PAPER

AWS Insider.net
THE INDEPENDENT RESOURCE FOR
AMAZON WEB SERVICES

PRIVACY AND ONLINE DATA: CAN WE HAVE BOTH?

By Peter Varhol

actian™
activate your data

www.actian.com

Significant change has arrived in how computing and storage consumes data concerning individuals. Merchants, data collection and aggregation companies are legally required to protect personal data. Those legal requirements are increasing this year with the enforcement of the EU General Data Protection Regulation (GDPR), which gives individuals the ability to better control their personal data.

Personal data or privacy data covers a wide range of topics including physical address, income, unique identifiers, credit instruments, bank details, and payment data. In addition to what is generally considered privacy data, it encompasses both professional and public data, such as a photo, an email address, posts on social networking websites, and medical information. This makes it far more encompassing than simple financial and contact information.

Merchants who process data have to handle it in ways that protect it. Those who store it, and those who aggregate it, and those who transfer it are all subject to provisions under the regulations. While it applies specifically to EU citizens, any company that collects or possesses personal data of an EU citizen is legally obligated to provide the same protections and services.

The GDPR was passed in April 2016, with a two-year transition period. It fully takes effect in May 2018. Once it does, companies can be punished for violations, in some cases with fines up to one percent of worldwide annual turnover.

GDPR and similar regulations have the potential to be the most drastic and far-reaching changes to personal data collection and management, ever. Hacks and other data breaches can lead to significant fines, and open the door to lawsuits from those whose data have been violated. There are many incidents in recent years where companies have become lackadaisical in safeguarding personal data, and may be in for a rude shock in the future.

YOU HAVE THE RIGHT TO BE FORGOTTEN

GDPR is an extension and generalization of the EU rule passed in 2006 known as the “right to be forgotten.” This right enabled EU citizens to petition to have specific personal data removed from various search engines and websites, even if that data were true. The EU

Cover Image: Blackboard/Shutterstock.com

recognized that not everything in the lives of its citizens should be searchable on the Internet, and that individual citizens should be able to make that decision. Of course, the reality is more difficult, as data flows to different places easily, but pulling it from search engines makes it far more difficult to find.

PAYMENT CARD INDUSTRY (PCI) REGULATIONS

Possibly the most important personal data for any individual is financial and payments data. There are few types of data that can cost individuals more by a loss. Money and credit information can be stolen and misused, leaving the individual with a big bill and a time-consuming journey to reclaim their identify.

THE GDPR FULLY TAKES EFFECT IN MAY 2018. ONCE IT DOES, COMPANIES CAN BE PUNISHED FOR VIOLATIONS, IN SOME CASES WITH FINES UP TO ONE PERCENT OF WORLDWIDE ANNUAL TURNOVER.

The notice requirements for data are expanded in GDPR. They include the retention time for personal data, and contact information for data controller and data protection officer has to be provided. People also have the right to opt out of certain types of data collection activities if they are not vital to the transaction. In all cases, they must be given the opportunity to explicitly opt in.

Citizens have rights to question and fight significant decisions that affect them that have been made on a solely-algorithmic basis. While it's not yet clear how this right will work in practice, in theory it provides an appeal process in case of biased data and poor algorithms. Further, it provides consumer protections far beyond what are available in most cases today.

The payment card industry consists of all the organizations which store, process and transmit cardholder data. PCI regulations are security standards to ensure that companies that accept, process, store or transmit credit card information maintain a secure environment. PCI in general refers to debit, credit, prepaid, e-purse, ATM, and POS cards and associated businesses that rely on non-cash means of settling debts.

The purpose of the PCI organization and regulation is primarily security; how will merchants and processers protect the private financial data of individuals? And if companies cannot secure personal financial data, what hope is there of protecting any data at all?

The security standards are developed by the Payment Card Industry Security Standards Council, which creates and promotes the PCI Data Security Standards used throughout the industry. The goal of PCI is to protect credit card numbers, transactions, and data from identity theft; and to ensure that merchants are held accountable for protecting data. The PCI group has enforcement responsibility, and annually hires outside firms for compliance verification.

PCI Data Security Standard specifies requirements for compliance that are organized into six logical groups called control objectives. These six groups are:

1. Build and Maintain a Secure Network and Systems
2. Protect Cardholder Data
3. Maintain a Vulnerability Management Program
4. Implement Strong Access Control Measures
5. Regularly Monitor and Test Networks to ensure they are functional and secure
6. Maintain an Information Security Policy

While each of these is complex in terms of definition and implementation, they are conceptually easy to understand. In sum, the PCI standards attempt to make sure that everyone who collects, stores, and processes protects data unequivocally, and through a separate secure network that is tested on a regular basis. The data requires strong access controls to protect it from unauthorized access and use, and all companies have to have policies that define the scope of data they collect and/or aggregate, and how that data is stored and managed.

REPORTING AND DATA BREACHES

Further, there is a requirement for an Information Privacy Officer, located at the company headquarters or principal city of operation in the EU. This person is responsible for managing the privacy program, including development of policies and procedures and reporting of breaches. Reporting policies have yet to be established, so it's not yet clear if there is to be ongoing reporting or simply exception reporting.

And speaking of breaches, any unauthorized access to privacy data has to be reported promptly, usually within 24 hours. The days of companies waiting months or longer to report a breach of privacy data without consequence will be in the past. In addition, in the past many breaches have not been reported at all, and hackers have been paid off to return stolen data.

The intent is to provide consumers with positive control over their privacy data. GDPR establishes that individuals own their own data, not collectors or aggregators, and can determine what data is stored by which entity. While there are many circumstances where companies should be able to possess privacy data for legitimate purposes, GDPR shifts control back to the individual, whose data it is to

begin with. In contrast, in the United States, data collectors and aggregators claim ownership over the data.

The responsibility of the company is to provide clear and easy-to-use opt-in and opt-out options for the individual on data held by the company, and to explain how the data will be used. Any attempt to circumvent either directive will run afoul of GDPR.

sounds straightforward enough, the potential for errors in selecting data, encrypting data, transferring data, and decrypting data for processing makes processing data both complex and highly susceptible to error or breach.

Even changing or deleting records can be affected by GDPR. Deletion typically follows when the individual has opted out

GDPR DOESN'T PREVENT THE USE OF ANALYTICS ON PERSONAL DATA, BUT IT DOES REQUIRE THAT INDIVIDUAL DATA NOT BE IDENTIFIABLE IN THE PROCESS.

PROCESSING PERSONAL DATA

Last, there is the question of analyzing personal data without exposing the individual involved. GDPR doesn't prevent the use of analytics on personal data, but it does require that individual data not be identifiable in the process. In practice, this may require masking the name field along with other identifiable information, or it may require querying the primary database and moving results to a special analytic engine. The data can be processed, as long as individual records aren't revealed.

What makes this even more of a challenge is that data is typically pulled in from multiple sources that are also protected under GDPR. Transferring data between data sources requires both secure channels and encryption. While it

of a particular service or database. In these cases, all copies of the full record must be deleted, and none of the record's data can be used in further processing. This means that the record copy, or component of a record, must be confirmed as being deleted.

In many cases, this might require a formal audit to ensure that all copies and related data are actually deleted and no longer available. It is probably prohibitive to have an audit to remove individual records, so additional technology will likely be required to automate and confirm deletion.

PROBLEMS AND SOLUTIONS

Giving individuals control over their privacy data, and expanding the scope of privacy data, will have serious consequences on how organizations

store, manage, and use that data. Traditional methods and technologies of data management simply will not be effective at complying with GDPR. Adding, storing, transferring, processing, and encrypting data will be a significant challenge using existing data solutions.

ORGANIZATIONS COLLECTING OR HOLDING PRIVACY DATA HAVE TO RETHINK THEIR ENTIRE WORKFLOW AND TOOL CHAIN FOR THIS DATA.

Companies in the business of collecting, aggregating, or holding privacy data are going to have to consider storage solutions that are more amenable to adjustment and modification. In most cases, the commonplace SQL-based relational database alone may not be a viable solution. At the very least, significant modifications to technology and workflow will be required.

In practice, there are a number of ways that companies can manipulate and analyze data without violating the principles of GDPR. A company may be required to change or add to an existing technology stack in order to meet the strict requirements embodied by GDPR. For example, it is possible to perform analytics without moving data away from the main transactional database. This approach obviates the need to encrypt data to move it from one storage system to another.

Organizations collecting or holding privacy data have to rethink their entire workflow and tool chain for this data. At front and center is the database used to store and process this information. Complying with GDPR will be much easier with the right software tools in place.

In particular, the database needs to be able to have the ability to store diverse types of data, beyond numbers or text. A lot of privacy data can be unstructured, such as a photo or other binary object. Being able to manage these as a part of an individual record, rather than in separate entities and storage locations, reduces complexity and makes the correct outcome more likely.

Performance should also be a consideration. Databases that offer fast analytics can keep data for less periods of time, improving overall security of that data. And while performance can also provide other advantages, such as faster access to information, fast processing is definitely helpful in building a compliant solution.

ACTIAN DATABASE SOLUTIONS AND GDPR

Some databases are adapting to the different requirements of GDPR. Actian X increases control by allowing organizations to store data in traditional relational

MANY HADOOP BASED BIG DATA STORES DON'T DO A GOOD JOB OF SUPPORTING THE ABILITY TO EASILY DELETE RECORDS AND CONFIRM THE DELETION, WHICH IS A VITAL PART OF ANY GDPR SOLUTION.

row based tables and analyze the data at high speed using the built-in Actian Vector analytical database engine. The University of Oxford's Clinical Trial Service Unit does just this by tracking 5,000 data points for 500,000 people and analyzing the resulting billions of data points without moving the base data, increasing control and protecting the privacy of personal data. Vector processing enables an entire column of data to be analyzed in a single CPU instruction as a stream.

Many Hadoop based big data stores don't do a good job of supporting the ability to easily delete records and confirm the deletion, which is a vital part of any GDPR solution. These databases simply flag deleted records and continue to retain them. Actian Vector helps here too. To support the right to be forgotten, records can be more definitively deleted. When a record is fully deleted in Vector, it is erased.

If your company is looking for ways to increase data control while gaining analytics performance in order to comply with GDPR, give Actian Vector a test drive. You can download a free copy by visiting www.actian.com/vce.