telecom

MDPI

Review

A Survey on User Profiling, Data Collection, and Privacy Issues of Internet Services

Dardan Maraj *, Marin Vuković and Petar Hotovec

Faculty of Electrical Engineering and Computing, University of Zagreb, 10000 Zagreb, Croatia; marin.vukovic@fer.hr (M.V.); petar.hotovec@fer.hr (P.H.)
* Correspondence: dardan.maraj@fer.hr

Abstract: Users are usually required to share several types of data, including their personal data, as different providers strive to offer high-quality services that are often tailored to end-users' preferences. However, when it comes to personalizing services, there are several challenges for meeting user's needs and preferences. For content personalization and delivery of services to end users, services typically create user profiles. When user profiles are created, user data is collected and organized to meet the personalization requirements of the services. In this paper, we provide an overview of current research activities that focus on user profiling and ways to protect user data privacy. The paper presents different types of data that services collect from users on examples of commonly used Internet services. It proposes data categorization as a prerequisite for controlled data sharing between users and Internet services. Furthermore, it discusses how data generalization can be used for anonymization purposes on examples of the proposed data categories. Finally, it gives an overview of the privacy framework being developed and gives guidelines for future work focusing on data generalization methods in order to reduce user privacy risks.

Keywords: user profiling; personalization; internet services; privacy issues; data protection; data collection

check for updates

Citation: Maraj, D.; Vuković, M.; Hotovec, P. A Survey on User Profiling, Data Collection, and Privacy Issues of Internet Services. *Telecom* **2024**, *5*, 961–976. https://doi.org/10.3390/telecom5040048

Academic Editor: Petros Bithas

Received: 17 August 2024
Revised: 17 September 2024
Accepted: 23 September 2024
Published: 1 October 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

To ensure that services are provided to the right people, in the right form, and at the right time, service providers need to know the needs, preferences, and behavior of their users. Obtaining this knowledge about the user, which is based on a variety of independent characteristics and information, is a difficult task. It is necessary to combine and integrate the user's current situation, history, and social environment in order to fully understand the user context, previous interactions, activities, and relationships with other people. On the other hand, the increase in information on the Internet and the enormous diversity of users gives high priority to personalization in order to provide better services.

With the development of new technologies, privacy issues are becoming a major concern for users of these systems. With the advances in information and communication technologies, a clear need for personalized information systems has emerged. These systems aim to customize the functionality of information exchanges to the specific interests and requirements of their users. The term personalized information system is closely related to user profiling, where a range of data and parameters are collected to create comprehensive profiles for users. In general, users provide a lot of data when creating their profiles on various Internet platforms, such as web and mobile applications. In some cases, the data that users provide in exchange for creating their profiles may be personally identifiable, while there are cases where this information is not sensitive in this sense. Generally, the parameters that may be collected in user profiling are demographic data, behavioral data, device and technology data, communication data, contextual data, medical data, accessibility data, user feedback data, etc. [1].

Nowadays, users have access to a wide range of Internet services across different devices. User profiles are crucial for service providers to successfully personalize their content in this competitive market. Therefore, the main goal of personalized services is to collect and analyze users' personal information. The success of these services depends on how well the service provider understands the users and how well this is reflected in the services. In this sense, user profiles are the result of the user profiling process and serve as a representation of the users tailored to the specific service [2].

Depending on the area of application, the content of the user profile and its scope may change. Regardless of the data, the completeness of the user profile is determined by the methods used to collect and organize the data about the user and how well this data represents the individual. In the literature, two main methods of collecting data about users are distinguished. These methods are referred to as explicit or implicit data collection methods. In explicit methods, users voluntarily provide the system with information about their interests and preferences. In contrast, in the implicit methods, user data are collected dynamically by automatically monitoring the user's interactions with the system, usually through cookies on the websites, APIs, various types of sensors, and so on [2].

In this paper, we will provide an overview of existing services and the data they collect and process to personalize their services for users. In addition, we will highlight the current state-of-the-art in preserving user privacy in the context of personalizing services and propose guidelines for further work in the area of anonymizing users for personalizing services. In this context, our goal is to determine the extent of similarity of collected data across services and provide a framework for categorizing anonymous user data according to the type of services used.

As this is a survey paper that aims to categorize user-related data gathered by common Internet services, specific methods for generalizing user data to enhance privacy protection are presented on examples and will be more formally defined and implemented in future work. The main hypothesis of the paper is that services need only specific user data in order to personalize the provided content. By grouping the data, we can define which groups of data the services are allowed to access. Furthermore, the generalization ability enables blurring of user data in order to make it less personally identifiable and, thus, preserving privacy. In this sense, the goal is to categorize user data and offer generalization ability in order to (i) provide only data that services really need and (ii) blur the provided data to an extent that is acceptable to both users and services. The trade-off between personalization and privacy is evident and quite clear because if users want to receive more personalized services they must share more personal data, whereas maintaining higher privacy level, requires 'sacrificing' some degree of personalization.

This paper is organized as follows: After the introduction part, Section 2 presents the most popular Google services and the user data they collect. Section 3 provides an overview of related work by various researchers on privacy and data personalization in user profiling. Section 4 proposes a method for data collection and its classification that includes different types of services. Finally, conclusions and guidelines for future work are given in Section 5.

2. Theoretical Background

An Overview of Commonly Used Services and Collected Data from Users

Nowadays, users of various Internet platforms exchange different types of data, including their personal information, such as name, age, gender, location, etc., in order to use their preferred services.

Some of the users are aware of the potential risks associated with privacy breaches, but there are also users who are not concerned when sharing their data or are even aware of the possible consequences of data leaks. For example, Google collects various types of data that help the company to personalize and target advertising to users, especially data linked to the user's identifier. Google may also use some data to monitor the continued functionality of its applications. This may include diagnostic and crash data that inform

the company or organization on why a particular application stops working on certain user devices at different times.

On the other hand, when capturing user activity, such as a shopping session, a user may use Google to search for a book, then visit the Wikipedia page to learn more about it, compare prices on Amazon, and then complete the purchase on the website. Later, the same user might start experimenting with technology after seeing a post about the latest iPad model on a friend's Facebook page. This sequence of an individual's online activities comprises visits to various websites across different categories, reflecting a combination of the user's various interests and behaviors [3].

In this survey paper, we have analyzed some of the most widespread Internet services aimed to see how many users use them. We selected some Google services and some other services, based on their popularity and from usage statistics presented in the paper [4]. According to statistics, Google and its services dominate the global search market, holding a significant majority share. These statistics highlight their significant impact, which is why we mostly focused on them. Besides the most common parameters, we have also introduced accessibility parameters because this is something specific that recent privacy regulations have brought to our attention as a new factor. These parameters are becoming more relevant due to their importance for meeting the needs of people with various disabilities, even though it is not yet so widespread. In this sense, we will describe some of the most popular Google services and the data collected by each service for personalization purposes. The commonly used Google services that we focus on include YouTube, Gmail, Google Search, Google Assistant, Google Play Store, Google Maps, and Google Calendar. On the other hand, we also examine other popular Internet services to compare the data they collect with that of Google services and to identify similarities in their data collection and processing practices.

Google uses several methods to collect user data. The most obvious are the explicit ones, where the user actively and knowingly provides information to Google, e.g., when logging in to one of the popular services, such as YouTube, Gmail, Search, etc. Implicit methods are less obvious ways for Google to collect user data, where an application collects information while it is being used, possibly without the user's understanding. Google's passive data collection methods arise through platforms (e.g., Android and Google Chrome) and applications (e.g., Google Search, YouTube, Google Maps, Google Analytics, AdSense, Ad Mob, AdWords, etc.) [5].

In general, Google uses a wide range of parameters in its various services to provide personalized experiences to its users. For example, Google Search uses various parameters to personalize search results, including search history, location, language, device type, etc. These types of parameters used by Google Search can be categorized into location parameters and user preference parameters.. On the other hand, Google Maps uses similar parameters to Google Search, including search history, location, time of day, etc., to provide personalized recommendations and directions. YouTube uses parameters, such as viewing history, search history, and location, to recommend videos and personalize the home page for each user. Google Assistant uses parameters, such as speech, voice recognition, and search history, to provide personalized assistance and recommendations. On the other hand, the Google Ads service uses parameters, such as search history, location, and inferred interests, to personalize the ads displayed to users. Some of the parameters used in Google Ads are also used by Google News to personalize news content and recommendations [5].

Google Play Store uses parameters, such as app usage history, search history, and location, to recommend apps and personalize the home page for each user. Google Chrome uses parameters, such as browsing history, bookmarks, and location, to personalize the browsing experience and provide personalized recommendations to its users. Gmail uses several parameters to personalize email communication. Some of the parameters collected by the Gmail service to personalize users' emails are: first and last name, email address, location, job title, company name, etc. Generally, different webmail applications collect various types of data for user personalization with the aim of providing better services to

them. Even though the collection of the user's personal information raises privacy concerns, such data can be collected and used to provide personalized services and improve the user experience with the user's appropriate consent and transparency. Various parameters can be collected by a webmail or any other application for personalization purposes, such as:

- Email content: this type of data can be analyzed to identify topics, interests, and preferences of users. This can help in providing personalized recommendations, targeted ads, and content based on user interests [6].
- Contact lists: The contacts that users mostly interact with can be used to suggest connections, networking opportunities, and relevant events [6].
- Usage patterns: Data on the frequency and timing of emails can provide insights into user behavior, such as work hours, preferred communication methods, and time zones. This information can be used to optimize email delivery and improve user experience [7].
- Location data: If the user has provided permission to access location data, they can be used to provide local weather updates, news, events, and recommendations based on their location. Furthermore, it can enable location prediction and more personalized services [8].
- Device information: this kind of information can be used to provide personalized recommendations and optimize the application for the user's device [9].

Different groups of parameters, including location parameters, user demographics, accessibility requirements, user preference parameters, and behavioral parameters, are listed in Table 1. On the other hand, different data are provided depending on the parameter group. In this paper, we will also classify various data collected from different services and their relationship with parameters listed in Table 1.

Table 1. Specific parameters grouped according to proposed categories.

Location Parameters [10]	Demographic Parameters [11,12]	Accessibility Parameters [11,12]	Preference Parameters [13]	Behavior Parameters [13]
○ Country ○ Region ○ City ○ Town ○ GPS coordinates ○ IP address	○ Name ○ Surname ○ Age ○ Gender ○ E-mail ○ Income level ○ Education level ○ Occupation ○ Country	○ Visual category (e.g., low vision, blindness, color blindness, etc.) ○ Hearing impairment (e.g., captions, transcripts, etc.) ○ Motor impairment (e.g., speech recognition software) ○ Cognitive impairment (e.g., simple language, multiple formats, eye tracking web page, reading difficulties, difficulties in understanding, etc.) ○ Visual aids ○ Sign language ○ Text-to-speech ○ Voice recognition	○ User's browsing history ○ Search queries ○ Purchase history ○ Device types ○ Clicked hyperlinks	○ Browsing and search history ○ Frequently visited websites ○ Frequently page visited ○ Actions taken on websites ○ Likes ○ Shares ○ Comments ○ Posts

3. Related Work

This chapter presents key findings from different authors related to data collection from various Internet services. As stated earlier, our idea is to see what parameters the various Internet services use and somehow to group them later in the paper.

It is important to note that data collection practices may vary from one platform to another and are subject to privacy policies and regulations. Users often have the ability to manage their privacy settings and control the extent to which their data are collected and used for personalization purposes. Personalization on various Internet platforms involves the collection and analysis of user data to provide tailored experiences, recommendations, and targeted advertising. In general, there are numerous specific data that are collected

and used by different Internet platforms, but here are some common types of data that are collected for personalization. These data include user profiling data, browsing history, search queries, clickstream data, device information, location data, social media activity, user feedback data, etc.

Nowadays, it is common practice for various websites and multimedia content applications, such as YouTube, Netflix, Zoom, etc. to collect user data for various reasons and purposes. This can be performed to make recommendations to users, to tailor services to specific device/network capabilities or preferences, and to perform detailed analysis for service improvement, marketing, sales, and other purposes. On the other hand, the question often arises as to how the data collected may affect the privacy of the individual. Data privacy is a multidimensional concept that can be understood from various perspectives, including the purpose of data collection and privacy preferences of users. These aspects are crucial for developing comprehensive user privacy models and strategies [14]. On the other hand, data privacy also involves with the ways of collecting user's data, sharing those data with third parties, and deleting them in accordance with laws and privacy regulations [15].

With regards to the privacy issues of users and user-related data collected by various Internet services and users' concerns about their privacy, numerous research papers have been published. For example, considering Zoom as one of the most popular conferencing applications and its privacy issues and policies, it collects and stores numerous personal data about users. The user-related data collected by this platform include personal data, such as name and contact details, IP address and device identifiers, and user-generated data, arising from meetings, messages, files shared between participants, etc. The Zoom platform also includes a passive way of collecting user-related data via the use of cookies and tracking technology such as browser type, service provider, operating system etc. On the other side, Zoom does not store video and audio content of the user's meetings, unless the user has chosen this option in the user settings [16,17].

Regarding the collection of accessibility data, authors in [18] presented an evaluation of users of virtual classrooms about their accessibility needs in an e-learning system. The survey has been performed by 115 people, of whom 62 were professionals, including 53 students with visual, auditory, and mobility impairment, as well as non-disabled students. Through this evaluation, they collected evidence of user experience when establishing the different accessibility needs within an e-learning platform, such as visual, auditive, cognitive, motor, elderly, and linguistic impairments of the users. The parameters acquired in the presented study are used for creating an accessibility group of parameters.

In another example, if we consider the popular platform Netflix from the perspective of user data collection, we can see that user information is gathered in different ways. First, the registration process includes collection of information, including name, email address, payment methods, telephone number, etc. Second, automatic data collection during service usage is used to collect information about user interactions with the platform, accessed content, and information regarding user's devices, such as device ID, device and software characteristics (e.g., connection information, IP address from which it can gather the information about the location and so on [19].

Some relevant works regarding the types of user data collected for user profiling by Internet platforms, such as Facebook and Twitter and collection methods of the data are provided in Table 2:

Table 2. Some selected works on collected user data for user profiling in social networks.

Ref. No.	Types of User-Related Data	How Are Data Collected and Collection Mechanism?	Internet Platform
[20]	Demographic data (the user's personal information, such as age, gender, education, and personality)	The data are gathered from user's profiles from a Tweet's text through a Machine Learning models for detecting social bots on Twitter.	Twitter
[21]	User behavior data (e.g., personal profile in Facebook, liked posts, links, status tagging of other users in photos, etc.)	Personality traits prediction model based on user behaviors on social networks	Facebook
[22]	Demographic data (e.g., name, gender, age, native language, education, family members, etc.)	The data are collected through a hybrid profiling system (explicit and implicit user profile). The implicit user profile is created by collecting user's data based on their interests in terms of keyworks, while implicit user profile is created by monitoring user activities by storing the browsing history, the time spent on each page, printing, saving, etc.	Different Internet platforms
[23]	Preference data (e.g., news data and political preferences of users)	The data are collected automatically through algorithms used by Facebook. The number of liked pages related to news and politics assigned is highly correlated with the number of ad topics related to news and politics assigned to each user. These topics do not represent ads that individuals have seen, but topics that may be interested in.	Facebook
[24]	Demographic and preference user data (e.g., user tweets, user follower's tweets, URLs posted by user followers, URLs posted by followers of followers, etc.) Also, a Deep Learning approach was used to extract status updates, profile photos, and page likes from a Facebook user's profile.	Data are collected through a Statistical Based Approach and Deep Learning approach. Also, a variety of information is extracted through a Neighborhood Approach, including user interests, the number of followers, the strength of the relationship between the person and their followers, and much more.	Facebook and Twitter

3.1. Privacy Issues Regarding Collected Data

When looking at user profiling and service personalization, related research shows that users are generally concerned about what happens with their personal data, how the data are stored, who has access to the data and what are the safeguards for protecting their data. This is partially due to the awareness raised by numerous data breaches of different Internet platforms that have happened over the years [25]. Privacy issues regarding the collection of data in the user profiling process can lead to various consequences related to the leakage of personal data, both for individuals and businesses. These consequences to individuals can be related to identity theft, financial loss, discrimination, reputational damage, unauthorized access to personal accounts, loss of privacy, and so on. On the other hand, organizations may face a lot of consequences as well, including loss of customer trust and reputation, infringement of intellectual property, disruption of operations, regulatory penalties, compliance violations, etc. The identification of such consequences provides a more thorough understanding of the privacy concerns and serves to emphasize the importance of data protection for individuals and organizations [26].

Some of the most well-known breaches are Yahoo (2013–2014), Equifax (2017), Marriot International hotels (2014–2018), Facebook–Cambridge Analytica (2013–2018), Capital One (2019), SolarWinds (2020), etc. In 2016, Yahoo informed the public that it had multiple data breaches between 2013 and 2014. The breaches compromised the personal information of more than 3 billion user accounts, including names, email addresses, telephone numbers, birthdates, hashed passwords, as well as the security questions of over 500 million users [27].

Equifax, one of the largest credit reporting agencies in USA, suffered a data breach that exposed the personal information of approximately 147 million people. The breaches compromised data, such as names, Social Security numbers, birth dates, addresses, and in some cases, drivers' license numbers [28].

Marriot International hotels announced in 2018 that unauthorized access happened to its database early in 2014, compromising the personal information of approximately 500 million users. The stolen data included names, passport numbers, email addresses, phone numbers, and other travel details [29].

Facebook–Cambridge Analytica (2013–2018) revealed that the political consulting firm Cambridge Analytica has misused personal data from around 87 million Facebook profiles without users' consent. This data were used for targeted political advertising during the 2016 US presidential election [30].

The Capital One data breach in 2019 impacted 106 million customers and the impact was similar with the data breach of Equifax. The breach of this bank included names, addresses, credit scores, and Social Security numbers [31].

The SolarWinds data breach was a highly sophisticated supply chain attack, which targeted US government agencies as well as private companies. This hack impacted an estimated 18,000 of its 300,000 customers worldwide. The breach allowed hackers to access and monitor sensitive information from the victim's systems [32].

These are some of the well-known examples of data breaches and, unfortunately, there have been many more incidents impacting numerous companies over the world. Besides the large data breaches reported by the global companies as mentioned in this section, there are numerous breaches of smaller websites and more specific services that usually remain unnoticed as the companies are typically not willing to go public due to reputation concerns. Nevertheless, data breaches highlight the importance of robust cybersecurity measures and the need for individuals and companies to protect their data.

3.2. Methods of Protection and Regulatory Aspects

Almost every country has approved data privacy laws, aiming to regulate how information is collected and what control companies have over the data during their transmission. Failure to follow data privacy regulations may lead to lawsuits or even prohibition of an application's use in certain jurisdictions. Adherence to these rules can be daunting, but all Internet applications should be familiar with the data privacy laws that may affect their users.

The most important general data protection legislation in Europe, currently, is the General Data Protection Regulation (GDPR). It establishes a set of rules for the collection, use, transmission, and security of data collected from residents of any of the 28 member countries of the European Union (EU) [25].

Under the GDPR, the types of data considered as personal, or personally identifiable, include name, address, photos, etc. GDPR extends the definition of personal data so that something like an IP address can be considered as personal data, as the user might be identified personally through the IP address and telecom operator. It also includes sensitive data, such as genetic and biometric data, which could be processed to uniquely identify a user [33].

On the other hand, the United States does not completely restrict cross-border data flows and traditionally has regulated data privacy at a sector level to cover certain types of data. Both the U.S. and EU countries are committed to supporting individual privacy

rights and ensuring the protection of personal data, including electronic data. However, there are essential differences in the treatment of data privacy between the U.S. and the EU, and security relations between the two continents have long been a discussion point. The GDPR highlights some of the main differences and challenges for U.S. companies operating in the EU. Since 2000, many entities used U.S.–EU negotiated agreements for cross-border data flows, but the EU's top court has invalidated those agreements due to concerns about U.S. surveillance laws. Many U.S. companies have made changes according to GDPR, such as revising and clarifying user terms of agreements and seeking explicit consent. Consequently, this creates more requirements for companies that collect user data and with respect to this issue, numerous experts in the field think that the GDPR may simplify compliance for U.S. companies, because the EU applies the same set of data protection rules [33].

It is a known fact that due to the large number of data breaches and hacks that occur, different user data, such as email address, password, social security number, or confidential health records, have been exposed on the Internet. Different companies are responsible for notifying the appropriate national bodies as soon as possible in order to ensure that EU citizens take appropriate measures to prevent their data from being abused [34].

There are evident challenges and limitations in implementing the privacy framework across different Internet services, especially regarding interoperability and compliance. Basically, users can provide all personal data to the services if they give their consent. The heterogeneity of the systems, where services can vary in terms of data collection and different protocols for data processing, storage, and transmission, is one of the key challenges and limitations in the implementation of privacy frameworks. On the other hand, differences between global privacy regulations can be considered another challenge in implementing any privacy framework and respectively of our privacy model. In this sense, different privacy regulations are supported by corresponding data protection laws, such as the GDPR in Europe, the CCPA in California, and other national regulations (e.g., the LGPD in Brazil). The overlapping of legal requirements between these privacy regulations can be challenging as well. Also, it is well known that these regulations grant users with specific rights, such as the right to access or delete their data. In this aspect, the most challenging part is related to the user's location and location of the user's data [35].

There are numerous examples of related works on storing data on various Internet platforms. The field of data storage is constantly evolving, as new technologies and methodologies emerge to address the challenges of managing and storing large-scale data in distributed environments [36].

In this regard, authors in [36] proposed a platform called PRIPRO ("Privacy Profiles"), designed to address user profiling and data storage concerns in smart environments while emphasizing privacy. This concept is crucial as smart environments often collect and utilize user data for various purposes, including automation, personalization, and analytics. This platform also includes mechanisms for securely storing user profiles and associated data. This includes encryption, access control, and data retention policies. The platform presented in this paper assists in defining the set of rules for the smart environment and updating the user's profile. Based on that, authors implemented a solution that is composed of a client application to run on the user's smartphone and a web server in order to manage the evolution of the user's profile and permissions.

Authors in [37] presented the middleware Ubiquitous Privacy (UbiPri) for the control and management of data privacy based on a generic model. This model includes several modules, such as Data Base, Controller module, Data module, Privacy Control Management User (PRICMU), Privacy Communication (PRICOM), Privacy Devices (PRIDEV), Privacy Profile (PRIPRO), Privacy Adaptation (PRIADA), Privacy Environment (PRIENV), Privacy Criteria (PRICRI), Privacy History (PRIHIS), and Privacy Security (PRISEC). The PRIHIS module stores and handles information related to the user's history, environment, devices, and other variables that can be added later to obtain contextual information. It is important to mention that all modules work independently, with characteristics that vary according

to the rules previously established in the environment. Therefore, for a module to be able to operate correctly, access is allowed to the data contained in the database.

When it comes to Internet services, we know that they collect our personal data on a daily basis. At the beginning, Internet services have collected only a few types of personal data, such as names, ages, birthdays, etc., but with the rapid spread of social networks and personalized services, more and more personal data have been collected. Furthermore, according to authors in [38], IoT devices are increasingly being adopted by consumers, making it possible for companies to capture personal data, including sensitive data, with much less effort and at very low cost. Recent systems architectures aim to collect, store, and process personal data in the cloud with very limited control available to end-users. Therefore, a platform called Personal Data Stores (PDS) has been proposed as an alternative architecture where personal data will be stored, giving the users complete control over their data. On the other side, data sovereignty is another concept to the PDS model, which is defined as the capability of users to have full control as well as to determine their restrictions about the usage of their data, such as access control, authorization, usage duration, etc., before sharing with third parties.

3.3. How Do Users Perceive Privacy?

It is expected that privacy awareness will be one of the most significant developments, with big pressure put on governments to implement data protection laws and to monitor how different companies handle an individual's data. However, there is still much work to be performed on this issue. According to the results of several surveys, the conclusion may be drawn that a lot of individuals still do not know how to protect their data and, thus, show distrust in the way their data are handled. On the other hand, most people today understand that it is beneficial and often necessary to provide certain personal data to companies and online applications they use in order to receive benefits for the services they want. At the same time, users are concerned about the privacy of their data [39].

Regarding these issues, the Cisco Consumer Privacy Survey conducted in 2019, shows different perspectives on what companies, governments, and individuals have done, and could do, to better protect data privacy. This survey involved 2600 adult respondents in twelve of the world's largest economies, where five of these were in Europe, four in Asia Pacific, and three in the USA. The profile of respondents was different in terms of age, gender, and income levels. Respondents were asked mostly about their attitudes and actions regarding their personal data, the products, and services they use as well as the impact of data privacy regulations on their behavior. Specifically, the survey explored different aspects of how people care about data privacy and what actions have been taken to protect it. When it comes to data protection, the big question has been whether they are willing to act, for example, by giving up certain benefits or paying more for stronger protection and control. On the other hand, user awareness has recently increased because of EU's GDPR, which requires disclosures to users about their rights, and constant headlines of data breaches affecting billions of data records and millions of users. With the Cisco Consumer Privacy Survey, the main purpose was to better understand user behaviors and how far they are willing to go to protect their privacy. Based on the results gathered from respondents, they identified that a large part of the population, that indicates it cares about data privacy, is willing to act and in fact has already acted. In terms of percentage, 84% of respondents stated that they care about data privacy, referring to both their own data as well as the data of other members of society, and they want more control over how the data are being used. From this group of respondents, 80% stated that they are willing to spend time or money in order to keep their data safe, because they think that data privacy is an important factor, which is directly related with buying decisions. On the other hand, this pushes them to pay more for services with better protection. Another important issue regarding the results obtained from the survey is that nearly half (48%) of respondents stated that they had already switched service providers or companies due to the fact that they were not satisfied with their data policies or data sharing practices [39].

To build a user profile, authors in [40] proposed an independent model for both long-term and short-term user preferences. The proposed framework includes a recent page history buffer for the short-term model and a taxonomic hierarchy for the long-term model. Long-term model was created from the Google Directory using topics of clicked streaming results, while short-term models by using a cache of recently clicked results. While authors in [41], proposed a model that gathers demographic parameters, such as name, age, gender, education level, occupation, etc., about the users as well as his/her actions within online communities.

The report presented in [42] presents a survey of online privacy concerns in the United Kingdom (UK), as well as how users feel about privacy and their data. The number of people in the UK who claim to be concerned about the issue of online privacy has fallen from 84% in 2012 to 75% in 2017. Among 18–24-year-old users, concern with online privacy falls to just 58%, down from 75% in 2012. At the same time, almost two-thirds of users are now happy with the amount of personal data they share with companies or organizations. It is well known that the GDPR seeks to balance the user's right to privacy with the legitimate interests of service providers wanting to serve them better [43]. Therefore, according to this report and conducted surveys, it shows that overall privacy concerns decreased since the last survey conducted in 2012. In general, the findings presented in this report provide an optimistic overview for the future data economy in the UK. According to the conducted survey, a significant proportion of users indicate that they would prefer to pay for online services rather than sharing any personal data. The highest preference for paid services is expressed in TV and in multimedia content platforms, such as Netflix, Amazon Video, etc. This may be related to the fact that these categories already provide optional paid subscription models to their users. On the other hand, online services that are normally free, such as email, messaging apps, and social networks, receive relatively lower preferences for paying over sharing personal information. However, a considerable number of respondents' state that they would be willing to pay also for these services rather than providing personal information. In order to come up with these results and statistics, a question was provided to a specific number of respondents, as follows: "For each of the following services, please state whether you would prefer to pay for the service or share personal information in exchange for free access to the service?" Figure 1 presents the results from the conducted survey based on responses from users for some specific Internet services:

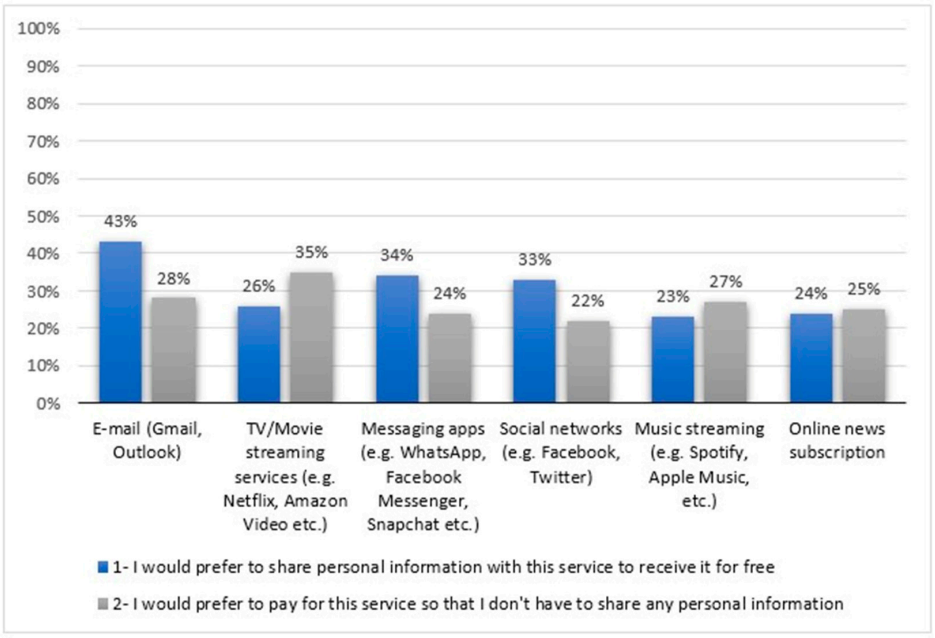


Figure 1. Respondents' responses regarding the sharing of their personal data with various Internet services [42].

4. Discussion and Proposed Privacy Model Based on Service Category

As users of today's technologies, we are amazed with many features that these technologies can provide. However, users often do not consider the implications with respect to privacy issues and sharing vast amounts of personal data, leading to several security threats and data breaches. These problems are more complex given the fact that users sometimes may compromise privacy and security in order to obtain access to the latest technologies and services [44]. With respect to this issue, numerous architectures are designed with the aim to overcome technical challenges related to delivering different kinds of services in the future Internet [45]. Generally speaking, security and privacy issues arise especially for clients who use intelligent end devices. Table 1 indicates sensitive data that can be stored on both the client side and in the operator network, such as history records, application usage, user mobility, contacts, location, etc. Therefore, this collected data has the potential to facilitate different services to personalize those kinds of data. Due to this fact, an appropriate level of security must be guaranteed, and a proper balance should be made between the level of security and performance as well.

Delivering "personalized" services to end users has a few challenges, whereby users may not be so keen on sharing certain personal data [46]. On the other hand, it should be taken into account that, nowadays, users generally tend to "accept all terms and conditions", for example, when using Google services and, thus, allowing Google access to a large number of user-related data [47].

Our proposal is to define a privacy framework as a central node that would be responsible for user data aggregation, handling, and sharing across services. When a user accesses a service that requires access to his/her data, the service would request certain data category or multiple categories, as defined in Table 3. In services nowadays, this is being performed directly through forms or implicit data acquisition, but in our proposal the idea is that the service would ask these data from the proposed privacy framework, in a similar manner to how it is being carried out in Single Sign On (SSO) [48] using the OAuth protocol. The privacy framework would then control users' dynamic consent and real-time privacy adjustments based on the current user input, where users can approve or reject one-time data sharing. The framework also provides information about the types of data that the service gathers from users and informs users about potential risks if the required data is leaked. An important aspect of our proposal is that users only give their data when using the service, which means that their data is shared on a per-session basis, rather than in its entirety and, thus, it is not being stored long-term in the services domain and, thus, minimizes the risks of data leaks from services handling the data.

Our approach focuses on taking the data collected by services, estimating the associated risks, and then categorizing them for category-based sharing and anonymization purposes.

Anonymization can be performed by generalization of the acquired data. For example, some services might provide well-personalized content even with generalized data (e.g., weather forecast), while some cannot (e.g., online deliveries from web shops). In our proposal, it is up to the service to define the minimum amount of necessary data and the granularity of the data, and it is up to a user to decide how much is he/she willing to disclose in order to get a more personalized service.

One example of how anonymization can be performed is by generalizing location data, as it is presented by authors in [49]. They proposed a location proxy to preserve user privacy by limiting the use of precise location data. The idea of the proxy is to somehow "blur" more precise user locations where such precision is unnecessary. Examples of generalization of user location data, such as GPS data, IP address, home address, country, etc., are based on previous research findings and proposed generalization methods for this data category. Another example could involve the anonymization of an accessibility dataset. In this regard, the anonymization method can be applied to all parameters within this dataset, but it will be more efficient by treating more sensitive data, for privacy preservation of users. An anonymization process of data, such as font size, font dyslexia settings, font color, screen reader usage, text spacing, magnification level, etc., typically involves masking or removing

any data that can identify users, while retaining the usefulness of accessibility data. The trade-off between personalization and privacy is evident and quite clear, because if users want to receive more personalized services, they must share more personal data, whereas maintaining a higher privacy level, requires "sacrificing" some degree of personalization.

Table 3. A survey of some common user's data collected from different Internet services.

Collected User Data	Internet Services											
	Google Services								Other Internet Services			
	YouTube	Gmail	Google Search	Google Assistant	Google Play Store	Google Maps	Google Calendar	Outlook Web Mail	Netflix	Amazone Prime	Zoom	eBay
Search history	√		√	√	√	√			√	√		√
Location information	√	√	√	√	√	√	√	√	√	√	√	√
Language	√	√	√	√	√	√	√	√	√	√	√	√
Device type	√	√	√	√	√	√	√	√	√	√	√	√
Time of the day	√	√	√	√	√	√	√	√	√	√	√	√
Voice recognition	√		√	√		√						
App usage history	√	√	√	√	√	√	√	√	√	√	√	√
Bookmarks	√					√			√			
Name, Surname, Age, etc.	√	√	√	√	√	√		√	√	√		√
User's patterns		√					√	√			√	
Calendar data							√				√	
Email content		√						√				
Sender/recipient address		√						√			√	
Contact details		√						√			√	
Watch history	√		√						√	√		√
Ratings	√				√				√			
Purchase history												√
Subscription information									√	√		
Meeting data							√				√	
Call logs											√	
Chat data		√									√	
Recording data											√	
File sharing information		√						√			√	
Transaction data												√

The scalability of the proposed privacy-preserving scheme will be addressed by using proved existing state-of-the-art methods to guarantee interoperability between services and scalability. In our next publication, we plan to describe one data service provider rather than having diverse service providers. In this sense, we aim to have a centralized data point like Google for example, which then gives data to other services. The data is managed by this centralized system like Google, Facebook, or other SSO services, which permits complete or partial sharing of the data with other services.

This section gives a more detailed overview of the data collection from users as well as their classification, including different kind of services. The table below lists all the parameters that services can require from the users, acquired from analysis of widely used services and related literature. Through this representation, we can see the types of data that are gathered by specific services as well as which services collect common user data. This can provide guidelines for categorizing various parameters according to different services and facilitating the generalization process of user profiles. On the other hand, grouping of common parameters will help us in developing general models for user profiling and anonymization.

The classification of parameters will be performed according to different data sets, such as location data (e.g., GPS and IP addresses), site accessibility data (e.g., font size, font color, and voice recognition), user preferences, language, device type, application usage history, demographic data, meeting data, search history, and so on.

Furthermore, bearing in mind that all Internet services have their own data about the users, this does not exclude the fact that most of the services also have common data between them. In this relation, according to specific parameters grouped and proposed categories, we can see that there exists some common data across all services, including name, surname, contact details, browsing history, country of origin, city, device type(s), actions taken, search queries, and so on. When it comes to data collection about users from specific services, it makes sense to gather data that are common across all services,

but it does not make sense to collect something specific for a single service. Regarding these issues, we will propose a framework for data exchange and data protection that will give to users most of the data that are common across all or most of the services. In our future work, this privacy framework will enable centralized and fine-tuned control of data privacy of users and require user consent for sharing their data with Internet services. The purpose of grouping personal parameters is connected to identifying data for the process of generalization. In this sense, some of the most popular services along with Google services, are indicated in the Table 3. In this table, we presented some personal data about the users, which they share with Internet platforms, in exchange for the services they want. Also, we highlighted some common user data between various Internet services, as will be shown below.

Table 3 shows the data collected from users when interacting with Internet services. Our proposal is related to grouping parameters or data collected from users, in order to make those data more general. For example, having exact user geocoordinates might be necessary for some services, but this kind of information can be generalized to city, country or even region, if such a location granularity is suitable for other services. In this way, services are still able to personalize their content but will be unable to personally identify each user and, thus, increasing the privacy of users.

From the table we can see also a list of parameters that specific services use for some kind of personalization. This representation enables us to see which data are gathered by specific services and which are mutual data collected by them. In this relation, we have identified some of the most popular services available on the Internet today.

Based on the aforementioned table and the classification of the parameters that were gathered from users, we can draw a conclusion that services used for similar purposes gather similar data from users. Therefore, different web mail services, video streaming services, online storage services, etc., collect similar data, such as other respective services, by providing enhanced services to the users.

In this regard, for Internet services, such as web mail, the necessary parameters include the user's communication patterns, email content, mail contacts, chat data, location parameters, etc. On the other hand, video streaming services typically require parameters, such as search history, location data, device type(s), time of the day, viewing history, ratings, purchase history, and so on. Online storage services collect various types of data from users to provide their services and improve their content and delivery. The specific data collected can vary from one service to another, but some common types of data that those services may collect include user account information, file metadata, file content, usage data, location data, payment information, device information, log data, communication data, user preferences, etc.

Considering the above summarization, user privacy aspects and their concerns, the following questions always arise: Does these services really need those data about the users? How do specific services use these kinds of data to provide better services to them? According to these conclusions, our future research will answer those questions by providing a privacy framework, where each service will only attain the data it really needs, with the granularity of the data necessary for each service provision and personalization. On the other hand, the users should provide consent related to the risk to their privacy when sharing their personal data to the services. In this sense, the goal of our proposed privacy framework is to estimate the privacy risks for each parameter, groups of parameters, and respective generalization of parameters. The generalization method will be applied to all data sets and their corresponding parameters and, therefore, the privacy concerns of users will be decreased.

5. Conclusions and Future Work

The paper presents a survey of widely used Internet services with a focus on the data that services gather from users and process in order to personalize the content and service delivery. Furthermore, we highlighted some current research activities with a focus on

user profiling and modelling in different types of services. In this relation, this paper also presents specific types of data that services collect from users using examples of commonly used Internet services.

The conclusion of this survey paper is three-fold. Firstly, we have identified the types of data that Internet services gather from users in the user profiling process. Second, the literature analysis and comparison has been conducted. Thirdly, we have identified the gaps and issues that need to be addressed in our future work by proposing generalization methods for tackling privacy issues and estimating privacy risks in user profiling and data collection.

In our future research work we will propose a privacy framework where each Internet service would attain only the data it really needs, but in a way that the user gives consent to some level of privacy disclosure based on service it is using. In this relation, we will offer additional materials and tools to make sure that users are fully aware of the implications of each "level of access", similar to the one presented in [50]. These methods will help in educating users about the possible consequences of revealing their personal information from third parties. One potential resource could be any privacy calculator as online web service that helps users in understanding the risk level associated with sharing different types of personal data. We will also present the working principle of such a framework, based on the data that different users possess. Currently, each service requires, takes, and processes (or sells) additional data it does not really need for service personalization. Therefore, for example, detailed location data, detailed browsing history, specific device data, unrelated behavioral data, personal preferences, and so on, may be collected by many Internet applications, even though only general activity tracking is required. In this relation, we will develop a framework where each user will give a "level access" to each data. The categorization will be performed depending on how much data the user is able to provide, in exchange for the required services. If the user gives just some basic data, this user will be classified in the lowest level in our model. Then, for a better service he/she can choose to "trade" his privacy for a better, more personalized, and probably free service in some cases. In this context, risk estimation will define a level of access. This estimation will be conducted using various existing models and analyzing aspects of the data, such as their type, size, duration, reputation of the service, and the sensitivity of the data [51].

With the existing approach, services can collect all types of data, whether necessary or not. With our proposed method, users will have the ability to choose and be informed about what data are shared, for how long, for what purpose, and the potential risks if the data are compromised (e.g., fraud scenarios).

Also, in the future we will address the validation of the proposed model, which will be performed by using similar methods to k-anonymization techniques. On the other hand, a de-anonymization method will be used to attempt to reverse the anonymized parameters with the aim of re-identifying individuals from anonymized datasets [52].

Author Contributions: D.M. and M.V. suggested the design of the study and wrote the methodology, supervised whole research; D.M. and M.V. searched the databases, prepared the tables, interpreted the results, visualized and wrote the original draft of the manuscript; D.M., M.V. and P.H. reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data supporting the conclusions of this article will be made available by the corresponding author on request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Eke, C.I.; Norman, A.A.; Henry, L.; Nweke, F. A Survey of User Profiling: State-of-the-Art, Challenges and Solutions. *IEEE Access* **2019**, *7*, 144907–144924. [CrossRef]
2. Cufoglu, A. User Profiling—A Short Review. *Int. J. Comput. Appl.* **2014**, *108*, 1–9. [CrossRef]

3. Trusov, M.; Ma, L.; Jamal, Z. Crumbs of the Cookie: User Profiling in Customer-Based Analysis and Behavioral Targeting. *Mark. Sci.* **2016**, *35*, 405–426. [CrossRef]
4. Tatar, A.; de Amorim, M.D.; Fdida, S. A survey on predicting the popularity of web content. *J. Internet Serv. Appl.* **2014**, *5*, 8. [CrossRef]
5. Schmidt, DataCollection in the Age of Surveillance Capitalism, Google Data Collection. August 2018. Available online: https://www.dre.vanderbilt.edu/~schmidt/PDF/Schmidt-Survelliance-Capitalism-v2.pdf (accessed on 1 August 2024).
6. Farid, M.; Elgohary, R.; Moawad, I.; Roushdy, M. User Profiling Approaches, Modeling, and Personalization. In Proceedings of the 11th International Conference on Informatics & Systems (INFOS 2018), Doha, Qatar, 18–21 November 2019.
7. Atote, B.S.; Saini, T.S.; Bedekar, M.; Zahoor, S. Inferring Emotional State of a User by User Profiling. In Proceedings of the 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I), Greater Noida, India, 14–17 December 2016.
8. Vuković, M.; Jevtic, D. Agent-based Movement Analysis and Location Prediction in Cellular Networks. *Procedia Comput. Sci.* **2015**, *60*, 517–526. [CrossRef]
9. Zhao, S.; Li, S.; Ramos, J.; Luo, Z.; Jiang, Z.; Dey, A.K.; Pan, G. User profiling from their use of smartphone applications: A survey. *Pervasive Mob. Comput.* **2019**, *59*, 101052. [CrossRef]
10. Chen, J.; Liu, Y.; Zou, M. Home location profiling for users in social media. *Inf. Manag.* **2016**, *53*, 135–143. [CrossRef]
11. Li, D.; Li, Y.; Ji, W. Gender Identification via Reposting Behaviors in Social Media. *IEEE Access* **2017**, *6*, 2879–2888. [CrossRef]
12. Dougnon, R.Y.; Viger, P.F.; Nkambou, R. Inferring User Profiles in Online Social Networks Using a Partial Social Graph. *J. Intell. Inf. Syst.* **2015**, *28*, 84–99.
13. Setthawong, R.J. User Preferences Profiling Based on User Behaviors on Facebook Page Categories. In Proceedings of the Conference: 2017 9th International Conference on Knowledge and Smart Technology (KST), Chonburi, Thailand, 1–4 February 2017.
14. 5 Things You Need to Know about Data Privacy [Definition & Comparison]. 2023. Available online: https://dataprivacymanager.net/5-things-you-need-to-know-about-data-privacy/ (accessed on 10 January 2023).
15. Ellingwood, J. 2017. Available online: https://www.digitalocean.com/community/tutorials/user-data-collection-balancing-business-needs-and-user-privacy (accessed on 26 September 2017).
16. Zoom Privacy Policy. Available online: https://zoom.us/privacy (accessed on 29 March 2020).
17. Archibald, M.M.; Ambagtsheer, R.C.; Casey, M.G.; Lawless, M. Using Zoom Videoconferencing for Qualitative Data Collection:Perceptions and Experiences of Researches and Participants. *Int. J. Qual. Methods* **2019**, *18*, 1609406919874596. [CrossRef]
18. Ascaso, A.R.; Boticario, J.G.; Finat, C. Setting accessibility preferences about learning objects within adaptive elearning systems: User experience and organizational aspects. *Expert Syst.* **2017**, *34*, e12187. [CrossRef]
19. Maddodi, S. Netflix Bigdata Analytics—The Emergence of Data Driven Recommendation. *Int. J. Case Stud. Bus. IT Educ. (IJCSBE)* **2019**, *3*, 41–51. [CrossRef]
20. Jeidari, M.; Jones, J.H., Jr.; Uzuner, O. Online User Profilling to Detect Social Bots on Twitter. *arXiv* **2022**, arXiv:2203.05966.
21. Baik, J.; Lee, K.; Lee, S.; Kim, Y. Predicting personality traits related to consumer behavior using SNS analysis. *New Rev. Hypermedia Multimed.* **2016**, *22*, 189–206. [CrossRef]
22. Shitole, P.; Potey, M. Focusing User Modeling For Age Specific Differences. *Int. J. Emerg. Trends Technol. Comput. Sci. (IJETTCS)* **2015**, *4*, 238–244.
23. Thorson, K.; Cotter, K.; Medeiros, M.; Pak, C. Algorithmic inference, political interest, and exposure to news and politics on Facebook. *Inf. Commun. Soc.* **2019**, *24*, 183–200. [CrossRef]
24. Kumbhar, M.F.; Rajput, H. An Efficient Approach for User Profiling Through Social Media Analytics. 2021. Available online: https://www.researchgate.net/publication/355167541 (accessed on 1 August 2024).
25. Data Privacy Laws: What You Need to Know in 2020. Available online: https://www.osano.com/articles/data-privacy-laws (accessed on 8 November 2020).
26. Senapati, K.K.; Kumar, A.; Sinha, K. Impact of Information Leakage and Conserving Digital Privacy. In *Malware Analysis and Intrusion Detection in Cyber-Physical Systems*; IGI GLOBAL: Hershey, PA, USA, 2023; pp. 1–23. [CrossRef]
27. Daswani, N.; Elbayadi, M. The Yahoo Breaches of 2013 and 2014. In *Big Breaches*; Apress: Berkeley, CA, USA, 2021; pp. 155–169. [CrossRef]
28. The Equifax Data Breach. 2018. Available online: https://www.ftc.gov/enforcement/refunds/equifax-data-breach-settlement (accessed on 5 June 2024).
29. Brusk, C.D.; Mee, P.; Brandenburg, R. The Marriott Data Breach. 2018. Available online: https://www.marshmclennan.com/content/dam/oliver-wyman/v2/publications/2018/december/Oliver_Wyman_Lessons_Learned_For_Boards_The_Marriott_Data_Breach.pdf (accessed on 20 June 2024).
30. Zinolabedini, D.; Arora, N. *The Ethical Implications of the 2018 Facebook-Cambridge Analytica Data Scandal*; The University of Texas: Austin, TX, USA, 2019.
31. Neto, N.N.; Madnick, S.; de Paula, A.M.G.; Borges, N.M. A Case Study of the Capital One Data Breach (Revised). 2020. Available online: https://ssrn.com/abstract=3542567 (accessed on 15 May 2024).
32. SolarWinds Data Breach Action Plan. 2020. Available online: https://hbr.org/podcast/2024/01/how-solarwinds-responded-to-the-2020-sunburst-cyberattack (accessed on 15 May 2024).

33. EU Data Protection Rules and U.S. Implications. Available online: https://fas.org/sgp/crs/row/IF10896.pdf (accessed on 17 July 2020).
34. Ducato, R. Data protection, scientific research, and the role of information. *Comput. Law Secur. Rev.* **2020**, *37*, 105412. [CrossRef]
35. Bakare, S.S.; Adeniyi, A.O.; Akpuokwe, C.U.; Eneh, N.E. Data privacy laws and compliance: A comparative review of the EU GDPR and USA regulations. *Comput. Sci. IT Res. J.* **2024**, *5*, 528–543. [CrossRef]
36. Cesconetto, J.; Silva, L.A.; Bortoluzzi, F.; Cáceres, M.N.; Zeferino, C.A.; Leithardt, V.R.Q. PRIPRO—Privacy Profiles: User Profiling Management for Smart Environoments. *Electronics* **2020**, *9*, 1519. [CrossRef]
37. Leithardt, V.R.Q.; Correia, L.H.A.; Borges, G.A.; Rossetto, A.G.M.; Rolim, C.O.; Geyer, C.F.R.; Silva, J.M.S. Mechanism for Privacy Management Based on Data History (UbiPri-His). *J. Ubiquitous Syst. Pervasive Netw.* **2018**, *10*, 11–19. [CrossRef]
38. Fallatah, K.U.; Barhamgi, M.; Perera, C. Personal Data Stores (PDS): A Review. *Sensors* **2023**, *23*, 1477. [CrossRef]
39. 100 Data Privacy and Data Security Statistics. Available online: https://dataprivacymanager.net/100-data-privacy-and-data-security-statistics-for-2020/ (accessed on 20 August 2020).
40. Li, L.; Yang, Z.; Wang, B.; Kitsuregawa, M. Dynamic adaptation strategies for long-term and short-term user profile to personalize search. In *Advances in Data and Web Management*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 228–240.
41. Fernandez, M.; Scharl, A.; Bontcheva, K.; Alani, H. User profile modelling in online communities. In Proceedings of the 3rd International Workshop on Semantic Web Collaborative Spaces, 13th International Semantic Web Conference (ISWC-2014), Riva del Garda, Italy, 8 November 2014.
42. Combemale, C. What the Consumer Really Thinks. Available online: https://dma.org.uk/uploads/misc/5a857c4fdf846-data-privacy---what-the-consumer-really-thinksfinal_5a857c4fdf799.pdf (accessed on 12 February 2018).
43. Taylor, I.B. White Paper on the General Data Protection Regulation (GDPR) and archives. *Archivar* **2022**, *70*, 184–193.
44. Iacob, B.; Marton, K. Streaming Video Detection and QoE Estimation in Encrypted Traffic. Seminar Future Internet WS2017/2018. 2018, pp. 1–6. Available online: https://www.net.in.tum.de/fileadmin/TUM/NET/NET-2018-03-1/NET-2018-03-1_01.pdf (accessed on 17 November 2023).
45. Meng, X.; Wang, S.; Shu, K.; Li, J.; Chen, B.; Liu, H.; Zhang, Y. Personalized Privacy-Preserving Social Recommendation. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), New Orleans, LA, USA, 2–7 February 2018.
46. Wang, Y.; Li, P.; Jiao, L.; Su, Z.; Cheng, N.; Shen, X.; Zhang, P. A Data-Driven Architecture for Personalized QoE Management in 5G Wireless Networks. *IEEE Wirel. Commun.* **2016**, *24*, 102–110. [CrossRef]
47. Peslak, A.; Kovalchick, L.; Conforti, M. A Longitudinal Study of Google Privacy Policies. *JISAR* **2020**, *13*, 54.
48. Rastogi, V.; Agrawal, A. All your Google and Facebook logins are belong to us: A case for single sign-off. In Proceedings of the Eighth International Conference on Contemporary Computing (IC3), Noida, India, 20–22 August 2015.
49. Vukovic, M.; Kordic, M.; Jevtic, D. Clustering Approach for User Location Data Privacy in Telecommunication Services. In Proceedings of the 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 30 May–3 June 2016.
50. Vukovic, M.; Skocir, P.; Katusic, D.; Jevtic, D.; Trutin, D. Estimating Real World Privacy Risk Scenarios. In Proceedings of the 13th International Conference on Telecommunications (ConTEL), Graz, Austria, 13–15 July 2015.
51. Silva, P.; Gonçalves, C.; Antunes, N.; Curado, M. *Privacy Risk Assessment and Privacy-Preserving Data Monitoring*; Elsevier: Amsterdam, The Netherlands, 2022; Volume 200, pp. 1–13.
52. Majeed, A.; Lee, S. Anonymization Techniques for Privacy Preserving Data Publishing: A Comprehensive Survey. *IEEE Access* **2020**, *9*, 8512–8545. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.