NUTS & BOLTS *of* USER PRIVACY

A primer for policymakers on user privacy

JULY 2019

presented by:

Engine

CHARLES
KOCH
INSTITUTE

INTRODUCTION

It seems like not a week goes by without a news headline or Hill hearing about consumer privacy. Whether it's the European Union's General Data Protection Regulation, or the looming California Consumer Privacy Act, policymakers across the world are grappling with what steps they can take to better safeguard consumers' online data while promoting competition and innovation.

As congressional leaders work on crafting a federal data privacy framework, policymakers must consider the importance of balancing pro-privacy regulations with feasible requirements that do not unduly burden startups on bootstrap budgets. Companies of all sizes rely on data collection to provide streamlined services to consumers. When considering how best to address data privacy concerns, policymakers must first gain a better understanding of the different types and uses of data that companies collect, store, and utilize.

Consumers and startups both support strong data privacy regulations that prohibit abusive data practices and ensure greater user choice and transparency. That's why Engine and the Charles Koch Institute hosted a "Nuts and Bolts of User Privacy" series to provide a more complete overview of what consumer data is, what it can do, and how startups in particular rely on collected user data to compete with more established industry players.

This report explores the state of data privacy today, including the buzzwords of user privacy, state efforts to protect consumer data, and how efforts like the GDPR and CCPA impact both consumers and startups. It will also examine the current legal landscape across the U.S. as well as the importance of separating "good" data collection and usage practices from efforts that unnecessarily sweep up consumers' personal information.

CONTENTS

GLOSSARY

Aggregation:

Combining records from individual users together, typically in such a way that an individual record cannot be inferred from the aggregated dataset.

Anonymization:

The process of removing direct and indirect identifiers and manipulating the data so that a record cannot be re-identified. In an anonymized data file, the data records cannot be linked back to a unique member of a dataset.

California Consumer Privacy Act (CCPA):

A state law passed in 2018, set to go into effect in 2020, that requires organizations to comply with certain new rights granted to California consumers when it comes to the sharing of their data.

Cookies:

A small text file stored on a user's computer, either temporarily or permanently, to provide a way for the website to recognize the user and track the user's website preferences.

Cookie Banner:

A small pop-up that appears when a website user visits the website. The cookie banner usually pops up during a user's first visit or after the user deletes his web history. Cookie banners give the website a way to get user consent to collect data from the website's cookies.

Consent:

The act of individuals agreeing to other parties collecting certain pieces of information.

Opt-in Consent:

When an individual expressly agrees to have his information collected. For example, a user must actively check a box to receive marketing emails from a website in order to satisfy opt-in consent.

Opt-out Consent:

When an individual has the option to decline having his information collected. For example, a user must actively uncheck a box to not receive marketing emails from a website to satisfy opt-out consent.

Data Breach:

An intentional or unintentional release of secured data to an untrusted group.

Data Interoperability:

The technological ability of a user to take his dataset from one online service in a way that preserves the meaning and the nature of the relationship between the data points so the dataset can be used by another online service.

Data Portability:

The technological ability of a user to extract his data from an online service in a machine-readable format in a timely fashion.

De-identification:

Removing or masking identifiers from a dataset, so that the remaining information does not uniquely identify a user and it is reasonably believed that the information cannot be used to re-identify a user.

General Data Protection Regulation (GDPR):

The General Data Protection Regulation is the European Union's law granting all EU residents certain rights regarding the collection and sharing of their data.

Controller:

Under GDPR, a data controller is the entity that collects data from consumers and determines how it will be stored and used.

Processor:

Under GDPR, a data processor is the third-party entity that processes consumer data at the direction of the controller that collected the data.

Identifier:

A piece of data that can be associated with an individual user.

Direct Identifier:

A piece of information that, on its own, uniquely identifies a user (ex: full name, social security number, etc.).

Indirect Identifiers:

A piece of information that, when combined with other information, can be used to uniquely identify a user (ex: gender, birth date, zip code, etc.).

Internet Protocol (IP) Address:

A unique string of numbers that identifies a specific device connected to a network such as the Internet.

Notice:

The mechanism used to inform individuals when their data is being collected, how that data is being used, and with whom that data is being shared.

Right to be Forgotten:

A right granted to European Union citizens where the EU user can request that an entity collecting, sharing, or publishing information about that user erase that user's personal information, subject to certain exceptions.

Third-Party Data Collection:

When a party collects data from any source other than directly from the user.

Direct Data Collection:

When a party collects data directly from the user.

VISUALIZING PRIVACY BUZZWORDS

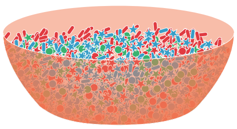
Aggregating and de-identifying data can be a good way to preserve user privacy while still using data to fuel innovation. During a "Nuts and Bolts of User Privacy" panel held in D.C. in May 2019, attendees used cupcakes and sprinkles to visualize aggregation and de-identification methods that let organizations glean important information to inform business decisions. The following is a description of that interactive demonstration.

Imagine you and your team are trying to launch your website in Country X in Asia. You know that first-time visitors tend to quickly look for the "Contact Us" button on your website, so you want to make sure the potential viewers in the new country can easily find your "Contact Us" button. You want to use information you already have about how your current users around the world find the "Contact Us" button to make sure the website that launches in the new country is designed intuitively.

You have three separate buckets of data, signified here as three containers of sprinkles. Each container of sprinkles is a different color. Blue sprinkles represent user traffic patterns in North America, red sprinkles represent user traffic patterns in Europe, and green sprinkles represent user traffic patterns across Asia.

Each container of colored sprinkles contains sprinkles of varying shapes, with each one representing a different, common user traffic pattern. Cylindrical sprinkles represent users who clicked directly on the "Contact Us" button from the home page. Spherical sprinkles represent users who clicked on the "About Us" page before finding the "Contact Us" page. The star sprinkles represent users who found the "Contact Us" page from a page of search engine results. These datasets, or containers of sprinkles, are largely de-identified; for each data point, you simply know the continent in which it was generated and what traffic pattern brought a user to the "Contact Us" page.

Once you combine the sprinkles—or aggregate the data—your team can sample the data by dipping your cupcake in the combined sprinkles. What can you infer from this de-identified, aggregated data? In this instance, it's that visitors to your website from Europe and North America largely find the "Contact Us" page directly from the home page, while visitors to your website from Asia are split between clicking on the "About Us" page first or finding the "Contact Us" page through a search engine. This could signify that your website, and the location of the "Contact Us" page specifically, is not designed in a way that users in Asia find intuitive.



When you go to launch the website in Country X, you might make the link to the "Contact Us" page larger on your homepage or put the information to contact your organization at the top of the "About Us" page. In doing so, you've used de-identified and aggregated data to make an improvement that will benefit your website's visitors.

CCPA VS. GDPR

GDPR: After several years of debate, the General Data Protection Regulation—Europe's set of sweeping privacy rules—went into effect in May 2018. The regulation codifies and unifies fundamental privacy rights for residents across the European Union when it comes to the collection, processing, sharing, and security of consumer data.

CCPA: In an attempt to stave off a ballot initiative, the California legislature quickly took up and passed the California Consumer Privacy Act in 2018. The law is set to go into effect in 2020. It requires businesses to comply with newly created rights for California consumers around the sharing and security of consumer data.

Under **GDPR**, personal information is defined as information relating to an identified or identifiable data subject, including name, identification number, location data, online identifier or data about the physical, physiological, genetic, mental, economic, cultural or social identity of a data subject.	ℹ	Under **CCPA**, personal information is defined as information that identifies, is about, describes, "is capable of being associated with, or could reasonably be linked, directly or indirectly," with a consumer or household, including name, address, IP address, location information, and Internet activity.
GDPR creates the following rights for European Union residents: • The right to access the personal information a business collected on a data subject. • The right to rectify any inaccurate personal information a business has on a data subject. • The right to be forgotten, or the right to erase personal information a business has collected about a data subject. • The right to port, or transfer, their personal information. • The right to restrict the processing of personal information. • The right to avoid decisions made by automated processing of personal information.	⚖	CCPA creates the following rights for consumers in California: • The right to know what a business is doing with their personal information. • The right to know what third parties have access to their personal information. • The right to access the personal information a business has collected on a consumer. • The right to delete the personal information a business has collected on a consumer. • The right to opt-out of the sale of personal information.
Under GDPR, user consent is one of the six legal bases under which businesses may process personal information. Opt-in consent must be informed, specific, and freely given.	☑	Under CCPA, consumers can opt-out of the sale of personal data, and organizations need affirmative, opt-in consent from consumers under the age of 16 to sell their data.
GDPR is enforced by EU countries' data protection authorities.	🛡	CCPA is enforced by the state Attorney General as well as through private lawsuits brought by consumers who have been impacted by data breaches.
Penalties under GDPR can go up to 20% of annual global revenue or €20 million, whichever is greater.	$	Penalties brought by the state Attorney General are up to $7,500 for intentional violations and $2,500 for unintentional violations. If a company suffers a data breach and is sued, it can be subject to damages of between $100 and $750 per user, per incident.
GDPR requires that businesses maintain appropriate security measures to protect personal information and notify authorities of a data breach within 72 hours.	🔒	CCPA requires that businesses maintain reasonable data security practices and allows consumers to bring lawsuits if a business suffers a data breach.

PRIVACY RULES AND PROPOSALS IN THE STATES

CALIFORNIA

The California Consumer Privacy Act (AB 375; SB 1121)

Passed in 2018, goes into effect in 2020

Summary: The CCPA requires organizations to comply with newly created privacy rights for California consumers, including a user's right to access and delete the information an organization has about that user and a user's right to opt out of many types of sharing of user data. It allows the state Attorney General to bring fines starting at $2,500 per violation of the law, and establishes the ability for consumers to sue organizations that suffer a data breach. The law includes a small business exemption for organizations that have data on fewer than 50,000 users or devices, less than $25 million in annual revenue, and that derive less than half of their revenue from selling California consumers' personal information.

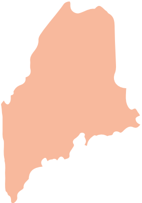
In 2018, California was the first state to adopt a comprehensive consumer privacy law, accelerating the national consumer privacy debate. Since the CCPA's passage, many states have followed suit. As of July 2019, 12 states have introduced comprehensive consumer privacy laws, similar to the CCPA, while others have strengthened consumer protections in certain areas. These laws and proposals, in addition to the proposals states are likely to introduce in the coming years, will create a patchwork of varying, and potentially even conflicting, requirements that will be impossible for small startups to navigate.

COLORADO

A Bill Concerning Strengthening Protections for Consumer Data Privacy (HB 18–1128)

Passed in 2018

Summary: This law amended the state's consumer data privacy law to require that companies must develop a written policy to destroy or dispose of user electronic and paper data that is no longer necessary to their business operation. Additionally, the Colorado law strengthens its data breach notification laws by requiring that covered entities must notify affected parties within 30 days, and that any breach reasonably believed to have affected 500 or more Colorado residents must be reported to the Colorado Attorney General within 30 days.

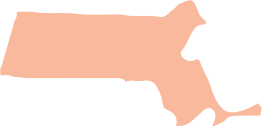


MAINE

An Act To Protect the Privacy of Online Customer Information (LD No. 946)

Passed in 2019, goes into effect in 2020

Summary: Mirroring privacy rules written by the FCC in 2015 and repealed by Congress in 2017, Maine's law prohibits all internet service providers (ISPs) doing business in Maine from selling any personal data without the affirmative consent of the individual. The law defines personal data as any information that can be identified or linked to an individual, including information on an individual's use of the Internet, such as browsing history and application usage history.

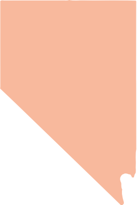


MASSACHUSETTS

An Act Relative to Consumer Data Privacy (S.120)
Introduced in 2019
Summary: This bill grants Massachusetts consumers the right to be notified before a business collects their personal information, the right to access their personal information held by a business, the right to delete their personal information held by a business, and the right to opt-out of the sharing of their data with third parties. Additionally, the bill allows consumers to sue a company that violates the law.

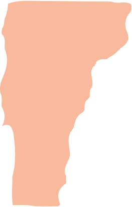
NEBRASKA

An Act relating to consumer protection (LB 757)
Passed and went into effect in 2018
Summary: This law requires that companies that have personal information of Nebraska residents implement and maintain reasonable security procedures and ensure that third parties with access to personal information also comply with the security procedures to protect the personal information. Failure to comply can subject the business to the Nebraska Attorney General's enforcement.



NEVADA

An Act relating to Internet Privacy (SB 220)
Passed and will go into effect in 2019
Summary: This law amends Nevada's original Internet privacy law to require that businesses provide some online mechanism, or toll-free phone number, that consumers who interact with a business for personal purposes can use to opt out of the sale of their personal information.



VERMONT

Vermont Data Broker Regulation (9 V.S.A. §§ 2430, 2433, 2446–47)
Passed in 2018, went into effect in 2019
Summary: This law requires that data brokers register with the Vermont Secretary of State and pay a $100 registration fee, annually disclose certain information about their activities, and maintain data security procedures to protect the data they hold. The law defines a data broker as an organization that knowingly collects, sells, or licenses "brokered personal information" to a third party, and does not have a direct relationship to the Vermont resident whose information is collected.

WASHINGTON

An act relating to the management and oversight of personal data (2SSB 5376)
Introduced in 2019, passed the state Senate but didn't move in the House
Summary: Modeled heavily after the European Union's GDPR, this bill would grant eight rights to Washington consumers regarding a controller's use of personal data. The law defines personal data as information that can identify a person, including sensitive data, but explicitly excludes de-identified data. The act explicitly lays out provisions for facial recognition technology, providing heightened requirements that controllers and processors that deal with this technology must follow.

DATA FOR GOOD

Companies of all sizes, especially startups, rely on consumer data every day to operate, advertise, and improve their services. Often, that data is anonymized, aggregated, or in some way altered to protect user privacy, and the uses of that innocuous data are intended to create benefits for consumers.

As policymakers considering enacting enhanced privacy protections, it's important that they consider some of the beneficial and necessary aspects of data usage that small businesses rely on to provide their services and even protect their users. Startups stand to lose the most in the current privacy debate. The large Internet companies that have already amassed large amounts of user data and have large budgets and legal teams will be best equipped to navigate the regulatory and business landscape that could result from reactionary policymaking in this space.

The following examples provide additional context as to how some companies use and collect data to innovate, grow, and thrive.

All the reviews that are fit to print

Consumers rely on review platforms to make decisions every day, from which restaurant to eat at, to which foreign city to visit next, to which home appliance to purchase. These review platforms help consumers make informed decisions about how to spend their hard-earned money, and platforms go to great lengths to ensure the reviews they host are accurate and relevant. Those efforts often include finding and removing "reviews" posted by fraudulent accounts created by anyone who stands to profit from positive reviews. For instance, if a hotel is receiving negative reviews on a review website, the owner might create several accounts on that website to post fake positive reviews and drown out the organic negative feedback the establishment is receiving. In that case, the review website can collect and retain information about its users, such as IP addresses, to identify the hotel posting from multiple accounts and remove the fake reviews.

You've got mail (to send)

Small businesses, including startups, rely on vendors to manage their users' contact information and contact their users as the businesses grow. Those vendors, in turn, can gather information about the ways customers use a startup's services to make sure the service is working as intended—whether that's taking steps to prevent fraud and abuse, or making changes to increase functionality. Take, for example, a vendor providing a customer management software tool that lets companies send mass emails to their consumers. Because of the number of emails that vendor sends on behalf of client companies, the vendor can tell that emails sent on Friday afternoons get the lowest open rates. As companies go to send emails using the vendor's software on Friday afternoons, the vendor can flag that emails sent on that day and time often get low open rates, and prompt the companies to schedule their emails for a later time. In this example, the vendor is using high-level data about its client companies, and the ways in which their users interact with emails, to help clients use their software more efficiently.

Managing misinformation mischief

Popular social media platforms can be an attractive breeding ground for misinformation campaigns. From spreading conspiracy theories to foreign election interference efforts, platforms are forced to spend time figuring out which users are disseminating misinformation in violation of their terms of service. One effective method to finding and containing users behind widespread misinformation campaigns is using IP addresses and geolocation information to map out networks of servers that are being used to push out misleading social media content. If, for instance, a platform can identify a network of users in a foreign country that starts by sharing videos about the moon landing being staged, but then escalates to sharing false information about U.S. politicians, the company can begin to predict where the most dangerous misinformation is coming from early on, and then remove it from the platform. By connecting the dots between specific users and networks that create and share misinformation, platforms can more efficiently identify and shut down misinformation campaigns.

To prevent a predator

Dating apps have become an incredibly popular way to connect in the digital age, but the companies operating those apps have faced several complaints from users who receive unwanted and harassing communications from other users. Many of the ways that dating apps crack down on negative behavior on their platforms involve collecting, analyzing, and retaining information about users and the contents of the communications they send. For instance, if a user of a dating app has been flagged by several other users for sending out unsolicited graphic imagery, the app can flag that user's account and limit his access to the app's features, as well as use the problematic images to train algorithms to better recognize other problematic images sent by future users. In that way, the dating app is using one user's data to protect other users from future unwanted and harassing communications.

Helping supply meet demand

Ecommerce delivery platforms that connect consumers with independent merchants have been a helpful way for brick-and-mortar small businesses to grow. In addition to steering more customers to these brick-and-mortar retailers, these ecommerce delivery platforms can also gather and analyze trend data from consumers and retailers to help businesses better respond to consumer demand. As an example, an ecommerce delivery platform could look at orders over the course of a week and notice that users tend to order more tacos on Tuesday nights. The platform can inform local taco restaurants of that trend, and the restaurants can stock their shelves to be able to fill the increased number of taco orders. In that case, the collection and use of anonymized, aggregated data benefits a local brick-and-mortar business and its customers, as well as the ecommerce delivery platform.

Engine

Engine was created in 2011 by a collection of startup CEOs, early-stage venture investors, and technology policy experts who believe that innovation and entrepreneurship are driven by small startups, competing in open, competitive markets where they can challenge dominant incumbents. We believe that entrepreneurship and innovation have stood at the core of what helps build great societies and economies, and such entrepreneurship and invention has historically been driven by small startups. Working with our ever-growing network of entrepreneurs, startups, venture capitalists, technologists, and technology policy experts across the United States, Engine ensures that the voice of the startup community is heard by policymakers at all levels of government. When startups speak, policymakers listen.

CHARLES
KOCH
INSTITUTE

For more than five decades, Charles Koch's philanthropy has inspired bold new ideas to improve American lives. Inspired by a recognition that free people are capable of extraordinary things, the Charles Koch Institute supports educational programs and dialogue to advance these principles, challenge convention, and eliminate barriers that stifle creativity and progress. We offer educational programs, paid internships, and job placement assistance to students and professionals, and encourage civil discussion about important issues like free speech, foreign policy, and criminal justice reform. In all of our programs, we are dedicated to identifying new perspectives and ideas that help people accomplish great things for themselves and others.