Cloud App Data Privacy to Comply with GDPR

Eugene Alooeff[1]*

[1]*ATEK, R&D Department, Wroclaw, Poland*

Abstract

In this paper we analyzed current law for data protection, requirements and trends in cybersecurity. We developed the data schema and algorithms to let the cloud applications comply with data protection regulations in the European Union. This solution had practical implementation on the health center cloud infrastructure. Security assessment has been passed and the system was successfully launched in production.

Keywords: *GDPR, PII, Data Privacy, Security, Compliance.*

1 Introduction

The EU General Data Protection Regulation (GDPR) is enforced across all EU Members from 2018, is a landmark in the evolution of the European privacy framework [1]. This law covers the personal data of all EU residents, regardless of their processing location. Personal data is information that, directly or indirectly, can identify an individual and specifically includes online identifiers such as IP addresses, location data. GDPR aims to bring a single standard for data protection among all EU member states, and applies to entities that operate in the EU or deal with the data of any EU resident, regardless of where the data is processed. This is much wider than the concept of personally identifiable information (PII) in US privacy law [2].

Depending on where your organization operates and what data stores, you may need to adhere to location-specific privacy laws. For example, if your organization operates in Turkey, you are subject to the Law on the Protection of Personal Data [3] and Regulation on Deletion, Destruction or anonymization of Personal data [4].

Globally, there are two types of privacy laws: comprehensive (applicable to all industries and sectors) and sectoral (applicable to specific industries or sectors). In the US, the federal government has historically taken a sectoral approach. For example, there's the Health Insurance Portability and Accountability Act (HIPAA) [5], which is the US healthcare privacy law protecting data that reveals the health status of an individual.

Cybersecurity compliance involves meeting various controls to protect the confidentiality, integrity, and availability of data. Regulatory compliance insists that the organizations should follow local, state, federal, and international laws and regulations relevant to its business function.

Many existing cloud operators process data submitted by customers for the purpose of providing online services to them. To fulfill these purposes, cloud operators may access the data to provide the services, to correct and adress technical issues. To show the compliance to the law, the can implement the Data Privacy Framework (DPF) program and publish the notice of certification under it [6].

2 Apply data privacy principles

There are many key privacy principles. In this work we take into account some of them:

Security principle - Anonymization technique - the process of protecting private or sensitive information by erasing, obfuscating (masking), or encrypting it. It removes all identifiers associated with a person [7]. If data is truly anonymized, then the data does not constitute personal data under the GDPR. However, the bar to be considered anonymous is high: It must be impossible for any individual to be identified from the data by any further processing or by combining it with other information.

Data Deletion and Retention - Organizations should only store personal data for as long as it's required and for the originally intended purpose. Organizations should not keep any personal data for an indefinite period even if it may be used in the future. Clear time frames should be established for when data is deleted with rationale for why the data

*Contact email: alooeff@gmail.com

is retained for that length of time. For example, you may need to retain security log files for certain periods of time to identify and track malicious adversary behaviors. However, the period still must remain finite, with supporting rationale. You should also be aware of data retention laws for specific types of data, such as legal documents, within the country where your organization provides that service.

3 GDPR and privacy by design

Privacy by Design is a key concept of the GDPR and is made a legal requirement. Privacy by Design means thinking about data privacy and its implications when you're developing products, features, even marketing campaigns based on personal data. It also means encouraging employees to ask themselves questions before collecting or using data:

- Do I need all the data I'm collecting here?
- Could I do this work without using personal data at all?
- Am I using the data in a way a user may not expect?
- And do I have a plan to delete this data once myself or my team no longer need it?

The GDPR also encourages organizations to document key privacy decisions they make around the collection, use and storage of personal data. Documenting compliance with the GDPR may be one of the most challenging and time consuming aspects of this law. There are several ways an organization can demonstrate and document compliance. Your organization may need to complete a privacy review process of products or features to ensure GDPR compliance before they go live [8].

This is often referred to as a data protection impact assessment or DPIA. DPIA's help document key decisions within an organization that have a privacy impact. companies should also inventory the personal data it stores and collects. Also companies should. update its existing policies and procedures or develop new ones that outline how personal data will be protected, deleted and processed.

4 Purpose

Since companies work with customer's data, they are personal data processors. They perform such operations on personal data, as collection, storage, alteration, retrieval, erasure or destruction. From the moment when any personal data gets uploaded to the company databases, they become a Processor of personal data. Under the GDPR, it is also mandatory for processors to designate a Data Protection Offices (DPO) [9].

To let Processors to comply with security requirements, we developed objects schema and the algorithms, implemented and tested them on practice.

5 Algorithm

To let Processors to comply with security requirements, we developed objects and the methods and tested them on practice.

Developed algorithm and objects can be implemented either as a part of internal system or as external system and work through API. To implement GDPR requirements in part of data retention and deletion / anonymization, we have to use 2 objects:

- Wipeout Policy object to store policies for each object in the database we have to apply the logic.
- Wipeout Policy Count object to store the history of the number of records affected by Wipeout Policy.

The records in the first object (Table 1) are being created and updated by the system administrator in accordance with the current database used in company and do not depend on the IT infrastructure.

Table 1. Wipeout Policy object definition.

Field Name	Type	Value Example
Name	text	Customer wipeout policy
Query	text	CLEAR UserName, Phone, Email FROM Appointment WHERE (Status = 'Closed') AND End < LAST_N_DAYS:180 AND End > LAST_RUN
Object Name	text	Customer
Wipeout Action	picklist	Clear_Fields / Deletion
Fields	Text	Name, Phone, Email, Priority
Values	text	null, null, null, normal
Retention Period	number	180
Wipeout Criteria	text	Status = 'Closed'
Active	boolean	true
Bypass Trigger	boolean	true

Each record is a rule, used by wipeout logic and corresponds to one Object (**Object Name** field). If

Wipeout Action is 'Clear_Fields', then the '**Fields**' and '**Values**' have text (comma separated values) about what object fields should be cleared (filled with null or blank values) or set to default value in case of picklist. If **Wipeout Action** is 'Deletion', then the records are being deleted. '**Wipeout Criteria**' field stores the rule to select the processing records by specific criteria. For example, for closed accounts, finished orders or paid bills. The format of this criteria should be just the WHERE condition for SQL query. In addition, '**Retention Period**' field shows the period when object records should remain unwiped. It needs to have the ability to control or audit closed deals for retention periods. **Query** field is generated on the policy criteria and retention period in SQL style. It is used on the dashboard to show how the set of records is selected. Finally, '**Bypass Trigger**' field shows whether the object trigger should be switched off during the object record processing (deleting or updating operation) to avoid additional logic run, as standard field validation or sending notifications.

The records in the second object (Table 2) are being created by Wypeout Scheduler and updated by Wypeout Batch. Each record is information about the daily run of Wipeout Batch and the result of its work. **Query** field is a copy of Wipeout Policy Query field in the moment of running the Wipeout Batch for future analysis if needed. **Before Count** and **After Count** fields show the number of fields in the **Query** before wipeout action and after. In the ideal situation (when there are no any errors) **After Count** should be always zero. **Start Time** field allows to keep the last run time to eliminate processing the records already processed previously. Dashboard uses the information from this object to show the history of batches that were run for the last days and create a chart for brief overview.

Table 2. Wipeout Policy Count object definition.

Field Name	Type	Value Example
Name	text	Customer wipeout policy
Policy Id	Id	
Batch Job Id	Id	
Query	text	CLEAR UserName, Phone, Email FROM Appointment WHERE (Status = 'Closed') AND End < LAST_N_DAYS:180 AND End > LAST_RUN
Object Name	text	Customer
Wipeout Action	picklist	Clear_Fields / Deletion
Before Count	number	1234
After Count	number	0
Start Time	datetime	
End Time	datetime	

To implement GDPR/HIPAA requirements in part of data retention and deletion / anonymization, we use 2 code blocks.

Wipeout Scheduler is the algorithm, which is scheduled to run daily. It gets all the active Wipeout Policies, and do such steps for all of them:

- Find the newest record in 'Wipeout Policy Count' object for current policy Id. Record should have Batch Job Id and Run Time End value.
- Create filter for the records to process by merging the criteria in 'Wipeout Policy' with the Start Time in 'Wipeout Policy Count' record.
- Get the current number of records in the DB which match the policy criteria.
- Create a new 'Wipeout Policy Count' record with 'Query', 'Before Count' and 'After Count' values, received in the previous step.
- Run Wipeout Batch for current policy.
- Repeat the logic for the next policy

Wipeout Batch is the algorithm, which process records in the next steps:

- Get the chunk of records based on policy and retention period criteria.
- Deactivate object trigger if needed by policy.
- Run the logic depending on Wipeout Action (clear fields or delete records). – Get 'Wipeout Policy Count' record related to this Batch Job Id.
- Update 'Run Time End' field with actual date/time and 'After Count' field with the number of the records failed during clearing or deleting.
- Repeat the logic for the next chunk of records.

6 Results

System administrator has access to the dashboard (Fig. 1), which shows the list of wipeout policies with SQL to retrieve the record to process (Description field) and the chart for the last month. On this chart we can see 2 graphs for 'Before Count' and 'After Count' values. In an ideal situation, the last graphs should always show zero values (all the records were processed successfully).

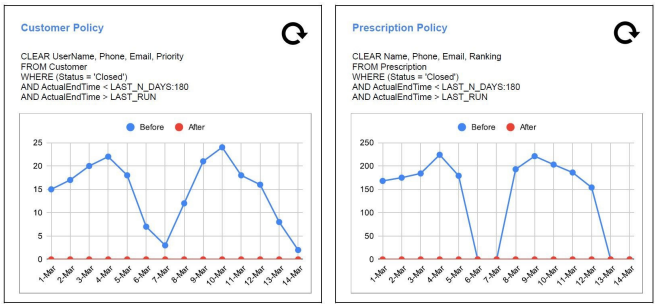

Fig.1. Dashboard

Retention period and Fields for wiping out are easily accessible to updating by system administrator and there is no need to change or reschedule the code logic. Being scheduled once, Wipeout Scheduler runs Wipeout Batches daily and puts the result into the Wipeout Policy Count object for easy monitoring by Dashboard.

7 Implementation in practice

Using a cybersecurity assessment and a structured approach to identifying regulations and implementing controls we have successfully implemented algorithms described above and tested them on the Health Care center. Internal audit of implementation and half-a-year run showed that after the retention period all obsolete sensitive information in the company database is being erased successfully in accordance with the internal documents related to GDPR. Dashboard shows the number of records processed and is easy to read and monitor the work of algorithms. As a result, implemented solution meets cybersecurity regulations and safeguard customers against cybersecurity attacks.

8 Trends in cybersecurity compliance

Trends in cybersecurity compliance and regulation are affecting organizations everywhere. Today's regulatory environment is more challenging than ever.

Cybersecurity Compliance Is Not Just an Information Technology Issue. Many fear cybersecurity compliance as an amorphous issue that only the information technology department handles. The reality is that the financial, legal, and reputational ramifications that arise from a data breach affect the entire organization.

If standards, regulations, or rules are too burdensome, organizations and their employees may engage in insecure workarounds to meet business needs. It's important to communicate to staff how to attain security without compromising business goals.

As guidance and regulations evolve quickly, organizations may not have the resources to keep up with changes year over year. Cybersecurity is a fast-moving sector, as both attackers and security providers vie to outsmart each other.

Gone are the days where organizations used a spreadsheet to document their controls and then had to copy, paste, and repeat against other regulations. Today, organizations use automation to document once, and then comply against many different standards. They use automated tools to collect evidence—such as vulnerability scan results and analysis, log correlation, and more—to assess their security posture, add workflow to task out items such as patching, and report to senior executives.

Automation helps streamline compliance especially with engineers using compliance as code where they design system and products to automatically meet control objectives. One example of automating compliance is the National Institute of Standards and Technology's (NIST's) Open Security Controls Assessment Language (OSCAL) [10].

The recruitment, retention, and training of the right security personnel to help meet compliance requirements is key. Ideally a combination of compliance analysts, security operations specialists, and software engineers make up these teams. Because the burden of regulatory compliance on organizations can be huge, organizations need to create focused cybersecurity teams that work in tandem with their internal risk teams. Organizations must hire more staff who are well-trained on the security aspects related to their industry.

With the economic effects of the COVID-19 pandemic still in full force, organizations are looking to contain costs by working with third-party vendors and partners on an ongoing basis.

One area organizations may look to outsource is compliance. This may be especially true to help perform gap assessments and meet reporting requirements, particularly if a company is small, doesn't have the right staff in house, or needs a small increase in effort for a short time to meet a new regulation or answer a point-in-time audit. As outsourcing compliance becomes more prevalent, frameworks and vendor assessments will also need

to raise the bar for third-party vetting to ensure a degradation in security doesn't occur [11].

References

[1] REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679

[2] Guidance on the Protection of Personal Identifiable Information https://www.dol.gov/general/ppii

[3] Law on the Protection of Personal Data N 6698 https://www.kvkk.gov.tr/Icerik/6649/Personal-Data-Protection-Law

[4] KİŞİSEL VERİLERİN SİLİNMESİ, YOK EDİLMESİ VEYA ANONİM HALE GET?R?LMES? HAKKINDA Y?NETMEL?K ttps://www.resmigazete.gov.tr/eskiler/2017/10/20171028-10.htm

[5] Health Insurance Portability and Accountability Act https://www.hhs.gov/hipaa/index.html

[6] Data Privacy Framework (DPF) Program https://www.dataprivacyframework.gov

[7] Anonymization. Imperva Learning Center https://www.imperva.com/learn/data-security/anonymization/

[8] Nishant Bhajaria. Data Privacy. A runbook for engineers. 2022 https://www.manning.com/books/data-privacy

[9] Guidelines on Data Protection Officers https://ec.europa.eu/newsroom/article29/items/612048

[10] OSCAL: the Open Security Controls Assessment Language https://pages.nist.gov/OSCAL/

[11] Forbes: Cybersecurity Compliance Trends in a Post-Pandemic World https://www.forbes.com/sites/forbestechcouncil/2020/05/11/cybersecuritycompliance-trends-in-a-post-pandemic-world/?sh=4a62bd6f4ae2