Washington Journal of Law, Technology & Arts

Volume 14 | Issue 2 Article 3

4-1-2019

Science and Privacy: Data Protection Laws and Their Impact on Research

Mike Hintze

WASHINGTON JOURNAL OF LAW, TECHNOLOGY & ARTS
VOLUME 14, ISSUE 2 SPRING 2019

SCIENCE AND PRIVACY: DATA PROTECTION LAWS AND THEIR IMPACT ON RESEARCH*

*Mike Hintze***

CITE AS: 14 WASH. J.L. TECH. & ARTS 103 (2019)

http://digital.law.washington.edu/dspace-law/bitstream/handle/1773.1/1877/14WJLTA103.pdf

TABLE OF CONTENTS

** Mike Hintze, Partner, Hintze Law PLLC; Affiliate Instructor of Law, University of Washington School of Law; and Senior Fellow, Future of Privacy Forum. Portions of this article were developed as part of a project funded by the Future of Privacy Forum through a grant from the Alfred P. Sloan Foundation. The views expressed in this article are my own and do not necessarily reflect the positions of any current or former employer or client.

INTRODUCTION

While privacy laws differ in their scope, focus, and approach, they all involve restrictions on the collection, use, sharing, or retention of information about people. In general, privacy laws reflect a societal consensus that privacy violations can lead to a wide range of financial, reputational, dignitary, and other harms, and that excessive collection and harmful uses of personal information should therefore be constrained. These laws require organizations to comply with a number of obligations concerning personal information.[1] In practice, these requirements can lead organizations to refrain from collecting certain data, only use data with the consent of the individual, or to delete data after a certain timeframe or at the request of the individual. Further, the global trend is toward both more and stricter privacy laws.

At the same time, scientific research is increasingly using the tools of data analytics and machine learning. These tools rely on "big data" and the idea that powerful computers and sophisticated analytical tools can examine very large data sets to reveal new insights and discoveries. Scientists believe this data-driven approach to research will lead to stunning breakthroughs in medicine, education, and many other fields that can dramatically advance human knowledge and well-being.

The tension between these two trends is clear. Most privacy laws acknowledge and address that tension. While privacy laws aim to restrict harmful data practices, they typically also are designed to allow for, or even encourage, uses of personal information that are beneficial and valuable to the individual or society. The inherent tension is often resolved by including reasonable exceptions in the laws to allow for necessary or beneficial data uses. But these

[1] For the sake of consistency, this article generally uses the term "personal information." Some of the laws it discusses use variations on that term, including "personally identifiable information" or "personal data," and those alternatives terms may be used when referring to a specific law.

exceptions are not complete exemptions from privacy obligations; even such beneficial uses of personal information typically remain subject to other protections in privacy laws such as an obligation to maintain the security of the data.

Protecting individual privacy is an important part of any use of personal information for research purposes. Organizations that collect, retain, use, or share personal information to advance scientific research should always handle that information with care, protect it from inadvertent disclosure or misuse, and be transparent about the use and protection of that data. But if privacy laws do not take into account and make allowances for the beneficial uses of personal information for research, the advancement of science, the expansion of knowledge, and the realization of new discoveries can be seriously impaired.

This article addresses how privacy laws can and should allow for scientific research while still providing meaningful protections for personal information. Part I discuses key principles found in many privacy laws and how each can potentially impact scientific research. Part II describes several prominent privacy laws across different jurisdictions and how each addresses research as a type of data use. Part III briefly discusses the distinction between academic or public-interest research and commercial research. Finally, Part IV provides specific recommendations to lawmakers and regulators on how privacy law should address and accommodate scientific research.

PART I: HOW PRIVACY LAW PRINCIPLES AFFECT RESEARCH

Privacy laws around the world reflect a common set of principles, frequently referred to as "fair information practice principles" or FIPPs. While there are different iterations of the FIPPs,[2] they generally include the concepts of transparency, data

[2] *See, e.g.*, U.S. Department of Health, Education and Welfare (HEW), Records, Computers, and the Rights of Citizens (1973), at 41, https://www.justice.gov/opcl/docs/rec-com-rights.pdf; OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data (1980), http://www.oecd.org/internet/ieconomy/oecd guidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm);

minimization, choice or consent, data access, and data security. Some or all of these principles are found in virtually every privacy law. A number of privacy laws have expanded on these core principles. Some have incorporated a right to data deletion as an extension of the right to data access.[3] Others have addressed the idea of de-identification as an extension of the principles of data minimization and data security.[4]

Some of these principles and the resulting legal obligations have little negative impact on the use and sharing of personal information for research purposes. Transparency and data security are prime examples, and applying them to the context of research makes good sense and provides important protections for personal information as it is being used for research purposes. However, other principles, such as consent and data deletion, can create significant obstacles to scientific research if not drafted with sensible exceptions. The following discussion addresses each of these principles in turn.

A. Transparency

Nearly every privacy law includes some type of transparency obligation. Often, laws set out specific notice obligations and list the types of information that must be disclosed to individuals from whom personal information is collected. Depending on the applicable law, required disclosures may include descriptions of:

- the categories of personal information collected,[5]

U.S. Federal Trade Commission (FTC), Privacy Online: A Report to Congress (1998), at 7, https://www.ftc.gov/sites/default/files/documents/reports/privacy-online-report-congress/priv-23a.pdf); Asia-Pacific Economic Cooperation (APEC), Privacy Framework (2005), https://www.apec.org/-/media/APEC/Publications/2005/12/APEC-Privacy-Framework/05_ecsg_privacyframewk.pdf).

[3] *See, e.g.*, 16 C.F.R. § 312.6(a)(2) (2019) (right of a parent to request deletion of personal information collected from a child); Council Regulation 2016/679, art. 17, 2016 O.J. (L 119) 1, 43 (EU) [hereinafter GDPR] (right of erasure).

[4] *See, e.g.*, GDPR, *supra* note 3 at 33 (definition of pseudonymisation); CAL. CIV. CODE § 1798.140(h) (definition of deidentified).

[5] California Business and Professions Code § 22575(b); GDPR, *supra* note 3 at 41, 43.

- the sources of personal information collected,[6]
- the intended uses of personal information,[7]
- the categories of third parties to whom personal information is disclosed,[8]
- how long the personal information is retained,[9] and
- a description of the rights that individuals have with respect to personal information, such as the right to access.[10]

These privacy notice disclosures are required whether or not personal information is used or shared for research purposes. An organization that intends to use or share data for research merely needs to state this intention in its privacy statement. Specifically, "research" should be included as one of the categories of data use, and academics and other researchers should be included among the categories of third parties to whom personal information is disclosed.

With regard to data retention, it is generally sufficient to describe the criteria used to determine retention timeframes, rather than necessarily having to list specific retention schedules in a privacy notice.[11] So, it may be useful to state one of the criteria as the period necessary to carry out legitimate scientific research.

Requiring organizations to be transparent about their use and sharing of personal information, including sharing information with researchers, is an important privacy protection and can promote greater accountability.[12] But these requirements do not themselves create any significant barrier to research. At most, they require organizations to consider and include research uses of data in the types of data use and sharing described in the organization's privacy statement.

---

[6] GDPR, *supra* note 3 at 42, 43.

[7] GDPR, *supra* note 3 at 40, 41, 43.

[8] California Business and Professions Code § 22575(b); GDPR, *supra* note 3 at 41, 43.

[9] GDPR, *supra* note 3 at 41, 42, 43.

[10] *Id.*

[11] *Id.*

[12] *See generally* Michael Hintze, *In Defense of the Long Privacy Statement*, 76 MD. L. REV. 1044 (2017).

B. Consent or choice

Many privacy laws require consent to collect, use, or share personal information. Depending on how the law is drafted, interpreted, and enforced, the required consent may be either explicit, such as requiring an affirmative opt-in choice, or implicit, implying consent based on a failure to opt-out or on some other basis. Implicit consent may be a fairly low bar. For example, it may be achieved merely by informing individuals of a particular data use and implying consent based on their continued use of a product or service. Increasingly, privacy laws are raising the bar for adequate consent, requiring explicit consent in some cases. However, obtaining explicit consent can be difficult; and in research, big data analytics, and machine learning scenarios, obtaining explicit consent may be impractical or impossible.

Furthermore, there is strong evidence that seeking and obtaining consent to have personal information included in a research study can result in a biased data sample and affect the outcome of the research.[13] An opt-in or explicit consent requirement is very likely to result in a non-representative sample and therefore bias the results of the research. But in many cases, even providing the ability to opt-out could significantly affect the data in undesirable ways.

Separate from the concerns regarding consent bias, researchers have also raised concerns that stringent requirements to obtain consent for accessing data for research purposes can lead to insufficient sample sizes, delay, and other costs that can interfere with efforts to produce timely and useful research results.[14]

Thus, the potential for consent requirements to negatively affect scientific research is quite high. Fortunately, as described in Part II, many privacy laws provide exceptions or alternatives to consent when it comes to using data for research.

---

[13] *See* Khaled El Emam et al., *A Review of Evidence on Consent Bias in Research*, 13 THE AMERICAN JOURNAL OF BIOETHICS, 42 (2013), https://www.tandfonline.com/doi/abs/10.1080/15265161.2013.767958; Michelle E. Kho et al., *Written Informed Consent and Selection Bias in Observational Studies Using Medical Records: Systematic Review*, BMJ 338:b866 (March 12, 2009), https://doi.org/10.1136bmj.b866.

[14] *See, e.g.*, Douglas B. McCarthy et al., *Medical Records and Privacy: Empirical Effects of Legislation*, 34 HEALTH SERVICES RESEARCH 417 (April 1999).

C. Right to Access

Another principle that is commonly reflected in privacy laws is an individual's right to access personal information about them being held by an organization. This right is closely related to the principle of transparency, as it enables an individual to learn not just what personal information the organization collects in general, but to know specifically what personal information the organization has. But because the right of access can result in specific information about an individual being released, it is important that the organization be able to authenticate and verify that the person making the request is the correct person to whom the data relates.

The right of access typically would not directly interfere with research uses of data, so the potential impact is relatively low. But like other privacy principles, it can add compliance obligations and resulting costs and overhead; organizations that use or share personal information for research purposes must take this into account.

For instance, while some examples of the right to access are limited to accessing the personal information itself, others give individuals the right to obtain other details, such as the third parties with whom personal information is shared. Under most privacy laws that include such requirements, limiting that disclosure to the *categories* of third parties will suffice.[15] In such cases, it will be quite easy for organizations that share personal information for research purposes to add "academic researchers" to their list of third-party recipients of personal information. This is similar to what should be done to meet transparency obligations.

But a few laws may require more granular disclosures to individuals who make an access requests, at least under certain circumstances. For example, under the newly-enacted California Consumer Privacy Act (CCPA), if the sharing of personal information with a researcher could be characterized as a "sale," a response to an access request may need to be more specific. Rather than just stating a category of third-party recipients, the response to

[15] *See, e.g.*, GDPR, *supra* note 3 at 43 ("[T]he recipients or categories of recipient to whom the personal data have been or will be disclosed").

the access request would need to be clear about whether or not personal information about the specific requesting individual has been shared with researchers, and/or identify the specific third parties that have received the individual's personal information.[16] Thus, organizations subject to such requirements should track and documents whose information is shared for research purposes and which third-party researchers or research organizations have received it, and be able to produce that data upon request.

Finally, because an individual's right to access can be asserted against recipients of personal information in many cases, researchers receiving the data could be subject to access requests. These data–receiving organizations should be prepared to respond to access requests under the applicable laws.

D. Right to delete

While less common than a right to access, an increasing number of privacy laws have added a right to delete, enabling individuals to request that organizations delete personal information about them.[17] Similar to the potential problems described relating to consent, an unchecked right to delete can create serious barriers to research.

As with consent, enabling individuals to selectively remove themselves from research by deleting their personal information can cause bias in the data sample, lead to smaller sample sizes, and impose additional costs and burdens on research.

For example, if a study is underway and one or more individuals demand removal of their personal information that is part of that study, that research could be derailed. Further, even if the deletion could be delayed until the completion of the study, it could still harm scientific research. A hallmark of good science is that results be testable and replicable. However, if individuals are able to later

---

[16] CAL. CIV. CODE § 1798.115(a)(2) ("The categories of personal information that the business sold about the consumer and the categories of third parties to whom the personal information was sold, by category or categories of personal information *for each third party* to whom the personal information was sold.") (emphasis added).

[17] GDPR, *supra* note 3 at 43-44 (EU); CAL. CIV. CODE § 1798.105 (California Consumer Privacy Act of 2018); 34 C.F.R. § 312.6(a)(2), 78 Fed. Reg. 4,012 (January 17, 2013) (giving parents the right to request deletion of personal information collected online from children under the age of thirteen).

remove their personal information from a data set that underlies a study, the ability to test and replicate that study could be significantly undermined.

Most privacy laws that include a right to delete include a number of exceptions, including where the data is needed for research. Such exceptions are essential to enable beneficial uses of personal information for scientific research.

E. Data Minimization

Data minimization is a somewhat amorphous principle that encompasses several ideas. One aspect of data minimization is the idea of collection limitation—don't collect more personal information than is reasonably needed to accomplish the purpose(s) for which it is collected. Another aspect is retention limitation—don't keep data longer than is reasonably needed to accomplish the purpose(s) for which it was collected. A related concept is that data can and should be made less sensitive through techniques such as de-identification (e.g., removing, masking, or altering data elements that can identify an individual) or reducing the precision of data (such as converting a precise GPS location point to zip-code level data).

Many privacy laws include some notion of data minimization.[18] This principle creates obvious tensions with the big data analytics and machine learning that underlie much scientific research today. The promise of big data is that applying massive computing power to very large data sets can reveal unexpected patterns, correlations, and connections within the data and result in surprising new insights and discoveries. At least in theory, the more data the researchers have, the more unanticipated breakthroughs are likely to emerge. The combination of the ideas that "more data is better" and that the outcomes are unknowable until the research occurs make it difficult to apply legal principles that suggest less data is better and that the purposes of the data should be established up front.

But this is a tension, not necessarily a conflict. As with the transparency principle, it is important to articulate up front, at least

---

[18] GDPR, *supra* note 3, arts. 5(1)(c) (collection limitation), 5(1)(e) (retention limitation) at 35-36 (EU); 34 C.F.R. § 312.7 (collection limitation), § 312.10 (retention limitation), 78 Fed. Reg. 4,012 (January 17, 2013).

in general terms, that research is one of the anticipated purposes for the data. By establishing this purpose at the outset, the amount of data collected, the length of retention, and the level of precision and identifiability maintained can take the anticipated research purposes into account. Beyond that, organizations should still be thoughtful about the data they collect and retain. If data is stale or unreliable, and as a result very unlikely to be useful for research, it should not be retained. If some level of de-identification is compatible with the research uses of the personal information, the compatible de-identification should be employed.

As with other aspects of data protection, data minimization can be thought of as a risk reduction principle and should be considered carefully along with other compliance measures. It is likely inevitable that any application of data minimization could potentially have some negative impact on research, but if carefully employed, an appropriate balance can be found and those impacts may be limited to the margins. To enable that balance, data minimization should be reflected in privacy laws as a flexible concept that can allow for broad use of data for scientific research.

F. Data Security

Another common element among privacy laws is an obligation to implement appropriate data security measures.[19] Typically (and ideally), data security obligations are drafted to create a flexible set of requirements that are not overly proscriptive and take into account the context, nature, and sensitivity of the personal information in question.

Responsible organizations that handle large amounts of personal information typically invest heavily in data security for their core systems, such as firewalls, access controls, encryption, and intrusion detection. But research uses of data may involve moving data into different systems in a different environment such as a university or a research lab. When using or sharing personal information for

---

[19] GDPR, *supra* note 3 at 51-52 (EU); Health Insurance Portability and Accountability Act of 1996 (HIPAA), Public Law 104-191 (Security Rule at 45 C.F.R. Part 160 and Subparts A and C of Part 164); 34 C.F.R. § 312.8, 78 Fed. Reg. 4,012 (January 17, 2013).

research purposes, it is important to ensure that all systems on which the data is stored maintain the appropriate level of data security. Doing so may require the performance of due diligence to ensure the destinations systems are secure, and if those systems fall short, there will likely be additional costs incurred from upgrading existing security protections or finding alternative data storage arrangements. But those are steps that should be taken in any case, and the costs are likely to pale in comparison to the costs of a security breach affecting personal information.

Thus, while data security requirements can create costs, those costs should be seen as prudent investments, and the requirements do not themselves create a legal barrier to using or sharing data for research purposes. They are an appropriate and important part of any privacy law.

G. De-identification

De-identification is a process by which personal information is manipulated in ways designed to make it more difficult to subsequently re-identify an individual from that data. It can involve a wide range of techniques. And there is a range of resulting strengths of de-identification, from relatively weak methods where the risk of re-identification is high, to very strong methods where the data can be considered irreversibly anonymized.

De-identification is relevant for compliance with nearly every privacy law. In some cases, de-identification is not explicitly mentioned, but it is implicit in the scope of the law. Most privacy laws are scoped to a defined set of "personal information." And if data can be strongly de-identified to the extent it no longer meets the law's definition of personal information, it will almost always fall outside the scope of the privacy law. As a result, research on strongly de-identified data can typically proceed without further privacy compliance obligations.

A number of privacy laws explicitly incorporate notions of de-identification as a way to meet some or all of the law's requirements. For instance, the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule includes a detailed set of criteria regarding de-identification that enable de-identified data to fall

outside the scope of the Rule.[20] In Europe, the GDPR recognizes a range of de-identification methods and strengths and provides regulatory incentives for different levels of de-identification.[21]

But while de-identification is included in a number of privacy laws, none have a blanket requirement that data be de-identified. Such a blanket rule would be unworkable, because there are many contexts in which de-identification of data would be incompatible with the intended uses of the data. And as a general rule, the stronger the de-identification, the lower the utility of the data. With some research, de-identification can be a practical and advisable option, but not in all cases.

Rather than absolute rules, privacy laws typically create incentives to de-identify data, and thereby encourage the use of techniques that can reduce risk in context where de-identification is a practical option. When properly drafted, these laws can encourage best practices to reduce privacy risks, while not imposing barriers to scientific research.

H. Summary

The following chart summarizes the likelihood of negative impacts on scientific research for each of these principles.

Principle	Likelihood of Negative Impact
Transparency	Low
Consent or choice	High
Right to access	Low
Right to delete	High
Data minimization	Medium
Data security	Low

[20] 45 C.F.R. § 164.514(a)–(b) (2018).

[21] GDPR, *supra* note 3, Recital 26 at 5 (EU) (discussing concepts of personal data and anonymous data); *id.* arts. 4(1) (definition of personal data), 4(5) (definition of pseudonymization) at 33; *id.* art. 11 (processing which does not require identification) at 39. *See* Michael Hintze, *Viewing the GDPR through a De-Identification Lens: A Tool for Compliance, Clarification, and Consistency*, 8 INT'L DATA PROTECTION L. 86 (2018), https://ssrn.com/abstract=2909121.

De-identification	Low

PART II: HOW EXISTING PRIVACY LAWS TREAT RESEARCH

As suggested in Part I above, different privacy laws treat research uses of personal information differently. Some address research in a thoughtful and flexible manner, creating appropriate exceptions for research that allow and encourage scientific research to flourish. Others, either by design or oversight, have been less friendly to research. In this Part, several prominent privacy laws are examined in terms of how they treat scientific research.

A. Europe: General Data Protection Regulation (GDPR)

The General Data Protection Regulation (GDPR) is a comprehensive privacy law that applies across the European Union (EU) and the European Economic Area (EEA). It also has been highly influential globally, with countries around the world adopting privacy laws based on the European model.

The GDPR incorporates the full range of principles discussed in Part I of this article, including data minimization, consent, and deletion rights, all of which can cause problems for beneficial research uses of personal data. However, the drafters of the GDPR included thoughtful exceptions and allowances for research that enable personal data to be used and shared for such purposes without serious impediment.

Consent is addressed in the GDPR in a unique way. The law sets a very high bar for consent. Under the GDPR, consent must be "freely given, specific, informed and unambiguous" and manifested through "a statement or by a clear affirmative action" indicating the data subject's agreement.[22] Obviously, obtaining such consent would create a significant barrier for research uses of personal data. However, consent is not always required. Rather, under the GDPR, processing personal data requires a legal basis.[23] There are several different legal bases available under the law, only one of which is

[22] GDPR, *supra* note 3 at 34 (EU).
[23] *Id.* art. 6 at 36.

the consent of the data subject.[24] Other common legal bases include where the processing is "necessary for the performance of a contract to which the data subject is party,"[25] or if the "legitimate interests" of the data controller or a third party outweighs the interests or rights of the data subject.[26]

Further, the GDPR specifies that if there is a legal basis for the original purpose of collecting the personal data, a secondary use of that data need not have a separate legal basis if that use is "compatible" with the purpose for which the data was collected. Recital 50 of the GDPR helpfully notes that scientific research is such a compatible purpose:

> The processing of personal data for purposes other than those for which the personal data were initially collected should be allowed only where the processing is compatible with the purposes for which the personal data were initially collected. In such a case, no legal basis separate from that which allowed the collection of the personal data is required . . . . Further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes should be considered to be compatible lawful processing operations.[27]

This language is reflected in Article 5 which provides that:

> [Personal data shall be] collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ('purpose limitation').[28]

---

[24] *Id.* art. 6(1)(a) at. 36.
[25] *Id.* art. 6(1)(b) at 36.
[26] *Id.* art. 6(1)(f) at 36.
[27] *Id.* Recital 50 at 9.
[28] *Id.* art. 5(1)(b) at 35.

The reference to Article 89(1) suggests that there are certain conditions that must be met in order for scientific research to be considered a compatible purpose, requiring no additional legal basis. Those conditions come down to there being "appropriate safeguards" in place designed to protect privacy and individual rights:

> Processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, shall be subject to appropriate safeguards, in accordance with this Regulation, for the rights and freedoms of the data subject. Those safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner. Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner.[29]

Likewise, Article 6(4) re-iterates the need for "appropriate safeguards" where the secondary use is based on the use being "compatible" with the original purpose of collection:

> Where the processing for a purpose other than that for which the personal data have been collected is not based on the data subject's consent [or a legal requirement], the controller shall, in order to ascertain whether processing for another purpose is compatible with the purpose for which the personal data are initially collected, take into account, inter alia . . . the possible consequences of the intended further processing for data subjects [and] the existence of appropriate safeguards, which may include encryption or pseudonymisation.[30]

---

[29] *Id.* art. 89(1) at 84.
[30] *Id.* art. 6(4) at 37.

There are conflicting readings of some of the relevant provisions cited above, particularly with respect to scenarios where the original legal basis for collecting and using the data was the consent of the data subject. One potential issue arises from the fact that where the original legal basis is consent, the GDPR gives the data subject the right to withdraw consent;[31] if this right is exercised, the compatibility analysis may no longer be available. However, the withdrawal of consent is not retroactive, so at most it would affect only future research that commences after the withdrawal of consent for the original purpose. The European Commission published brief (and non-binding) guidance that goes even further, suggesting that the compatibility analysis for secondary uses of data is never allowed where the original legal basis is consent,[32] however that conclusion appears to be based on a misreading of Article 6(4), quoted above.[33]

---

[31] *Id.* art. 7(3) at 37.

[32] *See* Eur. Comm'n, *Can We Use Data for Another Purpose?*, https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/principles-gdpr/purpose-data-processing/can-we-use-data-another-purpose_en (last visited Apr. 14, 2019) ("If your company/organisation has collected data on the basis of **legitimate interest**, a **contract** or **vital interests** it can be used for another purpose but only after checking that **the new purpose is compatible with the original purpose**. . . . If your company/organisation has collected the data on the basis of **consent or following a legal requirement**, no further processing beyond what is covered by the original consent or the provisions of the law is possible. Further processing would require obtaining new consent or a new legal basis.").

[33] Article 6(4) is a long sentence and a bit difficult to parse on an initial reading. But the subject of the sentence is "processing for a purpose other than that for which the personal data have been collected" – in other words a "secondary purpose." So that provision says, in effect, "Where [a secondary purpose] is not based on [consent or a legal requirement], the controller shall, in order to ascertain whether [the secondary purpose] is compatible with the [primary purpose], take into account . . ." In other words, unless the secondary purpose has a legal basis of consent or is a legal requirement, it is necessary to do the compatibility analysis. Conversely, if a secondary use is based on consent or a legal requirement, it is not necessary to do a compatibility analysis. That conclusion makes sense and reflects a sound policy. Conversely, Article 6(4) does not appear to say, as the Commission guidance seems to suggest, that if the collection and primary use is based on consent or a legal requirement, the compatibility analysis is unavailable or irrelevant, and any secondary use is prohibited unless there is a separate legal basis. Such a surprising conclusion does

Regardless, even if a separate legal basis is required for a secondary research use of personal data, the GDPR also offers a practical solution with the availability of "legitimate interests" as a legal basis. This legal basis is available where the "processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data."[34] Thus, relying on the legitimate interests basis for scientific research requires balancing the benefits and interests in doing the research against the risks to the rights and freedoms of the individual. The benefits side of the equation is not limited to the interests of the organizations using or sharing the personal information for research purposes, but can also take into account the interests of any third party—including the public interest and benefits of the research. And on the risk side of the equation, unlike uses of data that directly impact the data subject, such as marketing, advertising, personalization, or other individualized decision-making, research typically does not have a direct impact on the individual data subjects. Moreover, if additional privacy protections are in place, such as the "appropriate safeguards" noted above, the balancing test will almost always come out in favor of being able to rely on legitimate interests as a legal basis for using personal data to conduct scientific research.

Based on these provisions, scientific research is almost certain to be considered a purpose "compatible" with the purpose(s) for which the data was originally collected, requiring no additional legal basis. To the extent that an additional legal basis may be needed, scientific research is almost certain to be eligible for the legitimate interests basis. Both of those approaches, however, require that appropriate safeguards be in place to protect the privacy and rights of the individuals whose personal data is being used.

The references to "appropriate safeguards" include de-identification as a key example of the measures organizations engaged in research can take to protect privacy and the rights of the data subjects. But it is not a mandate. Rather, de-identification need

---

not seem to be supported by the text of the law.

[34] GDPR, *supra* note 3 at 36.

only be applied to the extent compatible with the research needs. But if de-identification would reduce the utility of the data for the intended research purpose, it need not be applied (but other safeguards, such as strong security, should be in place). In any case, the organizations involved in research need to demonstrate that they have applied "appropriate safeguards" and that the likelihood of negative consequences on the data subject is low.

The GDPR also includes data minimization principles, including retention limitations which may be in tension with the idea that researchers need to gather and retain large volumes of data to conduct big data analytics tools and machine learning. However, the retention limitation principle in the GDPR includes a specific carve-out for research:

> Personal data shall be: . . . kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject.[35]

Finally, the GDPR provides for a limited right for individuals to request the deletion of personal data.[36] But that right only applies under certain specified instances. One instance is when the data subject withdraws consent and there is no other legal ground for processing.[37] Thus, if the research is based on the consent of the data subject (or if it is based on being compatible with the original purpose of collection, which was based on the consent of the data subject), and the data subject subsequently withdraws that consent, the data subject may have a right to have the data deleted. However, the right to delete in that case would apply only if there is no other

---

[35] *Id.* art. 5(1)(e) at 36.
[36] *Id.* art. 17 at 43.
[37] *Id.* art. 17(1)(b) at 44.

legal basis for continued processing. As discussed above, scientific research is very likely to have a separate legal basis of legitimate interests. But more importantly, the exceptions to the right of erasure specifically reference research uses of data:

> Paragraphs 1 and 2 shall not apply to the extent that processing is necessary: . . . (d) for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) in so far as the right referred to in paragraph 1 is likely to render impossible or seriously impair the achievement of the objectives of that processing.[38]

Thus, while research uses of personal data do not create an absolute exception to the right of erasure, a data processor can refuse an erasure request following a withdrawal of consent if it can either (1) establish an alternate legal basis, such as legitimate interests, or (2) demonstrate that deletion of the data related to the data subject will seriously impair the research objective.

Thus, while the GDPR includes principles such as consent, data minimization, and a right to delete that can potentially impede research uses of data, it also provides flexibility and exceptions that should allow research to flourish. Where obtaining explicit consent from each individual data subject is often impractical and could undermine the statistical validity of outcomes, the GDPR provides practical alternatives to consent. Data controllers conducting research on data have a strong case under the GDPR for relying on a legal basis other than consent, such as legitimate interests. Alternatively, data controllers need not have an additional legal basis at all if the secondary purpose is "compatible" with the original purpose of collection, and the GDPR specifies that research is a compatible purpose. And research is called out explicitly as an exception to retention limitation and the right to delete.

B. Canada: Personal Information Protection and Electronic Documents Act (PIPEDA)

---

[38] *Id.* art. 17(3) at 44.

In Canada, PIPEDA is another comprehensive privacy law that includes a range of principles including transparency, retention limitation, right of access, and data security. It generally requires notice and consent for uses and disclosures of personal information.[39] However, the law permits disclosures of personal information for "scholarly study or research" without notice or consent if:

- the purposes cannot be achieved without the use or disclosure of the information,
- it is impracticable to obtain consent, and
- the organization provides prior notice to the Privacy Commissioner of Canada.[40]

The exemption from consent requirements for research is common in privacy laws, but the requirement to inform the regulator when an organization seeks to use that exemption is unique to PIPEDA.

The retention limitation principle in PIPEDA is tied to the transparency principle. It provides that "[p]ersonal information that is no longer required to fulfil the identified purposes should be destroyed, erased, or made anonymous."[41] Thus, if research is one of the identified purposes, the retention limitation provision should allow for the retention of data for as long as needed for research.

C. United States: Health Insurance Portability and Accountability Act (HIPAA)

The HIPAA Privacy Rule applies to protected health information (PHI), and provides detailed regulations reflecting a broad range of principles. It also specifically addresses the use and

---

[39] Personal Information Protection and Electronic Documents Act, S.C. 2000, c 5 sch 1 cl 4.3 (Can.).

[40] *Id.* s 7(3)(f); *see also id.* s 7(2)(c) (outlining a similar exception to notice and consent for uses of personal information for research purposes, which includes the same three conditions that apply to disclosures, plus that the information be "used in a manner that will ensure its confidentiality.").

[41] *Id.* sch 1 cl 4.5.3.

disclosure of PHI for research purposes.[42]

The HIPAA Privacy Rule defines research fairly broadly as "a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge."[43] Notably, it does not limit research to medical or health research.

PHI can be used for research purposes under several different circumstances. First, if the data to be used for research meets the HIPAA standards for de-identification (based either on the safe harbor method or the expert determination method), then the data is no longer PHI subject to the Privacy Rule and it can be freely disclosed for research (or any other) purposes.[44]

Second, PHI can be used or disclosed for research purposes with the individual's authorization (i.e. with consent).[45]

Third, PHI can be used or disclosed for research purposes without the individual's authorization if there is a documented waiver approved by an Institutional Review Board (IRB) or Privacy Board.[46] The IRB or Privacy Board can be that of the covered entity making the disclosure of PHI, the recipient researcher, or an independent board.[47]

Fourth, limited sets of PHI (which means that certain direct identifiers have been removed, although the data is not fully de-identified as defined by the Rule) can be disclosed to a researcher

---

[42] A useful overview of how HIPAA addresses research uses of personal information can be found at: https://www.hhs.gov/hipaa/for-professionals/special-topics/research/index.html.

[43] 45 C.F.R. § 164.501 (2018).

[44] *Id.* § 164.514(a)-(c).

[45] *See generally id.* § 164.508 (regarding individual authorization for uses or disclosures of PHI); *id.* § 164.508(b)(3)(i) (permitting authorization for research purposes to be combined under certain circumstances); *id.* § 508(b)(4)(1) (allowing a covered health care provider to "condition the provision of research-related treatment on provision of an authorization for the use or disclosure of protected health information for such research.").

[46] *Id.* § 164.512(i)(1).

[47] *See* U.S. Dep't of Health & Hum. Serv., *HIPAA FAQs for Professionals*, https://www.hhs.gov/hipaa/for-professionals/faq/310/does-hipaa-require-a-covered-entity-to-create-an-irb-or-privacy-board/index.html (last visited Apr. 14, 2019).

pursuant to a data use agreement.[48]

Additionally, the HIPAA Privacy Rule also allows the use and disclosures of PHI of decedents for research purposes.[49] The Rule also allows for researchers to access PHI (including remote access) for purposes that are preparatory to research (such as designing a study or preparing a research protocol).[50]

Finally, certain disclosures of PHI for research purposes are subject to an individual's right to receive an accounting of such disclosures that occurred over the last six years. Thus, covered entities must ensure that such disclosures are well-documented in a way that would enable them to respond to these types of requests from individuals.

The level of specific detail regarding research uses and disclosures of data in HIPAA is well beyond that found in any other privacy law. Despite the rigor of these requirements, and the sensitivity of the personal information involved, they provide significant flexibility to enable scientific research. As a result, some elements of the HIPAA privacy rule may be useful to consider in using personal information for research beyond the health context.

D. United States: Family Educational Rights and Privacy Act (FERPA)

The Family Educational Rights and Privacy Act is a U.S. federal law that applies to schools that receive federal funding.[51] It includes several privacy principles that apply to student records. FERPA gives parents (and students who are 18 years old or over) the right to access student records, and it requires their consent for certain disclosures of educational records.

However, consent is not always required. Under FERPA,

---

[48] 45 C.F.R. § 164.514(e) (2018). Templates for such data use agreements are available from various sources. For example, the Health Care Systems Research Network (HCSRN) has published a tool kit and templates for data use agreements at http://www.hcsrn.org/en/Tools%20&%20Materials/GrantsContracting/.

[49] *Id.* § 164.512(i)(1)(iii).

[50] *Id.* § 164.512(i)(1)(ii).

[51] Useful information and resources related to FERPA can be found at https://ed.gov/policy/gen/guid/fpco/ferpa/ and https://ferpasherpa.org/.

schools and other educational institutions can disclose personal information related to students for research purposes under any of the following conditions:

- With the consent of the parent—or the consent of the student if the student is 18 years old (or older) or attending a post-secondary education institution.[52]
- The personally identifiable information is limited to "directory information."[53]
- The personally identifiable information is de-identified through the "removal of all personally identifiable information" and a "reasonable determination" that no student is identifiable from the data alone or in combination with "other reasonably available information."[54]
- The personally identifiable information is de-identified through key-coding (i.e. a reversable method of de-identification) if the purpose of the research is "education research."[55]
- The disclosure to a third-party organization conducting studies, *on behalf of* educational agencies or institutions and pursuant to a *written agreement*, designed to (A) develop, validate, or administer predictive tests; (B) administer student aid programs; or (C) improve instruction.[56]

The written agreement required for the "studies exception" must include certain specific terms as follows:

(1) Specifies the purpose, scope, and duration of the study

---

[52] 34 C.F.R. § 99.30 (2018).

[53] *Id.* §§ 99.31(a)(11), 99.37. "'Directory information' means information contained in an education record of a student that would not generally be considered harmful or an invasion of privacy if disclosed. . . . Directory information includes, but is not limited to, the student's name; address; telephone listing; electronic mail address; photograph; date and place of birth; major field of study; grade level; enrollment status (e.g., undergraduate or graduate, full-time or part-time); dates of attendance; participation in officially recognized activities and sports; weight and height of members of athletic teams; degrees, honors and awards received; and the most recent educational agency or institution attended." 34 C.F.R. § 99.3.

[54] *Id.* § 99.31(b)(1).

[55] *Id.* § 99.31(b)(2).

[56] *Id.* § 99.31(a)(6).

or studies and the information to be disclosed;

(2) Requires the organization to use personally identifiable information from education records only to meet the purpose or purposes of the study as stated in the written agreement;

(3) Requires the organization to conduct the study in a manner that does not permit personal identification of parents and students, as defined in this part, by anyone other than representatives of the organization with legitimate interests; and

(4) Requires the organization to destroy or return to the educational agency or institution all personally identifiable information when the information is no longer needed for the purposes for which the study was conducted and specifies the time period in which the information must be returned or destroyed.[57]

These exceptions to consent for certain research purposes under FERPA are narrower than a general research exception. Disclosures for "education research" purposes are given more favorable treatment. But research for other purposes can occur if, for example, the student records are strongly de-identified.

E. United States: Children's Online Privacy Protection Act (COPPA)

COPPA is another U.S. federal privacy law that applies to personal information collected from children.[58] It requires operators of online services to obtain verifiable parental consent prior to the online collection, use, or disclosure of personal information from children under the age of thirteen.[59] It gives parents the right to access and delete children's personal information.[60] And it includes requirements for data minimization (collection limitations),

---

[57] *Id.* § 99.31(a)(6)(iii)(C).

[58] 15 U.S.C. §§ 6501–6506 (Pub.L. 105–277, 112 Stat. 2681-728, enacted October 21, 1998), implementing regulations at 16 C.F.R. Part 312.

[59] 16 C.F.R. § 312.5(a) (2019), 78 Fed. Reg. 4,011 (January 17, 2013).

[60] *Id*. § 312.6, at 4,012.

transparency, and reasonable security procedures.[61] COPPA is unusual among privacy laws in that it does not have any particular carve-out for uses and/or disclosures of personal information for research purposes.

Prior to 2013, the impact of COPPA on research would have been lower due to the fact that it had a relatively narrow definition of "personal information."[62] Thus, through reasonable de-identification, data could be taken outside the scope of personal information as defined by COPPA, and therefore no longer subjected to its requirements, including parental consent and the right to delete. But through its rulemaking authority, the Federal Trade Commission expanded the definition of personal information in 2013 to include, among other data types, any persistent identifier "that can be used to recognize a user over time and across different website or online services;" "a photograph, video, or audio file, where such file contains a child's image or voice;" and precise geolocation information.[63] This broader definition of personal information means that it will be more difficult to de-identify the data such that it is no longer within the scope of the law.

The rigor of the parental consent requirement under COPPA, and the lack of an exception for research, undoubtedly reflects the notion that higher levels of privacy protections are appropriate when it comes to young children. However, the downside of this approach is that it could impair scientific research that could benefit children.

---

[61] *Id*. § 312.7, at 4,012 (collection limitations), § 312.4, at 4,010 (transparency), § 312.8, at 4,012 (security).

[62] "Personal information means individually identifiable information about an individual collected online, including: (a) A first and last name; (b) A home or other physical address including street name and name of a city or town; (c) An e-mail address or other online contact information, including but not limited to an instant messaging user identifier, or a screen name that reveals an individual's e-mail address; (d) A telephone number; (e) A Social Security number; (f) A persistent identifier, such as a customer number held in a cookie or a processor serial number, where such identifier is associated with individually identifiable information; or a combination of a last name or photograph of the individual with other information such that the combination permits physical or online contacting; or (g) Information concerning the child or the parents of that child that the operator collects online from the child and combines with an identifier described in this definition." 16 C.F.R. § 312.2 (2019), 64 Fed. Reg. 59,912 (November 3, 1999).

[63] 16 C.F.R. § 312.2 (2019), 78 Fed. Reg. 4,009 (January 17, 2013).

Organizations that wish to use or disclose personal information collected from children under the age of thirteen will have to obtain parental consent. Thus, research uses and disclosures should be included in the scope of consent obtained from parents with the original collection of the personal information.

An additional complication with regard to getting consent for research purposes is that COPPA also includes a requirement that parents are offered the opportunity to provide consent to the collection and use of personal information, but not to the disclosure of such information.[64] This "limited consent" option means that if an organization wishes to share with researchers the personal information collected from children under thirteen, it cannot combine the parental consent for such sharing with the parental consent for the original collection—or at least it must offer the parents to option to opt-out from the sharing of the personal information.

F. California Consumer Privacy Act (CCPA)

The new California Consumer Privacy Act (CCPA) will come into effect on January 1, 2020. Many critics have raised concerns regarding the numerous ambiguities and inconsistencies in the law that have created great uncertainty about how the CCPA will be interpreted and enforced. There is some hope that the statute will be further amended to clarify certain aspects of the law or that the California Attorney General will develop regulations or guidelines that will increase clarity. Nevertheless, significant ambiguity is likely to remain for the foreseeable future, and despite those problems, companies need to take steps to attempt to come into compliance the best they can by the end of 2019.

In contrast to the GDPR, the CCPA does not require consent or an alternate legal basis for all collection and use of personal information. Nor does it have a general data minimization obligation. Instead, it is focused primarily on giving consumers the right to opt-out from "sales" of personal information,[65] giving users

---

[64] 16 C.F.R. § 312.5(a)(2) (2019).

[65] CAL. CIV. CODE § 1798.120.

broad data access, correction, and deletion rights,[66] and imposing additional transparency obligations on companies.[67] Thus, there is no general obligation on companies to get consent or establish an alternative basis for sharing data with researchers—unless that sharing could be characterized as a "sale" of data.

However, if the transfer of data for research purpose *is* considered a "sale" of personal information, that transfer will be highly regulated and the CCPA will impose a number of proscriptive obligations. The definition of "sale" is very broad and includes any transfer of data for "consideration" (which can be interpreted to cover many or most commercial transactions involving data transfers).[68] The breadth of that definition puts at least some data sharing for research purposes at risk, particularly where a commercial entity is sharing data and expects to obtain some commercial benefit from the resulting research. Other sharing of personal information for research purposes would likely not cross the line into being a data "sale," particularly if the sharing is with academic researchers and there is not a direct commercial benefit expected. Thus, the risk of the sharing being deemed a "sale" is a factor worth considering when making arrangements to share data for research purposes, and it may be worth explicitly stating in the data sharing agreement that no consideration is being provided for the transfer of the data.

The CCPA also includes a right for individuals to request the deletion of personal information.[69] The impact of this right is likely to be quite limited, however, because there are many exceptions such that companies will be able to decline a request to delete information in most cases. For instance, personal information can be retained, despite a deletion request from an individual, if the information is necessary to provide a service requested or reasonably anticipated by the individual,[70] for security purposes,[71]

---

[66] *Id*. §§ 1798.105, 1798.110, 1798.115.

[67] *Id*. §§ 1798.100(b), 1798.105(b), 1798.110(c), 1798.115(c), 1798.130(a)(5), 1798.135(a)(2).

[68] *Id*. § 1798.140(t).

[69] *Id*. § 1798.105.

[70] *Id.* § 1798.105(d)(1).

[71] *Id.* § 1798.105(d)(2).

or to use the data internally for purposes that are "reasonably aligned with the expectations of the consumer based on the consumer's relationship with the business"[72] or "compatible with the context in which the consumer provided the information."[73] It is hard to imagine a scenario in which a business couldn't credibly claim that one or more of those exceptions apply. And if the personal information can be retained for one of those broad purposes, there is nothing in the CCPA preventing that retained information from also being used for research purposes.

It is worth noting that in addition to those broad exceptions to the right to delete, there is also a very narrow exception that is specific to certain types of scientific research. But it is so narrow and confusingly drafted that it is less likely to be useful for research than the broader, more general exceptions. Specifically, the research exception to the right of deletion only applies to research that is:

> [P]ublic or peer-reviewed scientific, historical, or statistical research in the public interest that adheres to all other applicable ethics and privacy laws, when the businesses' deletion of the information is likely to render impossible or seriously impair the achievement of such research, if the consumer has provided informed consent.[74]

Further, "research" is a defined term under the CCPA, and the definition reinforces the idea that it must be for the public interest. And it also sets out several other (somewhat redundant) criteria, including that the research be for non-commercial purposes and that the data must be de-identified:

> "Research" means scientific, systematic study and observation, including basic research or applied research that is in the public interest and that adheres to all other applicable ethics and privacy laws or studies conducted in the public interest in the area of public health. Research with personal information that may have been collected from a consumer in the course of the consumer's interactions with

---

[72] *Id.* § 1798.105(d)(7).
[73] *Id.* § 1798.105(d)(9).
[74] *Id.* § 1798.105(d)(6).

> a business's service or device for other purposes shall be:
>
> (1) Compatible with the business purpose for which the personal information was collected.
>
> (2) Subsequently pseudonymized and deidentified, or deidentified and in the aggregate, such that the information cannot reasonably identify, relate to, describe, be capable of being associated with, or be linked, directly or indirectly, to a particular consumer.
>
> (3) Made subject to technical safeguards that prohibit reidentification of the consumer to whom the information may pertain.
>
> (4) Subject to business processes that specifically prohibit reidentification of the information.
>
> (5) Made subject to business processes to prevent inadvertent release of deidentified information.
>
> (6) Protected from any reidentification attempts.
>
> (7) Used solely for research purposes that are compatible with the context in which the personal information was collected.
>
> (8) Not be used for any commercial purpose.
>
> (9) Subjected by the business conducting the research to additional security controls [that] limit access to the research data to only those individuals in a business as are necessary to carry out the research purpose.[75]

There is significant ambiguity in this definition, indicative of the generally poor drafting of the CCPA. For instance, in the sentence preceding the nine conditions included in the definition, the subject of that sentence is "research." But several of the nine conditions that apply to that subject make little sense as a condition on research, but rather seem to have been drafted as a condition on "personal information." For instance, requiring that personal information be de-identified makes sense, but that's not what this definition requires. Instead, it says that the "research . . . shall be . . . subsequently pseudonymized and deidentified . . . [and] protected from any reidentification attempts."[76] Does this mean that

---

[75] *Id.* § 1798.140(s).

[76] *Id.* § 1798.140(s)(2), (6).

the input (i.e. the personal information) need not be de-identified, but that the output of the research must be? The clumsy drafting makes this a legitimate question for which there is no clear answer.

Furthermore, one of the nine conditions in the definition of "research" is that the research "not [be] used for any commercial purpose."[77] On its face, such a condition is absurd. Taken literally, that means that the scientific research involving the use of personal information could never be used for any commercial purpose. If a new drug is developed from such research, it could not be produced or distributed by any company. If a new security method or technology is developed from such research, private sector entities could not use it to protect their data or that of their customers. If a new clean energy source is developed from such research, automakers could not incorporate it into their vehicles and manufacturing processes. This condition, in effect, means that the definition of "research" will apply to virtually no real-world research.

Moreover, this condition conflicts with another definition within the CCPA itself. The defined term "research" is used within the definition of "business purpose." Under the CCPA, "business purpose" means a use of personal information for certain "operational purposes, or other notified purposes,"[78] and the definition provides several examples of business purposes including "undertaking internal research for technological development and demonstration."[79] So, "internal research" is a "business purpose," but "research" is defined in such a way that it must be non-commercial. But a business purpose is something carried out by a business, so it's inherently commercial. Thus, this example of a business purpose only applies to non-commercial research carried out for a commercial purpose. Such conflicts make the CCPA definitions section remind one of a paradox Captain Kirk would give to a malevolent computer to make it self-destruct.[80]

Needless to say, the CCPA is not the ideal model for how a

---

[77] *Id.* § 1798.140(s)(8).

[78] *Id.* § 1798.140(d).

[79] *Id.* § 1798.140(d)(6).

[80] *See, e.g.*, the *Star Trek* episodes *Return of the Archons* (NBC television broadcast Feb. 9, 1967), *I, Mudd* (NBC television broadcast Nov. 3, 1967), and *The Ultimate Computer* (NBC television broadcast Mar. 8, 1968).

privacy law should define and address scientific research. The definition of research is unclear, contradictory, and so narrow that it would apply to little or no real-world research.

Fortunately, that poorly-drafted definition appears to have little practical effect. The term "research" is used only twice in CCPA. First, it creates an exception to consumers' right to delete personal information that is at best extraordinarily narrow and at worst nonsensical. But there are other exceptions to the right to delete that are extremely broad and are likely to make having to delete data needed for any research purposes very rare. Second, it creates a contradiction within the definition of "business purpose," but which has little or no effect on the use or sharing or personal information for research purposes.

At the end of the day, the exceedingly narrow statutory definition of "research" under the CCPA does not result in any prohibition or restriction on the use or disclosure of personal information for research purposes. The only case where such uses and disclosures would be restricted is when a disclosure is deemed a "sale" of personal information, a factor companies will want to consider and perhaps take steps to design the data sharing arrangement with an eye toward avoiding it being considered a sale.

Nevertheless, the numerous problems with how the "research" definition is drafted could cause problems in the future if CCPA is ever expanded to include things like a broad consent obligation for data sharing, or if that definition is adopted or used in another context that would impose significant restrictions in data uses and disclosures that do not meet that definition.

PART III: COMMERCIAL VS. ACADEMIC RESEARCH

While privacy laws address and affect scientific research in different ways, most have some provisions that make reasonable allowances for research. Questions may arise regarding the applicability of these provisions based on the purposes and nature of the research being conducted. For example, where personal information is used for research, there are a variety of possible research purposes and a broad spectrum of uses ranging from the purely academic to the purely commercial. Some research may be

focused on promoting some widely–supported public interest objective. Some academic research may simply be aimed at advancing scientific knowledge. Some may be designed to develop new technologies. Some may focus on developing or improving commercial products or services. Some may be to study the safety of a product or service. And some may be aimed at improving the effectiveness of information or marketing messages provided to consumers.

Common sense and practical experience may lead organizations to conclude that regulators are likely to look more favorably upon research purposes that are closer to the "purely academic" end of the spectrum, or where a strong public benefit to the research can be demonstrated. And in fact, the text of several privacy laws suggests that lawmakers have shown a preference for research in the public interest. For instance, under the CCPA, the research exception to the right to delete requires that the research be in the public interest and not for commercial purposes.[81] And under the GDPR, the higher level of restrictions on the processing of special categories of sensitive personal data (including health data) don't apply where the "processing is necessary for reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health or ensuring high standards of quality and safety of health care and of medicinal products or medical devices."[82]

However, it also appears that even within these laws that give preference to research concerning issues of general public welfare, the lines between academic and commercial research are not determinative. As noted in Part II above, the problematic CCPA definition that says research must be non-commercial conflicts with how the term "research" is used elsewhere in CCPA in a purely commercial context, and in any event, it has little practical effect. In Europe, the GDPR includes a helpful recital that says: "For the purposes of this Regulation, the processing of personal data for scientific research purposes should be interpreted in a broad manner including for example technological development and demonstration, fundamental research, applied research and privately

---

[81] CAL. CIV. CODE §§ 1798.105(d)(6), 1798.140(s)(8). But see the problems with that definition discussed in Part II.

[82] GDPR, *supra* note 3 at 38.

funded research."[83]

Thus, the conclusions regarding permissible use, sharing, and retention of personal information for research purposes may not be fundamentally different for commercial research vs. purely academic research under the laws addressed in this paper.[84] Most privacy laws provide leeway to allow for use of personal information for scientific research, and in most or all cases, that leeway extends to at least some commercial research.

That flexibility makes sense. After all, the lines between academic or public-interest research and commercial research often are not clear or obvious. Much academic research is leveraged for beneficial commercial applications. And often, there are close academic-commercial collaborations that make it possible to turn research into, for example, mass-market drugs. Conversely, much research performed by commercial entities promotes public interests, such as advancing scientific knowledge and furthering public health. Privacy law should allow for, and encourage, all such research.

PART IV: RECOMMENDATIONS AND CONCLUSION

As policymakers draft new privacy laws, update existing privacy laws, or develop guidance applying or interpreting privacy laws, they must consider carefully how their laws, guidance, and interpretations impact scientific research. Privacy laws should allow for and encourage responsible use of personal information to advance scientific knowledge and innovation. To overly burden the use of personal information for research purposes is to foreclose the possibility of such research and the many benefits that flow from enabling it.

Particularly when incorporating principles such as consent, a

---

[83] *Id.* Recital 159 at 30.

[84] It is worth noting, however, that many uses of data flowing from research for commercial purposes will raise additional legal obligations. For example, if an output of research is a better algorithm for tailoring marketing messages to consumers, the company wishing to send those tailored marketing messages will need to comply with all the legal obligations that apply to direct marketing, including initial consent, providing users the ability to stop receiving such messages at any time, etc. Thus, organizations need to be aware of the regulatory obligations applicable to all subsequent data uses.

right to delete personal information, and notions of data minimization, it is important that privacy laws include reasonable and realistic exceptions for research. To that end, policymakers should adopt the following recommendations:

- If a privacy law requires consent for certain uses or disclosures of personal information, there must also be practical alternatives to consent available.
- If a privacy law includes a right for individuals to request the deletion of personal information, there must be an exception available if the personal information is needed for research purposes.
- If a privacy law includes a collection limitation principle, collecting data needed for research purposes must be considered an appropriate ground for collecting personal information.
- If a privacy law includes a retention limitation principle, retaining data needed for research purposes must be an appropriate ground for retention.
- If a law addresses de-identification, the law should provide incentives for its use of rather than imposing a mandate. Further, such incentives should acknowledge that de-identification may not always be appropriate or compatible with certain uses of personal information, including certain research uses.
- Definitions of "research" and the ways the law addresses research uses of personal information should not limit the allowances for research to just "non-commercial" research, but should instead be flexible and acknowledge that commercial or privately-funded research can lead to the same desirable outcomes as publicly-funded or purely academic research.

Making such allowances for scientific research in privacy law does not mean that privacy need be sacrificed. Other privacy principles can provide protection while still allowing for research uses of the data. Transparency can help create understanding of such uses and provide for organizational accountability. De-identification, when appropriately applied in a manner compatible with the research purpose, can protect personal information and reduce privacy risks. Data security is essential and can help ensure

that personal information in a research context is protected from unauthorized access and misuse.

Existing laws, such as the GDPR, include these types of allowances for scientific research while still providing robust privacy protections. By following these recommendations, policymakers can strike an appropriate balance that enables and encourages socially beneficial uses of personal date while protecting the privacy of individuals.