CLOUDERA

WHITE PAPER

DATA PRIVACY AS A CORE BUSINESS PROCESS

What compliance requires from an enterprise data platform

PRIVACY BY DESIGN

Table of Contents

Ultimately, organizations today have no choice but to make "data privacy" a core business process and a key tenet that is critical to their successful functioning as a business.

Introduction

Besides acting, what do Arnold Schwarzenegger, Drew Barrymore, Leonardo DiCaprio, and Tom Hanks have in common? They've all had their health records accessed by the prying eyes of a disgruntled cardiothoracic surgeon fired from his researcher job at the UCLA school of medicine[1]. They are not alone. Earlier this year and following a suspected social engineering attack, hundreds of German politicians had sensitive personal data leaked online; Chancellor Angela Merkel's email address and several letters were published while Greens leader Robert Habeck's private chats with family and credit card details were exposed[2]. You may not enjoy the fame of a celebrity or a German politician, but if you're one of the 87 million Facebook users whose data ended up in the hands of Cambridge Analytica in 2018, information about you has been bought and sold to deliver targeted messages to unknowing audiences[3].

Data is the new gold, and mining and selling it is today's equivalent of 1849's Gold Rush. Consider that, in 2019, a collection of 2.7 billion identity records, consisting of 774 million unique email addresses and 21 million unique passwords, was posted on the web for sale[4]. As personal data grows—especially personally sensitive or identifiable data—so has the market for it. And as a result, so have the number and frequency of global regulations meant to give more control over data to the individual and ensure compliance by corporations.

With privacy regulations on the increase, corporations are obligated to ensure personal data is exposed to only the right people and in the right way and kept safe at the same time. And they must do so knowing that, as they manage more personal data, regulations will accelerate, and enterprise data systems must evolve to keep pace. Ultimately, they have no choice but to make "data privacy" as core a business process as they do for sales and marketing, or accounting, or customer service—or any other process that is critical to successful functioning as a business.

This paper will help IT leaders and compliance officers understand how data regulations are becoming increasingly important, explore how they are causing all sorts of institutions to adapt, review your responsibilities around privacy, and outline what you need in a data platform to remain compliant.

Why regulations have become increasingly important

A short chronology of privacy regulations makes clear that their frequency is quickening. More data, more data sources, more devices connected to the internet, and more publicity around privacy violations has led the public to become increasingly aware of the data they generate and welcoming of regulations that guard their privacy. Industries like financial services, insurance, health care, telco, utilities, and more have become highly regulated. And with increasing frequency, as this list of privacy acts enacted over the last decade or so—though not comprehensive—illustrates:

- **1970 German Bundesdatenschutzgesetz (BDSG):** A German federal data protection act, that together with the data protection acts of the German federal states and other area-specific regulations, governs the exposure of personal data, which is manually processed or stored.
- **1988 Video Privacy Protection Act:** Thanks to the 1987 incident with U.S. Judge Robert Bork's exposed video rentals[5], this made it illegal to share video lists without the customer's written consent.
- **1988 Privacy Act of Australia:** Regulates how personal information about individuals is collected, used, stored, and disclosed by most Australian, Australian Capital Territory, and Norfolk Island public sector agencies, large businesses, health service providers, and some small businesses and non-government organizations.

Though there are differences in the details of privacy regulations around the world, their intent and implications are comparable: to protect an individual's right to their own data; govern how institutions collect, process, store, manage, and secure that data; and correct or return ownership of it to the data subject when requested.

- **1996 Health Insurance Portability and Accountability Act (HIPAA):** Modernizes the flow of healthcare information and protects personally identifiable information (PII) from fraud and theft. Because of HIPAA, the celebrity-health-records-peeping former UCLA researcher mentioned above spent four months in jail and was fined $2,000.
- **2001 Canadian Personal Information Protection and Electronic Documents Act (PIPEDA):** Applies to private-sector organizations across Canada that collect, use, or disclose personal information in the course of a commercial activity. People have the right to access and challenge the accuracy of the personal information an organization holds and must give consent for the organization to use it for a purpose other than intended when originally collected.
- **2003 Act on Protection of Personal Information (APPI) in Japan:** Applying to business operators that hold the personal information of 5,000 or more individuals, the APPI requires them to specify the purpose for which they're using personal information and disclose that if data subjects request.
- **2012 Personal Data Protection Act (PDPA) Singapore:** Governs how any private organization—even those that are not physically located in Singapore—collects, uses, and discloses personal data.
- **2013 Protection Of Personal Information Act (POPIA), South Africa:** Ensures all South African institutions conduct themselves responsibly when collecting, processing, storing, and sharing others' personal information.
- **2018 General Data Protection Regulation (GDPR) in the EU:** Taking 10 years from inception to completion in the EU, GDPR protects EU citizens from privacy and data breaches and gives them the right to access their PII and request it be deleted.
- **2018 California Consumer Privacy Act (CCPA):** The U.S. adopts privacy regulation in California, empowering consumers to own, control, secure, and prevent corporations from selling or disclosing their personal information. The law takes effect in 2020.
- **2019 New York Privacy Act:** Introduced in May and similar to California's law, New York State's privacy law would give residents there more control over their personal information than in any other state. The law failed to pass during a June Senate session.

In addition to the regulatory acts above, global policy makers are exploring the use of and privacy around image recognition technology. For instance, King's Cross in London uses facial recognition to track visitors, and the Canary Wharf development plans to follow suit. If those plans are realized, over 240 acres of London will be covered by the technology[6]. Lawmakers across the U.S., however, are currently debating city governments' use of biometric and image recognition technology, claiming it violates civil rights and is prone to error when identifying women and people of color. Cities argue that image recognition helps them better understand traffic flows, parking violations, crime, and more. But when data is such a commodity, who wants an image of their face available on the open market?

Though there are differences in the details of privacy regulations around the world, their intent and implications are comparable: to protect an individual's right to their own data; govern how institutions collect, process, store, manage, and secure that data; and correct or return ownership of it to the data subject when requested. As news headlines increasingly report large data breaches (Wikipedia even has a page listing them), privacy regulations are more culturally accepted, supported, and encouraged, which has accelerated their evolution around the world.

Data judgment, ethics, and the risk of being "creepy"

Organizations—especially marketers—are faced with the challenge of how to appropriately use the information they have about customers (web pages visited, products bought, shopping carts abandoned) to continue engaging them without the experience seeming "creepy." As Gartner's Managing Vice President Carsten Casper has stated, "Most people worry about surveillance, and how to avoid being spied on. Individuals risk that something will be concluded about them that they didn't know about themselves."[15]

In other words, processes and technologies that ensure data privacy laws are adhered to don't always guarantee the people applying them are doing so ethically. Marketers and customer-facing employees must not only keep data private but also use good judgment in creating experiences and communications based on that data. What's legally allowed with data doesn't make it ethically right.

A case in point is the humorous and satiric video "Ordering Pizza in 2015." Originally produced in 2004 by the ACLU, the video went viral as the single most-downloaded piece of content the ACLU ever produced and illustrated just how creepy (and unethical) a pizza company (or any company) might be with the personal data it has available to them.

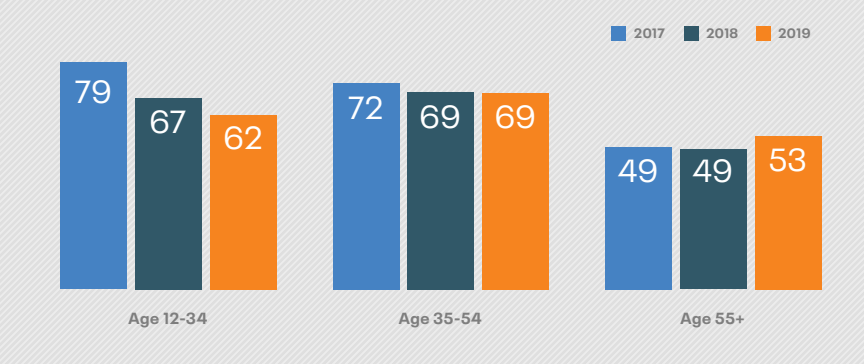
Remember, just because you can, doesn't mean you should.

What it costs companies for violating privacy

And while privacy laws accelerate, unprepared companies have much to be concerned about. If their companies suffer such data breaches, they not only risk fines from privacy laws, but also suffer damaged reputations, negative media coverage, lawsuits, and potential loss of customers. Facebook, for instance, has an estimated 15 million fewer users in 2019 than it did in 2017[7] after privacy violations surfaced regarding its sale of data to Cambridge Analytica in 2018. (This diagram from Vox illustrates exactly what happened.) The Federal Trade Commission (FTC) issued Facebook an approximately $5 billion fine[8]—its largest ever—for those privacy violations.

Facebook Usage

U.S. POPULATION | % USING FACEBOOK



Source: Edison Research[9]

Under Armour's data breach, announced in March 2019 and affecting an estimated 150 users of its food and nutrition application FitnessPal, cost the company a drop in share price of 3.8 percent[10]. It now faces a class action lawsuit for, among other charges, invasion of privacy[11].

One study[12] from the Journal of Business Research found that consumers "find violations of privacy expectations, specifically the secondary uses of information, to diminish trust in a website" and that offending companies are penalized twice: once for destroying trust in the website and then again for diminishing trust in the firm's integrity and ability to rebuild that trust. Cass R. Sunstein points out in his "The Ethics of Influence" book that people will react poorly to nudges if they no longer trust the entity creating them—even when they want the guidance.

As far back as 2015, Gartner's Risk and Security Survey[13] concisely articulated the four risks that companies worried about with privacy and security. Though a bit dated, the conclusions are relevant today:

1. Reputation and brand damage: 45% of companies were concerned about reputation damage from privacy risk, and 43% about loss of customers.
2. Compliance: One third of respondents were worried about fines, audits, and other enforcement.
3. Lost business opportunities: 33% were concerned about lengthy sales cycles due to privacy concerns or about missing marketing opportunities.
4. Infrastructure: 32% were concerned about maintaining unnecessary IT infrastructure to comply with national privacy laws.

And that was 2015. Gartner's latest quarterly "Emerging Risks Monitor Report" indicates 98 percent of senior executives across industries list "accelerating privacy regulation" as the top emerging risk that worries them in early 2019[14].

Whether they purchase personal data, sell it, or simply gather it, organizations must know what information they have on consumers, what has happened with it, where it is stored, how and when it is used or processed, and who is using it or has access.

Privacy as a core business process

Clearly, data privacy needs to be as much a core business process for companies to function successfully today as are sales and marketing, or customer service, or product development. The financial, reputational, and customer retention risks are too great for companies not to codify an approach to the people, processes, and technologies they use to store, process, retrieve, protect, delete, deliver, and otherwise manage the personal customer data they collect. Whether they purchase that personal data, sell it, or simply gather it, organizations must know what information they have on consumers, what has happened with it, where it is stored, how and when it is used or processed, and who is using it or has access.

So what does data privacy as a core business process look like? How are companies adapting and deploying people, processes, and enterprise data technology to prioritize and proactively embed data privacy principles? Consider the following use cases:

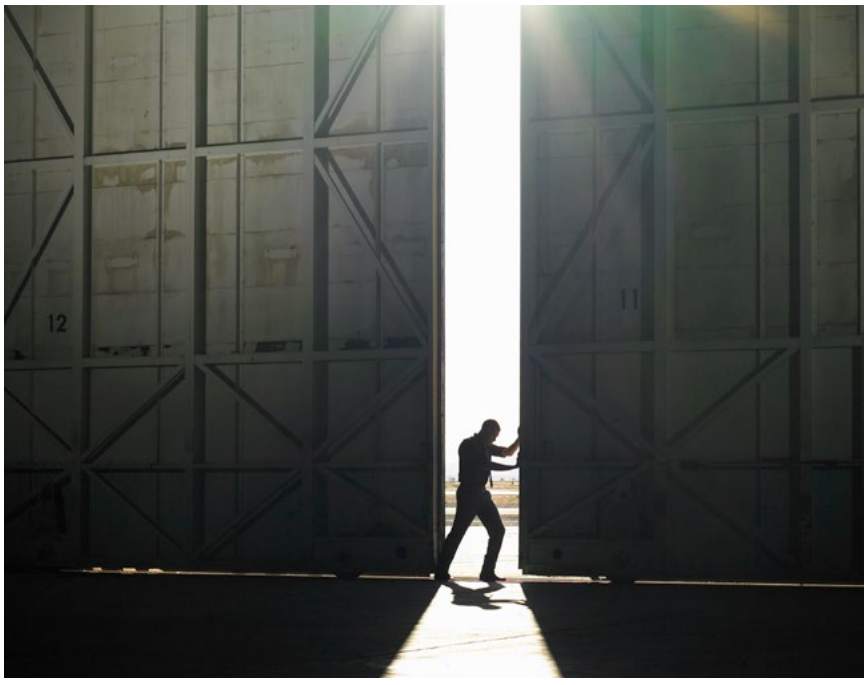
- In healthcare, retail, and financial services industries, some companies are basing an employee's access to sensitive data on their role and location and are regularly auditing for compliance. After all, the top three causes of breaches in early 2019 were attributable to employee error, unauthorized internal access, or accidental web/internet exposure[16], so internal process issues need attention equal to that given for keeping attackers out. Even a single person gone rogue can hack into systems and steal personal data—consider the former Amazon Web Services software engineer who obtained personal data on over 100 million people from a Capital One server[17]. This is a great example of why you must encrypt your data--had they encrypted, the hacker would not have been able to read the contents without the key.
- Forty-seven percent of business leaders admit that human error—such as the accidental loss of a device or document or response to a phishing attack—has caused a data breach[18]. Companies must establish breach notifications to help them quickly identify who had access to what data and what data may have been exposed so they can notify and protect customers and staff.
- Many regulations restrict the use of collected data to the original reason for which it was collected. By tagging and tracking data based on its nature (e.g., PII) or how you obtained it (e.g., registration for a webinar), you can enforce and audit access policies.
- Traditionally, data is analyzed in its original form, sometimes intact, and sometimes with key identifying fields removed. Even anonymized data can be dangerous, however. An MIT study found that anonymized data is a double-edged sword, e.g., anonymized data with geographic coordinates and time stamps could be merged with credit card transactions and other data to identify individuals. Companies have used tag-based security, auditing, field-level encryption, and datapoint obfuscation to reduce risk, analyze exposure, and audit analytics data flows—all while still enabling advanced applications of analytics and AI to improve lives. Doing so requires them to enforce the prevention of "toxic combinations" of data, where feasible, to reduce the risk of re-identification.

"Privacy by design" is a framework that "seeks to proactively embed privacy into the design specifications of information technologies, networked infrastructure, and business practices of an organization.

Privacy by design

"Privacy by design[19]" is a framework that "seeks to proactively embed privacy into the design specifications of information technologies, networked infrastructure, and business practices. " It was founded by Dr. Ann Cavoukian, LL.D. (Hon.), M.S.M., who is recognized as one of the world's leading privacy experts and is currently the distinguished expert-in-residence and leader of the Privacy by Design Centre of Excellence at Ryerson University. The framework incorporates seven foundational principles designed to ensure organizations gain a sustainable competitive advantage by preventing privacy infractions and data breaches from occurring, right from the outset. They are:

1. **Proactive, not reactive:** Privacy by design anticipates risks and prevents privacy invasion before it occurs. It comes "before the fact, not after."
2. **Privacy as default:** Privacy by design ensures personal data is automatically protected in any IT system or business practice, as the default. In other words, privacy is built into the system.
3. **Privacy embedded into design:** Privacy measures need to be essential, integral components of the core functionality, not bolted on as add-ons to the design and architecture of IT systems and business practices.
4. **Full functionality**—positive-sum, not zero-sum: Privacy by design avoids "either-or" dichotomies, such as privacy vs. security, where unnecessary trade-offs occur. It demonstrates that it is possible to have both.
5. **End-to-end security**—full lifecycle protection: Having been embedded into IT systems' design from the start, privacy by design extends security throughout the lifecycle of the data involved, from its collection and use right through to its destruction or removal.
6. **Visibility and transparency**: In a "trust but verify" approach, privacy by design ensures the data subject is fully aware of personal data being collected and why. All component parts remain transparent to users and providers.
7. **Respect for user privacy:** The goal of user-centered privacy requires architects and operators to keep the interests of the individual as paramount by offering strong privacy defaults, appropriate notice, and empowering, user-centric options.



Cloudera technology can help:

- Find Data
- Encrypt Data
- Manage Data Governance
- Manage Data Provenance
- Retrieve Data
- Manage Alerts

Four steps to privacy as a core business process

Making data privacy a core business process doesn't happen overnight, but the first step is defining a data strategy, one that includes plans for how you're managing data, for instance, in a hybrid cloud environment. Also important for your data strategy is defining exactly how the organization will embed a "privacy by design" approach

Your data strategy should also articulate what data your organization stores about regulated data subjects, where it's stored, how it is being used, when, by whom, what level of permission they have from the data subject for usage, and how to remove, delete, or mask that data upon request. And then the data strategy needs to be evaluated for compliance with privacy regulations.

Next, you'll need to evaluate and adjust your data architecture where needed, and ensure the architecture supports appropriate use cases for your company, e.g., marketers needing 360-degree views of customer data, fulfillment teams needing to manage supply chains, and your finance departments' compliance with financial regulations.

Last is implementing your data strategy on an enterprise data platform. Ideally, all your data should be in one central location (like a data lake), regardless of format or structure.

How Cloudera technology helps

Find it

Helps you search and locate all your data, regardless of whether it's stored in enterprise systems or applications. Because regulation covers all the data your company stores on individuals, not just the data it stores that's readily available.

Encrypt it

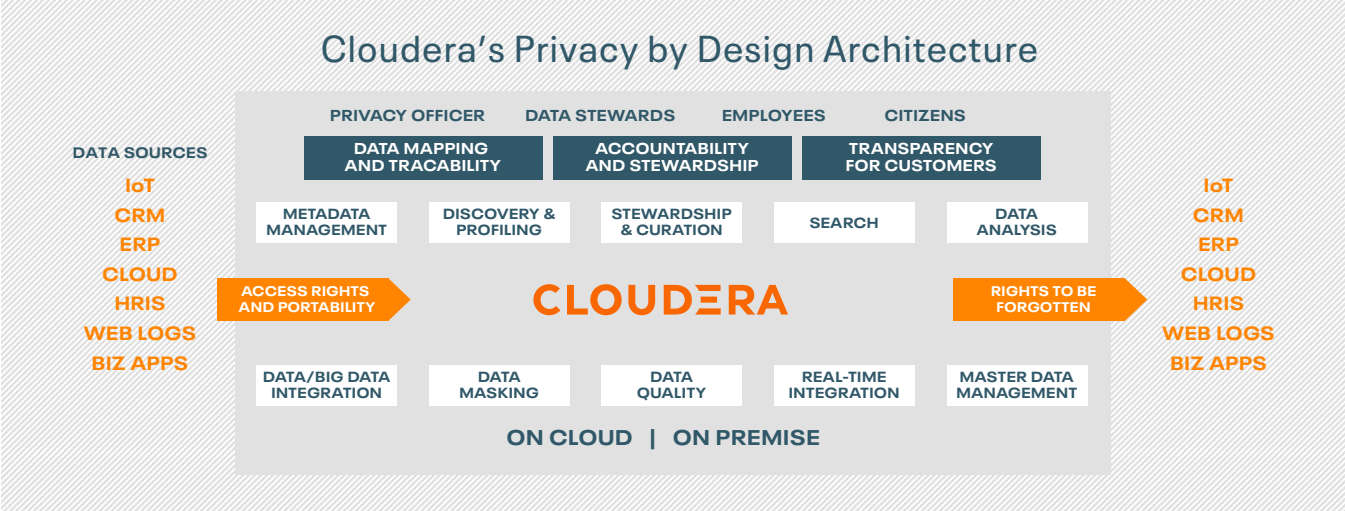
Always secures your data by tokenizing or encrypting it. So even if someone removes a hard disk from your cluster, your data will be safe because they won't have the token ID.

Manage it

Helps you know who has access to what data, why, and how and what they've done with it. Governance and security layers manage security, access, privilege, audits, and more.

Know how it moves

Ensures you know how data moves through the system—often called data lineage or data provenance—from where it was collected, how it was distributed, who changed it and why and when, where they moved it, and so on. The audit trail and access controls let you govern which tables people have a right to see.



About Cloudera
At Cloudera, we believe that data can make what is impossible today, possible tomorrow. We empower people to transform complex data into clear and actionable insights. Cloudera delivers an enterprise data cloud for any data, anywhere, from the Edge to AI. Powered by the relentless innovation of the open source community, Cloudera advances digital transformation for the world's largest enterprises.

Learn more at cloudera.com

Connect with Cloudera
About Cloudera:
cloudera.com/more/about.html

Read our VISION blog:
vision.cloudera.com
and Engineering blog:
blog.cloudera.com

Follow us on Twitter:
twitter.com/cloudera

Visit us on Facebook:
facebook.com/cloudera

See us on YouTube:
youtube.com/user/clouderahadoop

Join the Cloudera Community:
community.cloudera.com

Read about our customers' successes:
cloudera.com/more/customers.html

Retrieve it

Our search capability lets you quickly find what data you have and what's been lost, both structured and unstructured, images, texts, video, speech, and sensor data.

Alert you

You receive alerts if you've had a breach, since many regulations require you to notify people of breaches and tell them exactly what data was stolen.

Conclusion & next steps

The picture becomes very clear when you add the public acceptance of and increasing frequency of global privacy regulations to the monetary, legal, and reputational risks of breaches for companies and their brands. Enterprise companies must embed data privacy into the fabric of their operations and adopt it as a core function, with the people, processes, and enterprise data management technology to support it. Privacy must become a proactive, strategic function—not something that occurs only after a breach—and data management must ethically respect the individual's privacy as well as comply with regulations.

Cloudera's Data Platform (CDP) makes it faster, easier, and safer to build, deploy, and manage analytics and machine learning applications. By lowering costs and simplifying operations, CDP reduces the time required to onboard new data privacy use cases across the organization. The security and governance capabilities in CDP enable organizations to control data everywhere, in multiple public clouds, on-premises, and in a private cloud. Learn more about CDP at cloudera.com/cdp.

Sources

1 https://www.medprodisposal.com/20-catastrophic-hipaa-violation-cases-to-open-your-eyes
2 https://www.techworld.com/security/uks-most-infamous-data-breaches-3604586/
3 https://www.nytimes.com/2018/04/11/technology/facebook-privacy-hearings.html
4 https://gizmodo.com/mother-of-all-breaches-exposes-773-million-emails-21-m-1831833456
5 https://www.chicagotribune.com/news/ct-xpm-1987-11-20-8703270590-story.html
6 https://www.ft.com/content/8cbcb3ae-babd-11e9-8a88-aa6628ac896c?mod=djemCIO (subscription required)
7 https://www.cnet.com/news/facebook-lost-15-million-us-users-in-the-past-two-years-report-says/
8 https://www.nytimes.com/2019/07/12/technology/facebook-ftc-fine.html
9 https://www.edisonresearch.com/infinite-dial-2019/
10 https://www.cnbc.com/2018/03/29/under-armour-stock-falls-after-company-admits-data-breach.html
11 https://topclassactions.com/lawsuit-settlements/lawsuit-news/846290-under-armour-class-action-filed-myfitnesspal-data-breach/
12 https://www.sciencedirect.com/science/article/pii/S0148296317302965#ab0005
13 https://www.gartner.com/smarterwithgartner/avoid-crossing-the-creepy-line-with-the-internet-of-things/
14 https://www.gartner.com/en/newsroom/press-releases/2019-04-11-gartner-survey-shows-accelerating-privacy-regulation-returns-as-the-top-emerging-risk-worrying-organizations-in-1q19
15 https://www.gartner.com/smarterwithgartner/avoid-crossing-the-creepy-line-with-the-internet-of-things/
16 https://www.ryerson.ca/pbdce/about/
17 https://resources.infosecinstitute.com/common-causes-of-large-breaches/#gref
18 https://www.nytimes.com/2019/07/29/business/capital-one-data-breach-hacked.html
19 https://azbigmedia.com/business/technology/why-employee-negligence-is-the-main-factor-in-data-breaches/