cryptonumerics
Privacy for Data Science

The CCPA Creates the need for Privacy Automation

On January 1, 2020, the California Consumer Privacy Act (CCPA) comes into effect. The CCPA is a sweeping piece of legislation, aimed at protecting the personal information of California residents. It is going to force businesses to make major changes to how they handle their data.

Lots of the CCPA's regulations are built upon the bill's conception of "personal information." The CCPA defines personal information as *"any information that identifies, relates to, describes, is capable of being associated with, or could be reasonably linked, directly or indirectly, with a particular consumer or household."* In the CCPA era, how businesses handle this personal information will define whether or not they stay compliant and stay successful.

The Wild West days of monetizing secondary data are over

CCPA compliance is multifaceted. But here is one of the biggest challenges it creates for businesses: the need to implement organizational and technical controls to demonstrate that data used for secondary purposes – such as data science, analytics and monetization – is truly de-personalized.

Why? Because the CCPA says that if it classifies as personal data, it can't be used for secondary purposes. Unless each user gives their explicit consent for every different instance of use – an impossible undertaking for any business.

In effect, the real challenge here is that the cat is already out of the bag, and businesses need a way to catch it, and put it back in.

This is because until now, many data science, analytics, AI and ML projects have had open access to consumer data. The understandable appetite for greater business and customer access has led to a Wild West approach, where monetization teams have enjoyed limitless access to unprotected customer data. This is exactly the area of focus the CCPA wants to bring under control. The bill will force organizations to establish controls that protect consumer privacy rights.

These controls will require minimizing or removing personally identifiable data (PI) to prevent the risk of re-identification. This is the only way to put the cat back into the bag, and the only way not to violate CCPA.

But: organizations are desperate for their data to stay useful and monetizable. How can this balance be achieved?

How to de-identify data but retain its analytical value?

Governance processes and practices need to show CCPA compliance. But various forms of analytics, data science and data monetization are incredibly valuable to enterprises, powering precious business and customer insights.

If you apply data security encryption-based approaches and techniques like hashing, you will reduce or remove the analytical value of the data for data science. No good; you lose the value.

You want to de-identify the date, remove the risk of re-identification, and meet the legal specifications of de-identified data under CCPA. But you want to do this while preserving analytical value.

Both sides of the coin carry a financial imperative. The business motivation and incentive for organizations to ensure its customer data is de-identified is critical, as the overheads and restrictions of CCPA with regard to secondary use of data is prohibitive. Get it wrong, and what you believe to be de-identified data will in fact be in violation of CCPA.

This means you will be on the wrong side of the law. And under CCPA, intentional violations can bring civil penalties of up to $7,500 per violation, and consumer lawsuits can result in statutory damages of up to $750 per consumer per incident.
However, keeping your secondary data useable is financially essential, as correctly de-identified data meeting a CCPA specification will be considered outside the scope of CCPA. This means you can continue to use it for valuable analytics, data science and data monetization.

But you can only do this if the de-identification techniques you have used haven't rendered the data unusable.

This is the central challenge CCPA creates: How can I de-identify my data and meet the legal specifications outlined under CCPA, but still leverage my data for important organizational initiatives?

The CCPA creates the need for Privacy Automation

To be clear: the CCPA does not restrict an organization's ability to collect, use, retain, sell, or disclose a consumer's information that is de-identified or aggregated. However, the CCPA establishes a very high bar for claiming data is de-identified or aggregated.

In practice, the only way to meet this bar will be through Privacy Automation: state-of-the-art risk assessments tools for the risk of identification; advanced privacy protection actions that retain the analytical value of datasets; and audit reporting. These and other techniques make up what is being termed 'Privacy by Design' and 'Privacy by Default.'

In the CCPA era, a manual, 'two eyes' approach to assessing the risk of re-identification won't cut it. The scale and the legal significance of proving privacy compliance under the CCPA is too great. Effective de-identification can be broken into three focus areas. You must:

- **Use a 'state-of-the-art' de-identification method.** You need a process whereby consumer and personal data (as defined under CCPA) is transformed so that this data becomes de-personalized. This practice is at the heart of meeting, demonstrating and defending CCPA privacy compliance. This has to include cutting-edge privacy protection tools that retain the analytical value of the data for data science, rather than data encryption tools that break the analytical value of the data.
- **Assess for the likelihood of re-identification:** Research in 2000 proved that 87% of U.S. citizens can be re-identified on the basis of their gender, ZIP code and age. Just de-identifying direct identifiers alone still leaves an individual at risk of being identified from other information, whether within or without the dataset. Demonstrating the risk of re-identification using automated 'state-of-the-art' tools must be prioritized as organizations can no longer depend on manual processes.
- **Implement Segregation of Duties:** Companies need to ensure that customer data is only shared with departments and individuals who have a legitimate purpose in receiving the consumer personal information. They need to implement appropriate controls so that segregation of duties exists, and so that data required for secondary purposes is truly de-personalized and thus CCPA-compliant.

Instead, organizations must look to invest in automation and leverage new tools that instantly assess the risk of re-identification. These tools constitute an automated system that makes compliance watertight, while approaches also offer the starting point for transforming the data that retains that much needed analytical value for data science, but is a key component of a privacy governance framework to demonstrate privacy compliance but to also easily defend privacy compliance, especially using automated and systems based approaches to the risks of re-identification.

Post CCPA, almost all privacy programs will require updating and modifying to accommodate the imposed requirements relating to CCPA, and to leverage the availability of new automated and state-of-the-art tools and systems.

These tools and systems will be ones that instantly assess the risk of re-identification, and constitute an automated system that makes compliance watertight, while also enabling data science by not ruining the insight value of datasets.

CryptoNumerics Privacy Automation helps this CCPA dilemma by;

- Promoting a better understanding of how it is possible to de-identify a dataset and still preserve the analytical value of the data for data science and analytics.
- Leveraging systems-based technology to assess the risks of re-identification of an individual or individuals in datasets.
- Applying modern anonymization protection using privacy protection actions such as generalisation, hierarchies, and differential privacy techniques to demonstrate that datasets are fully anonymized.
- Build Privacy Automation into the heart of your data compliance plans and strategy.
- Through data protection by default and by design, make privacy by default and by design the heart of your compliance strategy and plan and defend compliance with PIAs and Audit reporting.

About CryptoNumerics

CryptoNumerics, have developed a range of Privacy and Data Science related software solutions including **Privacy Automation** and **Virtual Data Collaboration** solutions that fortune 1000 enterprises are deploying to address privacy compliance while still driving data science and innovation projects to obtain greater business and customer insights.

www.CryptoNumerics.com | info@CryptoNumerics.com