ALIGNING LEGAL DEFINITIONS OF PERSONAL INFORMATION WITH THE COMPUTER SCIENCE OF IDENTIFIABILITY

SCOTT JORDAN*

Abstract

The computer science literature on identification of people using personal information paints a wide spectrum, from aggregate information that doesn't contain information about individual people, to information that itself identifies a person. However, privacy laws and regulations often distinguish between only two types, often called personally identifiable information and de-identified information. We show that the collapse of this technological spectrum of identifiability into only two legal definitions results in the failure to encourage privacy-preserving practices. We propose a set of legal definitions that spans the spectrum.

We start with anonymous information. Computer science has created anonymization algorithms, including differential privacy, that provide mathematical guarantees that a person cannot be identified. Although the California Consumer Privacy Act (CCPA) defines aggregate information, it treats aggregate information the same as de-identified information. We propose a definition of anonymous information based on the technological possibility of logical association of the information with other information. We argue for the exclusion of anonymous information from notice and consent requirements.

We next consider de-identified information. Computer science has created de-identification algorithms, including generalization, that minimize (but not eliminate) the risk of re-identification. GDPR defines anonymous information but not de-identified information, and CCPA defines de-identified information but not anonymous information. The definitions do not align. We propose a definition of de-identified information based on the reasonableness of association with other information. We propose legal controls to protect against re-identification. We argue for the inclusion of de-identified information in notice requirements, but the exclusion of de-identified information from choice requirements.

We next address the distinction between trackable and non-trackable information. Computer science has shown how one-time identifiers can be used to protect reasonably linkable information from being tracked over time. Although both GDPR and CCPA discuss profiling, neither formally defines it as a form of personal information, and thus both fail to adequately protect against it. We propose definitions of trackable information and non-trackable information based on the likelihood of association with information from other contexts. We propose a set of legal controls to protect against tracking. We argue for requiring stronger forms of user choice for trackable information, which will encourage the use of non-trackable information.

Finally, we address the distinction between pseudonymous and reasonably identifiable information. Computer science has shown how pseudonyms can be used to reduce identification. Neither GDPR nor CCPA makes a distinction between pseudonymous and reasonable identifiable information. We propose definitions based on the reasonableness of identifiability of the information, and we propose a set of legal controls to protect against identification. We argue for requiring stronger forms of user choice for reasonably identifiable information, which will encourage the use of pseudonymous information.

* Scott Jordan is a Professor of Computer Science at the University of California, Irvine. He served as the Chief Technologist of the Federal Communications Commission during 2014-2016. Email: sjordan@uci.edu. Mailing address: 3214 Bren Hall, Department of Computer Science, University of California, Irvine, CA 92697-3435. Webpage: www.ics.uci.edu/~sjordan/
This material is based upon work supported by the Herman P. & Sophia Taubman Foundation and by the National Science Foundation under Grant No. 1956393.

Electronic copy available at: https://ssrn.com/abstract=3893833

Our definitions of anonymous information, de-identified information, non-trackable information, trackable information, and reasonably identifiable information can replace the over-simplified distinction between personally identifiable information versus de-identified information. We hope that this full spectrum of definitions can be used in a comprehensive privacy law to tailor notice and consent requirements to the characteristics of each type of information.

1. Introduction

The United States Congress has been devoting substantial attention in the last few years to the crafting of a comprehensive consumer privacy law. The two common starting points for a comprehensive consumer privacy law are the 2016 European General Data Protection Regulation (GDPR)[1] and the 2018 California Consumer Privacy Act (CCPA)[2]. The scope of the information subject to the law or regulation, often called personal information or personally identifiable information, is central to each. So is the scope of anonymous or de-identified information that is exempt from regulation.

The need for standardized definitions is great. Privacy policies often use non-standardized definitions of personal information that do not align with those in the GDPR or the CCPA or even with each other, leaving consumers confused about what constitutes personal information. Privacy policies often include assertions about the anonymity of personal information that exceed both the technical abilities and legal definitions of anonymization and of de-identification.

The academic literature does not provide much advice for statutory definitions of personal information and of de-identified information. The void in the academic literature has been filled by proposals from advocacy groups. The Mozilla Foundation, an advocacy group funded primarily by royalties from Firefox web browser search partnerships, proposes a broad definition of covered data that includes information that can be reasonably connected to either a person or a device.[3] The Information Technology & Innovation Foundation (ITIF), an advocacy group funded in large part by industry[4], argues for (but does not propose) a narrow definition of personally identifiable information that omits some types of linkable personal information, one.[5] It also argues for (but does not propose) a broad definition of de-identified data that includes pseudonymized data.[6]

Two of the most discussed bills in the last session of Congress were the Consumer Online Privacy Rights Act (COPRA)[7] sponsored by Sen. Cantwell, and the Setting an American Framework to Ensure Data

[1] General Data Protection Regulation, Regulation (EU) 2016/679 (as amended), (April 27 2016, as amended on May 5 2016) [hereinafter *GDPR*], *available at* https://eur-lex.europa.eu/eli/reg/2016/679/2016-05-04.

[2] California Consumer Privacy Act of 2018 (as amended by the California Privacy Rights Act of 2020) (November 3, 2020), [hereinafter *CCPA*], *available at* https://www.oag.ca.gov/system/files/initiatives/pdfs/19-0021A1%20%28Consumer%20Privacy%20-%20Version%203%29_1.pdf.

[3] Mozilla, *U.S. Consumer Privacy Bill Blueprint* (April 2019), https://blog.mozilla.org/netpolicy/2019/04/04/a-path-forward-rights-and-rules-to-protect-privacy-in-the-united-states/, at “Enforcement and Scope”, “Covered Data”.

[4] ITIF’s funders include Amazon, Apple, AT&T, Charter Communications, Comcast, CTIA, Facebook, Google, Microsoft, NCTA, T-Mobile, U.S. Telecom, and Verizon, among others; see https://itif.org/our-supporters.

[5] Alan McQuinn and Daniel Castro, *A Grand Bargain on Data Privacy Legislation for America* (January 2019), https://itif.org/publications/2019/01/14/grand-bargain-data-privacy-legislation-america, at 16.

[6] *Id.* at 18.

[7] Consumer Online Privacy Rights Act, S.2968, 116th Cong. (2019) [hereinafter *COPRA*], available *at* https://www.congress.gov/bill/116th-congress/senate-bill/2968.

Electronic copy available at: https://ssrn.com/abstract=3893833

Access, Transparency, and Accountability Act (SAFE DATA)[8] sponsored by Sen. Wicker. The COPRA bill includes a definition of *covered data* which includes information that is reasonably linkable to either an individual or a device.[9] The SAFE DATA bill includes a narrower definition of *covered data* which similarly includes information that is reasonably linkable to an individual, but only includes information that is reasonably linkable to a device if the device is itself reasonably linkable to an individual.[10] Both bills also include a definition of *de-identified data*, which includes information that is not reasonably linkable to either an individual or device.[11]

In Section 2, we discuss the limits imposed by using only two definitions (personal information and de-identified information) in addressing the full spectrum of identifiability of information. We look to the computer science literature to understand the abilities of various types of privacy-preserving algorithms and the spectrum of identifiability that they enable. We find that the GDPR's and the CCPA's definitions of personal information are too broad to differentiate between meaningful differences in identifiability within them, and thereby too broad to effectively encourage privacy-preserving treatment. We thus delineate between different types of personal information on the basis of whether the personal information is trackable and/or identifiable. We propose to divide personal information into three types: reasonably identifiable, pseudonymous, and non-trackable.

In Section 3, we discuss differences in consumer views of reasonably identifiable information, pseudonymous information, and non-trackable information. We give examples of collection, use, and sharing of these three types of personal information. These examples illustrate the need for notices that disclose these differences and the need for choice mechanisms that afford consumers different choices for different types of personal information.

In Section 4-6, we propose statutory definitions of reasonably identifiable information, pseudonymous information, and non-trackable information. In Section 4, we first propose a definition of *personal information*. Although our proposed definition draws from those in the GDPR and the CCPA, we address problems with the GDPR's and the CCPA's approaches to whether publicly available information, aggregate information, anonymous information, and/or de-identified information are included in, or excluded from, the scope of personal information. We then propose a definition of *reasonably linkable information*. The concept is widely used, but a formal definition has been lacking.

In Section 5, we differentiate between three different types of reasonably linkable information. Although both the GDPR and the CCPA are concerned with the issue of tracking, neither defines trackable information as an explicit subset of personal information. We look to the computer science literature for guidance on what makes information trackable, and based on that literature we propose a definition of *non-trackable information*. We then turn of the issue of whether information is pseudonymous or reasonably identifiable. Although both the GDPR and the CCPA attempt to address pseudonymization, neither defines pseudonymous information as an explicit of personal information. We look to the computer science literature for guidance on what makes information pseudonymous, and based on that literature we propose a definition of *pseudonymous information*. Finally, reasonably linkable information that is neither non-trackable nor pseudonymous is defined as *reasonably identifiable information*.

In Section 6, we propose definitions of de-identified information and anonymous information. The GDPR defines *anonymous information*, but not *de-identified information*. The CCPA defines *de-identified*

[8] SAFE DATA Act, S.4626, 116th Cong. (2020) [hereinafter *SAFE DATA*], *available at* https://www.congress.gov/bill/116th-congress/senate-bill/4626.
[9] *COPRA,* § 2(8).
[10] *SAFE DATA,* § 2(10).
[11] *COPRA,* § 2(10); *SAFE DATA,* § 2(10).

Electronic copy available at: https://ssrn.com/abstract=3893833

information, but not *anonymous information*. The two definitions are not the same. We again look to the computer science literature. Based on the differences in the ability of various algorithms to provide guarantees that information cannot be re-identified, we propose definitions of *de-identified information* and *anonymous information*.

Our proposed definitions of different types of information relies on characteristics of that information. However, it has often been recognized that in order for information to maintain those characteristics, legal controls are often required. In Section 7, we propose legal controls on de-identified information, non-trackable information, and pseudonymous information.

Finally, in Section 8, we investigate how advertising can be implemented using these different types of personal information. We first discuss how behavioral ads, the most privacy-invasive type of advertising, is based on reasonably identifiable information. We then show how audience segment ads could use pseudonymous information instead of reasonably identifiable information, or even non-trackable information. Finally, we show how contextual ads could use only de-identified information.

In conclusion, we argue that proper distinctions between different types of personal information could be incorporated into a choice framework in a manner that empowers consumers who desire forms of advertising that are more privacy-preserving.

2. Failures of the GDPR and the CCPA to Address the Spectrum of Identifiability

A. Limited Definitions in the GDPR and in the CCPA

Both the GDPR and the CCPA apply their choice frameworks to information related to an identifiable person, but not to information that is related to an unidentifiable person. The GDPR defines *personal data* (its version of personal information) as

> "any information relating to an identified or identifiable natural person".[12]

Under the GDPR, *personal data* does not include *anonymous information*, which it defines as

> "information which does not relate to an identified or identifiable natural person".[13]

Personal data is subject to the GDPR's choice framework, and *anonymous information* is not.

The CCPA defines *personal information* as

> "information that identifies, relates to, describes, is reasonably capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household".[14]

However, the CCPA also recognizes that there may be information that can be linked to a particular consumer or household, but for which the process of linking may be prohibitive due to the difficulty in finding other information with which it can be linked. In 2012, the Federal Trade Commission issued a report containing recommendations for businesses and policymakers.[15] It proposed that information be

[12] *GDPR,* Article 4(1).

[13] *Id.,* Recital 26.

[14] *CCPA,* Section 1798.140(v)(1).

[15] Federal Trade Commission, *Protecting Consumer Privacy in an Era of Rapid Change* (March 2012) [hereinafter *FTC Report*], *available at* https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf.

Electronic copy available at: https://ssrn.com/abstract=3893833

considered *de-identified information* if it is not reasonably linkable to a particular consumer or device.[16] In a similar vein, the CCPA defines *de-identified information* as

> "information that cannot reasonably be used to infer information about, or otherwise be linked to, a particular consumer …"[17]

Under the CCPA, *personal information* does not include *de-identified information*. *Personal information* is subject to the CCPA's choice framework, and *de-identified information* is not.

Both the GDPR and the CCPA thus classify any information relating to a person into one of two mutually exclusive sets (for the GDPR, *personal data* or *anonymous information*; for the CCPA, *personal information* or *de-identified information*) based on whether the person is identifiable.

Unfortunately, while this partition of information into only two sets is simple, it does not reflect the spectrum of identifiability of personal information. Within the category of information that the GDPR classifies as *personal data* and that the CCPA classifies as *personal information*, research has repeatedly shown that there are substantial differences in the degree of identifiability.

B. Lack of recognition of the benefits of pseudonymous information

Polonetsky (2016) presents a spectrum of identifiability of information. To differentiate degrees of identifiability, the paper uses the concepts of a *direct identifier* and of an *indirect identifier*.[18] Garfinkel (2015), a report by the National Institute of Standards and Technology, defines a *direct identifier* as "data that directly identifies a single individual".[19] Polonetsky (2016) somewhat similarly defines a *direct identifier* as "data that can be used to identify a person without additional information or with cross-linking through other information that is in the public domain".[20] Garfinkel (2015) then defines an *indirect identifier* as "information that can be used to identify an individual through association with other information".[21]

The most identifiable form of information is that relating to an identified person or household. It contains direct identifiers such as a person's name, personal telephone number, personal email address, driver's license number, or social security number. Polonetsky (2016) calls such information *explicitly personal data*, but we will use the term *reasonably identifiable information*. This type of information is classified as personal information under both the GDPR and the CCPA.

The second most identifiable form of information is information relating to a person or household that is identifiable but has not yet been identified, and that is tracked over time. It does not contain direct identifiers, and thus the person or household cannot be identified using a direct identifier. However, this type of information contains indirect identifiers, such as a device identifier or advertising identifier, that can be used to identify the person or household by combining the information with other information containing the same indirect identifiers. The indirect identifiers can also be used to track the person or household over time. Polonetsky (2016) calls such information *potentially identifiable*, but we will use the

[16] *Id.,* at 21.
[17] *CCPA,* Section 1798.140(m).
[18] Jules Polonetsky, Omer Tene, and Kelsey Finch, *Shades of Gray: Seeing the Full Spectrum of Practical Data De-identification*" (Polonetsky), 56 Santa Clara L. Rev. 593 (2016), *available at* http://digitalcommons.law.scu.edu/lawreview/vol56/iss3/3.
[19] Simson L. Garfinkel, *De-Identification of Personal Information*, National Institute of Standards and Technology Report NISTIR 8053 (October 2015) at 40, available at http://dx.doi.org/10.6028/NIST.IR.8053.
[20] Polonetsky, *supra* note 20, at 605.
[21] Garfinkel, *supra* note 21, at 41.

Electronic copy available at: https://ssrn.com/abstract=3893833

more common term *pseudonymous information*. This type of information is classified as personal information under both the GDPR and the CCPA, absent legal controls to prevent reidentification.

Neither the GDPR nor the CCPA differentiates between *reasonably identifiable information* and *pseudonymous information* in their choice frameworks. As a consequence, neither the GDPR nor the CCPA incentivize the use of pseudonyms in their choice frameworks.

C. Lack of recognition of the benefits of non-trackable information

A form of information that is less identifiable than *pseudonymous information* is information relating to a person or household that is identifiable but has not yet been identified, and that is *not* tracked over time. It does not contain direct identifiers. It may contain indirect identifiers, but these indirect identifiers cannot be persistent. An example of a non-persistent identifier is a randomized identifier that is only used in a single interaction with a consumer.[22] Apple is beginning to use such one-time identifiers in some of its applications. Polonetsky (2016) calls such information *pseudonymous*, but we will use the term *non-trackable information*. This type of information is classified as personal information under both the GDPR and the CCPA, absent legal controls to prevent reidentification.

Neither the GDPR nor the CCPA differentiate between *pseudonymous information* and *non-trackable information* in their choice frameworks. As a consequence, neither the GDPR nor the CCPA incentivize the use of one-time identifiers in their choice frameworks. However, the use of such one-time identifiers could eliminate tracking.

3. Differences in Consumer Views of reasonably identifiable information, pseudonymous information, and non-trackable information

The consumer views of *reasonably identifiable information*, *pseudonymous information*, and *non-trackable information* are quite different.

An example of *reasonably identifiable information* is a person's name paired with personal information about the person. The information can be used for behavioral advertising, since the personal information may provide valuable information about the person's interests. An ad broker can collect reasonably identifiable information and create a profile of the person, resulting in tracking. Furthermore, this profile is associated with the person's name.

An example of *pseudonymous information* is a device or advertising identifier paired with personal information about the person using the device. As with reasonably identifiable information, the information can be used for behavioral advertising and tracking. However, the profile is associated with the device or advertising identifier, not with the person's name, providing that device or advertising identifier is not associated with a person or household. As a result, the person seeing the advertisements may properly perceive that they are pseudonymous.

An example of *non-trackable information* is a one-time identifier paired with personal information. As with the other types of information, it can be used for behavioral advertising. However, it cannot be used for tracking. As a result, the person seeing the advertisements may properly perceive that they are pseudonymous and not tracked.

[22] *Polonetsky, supra note 20,* at 608.

Electronic copy available at: https://ssrn.com/abstract=3893833

These three types of personal information are summarized in Table 1. Although neither the GDPR nor the CCPA choice frameworks differentiate between these three types of personal information, consumers are likely to view their use very differently.

	reasonably identifiable information (I)	pseudonymous information (P)	non-trackable information (N)
example of personal information	name + personal information	device or advertising identifier + personal information	one-time identifier + personal information
example of a user's view of a use of personal information	behavioral advertising + tracking + associated with my name	behavioral advertising + tracking + associated with a pseudonym	behavioral advertising + no tracking + associated with a one-time identifier

Table 1. Examples of the three most identifiable types of personal information.

Table 2 presents some examples of the use and sharing of each type of personal information.

	reasonably identifiable information (I)	pseudonymous information (P)	non-trackable information (N)
use of personal information to provide turn-by-turn directions	Map app provides turn-by-turn directions based on name + precise geo-location	Map app provides turn-by-turn directions based on pseudonym + precise geo-location	Map app provides turn-by-turn directions based on rapidly resetting identifier + current precise geo-location
use of personal information to provide ads	Search provider displays ads based on name + audience segment	Search provider displays ads based on device identifier + audience segment	Search provider displays ads based on random rapidly resetting identifier + audience segment
sharing of personal information to provide ads	Website shares advertising identifier + audience segments with an ad broker	Website shares pseudonym + audience segments with an ad broker	Website shares one-time identifier + audience segments with an ad broker

Table 2. Examples of uses and sharing of various types of personal information.

Consider a map app that provides turn-by-turn directions. In order to determine directions, suppose the app collects the precise geo-location of the user. If the app pairs the precise geo-location with the user's name, then the combination constitutes *reasonably identifiable information*. Alternatively, if the app assigns the user a pseudonym, then the combination of the pseudonym and precise geo-location constitutes *pseudonymous information*. Finally, if the app assigns the user a random rapidly resetting identifier and collects only the current geo-location of the user (but not the location history), then the combination of the random rapidly resetting identifier and current precise geo-location constitutes *non-trackable information*.

Next consider a search provider that displays personalized ads aside search results. In order to determine which ads to display, suppose the search provider uses the search terms to place the user into audience segments (e.g., interested in tennis), and then immediately discards the search terms. If the search provider pairs the audience segments with the user's name, then the combination of the user's name and audience segments constitutes *reasonably identifiable information*. Alternatively, if the search provider pairs the audience segments with a device identifier, then the combination of the device identifier and audience segments constitutes *pseudonymous information*. Finally, if the search provider assigns the user a random

Electronic copy available at: https://ssrn.com/abstract=3893833

rapidly identifier, then the combination of the random rapidly resetting identifier and audience segment constitutes *non-trackable information*.

In addition to using personal information, service or app providers may also share personal information. Consider a website that wishes to display ads on one of its webpages. In order to determine which ads to display, suppose the website collects information about user interests, and places the user into audience segments. If the search provider discloses to an ad broker the audience segments paired with a user's advertising identifier, and does not limit how the ad broker uses this information, then the combination of the advertising identifier and audience segments constitutes *reasonably identifiable information*. The information is reasonably identifiable because the user corresponding to the advertising identifier is reasonably identifiable due to the lack of limitations on the ad broker's use of the information.

However, if the website discloses to an ad broker the same information pursuant to a written contract that prohibits the ad broker from identifying the person to whom the information relates, then the information constitutes *pseudonymous information*.

Finally, consider the case in which the website discloses to an ad broker the audience segments paired with a one-time identifier, pursuant to a contract that ensures that the information remains in a form in which there is no reasonable possibility of logical association with other information relating to the person or household to whom the one-time identifier relates. Then the combination of the one-time identifier and audience segment constitutes *non-trackable information*.

In each example in Table 2, some consumers may be comfortable if the app provider associates their personal information (e.g., their geo-location or their interests as described by an audience segment) with their name. However, others may wish to be pseudonymous and prefer that their personal information only be associated with a pseudonym. Yet others may wish not to be tracked and prefer that their personal information only be associated with a one-time identifier. Notices that disclose the differences between *reasonably identifiable information*, *pseudonymous information*, and *non-trackable information* would inform consumers of the degree of identifiability of their personal information. Choice frameworks that afford consumers choices over the form in which their personal information is used and shared would empower consumers to exercise their wishes.

4. Defining Personal Information and Reasonably Linkable Information

Notice and choice requirements typically apply only to information that is both personal and private. Privacy laws often call this type of information *personally identifiable information*, *personal information*, or *personal data*.

Many privacy policies lack any definition whatsoever of personally identifiable information. For example, Microsoft uses the term personal data, but does not define it.[23] Pinterest uses the term personal information, but does not define it.[24] Twitter interchangeably uses the terms personal information and personal data, but does define either of them.[25] By omitting a definition of personally identifiable information, the scope of such privacy policies is unknown, and we are left wondering what personally identifiable information is collected that the privacy policy fails to disclose.

[23] Microsoft, *Microsoft Privacy Statement* (last modified November, 2020), https://privacy.microsoft.com/en-us/privacystatement.

[24] Pinterest, *Privacy Policy* (last modified September 2, 2020), https://policy.pinterest.com/en/privacy-policy.

[25] Twitter, *Twitter Privacy Policy* (last modified June 18, 2020), https://www.twitter.com/en/privacy.

Electronic copy available at: https://ssrn.com/abstract=3893833

The GDPR defines *personal data* as

> "any information relating to an identified or identifiable natural person".[26]

The CCPA defines *personal information* as

> "information that identifies, relates to, describes, is reasonably capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household".[27]

In these definitions, both the GDPR and the CCPA combine the concept of personal information (e.g., information relating to a person) with the concept of identifiability (e.g., an identified or identifiable person). However, by combining these two concepts into a single definition, both the GDPR and the CCPA fail to address information that is personal but whose degree of identifiability falls short of relating to an "identifiable natural person".

Because of this conflation of personal and identifiable, the CCPA then goes back and separately defines other types of information -- including *publicly available* information, *aggregate consumer information*, and *de-identified information* – and proceeds to exclude each of these from personal information. In addition, the CCPA defines *pseudonymization*, but fails to address the relationship of pseudonymized information to personal information or to de-identified information.

The GDPR exhibits similar problems, but to a worse degree. The GDPR uses the terms *aggregate data* and *anonymous information*, both of which it excludes from personal data. In contrast to the CCPA, which excludes publicly available information from personal information, the GDPR uses (but not define) the term *public sector information*, which it appears to include in personal data. Finally, the GDPR defines the term *pseudonymisation,* and treats pseudonymized data as a subset of personal data, but it fails to apply any different notice and choice requirements to pseudonymized data than to other personal data.

Because of these problems, the next three subsections separately address personal information (i.e., information relating to a person), private information (i.e., information that is not public), and identifiable information (i.e., information relating to an identifiable person).

A. Is the information personal?

A consumer privacy bill is concerned with the privacy of people, not the privacy of organizations or businesses.

The GDPR limits *personal data* to "information relating to … [a] natural person".[28] The EU clarifies that a "natural person" means an individual, not a business, institution, or other entity.[29] The EU further clarifies that "relating to" means "information about a person" and that it includes not only "information pertaining to the private life of a person" but also "professional activities, as well as information about his or her public life".[30] As examples, the GDPR lists a "natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements".[31]

[26] *GDPR,* Article 4(1).

[27] *CCPA,* Section 1798.140(v)(1).

[28] *GDPR,* Article 4(1).

[29] European Union Agency for Fundamental Rights and Council of Europe, *Handbook on European data protection law* (2018), at 85 [hereinafter *EU Handbook*], *available at* https://www.echr.coe.int/Documents/Handbook_data_protection_ENG.pdf.

[30] *Id.,* at 83, 86.

[31] *GDPR,* Article 4(4).

Electronic copy available at: https://ssrn.com/abstract=3893833

The CCPA's list of terms used in its definition of personal information similarly includes "information that … relates to [or] describes … a particular consumer".[32] It is unclear whether the CCPA's addition of the word "describes" broadens its definition, since it is unclear whether there is any information that "describes", but does not "relate to", a particular consumer.

A consumer privacy law should define personal information and should require that privacy policies adhere to this definition. Today, privacy policies often deny that much information relating to a person is actually personal. Apple uses the term non-personal information to refer to "data in a form that does not, on its own, permit direct association with any specific individual".[33] Examples of non-personal information Apple collects and uses include occupation, location, and search queries.[34] However, the information is certainly personal, given that occupation, location, and search queries relate to a person.

Personal information should include, at a minimum, information which relates to an individual. However, there remains an important policy decision: should personal information also include information which relates to a household? Some identifiers used by services and apps to associate information identify a group of persons rather than a single person. Often, the group of persons constitutes a household. For example, a home postal address or home telephone number may be associated with a household rather than with a single person.

However, privacy policies are often unclear about whether they consider information relating to a household to be included in the scope of personal information. Indeed, providers of services and apps often argue that they are not. For example, the California Chamber of Commerce, representing a wide variety of businesses, argued that information associated with households should be excluded from the CCPA's scope of *personal information*.[35]

The ambiguity of whether information relating to household is included in privacy policy disclosures demonstrates the importance of clearly spelling out in a privacy law the role of information associated with a group of people such as a household, and the rights of individuals within such a group.

The GDPR seems to include information relating to households in its scope of personal data, since it states that the regulation "applies to controllers or processors which provide the means for processing personal data for such personal or household activities."[36] However, this should have been made clear.

The CCPA is more explicit. In its definition of personal information, it includes information that relates to either a consumer or a household.[37] A household is defined as a group of consumers who reside at the same address and share a common device or service. The CCPA exempts businesses from certain specified obligations insofar as they concern household data, but it is unclear whether these exemptions include notice and choice obligations.[38]

[32] *CCPA,* Section 1798.140(v)(1). In the following sections of the paper, we will consider the other phrases which we omitted from this quote.

[33] Apple, *Privacy Policy* (last modified December 31, 2019), https://www.apple.com/legal/privacy/en-ww/, at "Collection and Use of Non-Personal Information".

[34] *Id.* at "Collection and Use of Non-Personal Information".

[35] U.S. Chamber of Commerce, "*US Chamber of Commerce Privacy Comments on California Consumer Privacy Act Rulemaking*" (March 8, 2019), at 1-3, *available at* https://www.uschamber.com/comment/us-chamber-privacy-comments-california-consumer-privacy-act-rulemaking.

[36] *GDPR,* Recital 18.

[37] *CCPA,* Section 1798.140(v)(1).

[38] *Id.,* Section 1798.145(p).

Electronic copy available at: https://ssrn.com/abstract=3893833

A consumer privacy law should be explicit that information relating to a household qualifies as personal information. First, information relating to a household is clearly information relating to one or more natural persons in the household. Second, a household identifier has traditionally been treated as identification of a natural person, even if it is not sufficient to pin down which person within the household. For example, a home postal address and a home phone number are both always considered to be personal identifiers. For this reason, personal information should include information which relates to either an individual or a household.[39]

B. Is the information private?

A consumer privacy bill should be concerned with the use of private information, not with the use of publicly available information.

The CCPA excludes from the scope of personal information any information that is publicly available. It defines *publicly available* information to include information in government records, information about a consumer that a consumer him or herself made publicly available, information about a consumer that the consumer disclosed to a third party "if the consumer has not restricted the information to a specific audience", and information about a consumer that was made publicly available by "widely distributed media".[40]

The GDPR does not provide any similar exclusion from personal data for any type of publicly available information. It recognizes the existence of *public sector information*, which it does not define, but which appears by reference to consist of personal data that is held by a State, regional or local authority, by a body governed by public law, or by associations of such bodies.[41] Thus, unlike the CCPA, such public sector information remains a subset of personal data. The GDPR places the same notice requirements on public sector information as on other personal data, but it exempts public sector information from GDPR's choice requirements if public access to this information is provided for by EU or State law.[42]

The GDPR and the CCPA thus disagree on their approach to publicly available information. An intermediate approach would be in the public interest. As provided in the CCPA, information that a consumer has made publicly available should not be subject to notice and choice requirements, since the consumer has already decided to waive control over this information. However, CCPA's exemption goes beyond this. It also classifies information that a consumer has disclosed to a third party as *publicly available* if the consumer failed to restrict the third party's sharing of that information to a specific audience. This creates a chicken-and-egg situation. A consumer may wish to restrict sharing of personal information, but might not be accorded such a choice unless given this right by a privacy law. For this reason, the definition of *publicly available information* should not include such information.

In addition, even with respect to information in government records that are publicly available, the GDPR applies notice requirements, while the CCPA does not. While a consumer may benefit from transparency about a business's use of such publicly available information, applying notice requirements to information that is already publicly available goes beyond the mandate of a consumer privacy law that should be focused on private information.

Personal information should thus be defined as:

[39] However, there are peculiarities with other user rights, such as the right to inspect, when they concern household information.

[40] *CCPA,* Section 1798.140(v)(2).

[41] *GDPR,* Recital 154.

[42] *Id.,* Article 86.

Electronic copy available at: https://ssrn.com/abstract=3893833

> *The term 'personal information' means any information relating to a natural person or to a household, excluding publicly available information.*
>
> *The term 'publicly available information' means information relating to a natural person or to a household (a) in publicly available government records, (b) that the person or household to whom the personal information is related has made publicly available, or (c) that was made publicly available by widely distributed media.*

Personal information is thus personal and private.

C. Is the information reasonably linkable?

Having defined *personal information* as information that is both personal and private, we now turn to the issue of whether it is identifiable information (i.e., information relating to an identifiable person).

There are several methods by which a person may be identifiable. The most obvious method is the use of person's name. The GDPR specifies that a natural person may be identified "by reference to an identifier such as a name…".[43] The CCPA similarly specifies that a particular consumer may be identified using "a real name".[44] Other identifiers can also be used to reasonably establish a person's identity. For example, the CCPA specifies that a particular consumer may be identified using "a real name, … postal address, … email address, …, social security number, driver's license number, [and a] passport number".[45] Thus, under both the GDPR and the CCPA, it is clear that a natural person may be identifiable through, at a minimum, a person's name, personal telephone number, personal email address, and government issued individual identifiers (e.g., driver's license number, social security number, or passport number).

Many privacy policies limit the scope of personally identifiable information to an identifier that itself identifies a person. For example, Apple defines personal information as "data that can be used to identify or contact a single person".[46] Cox defines personally identifiable information as "subscriber name, service and mailing addresses, telephone numbers, social security number, driver's license number, email address, billing and payment records (including credit card and bank account numbers used to pay for our services), subscriber credit information, or other information that potentially could be used to identify, contact, or locate you".[47] Chase uses the term personal information to describe contact information, but excludes "usage and other information".[48]

However, often it is not the identifier itself that is personal. It is the information *associated* with an identifier that is personal. For example, a person may have a public telephone number listing, and hence that person's name and telephone number are public. However, a person's name and telephone number is often associated with information about that person's Internet browsing history, and it is the browsing history that is personal. By omitting information associated with an identifier from the scope of personally identifiable information, we are left wondering what personally identifiable information is collected that the privacy policy fails to disclose.

[43] *Id.,* Article 4(1).

[44] *CCPA,* Section 1798.140(v)(1).

[45] *Id.,* Section 1798.140(v)(1).

[46] Apple, *supra* note 35, at "Collection and Use of Non-Personal Information".

[47] Cox, *Your Privacy Rights as a Cox Customer and Related Information* (last modified August 1, 2020), https://www.cox.com/aboutus/policies/annual-privacy-notice.html, at "Your Information".

[48] JPMorgan Chase Bank, *Online Privacy Policy* (last modified January 15, 2020), https://www.chase.com/digital/resources/privacy-security/privacy/online-privacy-policy, at "Information we collect".

Electronic copy available at: https://ssrn.com/abstract=3893833

Other privacy policies limit the scope of personally identifiable information to an identifier that itself identifies a person and to information that the provider of that service or app links to that identifier. For example, Google defines personal information as "information that you provide to us which personally identifies you, such as your name, email address, or billing information, or other data that can be reasonably linked to such information by Google, such as information we associate with your Google Account."[49]

However, limiting the scope of reasonably linkable information to an identifier that itself identifies a person and to information that the provider of that service or app links to that identifier is severely underinclusive in two separate ways. Identifiers are often used that uniquely identify a person, but not by name, telephone number, or email address. For example, Google and Facebook assign their own identifiers to each person they profile. Such identifiers are then associated with personal information such as browsing history or social network posts. Cox considers contact information to be personally identifiable information, but considers "general location, demographics, …, usage, … and preferences" to be non-personally identifiable information unless it is directly linked to personally identifiable information.[50] Such definitions open up the possibility that these providers consider browsing history, social network posts, or usage information to be *excluded* from the scope of personally identifiable information, if not paired with an identifier that itself identifies a person, and thus not subject to disclosure requirements.

Although such privacy policies often then proceed to list categories of information that the service or app collects that do not fall into the severely limited scope of personally identifiable information as the provider defines it, the exclusion of information related to a person undermines the credibility that the privacy policy's disclosures are comprehensive.

In contrast, some privacy policies use definitions of personally identifiable information that either match or borrow language from those in the GDPR or the CCPA. AT&T uses the CCPA's definition of *personal information.*[51] Comcast defines personal information as "any information that is linked or reasonably linkable to you or your household"[52], which includes part of (but not the full) CCPA definition. Comcast states that personal information "can include information that does not personally identify you — such as device numbers, IP addresses, and account numbers" and "may also include information that does personally identify you, such as your name, address, and telephone number".[53]

Finally, some privacy policies use different terms and definitions depending on the privacy law that applies in the person's location. In its nationwide privacy policy, Facebook avoids use of the term personal information, but characterizes "information that personally identifies you" as "information such as your name or email address that by itself can be used to contact you or identifies who you are".[54] In contrast, in its California privacy policy, Facebook uses the term personal information, and adopts a definition similar to (but not exactly the same as) the CCPA's definition.[55]

[49] Google, *Google Privacy Policy* (last modified September 30, 2020), https://policies.google.com/privacy?hl=en-US, at "We want you to understand the types of information we collect as you use our services" in the pop-up window for "personal information".

[50] Cox, *supra* note 49, at "Your Information".

[51] AT&T, *AT&T Privacy Policy* (last modified June 19, 2020), https://about.att.com/csr/home/privacy/full_privacy_policy.html, at When this Policy applies".

[52] Comcast, *Our Privacy Policy explained* (last revised June 30, 2020), https://www.xfinity.com/privacy/policy, at "Introduction" in the popup window for *personal information*.

[53] *Id.*, "The Personal Information We Collect and How We Collect It".

[54] Facebook, *Data Policy* (last revised August 21, 2020), https://www.facebook.com/policy.php, at "Advertisers".

[55] Facebook, *California Privacy Notice* (last revised July 1, 2020), https://www.facebook.com/legal/policy/ccpa.

Electronic copy available at: https://ssrn.com/abstract=3893833

Both the GDPR and the CCPA also recognize that personal information may be used to establish a person's identity, even if the information lacks an identifier that itself establishes that identity. The GDPR specifies that a natural person may be identified "by reference to … location data … or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person".[56] The EU clarifies that "it is possible to categorise [a] person on the basis of socio-economic, psychological, philosophical or other criteria and attribute certain decisions to him or her".[57] Thus, data records that contain no personal identifiers still relate to an identifiable natural person, if the information in those records is "reasonably likely to be used", potentially in combination with other available information, "to identify the natural person" to whom the information relates.[58]

The CCPA takes a similar approach to the use of personal information to establish identity, albeit with different language. The CCPA's definition of *personal information* implies that a particular consumer may be identified using information that "is reasonably capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer …". The phrase "could reasonably be linked, directly or indirectly, with" is similar to that used often used by the Federal Trade Commission.

The concept of reasonable linkability is more familiar in the United States, and thus serves as a good starting point. However, the CCPA does not define the term.

Garfinkel (2015) defines *linkable information* as "information about or related to an individual for which there is a possibility of logical association with other information about the individual".[59] Adding a reasonableness test and leveraging the definition of *personal information* results in:

> *The term 'reasonably linkable information' means personal information for which there is a reasonable possibility of logical association with other information relating to the person or household to whom the personal information relates.*

Reasonably linkable information is thus personal, private, and reasonably identifiable.

5. Defining reasonably identifiable information, pseudonymous information, and non-trackable information

We now turn to differentiating between different types of reasonably linkable information.

A. Is the information trackable?

We start with the most privacy preserving form of reasonably linkable information: *non-trackable information*. Tracking is made possible by associating pieces of personal information with each other, even if they are not associated with a person by name.

In Section 2C, we discussed information relating to a person or household that is identifiable but has not yet been identified, and that is *not* tracked over time. Such personal information typically involves the use of non-persistent identifiers such as randomized one-time identifiers. Polonetsky (2016) states that, in such personal information, direct identifiers have been removed or transformed so that they cannot link back to

[56] *GDPR,* Article 4(1).
[57] *EU Handbook,* at 89, quoting an opinion issued by the Article 29 Data Protection Working Party.
[58] *GDPR,* Recital 26.
[59] *Garfinkel, supra* note 21*,* at 42.

Electronic copy available at: https://ssrn.com/abstract=3893833

any individual, but indirect identifiers may remain intact if they have "no life outside of the specific context in which it was used".[60]

Consumer privacy laws increasingly are concerned with whether personal information can be used to track a person and create a profile, even if the person's name is not associated with the profile. The CCPA defines *profiling* as "any form of automated processing of personal information … to evaluate certain personal aspects relating to a natural person and in particular to analyze or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behavior, location, or movements."[61] The CCPA distinguishes between profiling versus "[s]hort-term, transient use" of personal information.[62] For example, if a consumer opts-out of sharing of personal information, the CCPA prohibits a business from profiling that consumer, but allows the business to use the consumer's personal information for non-personalized advertising shown as part of the consumer's current interaction with the business.[63] The GDPR also extensively discusses profiling. It requires that privacy notices specifically include disclosure of profiling[64], and it gives consumers a right to opt-out of profiling used for direct marketing purposes[65].

However, neither the GDPR nor the CCPA defines trackable information as an explicit subset of personal information. Instead, they consider profiling as a particular use of personal information. As a result, while they include specific provisions related to profiling, neither require disclosure of whether personal information is stored and use in a trackable form, and neither incorporate tracking directly into their choice framework.

A spirited debate has occurred about whether personal information is trackable when non-persistent identifiers are used. Many identifiers used by service and apps to associate information relating to a person are resettable. Common examples of resettable identifiers include dynamic IP addresses, advertising identifiers that can be reset using mobile device settings, and cookies that can be cleared using browser settings.

Most privacy policies are unclear about whether they consider resettable identifiers and information associated with them to be included in the scope of personally identifiable information. Indeed, providers of services and apps often argue that they are not. Apple argued that information "identified by non-personally identifiable identifiers such as those that are random, resettable, or rotating" should *not* be included in the scope of *personal information* under the CCPA.[66] One common argument made by those opposed to classifying a household's IP address as a personal identifier is that IP addresses are often assigned to a household for only a limited period of time. The Network Advertising Initiative thus argued that resettable identifiers "do not in fact relate to any one unique consumer", and hence it proposed that

[60] *Polonetsky, supra* note 20*,* at 615.

[61] *CCPA,* Section 1798.140(z). The GDPR has an identical definition, except that it uses the term *personal data* instead of *personal information*; see *GDPR,* Article 4(4).

[62] *Id.,* Section 1798.140(e)(4).

[63] *Id.,* Sections 1798.135(f), 1798.140(e)(4).

[64] *GDPR,* Articles 13(2)(f), 14(2)(g).

[65] *Id.,* Article 21(2).

[66] Apple, *Comments of Apple Inc. in connection with the Office of the Attorney General Rulemaking regarding the California Consumer Privacy Act of 2018* (March 8, 2019), at 4.

Electronic copy available at: https://ssrn.com/abstract=3893833

probabilistic identifiers and information associated with them be *excluded* from the scope of *personal information*.[67]

The ambiguity of whether information relating to a person by reference to a resettable identifier is included in privacy policy disclosures demonstrates the importance of clearly spelling out in a consumer privacy law that resettable identifiers are a common method of tracking a person. The CCPA treats a resettable identifier, and the information associated with it, as *personal information* if it "can be used to recognize a consumer [or] a family … over time …".[68] The CCPA further explicitly states that an IP address qualifies as *personal information* if it could be reasonably linked with a particular consumer or household.[69] The GDPR takes a slightly different tack, classifying a resettable identifier, and the information associated with it, as a *personal data* if and only if it can be reasonably used to identify an individual or household.[70] EU guidance states than an IP address is *personal data* if there is additional information reasonably available that identifies the person to whom the IP address has been assigned.[71]

Dynamic IP addresses are usually assigned by an Internet Service Provider to a house's modem for at least a day at a time, and are usually renewed at the end of the IP address lease, so that a dynamic IP address is usually associated with a household for weeks or months at a time. Advertising identifiers and cookies are usually very persistent. In most situations, they are only cleared when a user explicitly does so.[72]

Many consumers have a higher sensitivity when their personal information is tracked over time than when it is used only in the current interaction with a business. However, whereas both the CCPA and the GDPR consider profiling to be a particular use of personal information, it is a cleaner approach to define a particular category of personal information that allows tracking to take place. The advantage of this approach is that *trackable information* takes it rightful place on the spectrum of identifiability, rather than being called out as a particular use of personal information. This helps guide an assignment of notice and consent obligations onto trackable information that is in the public interest and that is reasonable compared to the obligations placed onto other types of personal information.

Drawing on the CCPA's description of profiling as involving the linking of personal information from more than one interaction, and Polonetsky (2016)'s description of it as involving the linking of personal information from more than one context, *non-trackable information* can be defined as:

> *The term 'non-trackable information' means reasonably linkable information for there is no reasonable possibility of logical association of the information with other information relating to the person or household obtained from another context or another interaction with the person or household.*

The definition builds on the previous proposed definition of *reasonably linkable information*, but also requires that the logical association be not reasonably possible over time.

If *reasonably linkable information* fails to meet the definition of *non-trackable information*, then it remains trackable:

[67] Network Advertising Initiative, *Comments of the Network Advertising Initiative re Implementing Regulations for the California Consumer Privacy Act of 2018*, at 10, *available at* https://www.networkadvertising.org/sites/default/files/naicommentletterccpaimplementingregulations.pdf.

[68] *CCPA,* Section 1798.140(aj).

[69] *Id.,* Section 1798.140(v)(1).

[70] *GDPR,* Recital 26.

[71] *EU Handbook*, at 91-92.

[72] Less commonly, a user may have set a browser to automatically clear cookies upon exit.

Electronic copy available at: https://ssrn.com/abstract=3893833

> *The term ‘trackable information’ means reasonably linkable information that is not non-trackable information.*

B. Is the information reasonably identifiable?

We tun now to the second most privacy-preserving form of reasonably linkable information: *pseudonymous information*. In Section 2B, we discussed information relating to a person or household that is identifiable but has not yet been identified, and that is tracked over time. Such personal information typically involves the use of persistent identifiers such as device identifiers or advertising identifiers that can be used to track a person or household over time.

Almost all online services and apps collect device identifiers. These device identifiers very often include IP addresses, see e.g., the privacy policies of Chase, Uber, and United.[73] Apps that run on mobile devices also often collect the IMEI identifiers of mobile devices, see e.g., the privacy policies of Google, Microsoft, and Apple.[74] Often, privacy policies state that they collect device identifiers, but fail to specify which ones, see e.g., the privacy policies of AT&T, Comcast, Facebook, Pinterest, and Twitter.[75]

Almost all online advertising-supported service and apps also collect advertising identifiers. Such advertising identifiers are usually associated with a particular device, and thus serve as de-facto device identifiers. Most commonly, services and apps collect Apple and Android advertising identifiers, see e.g., the privacy policies of KAYAK, The Weather Channel, and Zillow.[76]

However, privacy policies differ in whether they include, in the scope of personally identifiable information, device identifiers and information that is associated with device identifiers. Many privacy policies limit the scope of personally identifiable information to an identifier that itself identifies a person, and perhaps to information that the provider of that service or app links to that identifier. Although such privacy policies almost always disclose that the service or app collects device identifiers, they do not typically discuss whether device identifiers are considered by the provider to qualify as a method of identification of a person. This leaves open the question of whether these privacy policies consider device identifiers, and information that is associated with device identifiers, to constitute personally identifiable information.

Indeed, providers of services and apps often argue that these device identifiers do *not* identify a person, and thus that information associated with device identifiers or advertising identifiers does *not* constitute personally identifiable information. Google argued that device identifiers are often *not* associated with a

[73] JPMorgan Chase Bank, *supra* note 158, at “Usage and Other Information” and “Chase Mobile”; Uber Technologies Inc., *Uber Privacy Notice*” (last modified October 27, 2020), https://www.uber.com/legal/en/document/?name=privacy-notice&country=united-states&lang=en, at III.A.2 “Device data”; United Airlines Inc., *Customer Data Privacy Policy*” (last modified January 15, 2016), https://www.united.com/ual/en/us/fly/privacy.html, at “Information we collect automatically” and “Information we collect through our mobile application(s)”.

[74] Google, *supra* note 31, at “Unique identifiers”; Microsoft, *supra* note 133, at “Personal data we collect”; Apple, *supra* note 143, at “What personal information we collect” and at “Cookies and Other Technologies”.

[75] AT&T, *supra* note 53, at “The information we collect”; Comcast at “The Personal Information We Collect and How We Collect It”; Facebook, *supra* note 57, at “Identifiers”; Pinterest, *supra* note 26, at “We also get technical information when you use Pinterest”; Twitter, *supra* note 27, at “You should read this policy in full, but here are a few key things we hope you take away from it” and at “Log Data”.

[76] KAYAK Software Corporation, *Privacy Policy* (last modified July 1, 2020), https://kayak.com/privacy, at “Information We Collect and Use”; The Weather Company, *Privacy Policy* (last modified December 29, 2019), https://weather.com/en-US/twc/privacy-policy, at 1.B; Zillow Group, *Privacy Policy* (last modified January 1, 2020), https://www.zillowgroup.com/zg-privacy-policy/, at “Device information”.

Electronic copy available at: https://ssrn.com/abstract=3893833

person's identity; and thus one should question whether Google's privacy policy considers information associated with a device identifier to constitute personally identifiable information.[77] The Internet & Television Association (NCTA), a trade association representing cable Internet Service Providers, argued that IP addresses *cannot* identify an individual on their own.[78] The Internet Advertising Bureau (IAB), a trade association representing Internet advertisers and ad brokers, argued that an "anonymous identifier" should *not* qualify as personally identifiable information.[79] The Network Advertising Initiative, a trade association representing Internet advertising companies, argued that IP addresses are not *personal information* under CCPA, unless a business "has linked it, or reasonably could link it, with additional pieces of information known by the business to identify a particular consumer or household, such as name or residential address."[80]

The ambiguity of whether information relating to a person by reference to a device identifier is included in privacy policy disclosures demonstrates the importance of clearly spelling out in a privacy law that device identifiers are a common method of linking information to a person. The GDPR classifies a device identifier, and any other information associated with it, as *personal data* if and only if it can be attributed to a natural person, including by the use of additional information.[81] So the question remains: can a device identifier be attributed to a person? EC guidance gives advertising identifiers as an example of *personal data* without limitation, but the question has likely not been definitely answered.[82] The CCPA is also less than clear on this issue. The original version of the CCPA explicitly included device identifiers in its definition of a *unique identifier*, which in turn implies that device identifiers are, without limitation, a form of identification of a person. However, the recently revised version of the CCPA may be interpreted to classify device identifiers and the information associated with it as *personally information* if and only if the device is "linked to" or "could be reasonably linked to" a consumer or family.[83]

The ability of a business to use a device identifier to establish the identity of a person depends on the nature of the device identifier and the availability of information that associates the device identifier with a natural person. Advertising identifiers are frequently shared by devices, and are shared widely within the advertising ecosystem. There is additional reasonably available information that associates an advertising identifier with a natural person. It should be presumed that a person's identity can be reasonably established using an advertising identifier, and thus that the combination of an advertising identifier with other personal information constitutes *reasonably identifiable information*. It is possible that a device identifier is shared by a device only in a pseudonymous fashion, and that subsequent user actions do not render that identifier sufficient to identify a person. However, in general, any persistent identifier that is shared widely within the advertising ecosystem will render that identifier sufficient to identify a person, because eventually that information will be associated with a person's identity, e.g., when a person registers with a website or purchases an item.

[77] Google, *Google comments on the Attorney General's draft regulations to govern compliance with the California Consumer Privacy Act* (December 6, 2019), at 3.
[78] Broadband Privacy Order, *Report and Order*, 31 FCC Rcd 13911 (October 27, 2016), para 94 n. 239 [hereinafter *FCC Order*], *available at* https://docs.fcc.gov/public/attachments/FCC-16-148A1.pdf,
[79] *Id.*
[80] Network Advertising Initiative, *Re: Modified Proposed Regulations for the California Consumer Privacy Act* (February 25, 2020), at 3.
[81] *GDPR,* Recital 26.
[82] European Commission, *What is personal data?* (accessed May 6, 2020), https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_en, at "Examples of personal data".
[83] *CCPA,* Section 1798.140(x); *CCPA,* Section 1798.140(ai).

Electronic copy available at: https://ssrn.com/abstract=3893833

The GDPR does not distinguish between persistent and non-persistent indirect identifiers in its definition of *pseudonymisation*:

> "the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person".[84]

Thus, under the GDPR, the use of persistent indirect identifiers might result in pseudonymous personal data used for tracking, whereas the use of non-persistent indirect identifiers might result in pseudonymous personal data not used for tracking. The GDPR considers both types of information to be personal data.

To avoid this confusing use of terminology, we use the term *non-trackable information* (as defined above) to describe the use of non-persistent indirect identifiers, and the term *pseudonymous information* to describe the use of persistent indirect identifiers. Drawing upon the GDPR's description of pseudonymous information as resulting in the inability to attribute the information to a specific person without the use of additional information, and adding a reasonableness test, *pseudonymous information* can be defined as:

> *The term 'pseudonymous information' means trackable information for which the related person or household is not reasonably identifiable using that personal information and other reasonably linkable information.*

The definition builds on the previous proposed definition of *trackable information*, but also requires that the information cannot be associated with other reasonably linkable information such that the combined information can be used to reasonably identify the related person or household.

If *trackable information* fails to meet the definition of *pseudonymous information*, then it remains reasonably identifiable:

> *The term 'reasonably identifiable information' means trackable information that is not pseudonymous information.*

6. Defining De-identified Information and Anonymous Information

The GDPR defines *anonymous information*, but not *de-identified information*. The CCPA defines *de-identified information*, but not *anonymous information*. The two definitions are not the same. In this subsection, we craft definitions of both.

A. Is the information anonymous?

The GDPR defines *anonymous information* as:

> "information which does not relate to an identified or identifiable natural person".[85]

This definition simply inverts the definition of *personal data*, and thus includes (a) information that is not personal and (b) information that is personal but whose degree of identifiability falls short of *personal data*. There are several problems with this definition. First, it needlessly includes information that does not relate to an individual or household, and thus conflates anonymous information with non-personal information. More critically, it fails to distinguish between anonymous information and de-identified information. The

[84] *GDPR,* Article 4(5).

[85] *Id.,* Recital 26.

Electronic copy available at: https://ssrn.com/abstract=3893833

GDPR simply excludes both from *personal data*, and then exempts them from notice and choice requirements.

Polonetsky (2016) describes anonymous information as personal information in which both direct and indirect identifiers have been removed or transformed so that they cannot link back to any individual, and in which the method for removal or transformation includes mathematical and technical guarantees that are sufficient on their own to distort the data so as to prevent reidentification.[86] As an example of an anonymization technique that can provide such guarantees, Polonetsky (2016) mentions differential privacy algorithms, which can hide whether or not an individual is present in a dataset.

Privacy policies often overreach in their claims that personal information is anonymous. AT&T defines Anonymous Information as "[i]nformation that doesn't directly identify and can't reasonably be used to identify an individual customer or user".[87] Anonymous Information is thus defined by AT&T as all information that AT&T does not consider to be Personal Information, which it defines as "[i]nformation that directly identifies or reasonably can be used to figure out the identity of a customer or user, such as your name, address, phone number and e-mail address."[88] AT&T then explains that "[w]e treat identifiers like cookies, advertising identifiers, device identifiers, and household identifiers as Anonymous Information except in circumstances where they can be used to identify you."[89] However, the information is almost certainly personal, and is presumably private. If it includes an identifier such as an advertising identifier or device identifier, then it is also linkable, not de-identified, and trackable. Thus, it certainly does not qualify as *anonymous information.* Under the GDPR, it almost certainly would be categorized as *personal data*, and under the CCPA as *personal information*.

KAYAK defines Anonymized Information as "information that cannot be linked to you or any other specific user using any means available to us, either because it was collected anonymously or has been subsequently anonymized."[90] KAYAK states that "[i]nformation that is anonymous or has been anonymized is no longer considered 'personal information'"[91] KAYAK appears to include in the scope of Anonymized Information, and thus exclude from the scope of Personal Information, information that it considers to be "de-identified usage data" that is associated with a mobile advertising identifier.[92] Indeed, KAYAK states that Anonymized Information "may be subsequently used for any purpose."[93] KAYAK's descriptions of Anonymized Information are inconsistent. If the information can truly not be linked to a person or household, including to a non-identifiable person or household, then it would qualify as *anonymous information*. However, if the information includes a mobile advertising identifier, then it neither qualifies as *anonymous information*, nor *de-identified information*, nor even as *non-trackable information*.

Amazon, in its privacy policy for its eero branded Wi-Fi products, defines Anonymous Data as "data that, either in its original form or as the result of anonymization procedures that we perform on Personal Data, is not associated with or linked to your Personal Data."[94] Amazon asserts that "Anonymous Data does not, by itself, permit the identification of individual persons."[95] Amazon explains "[w]e may create Anonymous

[86] *Polonetsky, supra* note 20*,* at 618.
[87] AT&T, *supra* note 53, at "Definitions".
[88] *Id.*
[89] *Id.*
[90] KAYAK, *supra* note 78, at "How we use your information".
[91] *Id.,* at "How we use your information".
[92] *Id.,* at "Our Advertising Cookies".
[93] *Id.,* at "How we use your information".
[94] Amazon, *Privacy for eero Devices, Applications and Services* (last modified February 28, 2020), https://eero.com/legal/privacy, at "Types of data we collect"
[95] *Id.*, at "Types of data we collect".

Electronic copy available at: https://ssrn.com/abstract=3893833

Data records from Personal Data by using various procedures to remove or obscure information (such as your name, email address, phone number or IP address) that makes the data personally identifiable to you", and then reserves the right to use and share Anonymous Data for any purposes, apparently without disclosure.[96] Amazon's description of Anonymous Data are too vague to allow classification. At best, Amazon's procedures to remove or obscure information may result in *anonymous information*, if the procedures include mathematical and technical guarantees that are sufficient on their own to prevent reidentification. However, Amazon's procedures to remove or obscure information may not have such guarantees, and may easily be *de-identified information, non-trackable information,* or *pseudonymous information*. Furthermore, given that Amazon only requires that Anonymous Data not "by itself" permit the identification of a person, it is possible that it is linkable to information that does permit the identification of a person, in which case the information is properly classified as *reasonably identifiable information*.

The ability, or lack thereof, to associate or link information to an individual or household features prominently in the distinction between anonymous information and other types of more identifiable personal information. Indeed, the CCPA's definition of *personal information* relies strongly on the concept: "information that … is reasonably capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household".[97]

Following the guidance in Polonetsky (2016) that anonymous information include mathematical and technical guarantees that are sufficient on their own to distort the data so as to prevent reidentification, we focus here on the technical ability of associating information with a particular consumer or household; in the following subsection, we consider whether the association can be reasonably made. Garfinkel (2015) defines linkable information as "information about or related to an individual for which there is a possibility of logical association with other information about the individual".[98]

This test can be adapted to the proposed definition of *personal information*:

> *The term 'anonymous information' means personal information for which there is no possibility of logical association with other information relating to the person or household to whom the personal information relates.*

If *personal information* is not anonymous, it should be classified it as *linkable information*, defined as:

> *The term 'linkable information' means personal information that is not anonymous information.*

Privacy laws sometimes also distinguish between anonymous information and aggregate information. Polonetsky (2016) considers aggregated anonymous information to be a subset of anonymous information. In aggregated anonymous information, the data are so highly aggregated that the aggregation itself serves as a mathematical and technical guarantee so as to prevent reidentification.[99]

The CCPA defines *aggregate consumer information* as "information that relates to a group or category of consumers, from which individual consumer identities have been removed, that is not linked or reasonably linkable to any consumer or household, including via a device", and then explains that it does not include "one or more individual consumer records that have been deidentified".[100] The CCPA then excludes

[96] *Id.*, at "Use of your Personal Data".
[97] *CCPA,* Section 1798.140(v)(1).
[98] *Garfinkel, supra* note 21*,* at 42.
[99] *Polonetsky, supra* note 20*,* at 618.
[100] *CCPA,* Section 1798.140(b).

Electronic copy available at: https://ssrn.com/abstract=3893833

aggregate consumer information from *personal information*. The GDPR similarly considers *aggregate data* to be a subset of *anonymous information*, and then excludes it from *personal data*.[101]

That said, we do not need to define in a consumer privacy law a category of aggregate information, because it is typically afforded the same treatment as other types of *anonymous information*.

B. Is the information de-identified?

Polonetsky (2016) distinguishes de-identified information from anonymous information based on the difficulty of associating the information with the person to whom it is related. In order to qualify as either de-identified information or anonymous information, both direct and indirect identifiers must have been removed or transformed so that they cannot link back to any individual. However, whereas for anonymous information the method for removal or transformation includes mathematical and technical guarantees that are sufficient on their own to distort the data so as to prevent reidentification, for de-identified information such mathematical and technical guarantees are absent and legal controls take the place of technological controls. As examples of de-identification techniques that remove both direct and indirect identifiers, but which cannot provide mathematical and technical guarantees, Polonetsky (2016) mentions suppression, generalization, perturbation, and swapping algorithms. Garfinkel (2015) provides an overview of these types of algorithms.[102]

Privacy policies often overreach in their claims that personal information is de-identified. KAYAK classifies as “de-identified usage data” mobile advertising identifiers, “anonymous device identifiers”, and cookies.[103] Such identifiers result in a reasonable possibility of logical association with other information relating to the person or household to whom the information relates, and thus it is not *de-identified information*. Furthermore, such identifiers are persistent, and thus the associated information is not *non-trackable information*.

The proposed definition of *anonymous information* already captures the subset of *personal information* in which reidentification is prevented solely using technological controls. If such technological controls are absent, the information remains classified as *linkable information*. It remains to distinguish between information for which the logical association can be reasonably made versus information for which the logical association is possible but not reasonable given the current state of technology, the availability of information with which it can be associated, and legal controls.

The CCPA defines *de-identified information* as

> “information that cannot reasonably be used to infer information about, or otherwise be linked to, a particular consumer …”[104]

The intent is to exempt from the CCPA’s definition of *personal information* a category of information that can be logically associated with an individual or household, but for which the association cannot be reasonably made due to a combination of technological and legal controls.

The CCPA’s definition can be adapted to build on the proposed definition of *linkable information*:

> *The term ‘de-identified information’ means linkable information for which there is no reasonable possibility of logical association with other information relating to the person or household to whom the linkable information relates.*

[101] *GDPR,* Recital 162.

[102] *Garfinkel, supra* note 21*,* at 20.

[103] KAYAK, *supra* note 78, at “Our Advertising Cookies”.

[104] *CCPA,* Section 1798.140(m).

Electronic copy available at: https://ssrn.com/abstract=3893833

7. Legal controls

The classification of various types of personal information relies on characteristics of that information. The CCPA recognizes that in order for information to maintain those characteristics, legal controls are often required.

A. Legal controls on de-identified information

The discussion is most developed in the context of de-identified information. The FTC Report proposed three legal controls, but it framed these three legal controls as a safe harbor. Specifically, it proposed that information should be considered "not [] reasonably linkable to a particular consumer or device" if the business possessing the information implements three legal controls.[105] In contrast, the CCPA first requires that the information actually be de-identified, i.e. that it cannot reasonably be used to infer information about, or otherwise be linked to, a particular consumer, and then in addition requires that a business that possesses *deidentified* information implement three legal controls.

The CCPA's higher level of protection is appropriate. If there is a reasonable possibility of logical association of *linkable information* with other information relating to the person or household to whom the *linkable information* relates, then it should not qualify as *de-identified information*, even if a business possessing that information implements the specified legal controls intended to prevent such association but fails to accomplish that goal. For this reason, a consumer privacy law should require legal controls on de-identified information:

> *In order to qualify as de-identified information, the entity possessing that information must implement controls (A1)-(D1) below.*

The first legal control in the FTC Report is that the business "must achieve a reasonable level of justified confidence that the data cannot reasonably be used to infer information about, or otherwise be linked to, a particular consumer, computer, or other device".[106] The CCPA somewhat similarly requires that a business possessing de-identified information "[t]ake reasonable measures to ensure that the information cannot be associated with a consumer or household".[107]

The FTC Report and the CCPA maintain this legal control for different reasons. In the FTC Report, a business's "reasonable level of justified confidence" that the information is de-identified is the principal element of the safe harbor. In the CCPA, however, the legal control is in addition to the requirement that the information actually be de-identified. There remains a good reason to add a similar legal control, because even if a business possesses de-identified information, it is in the public interest that the information not later be re-identified.

There is another difference between the FTC's phrasing and the CCPA's phrasing. The CCPA requires reasonable measures to ensure that the information "cannot be" re-identified, whereas the FTC requires reasonable measures to ensure that the information "cannot reasonably be" re-identified. The test should be "reasonably linkable", both to first qualify as *de-identified information* and also as a legal control.

Building on the proposed definition of *de-identified information*, the first legal control should be:

[105] *FTC Report,* at 21.
[106] *Id.*
[107] *CCPA,* Section 1798.140(m)(1).

Electronic copy available at: https://ssrn.com/abstract=3893833

> *(A1) It must take reasonable measures to ensure that the information remains in a form in which there is no reasonable possibility of logical association with other information relating to the person or household to whom the linkable information relates.*

The second legal control in the FTC Report is that the business "must publicly commit to maintain and use the data in a de-identified fashion, and not to attempt to re-identify the data".[108] For the FTC, this legal control enables the FTC to act under Section 5 of the FTC Act if the commitment is violated. In a privacy law, there would likely be other and stronger methods of enforcement. Nevertheless, such a public commitment is in the public interest, and the CCPA mirrors this legal control.[109] Thus, the second and third legal controls should be:

> *(B1) It must publicly commit to maintain and use the information only in a form in which there is no reasonable possibility of logical association with other information relating to the person or household to whom the linkable information relates.*
>
> *(C1) It must publicly commit to not attempt to associate the information with other information relating to the person or household to whom the linkable information relates.*

The third legal control in the FTC Report is that if a business makes de-identified information available to other companies, it must "contractually prohibit such entities from attempting to re-identify the data".[110] This legal control ensures that the direct recipient of de-identified information doesn't re-identify the information. The CCPA takes this a step further, requiring that a business possessing de-identified information "[c]ontractually obligates any recipients of the information to comply with all provisions of this subdivision".[111] Thus, in addition to prohibiting direct recipients from re-identification, it also requires recipients to make similar public commitments and to contractually prohibit any downstream recipients from re-identifying the information. This expanded legal control prohibits all downstream re-identification. The last legal control be:

> *(D1) It must contractually obligate any third parties to whom it discloses the information to implement controls (A1)-(D1).*

B. Legal controls on non-trackable information

As with de-identified information, the technological algorithm used to transform the personal information into non-trackable information is not sufficient to guarantee that tracking is not possible. There remains a need to add legal controls. Neither the GDPR nor the CCPA place legal controls on non-trackable information, since neither distinguishes such information from other types of personal data (under the GDPR) or personal information (under the CCPA). However, we can mirror the legal controls placed in the previous subsection on de-identified information:

> *In order to qualify as non-trackable information, the entity possessing that information must implement controls (A2)-(D2) below.*
>
> *(A2) It must take reasonable measures to ensure that the information remains in a form in which there is no reasonable possibility of logical association of the information with other information relating to the person or household obtained from another context or another interaction with the person or household.*

[108] *FTC Report*, at 21.
[109] *CCPA,* Section 1798.140(m)(2).
[110] *FTC Report*, at 21.
[111] *CCPA,* Section 1798.140(m)(3).

Electronic copy available at: https://ssrn.com/abstract=3893833

> *(B2) It must publicly commit to maintain and use the information only in a form in which there is no reasonable possibility of logical association of the information with other information relating to the person or household obtained from another context or another interaction with the person or household.*
>
> *(C2) It must publicly commit to not attempt to associate the information with other information relating to the person or household obtained from another context or another interaction with the person or household.*
>
> *(D2) It must contractually obligate any third parties to whom it discloses the information to implement controls (A2)-(D2).*

C. Legal controls on pseudonymous information

Pseudonymous information requires legal controls to ensure that the related person or household is not identified. Neither the GDPR nor the CCPA place legal controls on trackable information, since neither distinguishes such information from other types of personal data (under the GDPR) or personal information (under the CCPA). However, we can mirror the legal controls used above:

> *In order to qualify as pseudonymous information, the entity possessing that information must implement controls (A3)-(D3) below.*
>
> *(A3) It must take reasonable measures to ensure that the information remains in a form in which the related person or household is not reasonably identifiable using that personal information and other reasonably linkable information.*
>
> *(B3) It must publicly commit to maintain and use the information only in a form in which the related person or household is not reasonably identifiable using that personal information and other reasonably linkable information.*
>
> *(C3) It must publicly commit to not attempt to identify the person or household to whom the information is related.*
>
> *(D3) It must contractually obligate any third parties to whom it discloses the information to implement controls (A3)-(D3).*

8. Empowering Consumers Who Desire Privacy-Preserving Advertising

In this section, we investigate how advertising can be implemented using different types of personal information. The goal is to understand if and how differentiating between different types of personal information may affect consumers.

We give examples of advertising based on *reasonably identifiable information*, *pseudonymous information*, and *non-trackable information*. In each example, we consider the following entities:

- An ad venue, an entity which offers a venue in which ads appear, e.g., a website with ads on its webpages.
- An advertiser, an entity which offers ads to be published in ad venues, e.g., a business advertising a product.
- An ad broker, an entity which determines the ad venues on which a particular ad will appear, e.g., a business that contracts with both ad venues and advertisers and that determines the placement of each ad.

We presume that the advertiser and the ad broker have a contract under which the advertiser pays the ad broker to place an ad, and that the ad broker and the ad venue have a contract under which the ad broker

Electronic copy available at: https://ssrn.com/abstract=3893833

pays the ad venue to have the ad appear. We distinguish between the acts of "placing" and "publishing" an ad. *Placing* an ad is the function of determining the ad venues on which an ad appears; here we assume this is done by the ad broker. *Publishing* an ad is the technological function of causing the ad to appear; here we assume this may be done by any of the parties.

For each advertising example, we consider the types of personal information collected and used by each party, and the types disclosed or shared between parties. We start with a privacy-invasive example that is commonplace today, and then work our way through a sequence of increasingly less privacy-invasive examples.

A. Using reasonably identifiable information for behavioral ads published by an ad broker

We first consider the use of reasonably identifiable information to place behavioral ads. In this example, the advertiser chooses to advertise based on the behavior of people in the desired audience. We use the term *behavioral advertising* to describe this form.

For example, SmithLuxuryCars.com wishes to advertise to people who are interested in luxury automobiles. SmithLuxuryCars.com purchases a service from AbcAdBroker.com to place ads that will be seen only by people who are interested in luxury automobiles, based on detailed profiles of these people. AbcAdBroker.com contracts with websites (including CarReviews.com) that are often visited by people who are interested in luxury automobiles and that have detailed profiles of their website visitors.

When a person visits CarReviews.com, the website collects the person's email address and advertising id, and looks up a profile that was previously compiled based on the person's activity on the website. CarReviews.com shares the person's IP address, advertising identifier, and profile with AbcAdBroker.com, which shares this information with advertisers, and auctions off the ad. SmithLuxuryCars.com wins the auction, and AbcAdBroker.com tells CarReviews.com to redirect the website visitor to SmithLuxuryCars.com to obtain the ad. The ad is thus seen only be people whose profiles demonstrate that they are interested in luxury automobiles.

The collection, use, and sharing of personal information is shown in Figure 1. The combination of the person's IP address, email address, advertising identifier, and profile is *reasonably identifiable information*. The information shared with the ad broker and the advertiser remain *reasonably identifiable information*, presuming that the contracts between the ad venue, ad broker, and advertiser do not prohibit the ad broker or the advertiser from using the IP address and advertising identifier to identify the person.

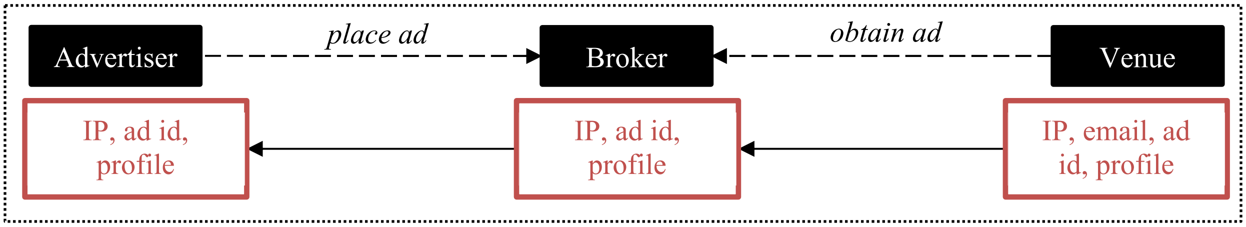

Figure 1. Behavioral ads

This type of advertising is common, but privacy-invasive since it uses the most identifiable form of information.

B. Using pseudonymous information for audience segment ads with tracking

We next consider the use of pseudonymous information to place audience segment ads. In this example, the advertiser chooses to advertise to people who fall into specified audience segments, based on prior tracking of these people.

For example, SmithLuxuryCars.com wishes to advertise to people who are interested in luxury automobiles. SmithLuxuryCars.com purchases a service from AbcAdBroker.com to place ads that will be seen only by people who fall into a luxury automobile audience segment, based on prior tracking. AbcAdBroker.com contracts with websites (including CarReviews.com) that are often visited by people who are interested in luxury automobiles and that can determine if its website visitors fall into the luxury automobile audience segment.

When a person visits CarReviews.com, the website collects the person's advertising id, and looks up a profile that was previously compiled based on the person's activity on the website. However, instead of sharing the person's profile with AbcAdBroker.com, CarReviews.com selects audience segments based on the profile, and only shares the person's IP address, advertising identifier, and audience segments. AbcAdBroker.com awards the ad to SmithLuxuryCars.com, who is the advertiser willing to pay the most to place an ad to a person in the luxury automobile audience segment. AbcAdBroker.com tells CarReviews.com to redirect the website visitor to AbcAdBroker.com to obtain the ad. AbcAdBroker.com generates summary statistics about its ad placements for SmithLuxuryCars.com, but it does not share information about the individual people who saw the ad.

The collection, use, and sharing of personal information is shown in Figure 2. Since a consumer may be reasonably identified using the consumer's IP address, the personal information is *reasonably identifiable information* if there are no legal controls preventing this identification. However, if the legal controls proposed in Section 7C were in place, then the information would be *pseudonymous*, and all entities using and sharing this information would commit to maintaining in a pseudonymous form. The advertiser only receives summary statistics, which qualify as *anonymous information*.

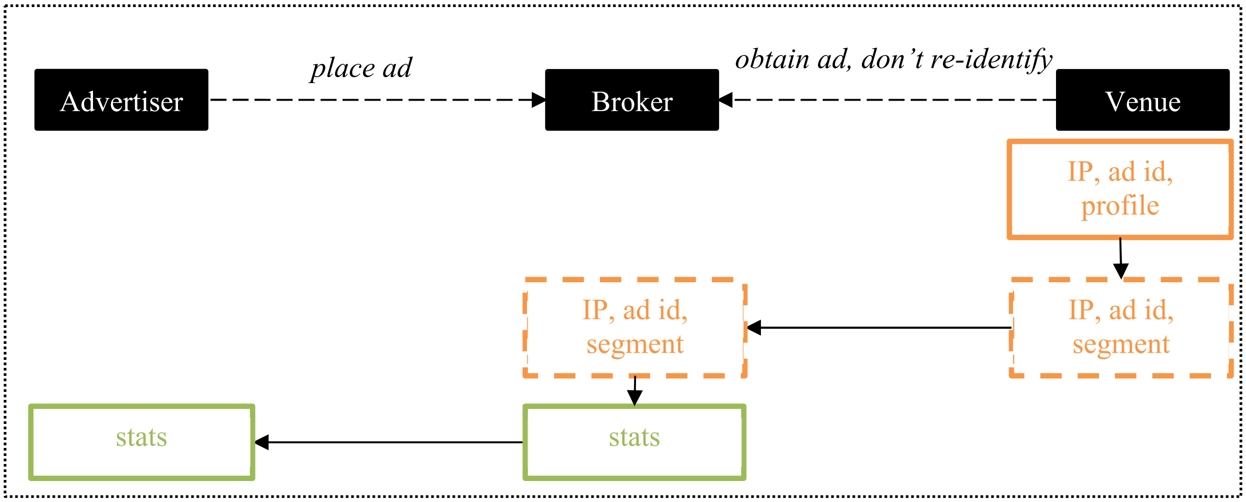

Figure 2. Audience segment ads with tracking

Audience segment advertising is common. Advertisers and ad brokers often claim the personal information used and shared is de-identified, but it is not. However, if legal controls were adopted, then the personal information would qualify as pseudonymous information, which would allow consumers to remain pseudonymous.

C. Using non-trackable information for audience segment ads without tracking

We next consider the use of non-trackable information to place audience segment ads. In this example, the advertiser chooses to advertise to people who fall into specified audience segments, based solely on the current interaction with these people.

Electronic copy available at: https://ssrn.com/abstract=3893833

For example, SmithLuxuryCars.com wishes to advertise to people who are interested in luxury automobiles. SmithLuxuryCars.com purchases a service from AbcAdBroker.com to place ads that will be seen only by people who fall into a luxury automobile audience segment, based solely on the current interaction with these people. AbcAdBroker.com contracts with websites (including CarReviews.com) that are often visited by people who are interested in luxury automobiles and that can determine if its website visitors fall into the luxury automobile audience segment based on the current website visit.

When a person visits CarReviews.com, the website collects the person's advertising id, and determines audience segments, based on the current website visit only. It generates a one-time identifier, and shares that one-time identifier and audience segments with AbcAdBroker.com, who awards the ad to SmithLuxuryCars.com, the advertiser willing to pay the most to place an ad to a person in the luxury automobile audience segment. AbcAdBroker.com tells CarReviews.com to publish SmithLuxuryCars.com's ad. AbcAdBroker.com generates summary statistics about its ad placements for SmithLuxuryCars.com, but it does not share information about the individual people who saw the ad.

The collection, use, and sharing of personal information is shown in Figure 3. The combination of the person's IP address, advertising identifier, and profile is *pseudonymous information*, if the ad venue implements the corresponding legal controls discussed in Section 7C (including not re-identifying the person). However, when the ad venue converts the profile information into audience segments and pairs it with a one-time identifier instead of an IP address, the information is transformed from *trackable* to *non-trackable*. Thus, the combination of the one-time identifier and audience segments shared with the ad broker are *non-trackable information*, if the contract between the ad broker and the ad venue commits the ad broker to implement the corresponding legal controls discussed in Section 7B (including maintaining the information in non-trackable form). The advertiser only receives summary statistics, which qualify as *anonymous information*.

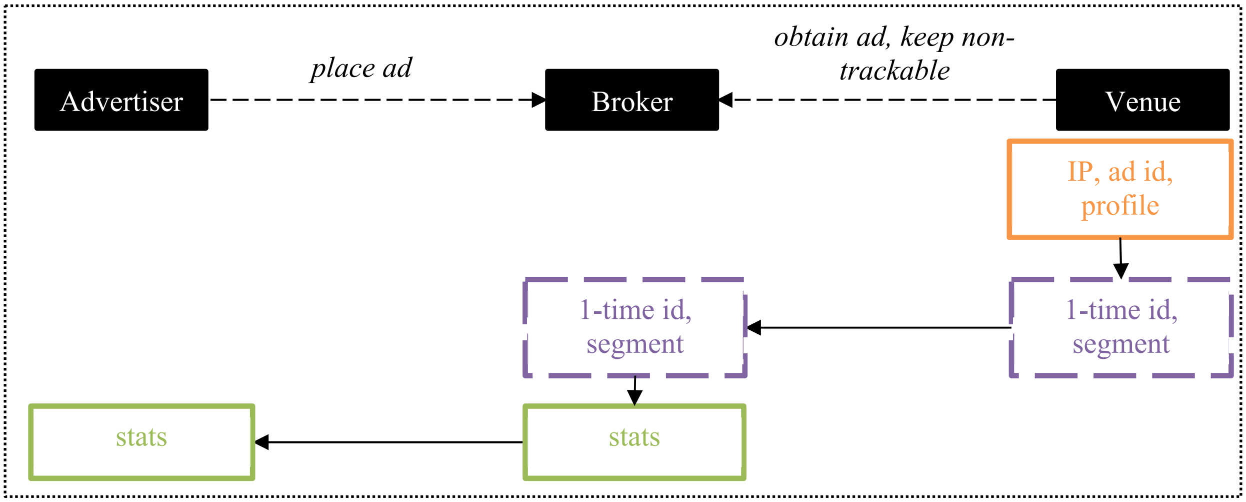

Figure 3. Audience segment ads without tracking

Audience segment advertising with tracking is common. The Do Not Track web browsing setting is commonly ignored by advertisers and ad brokers, due to a lack of incentive to honor these consumer requests. However, a consumer privacy bill's choice framework provided incentives for ad venues to implement the proposed legal controls on non-trackable information, then the corresponding personal information would remain non-trackable.

Electronic copy available at: https://ssrn.com/abstract=3893833

D. Using de-identified information for contextual ads

In our final example, we consider the use of de-identified information to place contextual ads. In this example, an advertiser advertises basely solely on characteristics of the ad venue. We use the term *contextual advertising* to describe this form.

For example, SmithLuxuryCars.com wishes to advertise on websites that are frequently viewed by people who are interested in luxury automobiles. SmithLuxuryCars.com purchases a service from AbcAdBroker.com to place ads on such websites. AbcAdBroker.com contracts with websites (including CarReviews.com) that provide summary statistics to show that they are often visited by people who are interested in luxury automobiles.

When people visit CarReviews.com, the website keeps track of the types of automobiles they are interested in, but it does not store any identifiers of its website visitors. In addition, it generalizes this information. Based on the generalized information, CarReviews.com generates summary statistics, including the percentage of its website visitors who are interested in luxury automobiles. It shares these statistics with AbcAdBroker.com, which auctions ads based on these statistics, and SmithLuxuryCars.com wins the auction. AbcAdBroker.com tells CarReviews.com to publish SmithLuxuryCars.com's ad.

The collection, use, and sharing of personal information is shown in Figure 4. The generalized information used by the ad venue may qualify as *de-identified information*, if the ad venue implements the corresponding legal controls proposed in Section 7A (including maintaining the information in de-identified form). The ad broker and the advertiser only receive summary statistics, which qualify as *anonymous information*.

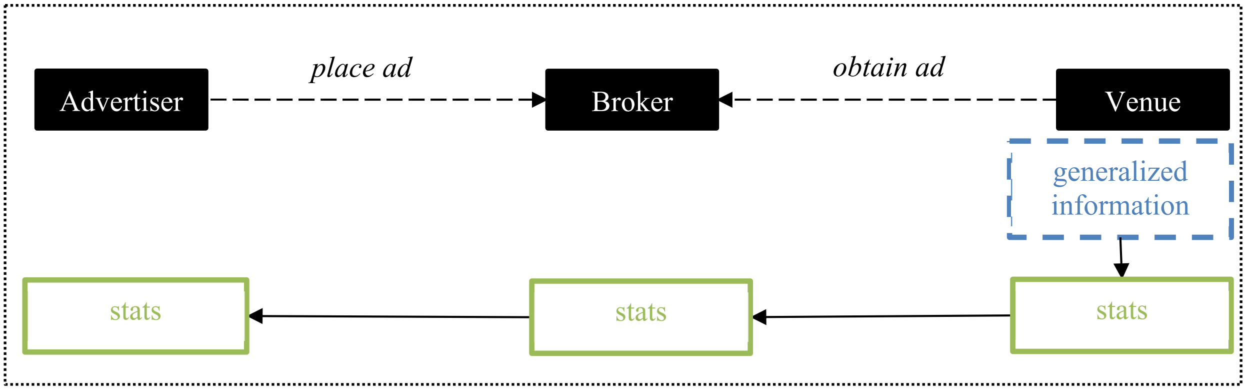

Figure 4. Contextual ads

Contextual advertising used to be the most common form before more targeted forms arose. A well-crafted definition of de-identified information supports this form of advertising.

9. Conclusion

Section 2 discussed the limits imposed by defining only two sets of information: personal information and de-identified information. Section 3 argued that consumers are likely to view the use and sharing of different types of personal information differently, based on whether the consumer is pseudonymous and whether the consumer is tracked.

Sections 4-6 developed proposed statutory definitions. These definitions, and the logical flow, is illustrated in Figure 5.

Electronic copy available at: https://ssrn.com/abstract=3893833

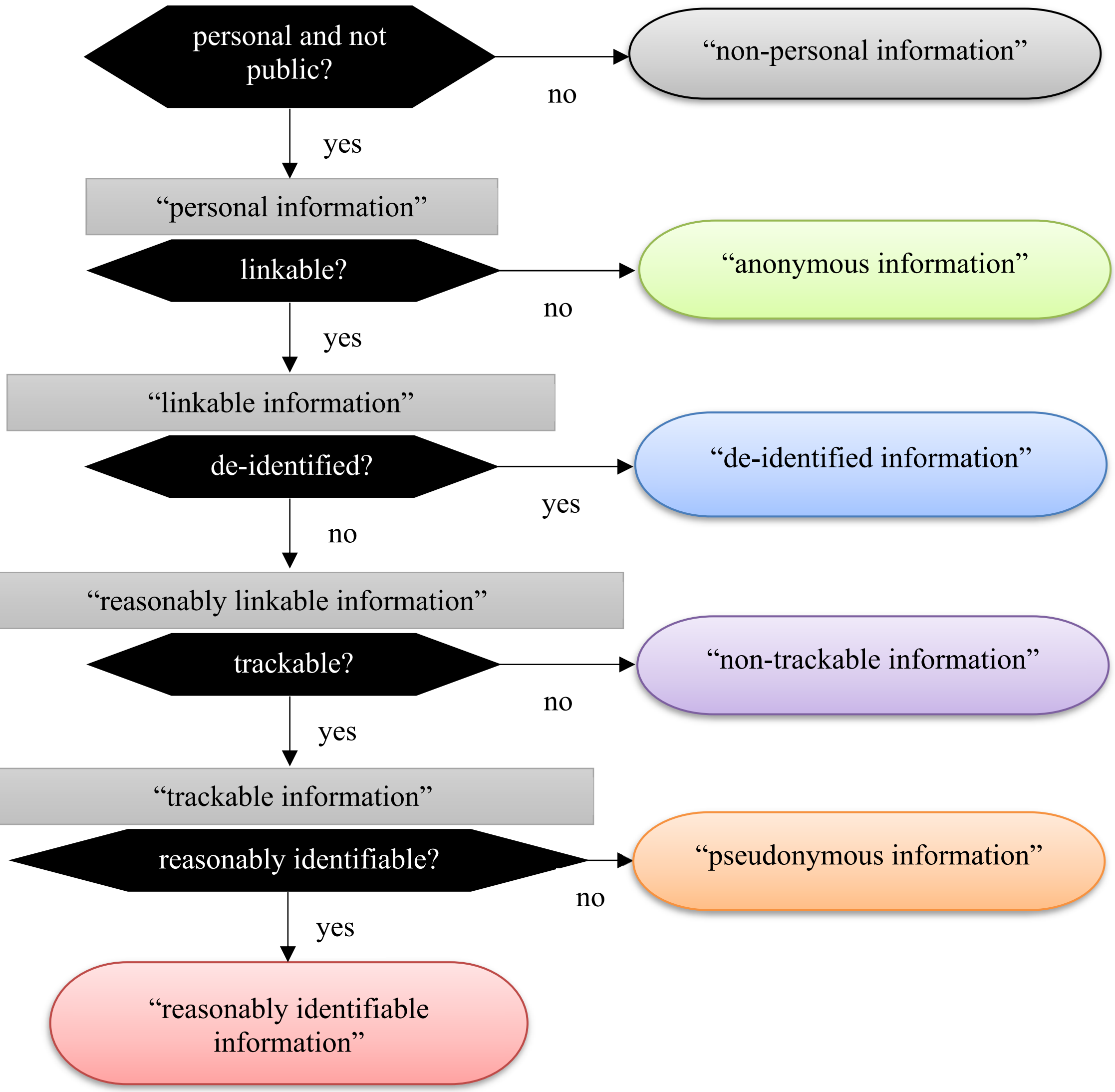

Figure 5. Classification of information.

The first question is whether information is both personal and not public. Information that is not personal would likely not be subject to regulation under a consumer privacy law.

The next question is whether information is linkable. Personal information qualifies as anonymous information if and only if there is *no possibility* of logical association with other information relating to the person or household to whom the personal information relates. There is little rationale for regulating anonymous information, since there is no privacy risk from the use or sharing of such information.

The next question is whether information is de-identified. Linkable information qualifies as de-identified information if and only if there is *no reasonable possibility* of logical association with other information relating to the person or household to whom the linkable information relates. Since legal controls are necessary to ensure that information remains de-identified, there should be notice requirements for such information. However, there are likely to be no choice requirements for such information.

The next question is whether information is trackable. Reasonably linkable information qualifies as non-trackable information if and only if there is no reasonable possibility of logical association of the

Electronic copy available at: https://ssrn.com/abstract=3893833

information with other information relating to the person or household *obtained from another context or another interaction* with the person or household qualifies as non-trackable information.

The last question is whether information is reasonably identifiable. Trackable information qualifies as pseudonymous information if and only the related person or household is not reasonably identifiable using that personal information and other reasonably linkable information. If not, it is classified as reasonably identifiable information.

These classifications of personal information may enable the creation of a choice framework that incentivizes the use of pseudonymous information instead of readily identifiable information, and that incentivizes the use of one-time identifiers and thereby reduces tracking. Neither the GDPR nor the CCPA incentivizes pseudonymization or disincentivizes tracking through their choice frameworks.

APPENDIX: DEFINITIONS

(1) ANONYMOUS INFORMATION.-- The term 'anonymous information' means personal information for which there is no possibility of logical association with other information relating to the person or household to whom the personal information relates.

(2) COLLECTION.-- The term 'collection' of personal information means access to personal information by any means, including but not limited to gathering, recording, storing, obtaining, receiving, buying, or renting.

(3) DE-IDENTIFIED INFORMATION.-- The term 'de-identified information' means linkable information for which there is no reasonable possibility of logical association with other information relating to the person or household to whom the linkable information relates, providing that the controller:

(A) takes reasonable measures to ensure that the information remains in a form in which there is no reasonable possibility of logical association with other information relating to the person or household to whom the linkable information relates,

(B) publicly commits to maintain and use the information only in a form in which there is no reasonable possibility of logical association with other information relating to the person or household to whom the linkable information relates,

(C) publicly commits to not attempt to associate the information with other information relating to the person or household to whom the linkable information relates, and

(D) contractually obligates any third parties to whom it discloses the information to implement controls (A), (B), and (C).

(4) LINKABLE INFORMATION.-- The term 'linkable information' means personal information that is not anonymous information.

(5) NON-TRACKABLE INFORMATION.-- The term 'non-trackable information' means reasonably linkable information for there is no reasonable possibility of logical association of the information with other information relating to the person or household obtained from another context or another interaction with the person or household, providing that the controller:

(A) takes reasonable measures to ensure that the information remains in a form in which there is no reasonable possibility of logical association of the information with other information relating to the person or household obtained from another context or another interaction with the person or household,

(B) publicly commits to maintain and use the information only in a form in which there is no reasonable possibility of logical association of the information with other information relating to the person or household obtained from another context or another interaction with the person or household,

Electronic copy available at: https://ssrn.com/abstract=3893833

(C) publicly commits to not attempt to associate the information with other information relating to the person or household obtained from another context or another interaction with the person or household, and

(D) contractually obligates any third parties to whom it discloses the information to implement controls (A), (B), and (C).

(6) PERSONAL INFORMATION.-- The term ‘personal information’ means any information relating to a natural person or to a household, excluding publicly available information.

(7) PSEUDONYMOUS INFORMATION.-- The term ‘pseudonymous information’ means trackable information for which the related person or household is not reasonably identifiable using that personal information and other reasonably linkable information, providing that the controller:

(A) takes reasonable measures to ensure that the information remains in a form in which the related person or household is not reasonably identifiable using that personal information and other reasonably linkable information,

(B) publicly commits to maintain and use the information only in a form in which the related person or household is not reasonably identifiable using that personal information and other reasonably linkable information,

(C) publicly commits to not attempt to identify the person or household to whom the information is related, and

(D) contractually obligates any third parties to whom it discloses the information to implement controls (A), (B), and (C).

(8) PUBLICLY AVAILABLE INFORMATION.-- The term ‘publicly available information’ means information relating to a natural person or to a household (a) in publicly available government records, (b) that the person or household to whom the personal information is related has made publicly available, or (c) that was made publicly available by widely distributed media.

(9) REASONABLY IDENTIFIABLE INFORMATION.-- The term ‘reasonably identifiable information’ means trackable information that is not pseudonymous information.

(10) REASONABLY LINKABLE INFORMATION.-- The term ‘reasonably linkable information’ means personal information for which there is a reasonable possibility of logical association with other information relating to the person or household to whom the personal information relates.

(11) TRACKABLE INFORMATION.-- The term ‘trackable information’ means reasonably linkable information that is not non-trackable information.

Electronic copy available at: https://ssrn.com/abstract=3893833