*, art120. [CrossRef]
30. Yao, A.C. Protocols for Secure Computations. In Proceedings of the 23rd Annual Symposium on Foundations of Computer Science (Sfcs 1982), Chicago, IL, USA, 3–5 November 1982; pp. 160–164. [CrossRef]
31. Goldreich, O.; Micali, S.; Wigderson, A. How to Play ANY Mental Game. In Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing, New York, NY, USA, 25–27 May 1987; pp. 218–229.
32. Micali, S.; Rogaway, P. Secure Computation. In Proceedings of the Annual International Cryptology Conference, Santa Barbara, CA, USA, 11–15 August 1991; Springer: Berlin/Heidelberg, Germany, 1991; pp. 392–404.
33. Liu, C.; Wang, X.S.; Nayak, K.; Huang, Y.; Shi, E. ObliVM: A Programming Framework for Secure Computation. In Proceedings of the 2015 IEEE Symposium on Security and Privacy, San Jose, CA, USA, 17–21 May 2015; pp. 359–376. [CrossRef]

34. Riazi, M.S.; Weinert, C.; Tkachenko, O.; Songhori, E.M.; Schneider, T.; Koushanfar, F. Chameleon: A Hybrid Secure Computation Framework for Machine Learning Applications. In Proceedings of the 2018 on Asia Conference on Computer and Communications Security, Incheon, Korea, 4–8 June 2018; p. 15.
35. Agrawal, R.; Srikant, R. Privacy-Preserving Data Mining. In Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, Dallas, TX, USA, 15–18 May 2000; pp. 439–450.
36. Lindell, Y.; Pinkas, B. Privacy Preserving Data Mining. In Proceedings of the Annual International Cryptology Conference, Santa Barbara, CA, USA, 20–24 August 2000; Springer: Berlin/Heidelberg, Germany, 2000; pp. 36–54.
37. Verykios, V.S.; Bertino, E.; Fovino, I.N.; Provenza, L.P.; Saygin, Y.; Theodoridis, Y. State-of-the-Art in Privacy Preserving Data Mining. *ACM Sigmod Rec.* **2004**, *33*, 50–57. [CrossRef]
38. Liang, F.; Yu, W.; An, D.; Yang, Q.; Fu, X.; Zhao, W. A Survey on Big Data Market: Pricing, Trading and Protection. *IEEE Access* **2018**, *6*, 15132–15154. [CrossRef]
39. Dai, W.; Dai, C.; Choo, K.K.R.; Cui, C.; Zou, D.; Jin, H. SDTE: A Secure Blockchain-Based Data Trading Ecosystem. *IEEE Trans. Inf. Forensics Secur.* **2020**, *15*, 725–737. [CrossRef]
40. Hayashi, T.; Ishimura, G.; Ohsawa, Y. Structural Characteristics of Stakeholder Relationships and Value Chain Network in Data Exchange Ecosystem. *IEEE Access* **2021**, *9*, 52266–52276. [CrossRef]
41. Karafiloski, E.; Mishev, A. Blockchain Solutions for Big Data Challenges: A Literature Review. In Proceedings of the IEEE EUROCON 2017-17th International Conference on Smart Technologies, Ohrid, Macedonia, 6–8 July 2017; pp. 763–768.
42. GG. *Basic Law for the Federal Republic of Germany (Grundgesetz—GG)*; GG: Germany, 2019.
43. GDPR. *General Data Protection Regulation (GDPR)—Official Legal Text*; GDPR: Germany, 2016.
44. CFR. *Charter of Fundamental Rights of the European Union*; CFR: Germany, 2012.
45. ECHR. *European Convention on Human Rights—Official Texts, Convention and Protocols*; ECHR: Germany, 2013.
46. BDSG. *Federal Data Protection Act (Bundesdatenschutzgesetz—BDSG)*; BDSG: Germany, 2019.
47. TMG. *Telemedia Act of the Federal Republic of Germany (Telemediengesetz—TMG)*; TMG: Germany, 2020.
48. De Hert, P.; Papakonstantinou, V.; Malgieri, G.; Beslay, L.; Sanchez, I. The Right to Data Portability in the GDPR: Towards User-Centric Interoperability of Digital Services. *Comput. Law Secur. Rev.* **2018**, *34*, 193–203. [CrossRef]
49. UrhG. *Act on Copyright and Related Rights (Urheberrechtsgesetz—UrhG)*; UrhG: Germany, 2018.
50. StGB. *German Criminal Code (Strafgesetzbuch—StGB)*; StGB: Germany, 2019.
51. Welp, J. Datenveränderung (§ 303a StGB)—Teil 1. *JurPC* **1988**, *1988*, 443–449.
52. UWG. *Act against Unfair Competition (Gesetz Gegen Den Unlauteren Wettbewerb—UWG)*; UWG: Germany, 2019.
53. BGB. *German Civil Code (Bürgerliches Gesetzbuch—BGB)*; BGB: Germany, 2013.
54. Van Panhuis, W.G.; Paul, P.; Emerson, C.; Grefenstette, J.; Wilder, R.; Herbst, A.J.; Heymann, D.; Burke, D.S. A Systematic Review of Barriers to Data Sharing in Public Health. *BMC Public Health* **2014**, *14*, 1144. [CrossRef] [PubMed]
55. Bietz, M.J.; Bloss, C.S.; Calvert, S.; Godino, J.G.; Gregory, J.; Claffey, M.P.; Sheehan, J.; Patrick, K. Opportunities and Challenges in the Use of Personal Health Data for Health Research. *J. Am. Med. Inform. Assoc.* **2016**, *23*, e42–e48. [CrossRef]
56. Bertino, E.; Ooi, B.C.; Yang, Y.; Deng, R.H. Privacy and Ownership Preserving of Outsourced Medical Data. In Proceedings of the 21st International Conference on Data Engineering (ICDE'05), Tokoyo, Japan, 5–8 April 2005; pp. 521–532.
57. Andreu-Perez, J.; Poon, C.C.Y.; Merrifield, R.D.; Wong, S.T.C.; Yang, G.Z. Big Data for Health. *IEEE J. Biomed. Health Inform.* **2015**, *19*, 1193–1208. [CrossRef]
58. Badawi, O.; Brennan, T.; Celi, L.A.; Feng, M.; Ghassemi, M.; Ippolito, A.; Johnson, A.; Mark, R.G.; Mayaud, L.; Moody, G.; et al. Making Big Data Useful for Health Care: A Summary of the Inaugural MIT Critical Data Conference. *JMIR Med. Inform.* **2014**, *2*, e22. [CrossRef]
59. Groves, P.; Kayyali, B.; Knott, D.; Van Kuiken, S. The 'Big Data' Revolution in Healthcare. *Mckinsey Q.* **2013**, *2*, 1–22.
60. Kostkova, P.; Brewer, H.; de Lusignan, S.; Fottrell, E.; Goldacre, B.; Hart, G.; Koczan, P.; Knight, P.; Marsolier, C.; McKendry, R.A.; et al. Who Owns the Data? Open Data for Healthcare. *Front. Public Health* **2016**, *4*, 7. [CrossRef]
61. Goldacre, B. Care.Data Is in Chaos. It Breaks My Heart | Ben Goldacre. *The Guardian*, 28 February 2014.
62. Godlee, F. What Can We Salvage from Care.Data? *BMJ* **2016**, *354*, i3907. [CrossRef]
63. Kish, L.J.; Topol, E.J. Unpatients—Why Patients Should Own Their Medical Data. *Nat. Biotechnol.* **2015**, *33*, 921–924. [CrossRef]
64. Liang, X.; Zhao, J.; Shetty, S.; Liu, J.; Li, D. Integrating Blockchain for Data Sharing and Collaboration in Mobile Healthcare Applications. In Proceedings of the 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Montreal, QC, Canada, 8–13 October 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–5. [CrossRef]
65. Cachin, C. Architecture of the Hyperledger Blockchain Fabric. In *Workshop on Distributed Cryptocurrencies and Consensus Ledgers*; Academic Press: Cambridge, MA, USA, 2016.
66. Androulaki, E.; Manevich, Y.; Muralidharan, S.; Murthy, C.; Nguyen, B.; Sethi, M.; Singh, G.; Smith, K.; Sorniotti, A.; Stathakopoulou, C.; et al. Hyperledger Fabric: A Distributed Operating System for Permissioned Blockchains. In Proceedings of the Thirteenth EuroSys Conference—EuroSys '18, Porto, Portugal, 23–26 April 2018; ACM Press: New York, NY, USA,2018; pp. 1–15. [CrossRef]
67. Azaria, A.; Ekblaw, A.; Vieira, T.; Lippman, A. MedRec: Using Blockchain for Medical Data Access and Permission Management. In Proceedings of the 2016 2nd International Conference on Open and Big Data (OBD), Vienna, Austria, 22–24 August 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 25–30. [CrossRef]

68. Ekblaw, A.; Azaria, A.; Halamka, J.D.; Lippman, A. A Case Study for Blockchain in Healthcare: "MedRec" Prototype for Electronic Health Records and Medical Research Data. In Proceedings of the IEEE Open & Big Data Conference, Vienna, Austria, 22–24 August 2016; Volume 13, p. 13.
69. Priisalu, J.; Ottis, R. Personal Control of Privacy and Data: Estonian Experience. *Health Technol.* **2017**, *7*, 441–451. [CrossRef]
70. Harding, A.; Harper, B.; Stone, D.; O'Neill, C.; Berger, P.; Harris, S.; Donatuto, J. Conducting Research with Tribal Communities: Sovereignty, Ethics, and Data-Sharing Issues. *Environ. Health Perspect.* **2012**, *120*, 6–10. [CrossRef]
71. Lin, L.C. Data Management and Security in Qualitative Research. *Dimens. Crit. Care Nurs.* **2009**, *28*, 132–137. [CrossRef]
72. Evans, B.J. Much Ado about Data Ownership. *Harv. JL Tech.* **2011**, *25*, 69.
73. Rodwin, M.A. The Case for Public Ownership of Patient Data. *JAMA* **2009**, *302*, 86–88. [CrossRef]
74. Coombs, C.R.; Doherty, N.F.; Loan-Clarke, J. The Importance of User Ownership and Positive User Attitudes in the Successful Adoption of Community Information Systems. *J. Organ. User Comput. JOEUC* **2001**, *13*, 5–16. [CrossRef]
75. Krimsky, S. Publication Bias, Data Ownership and the Funding Effect in Science: Threats to the Integrity of Biomedical Research. In *Rescuing Science from Politics: Regulation and the Distortion of Scientific Research*; Cambridge University Press: Cambridge, UK, 2006; pp. 61–85.
76. Kocharov, A. Data Ownership and Access Rights in the European Food Safety Authority. *EFFL* **2009**, *4*, 13.
77. Safran, C.; Bloomrosen, M.; Hammond, W.E.; Labkoff, S.; Markel-Fox, S.; Tang, P.C.; Detmer, D.E. Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper. *J. Am. Med. Inform. Assoc.* **2007**, *14*, 1–9. [CrossRef]
78. Cios, K.J.; Moore, G.W. Uniqueness of Medical Data Mining. *Artif. Intell. Med.* **2002**, *26*, 1–24. [CrossRef]
79. Lederman, J.; Taylor, B.D.; Garrett, M. A Private Matter: The Implications of Privacy Regulations for Intelligent Transportation Systems. *Transp. Plan. Technol.* **2016**, *39*, 115–135. [CrossRef]
80. Joy, J.; Gerla, M. Internet of Vehicles and Autonomous Connected Car-Privacy and Security Issues. In Proceedings of the 2017 26th International Conference on Computer Communication and Networks (ICCCN), Vancouver, BC, Canada, 31 July–3 August 2017; pp. 1–9.
81. Hoh, B.; Gruteser, M.; Hui, X.; Alrabady, A. Enhancing Security and Privacy in Traffic-Monitoring Systems. *IEEE Pervasive Comput.* **2006**, *5*, 38–46. [CrossRef]
82. Duri, S.; Gruteser, M.; Liu, X.; Moskowitz, P.; Perez, R.; Singh, M.; Tang, J.M. Framework for Security and Privacy in Automotive Telematics. In Proceedings of the 2nd International Workshop on Mobile Commerce, Atlanta, GA, USA, 28 September 2002; ACM: New York, NY, USA, 2002; pp. 25–32.
83. Rass, S.; Fuchs, S.; Schaffer, M.; Kyamakya, K. How to Protect Privacy in Floating Car Data Systems. In Proceedings of the Fifth ACM International Workshop on VehiculAr Inter-NETworking, San Francisco, CA, USA, 15 September 2008; ACM: New York, NY, USA, 2008; pp. 17–22.
84. Dötzer, F. Privacy Issues in Vehicular Ad Hoc Networks. In *International Workshop on Privacy Enhancing Technologies*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 197–209.
85. Stübing, H.; Jaeger, A.; Bißmeyer, N.; Schmidt, C.; Huss, S.A. Verifying Mobility Data under Privacy Considerations in Car-to-X Communication. In Proceedings of the 17th ITS World CongressITS JapanITS AmericaERTICO, Busan, Korea, 25–29 October 2010; p. 12.
86. Festag, D.A.; Baldessari, R.; Zhang, D.W. CAR-2-X Communication for Safety and Infotainment in Europe. *NEC Tech. J.* **2008**, *3*, 21–26.
87. Automobil Produktion. *Abwehr-Allianz: Daimler, BMW, Ford und Volvo kooperieren mit HERE und TomTom*; Automobil Produktion: Landsberg, Germany, 2019.
88. Ma, M.; Huang, Y.; Chu, C.H.; Wang, P. User-Driven Cloud Transportation System for Smart Driving. In Proceedings of the 4th IEEE International Conference on Cloud Computing Technology and Science Proceedings, Taipei, Taiwan, 3–6 December 2012; pp. 658–665. [CrossRef]
89. Zhang, J.; Wang, F.Y.; Wang, K.; Lin, W.H.; Xu, X.; Chen, C. Data-Driven Intelligent Transportation Systems: A Survey. *IEEE Trans. Intell. Transp. Syst.* **2011**, *12*, 1624–1639. [CrossRef]
90. Zheng, X.; Chen, W.; Wang, P.; Shen, D.; Chen, S.; Wang, X.; Zhang, Q.; Yang, L. Big Data for Social Transportation. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 620–630. [CrossRef]
91. Alic, A.S.; Almeida, J.; Aloisio, G.; Andrade, N.; Antunes, N.; Ardagna, D.; Badia, R.M.; Basso, T.; Blanquer, I.; Braz, T.; et al. BIGSEA: A Big Data Analytics Platform for Public Transportation Information. *Future Gener. Comput. Syst.* **2019**, *96*, 243–269. [CrossRef]
92. Darwish, T.S.J.; Abu Bakar, K. Fog Based Intelligent Transportation Big Data Analytics in The Internet of Vehicles Environment: Motivations, Architecture, Challenges, and Critical Issues. *IEEE Access* **2018**, *6*, 15679–15701. [CrossRef]
93. Neilson, A.; Indratmo; Daniel, B.; Tjandra, S. Systematic Review of the Literature on Big Data in the Transportation Domain: Concepts and Applications. *Big Data Res.* **2019**, *17*, 35–44. [CrossRef]
94. Hirtan, L.A.; Dobre, C. Blockchain Privacy-Preservation in Intelligent Transportation Systems. In Proceedings of the 2018 IEEE International Conference on Computational Science and Engineering (CSE), Bucharest, Romania, 29–31 October 2018; pp. 177–184. [CrossRef]

95. Yuan, Y.; Wang, F.Y. Towards Blockchain-Based Intelligent Transportation Systems. In Proceedings of the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), Rio de Janeiro, Brazil, 1–4 November 2016; pp. 2663–2668. [CrossRef]
96. Singh, M.; Kim, S. Blockchain Based Intelligent Vehicle Data Sharing Framework. *arXiv* **2017**, arXiv:1708.09721.
97. Singh, M.; Kim, S. Intelligent Vehicle-Trust Point: Reward Based Intelligent Vehicle Communication Using Blockchain. *arXiv* **2017**, arXiv:1707.07442.
98. Lei, A.; Cruickshank, H.; Cao, Y.; Asuquo, P.; Ogah, C.P.A.; Sun, Z. Blockchain-Based Dynamic Key Management for Heterogeneous Intelligent Transportation Systems. *IEEE Internet Things J.* **2017**, *4*, 1832–1843. [CrossRef]
99. Otto, B.; Lohmann, S.; Steinbuss, S.; Andreas, T. *IDS Reference Architecture Model. Industrial Data Space, Version 2.0*; Technical Report; International Data Spaces Association: Berlin, Germany, 2018.
100. Otto, B.; Auer, S.; Cirullies, J.; Jürjens, J.; Menz, N.; Schon, J.; Wenzel, S. Industrial Data Space: Digital Souvereignity Over Data. In *Fraunhofer Gesellschaft zur Forderung der Angewandten Forschung*; Office for Official Publications of the EU: Luxembourg, 2016, [CrossRef]
101. Otto, B.; Steinbuß, S.; Teuscher, A.; Lohmann, S. *IDS Reference Architecture Model*, Version3.0; Technical Report; International Data Spaces Association: Berlin, Germany, 2019.
102. BMWi. *GAIA-X: The European Project Kicks off the Next Phase*; Technical Report; Federal Ministry for Economic Affairs and Energy (BMWi): Berlin, Germany, 2020.
103. Biegel, F.; Bongers, A.; Chidambaram, R.; DE-CIX Management GmbH; Feld, T.; Garloff, K.; Ingenrieth, F.; Jochem, M.; Maier, B.; Marsch, C.; et al. *GAIA-X: Driver of Digital Innovation in Europe*; Technical Report; Federal Ministry for Economic Affairs and Energy (BMWi): Berlin, Germany, 2020.
104. DE-CIX Management GmbH; Eggers, G.; Fondermann, B.; Google Germany GmbH; Maier, B.; Ottradovetz, K.; Pfrommer, J.; Reinhardt, R.; Rollin, H.; Schmieg, A.; et al. *GAIA-X: Technical Architecture*; Technical Report; Federal Ministry for Economic Affairs and Energy (BMWi): Berlin, Germany, 2020.
105. Giaconi, G.; Gunduz, D.; Poor, H.V. Smart Meter Privacy with Renewable Energy and an Energy Storage Device. *arXiv* **2017**, arXiv:1703.08390.
106. Rusitschka, S.; Curry, E. Big Data in the Energy and Transport Sectors. In *New Horizons for a Data-Driven Economy*; Springer: Cham, Switzerland, 2016; pp. 225–244.
107. Balough, C.D. Privacy Implications of Smart Meters. *Chi.-Kent L. Rev.* **2011**, *86*, 161.
108. Asghar, M.R.; Dan, G.; Miorandi, D.; Chlamtac, I. Smart Meter Data Privacy: A Survey. *IEEE Commun. Surv. Tutorials* **2017**, *19*, 2820–2835. [CrossRef]
109. McHenry, M.P. Technical and Governance Considerations for Advanced Metering Infrastructure/Smart Meters: Technology, Security, Uncertainty, Costs, Benefits, and Risks. *Energy Policy* **2013**, *59*, 834–842. [CrossRef]
110. Verbong, G.P.; Beemsterboer, S.; Sengers, F. Smart Grids or Smart Users? Involving Users in Developing a Low Carbon Electricity Economy. *Energy Policy* **2013**, *52*, 117–125. [CrossRef]
111. Cavoukian, A.; Polonetsky, J.; Wolf, C. SmartPrivacy for the Smart Grid: Embedding Privacy into the Design of Electricity Conservation. *Identity Inf. Soc.* **2010**, *3*, 275–294. [CrossRef]
112. Chakravorty, A.; Wlodarczyk, T.; Rong, C. Privacy Preserving Data Analytics for Smart Homes. In Proceedings of the 2013 IEEE Security and Privacy Workshops, San Francisco, CA, USA, 23–24 May 2013; pp. 23–27. [CrossRef]
113. Efthymiou, C.; Kalogridis, G. Smart Grid Privacy via Anonymization of Smart Metering Data. In Proceedings of the 2010 First IEEE International Conference on Smart Grid Communications, Gaithersburg, MD, USA, 4–6 October 2010; pp. 238–243. [CrossRef]
114. Molina-Markham, A.; Shenoy, P.; Fu, K.; Cecchet, E.; Irwin, D. Private Memoirs of a Smart Meter. In Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building—BuildSys '10, Zurich, Switzerland, 3–5 November 2010; ACM Press: New York, NY, USA, 2010; p. 61, [CrossRef]
115. Aitzhan, N.Z.; Svetinovic, D. Security and Privacy in Decentralized Energy Trading Through Multi-Signatures, Blockchain and Anonymous Messaging Streams. *IEEE Trans. Dependable Secur. Comput.* **2018**, *15*, 840–852. [CrossRef]
116. Mengelkamp, E.; Notheisen, B.; Beer, C.; Dauer, D.; Weinhardt, C. A Blockchain-Based Smart Grid: Towards Sustainable Local Energy Markets. *Comput.-Sci.-Res. Dev.* **2018**, *33*, 207–214. [CrossRef]
117. DeKeles, J. Smart Cities Council | Our Vision, 2012.
118. Bajracharya, B.; Cattell, D.; Khanjanasthiti, I. Challenges and Opportunities to Develop a Smart City: A Case Study of Gold Coast, Australia. In *REAL CORP 2014–PLAN IT SMART! Clever Solutions for Smart Cities, Proceedings of 19th International Conference on Urban Planning, Regional Development and Information Society*; CORP—Competence Center of Urban and Regional Planning: Sitges, Spain, 2014; pp. 119–129.
119. Edelenbos, J.; Hirzalla, F.; van Zoonen, L.; van Dalen, J.; Bouma, G.; Slob, A.; Woestenburg, A. Governing the Complexity of Smart Data Cities: Setting a Research Agenda. In *Smart Technologies for Smart Governments*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 35–54.
120. Khan, Z.; Pervez, Z.; Abbasi, A.G. Towards a Secure Service Provisioning Framework in a Smart City Environment. *Future Gener. Comput. Syst.* **2017**, *77*, 112–135. [CrossRef]

121. Morozov, E.; Bria, F. Rethinking the Smart City. In *Democratizing Urban Technology*; Rosa Luxemburg Foundation: New York, NY, USA, 2018.
122. Bischof, S.; Karapantelakis, A.; Nechifor, C.S.; Sheth, A.P.; Mileo, A.; Barnaghi, P. *Semantic Modelling of Smart City Data*; Wright State University: Dayton, OH, USA, 2014.
123. CityPulse. *CityPulse: Real-Time IoT Stream Processing and Large-Scale Data Analytics for Smart City Applications*; Citypulse: Hong Kong, China, 2019.
124. FIESTA-IoT. FIESTA-IOT—Federated Interoperable Semantic IoT Testbeds and Applications, 2019.
125. Jara, A.J.; Serrano, M.; Gómez, A.; Fernández, D.; Molina, G.; Bocchi, Y.; Alcarria, R. Smart Cities Semantics and Data Models. In Proceedings of the International Conference on Information Technology & Systems (ICITS 2018), Libertad City, Ecuador, 10–12 January 2018; Springer International Publishing: Cham, Switzerland, 2018; Volume 721, pp. 77–85. [CrossRef]
126. Zhang, K.; Ni, J.; Yang, K.; Liang, X.; Ren, J.; Shen, X.S. Security and Privacy in Smart City Applications: Challenges and Solutions. *IEEE Commun. Mag.* **2017**, *55*, 122–129. [CrossRef]
127. van Zoonen, L. Privacy Concerns in Smart Cities. *Gov. Inf. Q.* **2016**, *33*, 472–480. [CrossRef]
128. Kakarontzas, G.; Anthopoulos, L.; Chatzakou, D.; Vakali, A. A Conceptual Enterprise Architecture Framework for Smart Cities - A Survey Based Approach. In Proceedings of the 11th International Conference on E-Business, Vienna, Austria, 28–30 August 2014; SCITEPRESS—Science and and Technology Publications: Setubal, Portugal, 2014; pp. 47–54. [CrossRef]
129. Molinari, A.; Maltese, V.; Vaccari, L.; Almi, A.; Bassi, E. Big Data and Open Data for a Smart City. In *IEEE-TN Smart Cities White Papers*; IEEE: Trento, Italy, 2014.
130. Al Nuaimi, E.; Al Neyadi, H.; Mohamed, N.; Al-Jaroodi, J. Applications of Big Data to Smart Cities. *J. Internet Serv. Appl.* **2015**, *6*, 25. [CrossRef]
131. Pramanik, M.I.; Lau, R.Y.; Demirkan, H.; Azad, M.A.K. Smart Health: Big Data Enabled Health Paradigm within Smart Cities. *Expert Syst. Appl.* **2017**, *87*, 370–383. [CrossRef]
132. Soto, J.Á.C.; Werner-Kytölä, O.; Jahn, M.; Pullmann, J.; Bonino, D.; Pastrone, C.; Spirito, M. Towards a Federation of Smart City Services. In *International Conference on Recent Advances in Computer Systems*; Atlantis Press: Berlin/Heidelberg, Germany, 2015.
133. Ramachandran, G.S.; Radhakrishnan, R.; Krishnamachari, B. Towards a Decentralized Data Marketplace for Smart Cities. In Proceedings of the 2018 IEEE International Smart Cities Conference (ISC2), Kansas City, MO, USA, 16–19 September 2018; pp. 1–8. [CrossRef]
134. Biswas, K.; Muthukkumarasamy, V. Securing Smart Cities Using Blockchain Technology. In Proceedings of the 2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), Sydney, Australia, 12–14 December 2016; pp. 1392–1393. [CrossRef]
135. Sharma, P.K.; Moon, S.Y.; Park, J.H. Block-VN: A Distributed Blockchain Based Vehicular Network Architecture in Smart City. *JIPS* **2017**, *13*, 184–195. [CrossRef]
136. Michelin, R.A.; Dorri, A.; Lunardi, R.C.; Steger, M.; Kanhere, S.S.; Jurdak, R.; Zorzo, A.F. SpeedyChain: A Framework for Decoupling Data from Blockchain for Smart Cities. *arXiv* **2018**, arXiv:1807.01980.
137. Sun, J.; Yan, J.; Zhang, K.Z.K. Blockchain-Based Sharing Services: What Blockchain Technology Can Contribute to Smart Cities. *Financ. Innov.* **2016**, *2*, 26. [CrossRef]
138. Rivera, R.; Robledo, J.G.; Larios, V.M.; Avalos, J.M. How Digital Identity on Blockchain Can Contribute in a Smart City Environment. In Proceedings of the 2017 International Smart Cities Conference (ISC2), Wuxi, China, 14–17 september 2017; pp. 1–4. [CrossRef]
139. Mühle, A.; Grüner, A.; Gayvoronskaya, T.; Meinel, C. A Survey on Essential Components of a Self-Sovereign Identity. *arXiv* **2018**. [CrossRef]
140. World Wide Web Consortium. *Verifiable Credentials Data Model 1.0: Expressing Verifiable Information on the Web*; World Wide Web Consortium: Cambridge, MA, USA, 2019.
141. Lagutin, D.; Kortesniemi, Y.; Fotiou, N.; Siris, V.A. Enabling Decentralised Identifiers and Verifiable Credentials for Constrained IoT Devices Using OAuth-Based Delegation. In Proceedings of the 2019 Workshop on Decentralized IoT Systems and Security, San Diego, CA, USA, 24 February 2019. [CrossRef]
142. Dehghani, Z. How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh. 2019. Available online: https://martinfowler.com/ (accessed on 21 August 2021).
143. Data Privacy and Ownership New.
144. KNOWN™ Platform | The next Chapter in Distributed Ledger Innovation.
145. Mine–The Future of Data Ownership.
146. LeapXpert Compliance & Data Ownership Focused Messaging Orchestration Platform.
147. Genomics Personalized Health, 2020.