Complying with European Data Protection Board Supplementary Measures with Anonomatic PII Vault

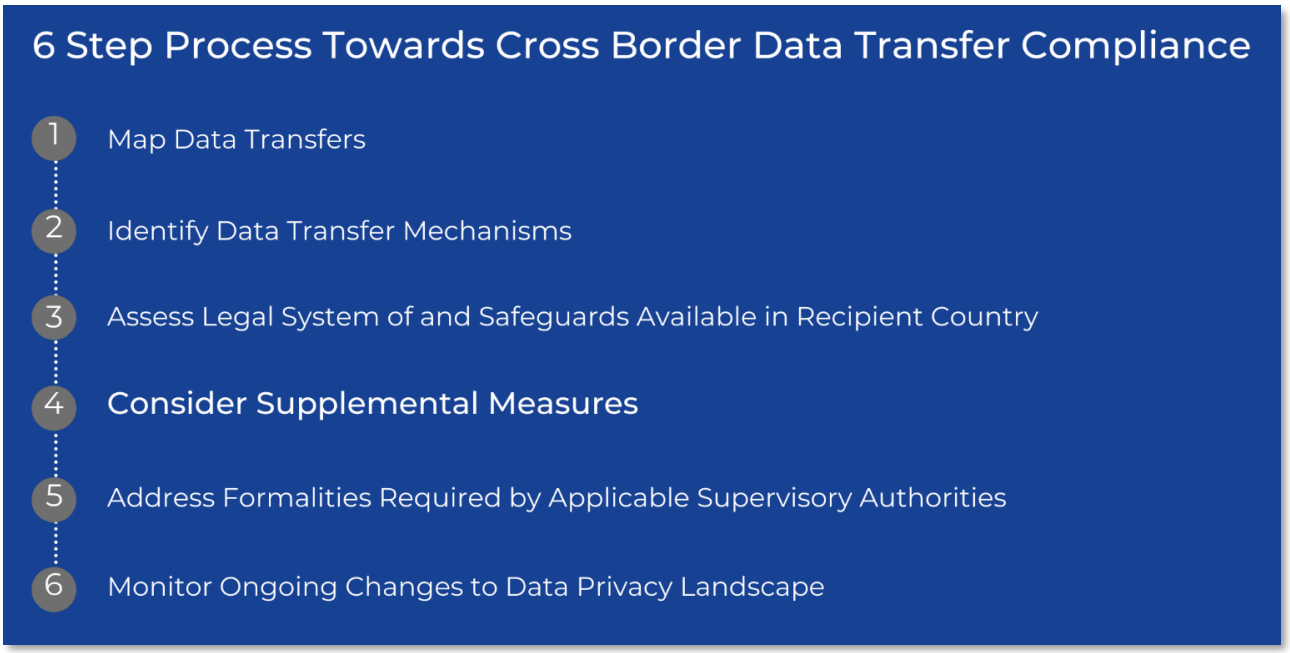
anonomatic

Table of Contents

Introduction

On July 16, 2020, the Court of Justice of the European Union (CJEU) sent shockwaves through the business community with their landmark ruling in the Schrems II case (C-311/18). With this judgement, the CJEU invalidated the EU-U.S. Privacy Shield framework. Almost a year later, on June 21, 2021, the European Data Protection Board ("EDPB") published their final recommendations on supplementary measures. This long-awaited documentation provides the requirements of international transfer safeguards, such as Standard Contractual Clauses ("SCCs"). The final recommendations provide a six-step process for compliance.

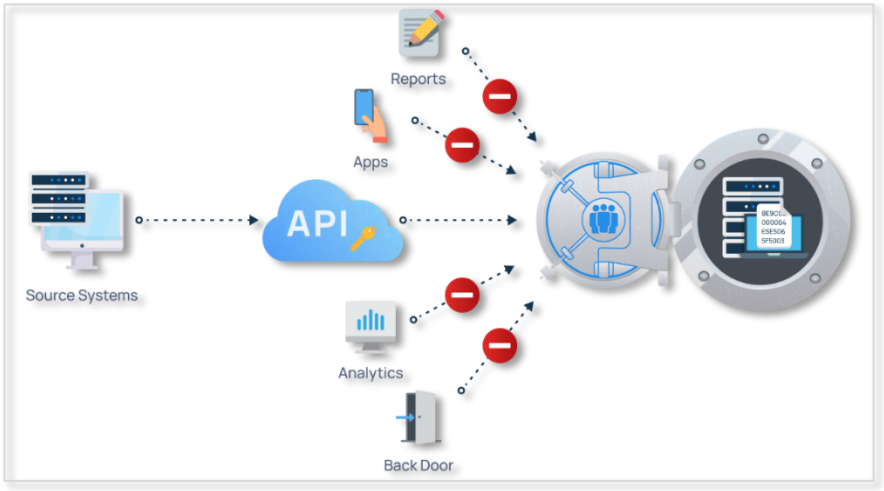
This white paper concentrates on #4 (Supplemental Measures) and how organizations can comply with these recommendations using Anonomatic® PII Vault™ and following best practices[1].



A Simple API Can Solve Data Privacy Issues

When introducing our technology to a new audience, Anonomatic has found it best to recommend the audience start by discarding everything they have ever read, seen, or done regarding how to implement data privacy. This is because we believe our approach is completely different from every other product in the market.

PII Vault is delivered as an Application Programming Interface (API) which can be utilized in a variety of different methods to achieve different levels of protection for identified data. In this document, we provide a guideline for how PII Vault can provide all the protection necessary for organizations within the EU to share data outside of the EU without the risk of loss or exposure of personal information or personal data (PII). We will also outline an approach by which standard business processing and data science operations can be done on the combination of EU-sourced data when matched with non-EU sourced data, in a manner that conforms to this same level of protection. Finally, we will demonstrate how the results of these operations can be returned to the source of the identified data, so organizations can maintain and update their own data repositories, with minimal risk of exposure of the identity of any of their data subjects.



Only one way to access PII Vault capabilities or data

As an API, PII Vault has no user interface, no reports, no analytics, and provides no user access to the data it protects. There is one and only one way to access any of the capability or data of the PII Vault and that is through authenticated API calls. There is no external access to any data in the PII Vault. The data in PII Vault is uniquely protected as described in the following sections.

PII Vault Beginnings

The Los Angeles Unified School District (LAUSD) is the second largest school district in the United States. It services over 600,000 students every year. Many of these students have little or no access to healthcare. As a result, LAUSD has over a hundred varied healthcare service facilities spread across their campuses. These facilities include a collection of School-Based Health Centers (SBHCs) operated by independent service providers.

The Los Angeles Trust for Children's Health (The Trust) was established in 1991 through a resolution of the Los Angeles Unified School District Board of Education to improve the health of LAUSD children and families and to provide support for the district's growing number of school-based health centers. The Trust is responsible for overseeing all the healthcare services provided through a variety of healthcare service providers who directly service the students of LAUSD. Starting in 2009, The L.A. Trust partnered with LAUSD to develop 17 Wellness Centers on the campuses of the highest-need LAUSD high schools and middle schools.

In early 2017, The Trust initiated a very ambitious project known as The Data xChange. At that time The Trust contracted with Anonomatic's founders to design and build a first-of-its-kind analytics solution with the goal of enabling researchers to identify potential links between the healthcare provided to students and their academic performance. While it may seem intuitive that a student who is not suffering from asthma or tooth-pain would have a higher attendance rate and high concentration levels, no one had ever been able to quantify a direct relationship correlating healthcare with academic performance.

To accurately identify relationships between healthcare services received and academic performance, detailed data from both the healthcare providers and the school district must be collected and merged at the individual level. Collecting and storing this level of extremely sensitive data was a daunting prospect, especially since the goal was to store this data indefinitely and to do so on an extremely limited budget.

THE L.A TRUST'S DATA XCHANGE IS DESIGNED TO MEASURE THE IMPACT OF WELLNESS CENTER INVESTMENTS, IMPROVE THE ALLOCATION OF HEALTH RESOURCES, AND CONNECT THE DOTS BETWEEN STUDENT HEALTH AND ACADEMIC ACHIEVEMENT.

The result was the development of version 1.0 what is now PII Vault. Following the successful implementation of this new technology for The Trust, Anonomatic has released version 2.0 to the general marketplace.

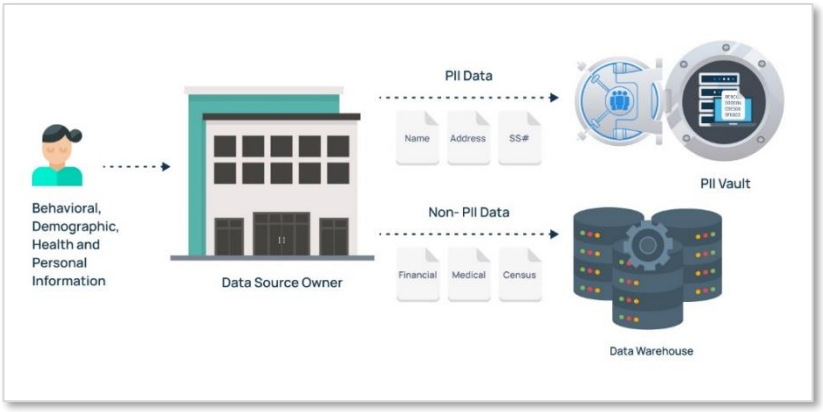
How PII Vault Operationalizes Identified Data While Protecting Privacy

Prior to PII Vault, solving the problem encountered by LAUSD would require collecting and correlating detailed and identified academic and healthcare data. Rather than mimic current approaches to this problem and inheriting the accompanying costs, limitations, and risks, all of which can be severe, PII Vault implements an entirely new, different, and cost-effective approach to achieving the goals of this project.

It starts with a simple observation: when identified data is split so that the identifiers are completely separated from the details they identify (fact data), the cumulative risk of both the identifiers and the underlying fact data is a fraction of the risk when those parts were combined. (A video describing this process is available on Anonomatic's website.)

Step 1– Data Separation

For the first part of the separation process, the identifiers, and only the identifiers, are collected by the PII Vault. PII Vault stores these details and returns a Poly-Anonymous Id (Poly-Id), as described in Step 2.



PII Vault collects only identified data

For example, imagine Jane Doe is a student who went to Planned Parenthood Los Angeles (one of the SBHCs) for reproductive advice or services. The details of Jane's visit, even that it ever occurred, are extremely sensitive and could materially harm Jane if ever exposed. However, if Jane's identifiers are separated from the underlying medical fact data, then the fact data of her medical visit is much safer because it is de-identified. Jane's identifying information is still sensitive but when there is no common, shared value that links her identifiers to the medical record, then there is no way, based on the data possessed by the data processor, to identify Jane, and thus this data has a much lower risk profile. This is what PII Vault does.

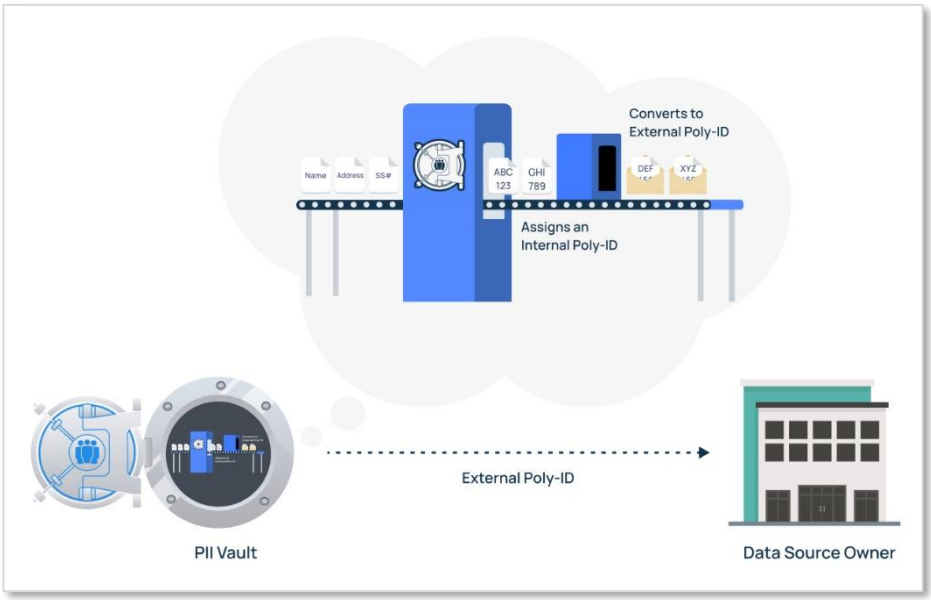
There are two ways to implement this data separation. The simplest is by using a patent-pending technology we refer to as Pass-Through Anonymization. With Pass-Through Anonymization all the fully identified data is passed through the PII Vault along with a definition of what data should be considered PII and which values should be redacted or masked. The result is pseudonymized fact data.

The second approach is for the users of PII Vault to manually separate the identifiers from the fact data, send those identifiers to PII Vault and exchange them, in the data, with the Poly-Id received from PII Vault. Either process achieves the desired result of fully pseudonymized fact data safe for sharing or use.

Step 2– Poly-Anonymization™

Poly-Anonymization™ is a patent-pending methodology that utilizes a Poly-Id, which is a multi-value, pseudonymous identifier. The concept behind Poly-Ids is that no specific value representing an individual is ever in more than one database at a time. This means that no matter how many databases might get breached in a concerted attack, no hacker could link the different records because there is no common value to link them.

When going back to Jane's medical record, the service provider sends her name and other identifiers to PII Vault, the vault stores those identifiers and assigns it an internal value such as "ABC123". It then converts the internal value to an external value such as "DEF456". The "DEF456" value is returned to the service provider and not stored in PII Vault. It should be noted that real Poly-Id values are much more complex and are known as Globally Unique Identifiers (GUIDs).[2]



PII Vault converts identified data into an internal and external Poly-Id

There are two very important parts of this process. First is the data processor never receives any actual identifiers. All the processor ever receives from any data source is completely pseudonymous data. Second, and just as critical, every data source will get a different Poly-Id value for the same individual. To continue the prior example when Jane's academic data is

pseudonymized, PII Vault will create both a new internal (ex: ”GHI567”) and external value for Jane (ex: ”XYZ789”).

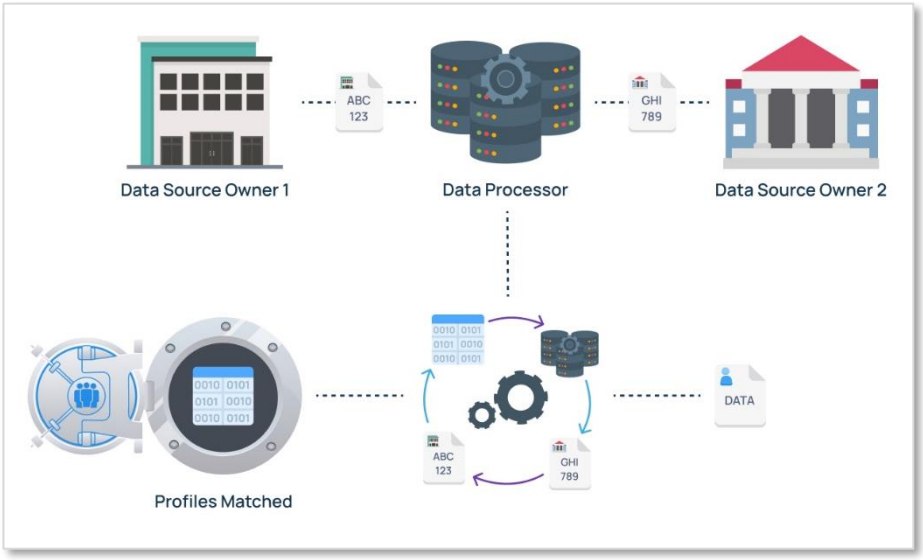
As a result, once the data processor receives the academic data, the processor will have two values for Jane, in this example “DEF456” and “XYZ789”. These two values are completely different from the values in PII Vault (“ABC123” and “GHI789”).

At this point, the data processor has data from two sources with data records that include Poly-Id values of “DEF456” and “XYZ789.” Because these two sources sent different values to represent the same person, there is no way for the data processor to link the data with these different Poly-Ids.

Step 3– Anonymous Data Matching

The goal of combining and operating on this pseudonymized, multi-sourced data is not possible until PII Vault is again used. PII Vault’s Anonymous Data Matching capability provides the means to link multi-sourced, pseudonymous data with Poly-Ids.

In this next step, the data processor makes a request to PII Vault to obtain a matching table. When this request is made, PII Vault verifies the requestor is permissioned, by the data sources, to request their Vault which will also exclude any individual who has requested their data not be used in this process (i.e., Right to be Forgotten).



PII Vault links multi-sourced, pseudonymous data with Poly-Ids

After performing these verification steps, PII Vault executes a series of matching procedures.

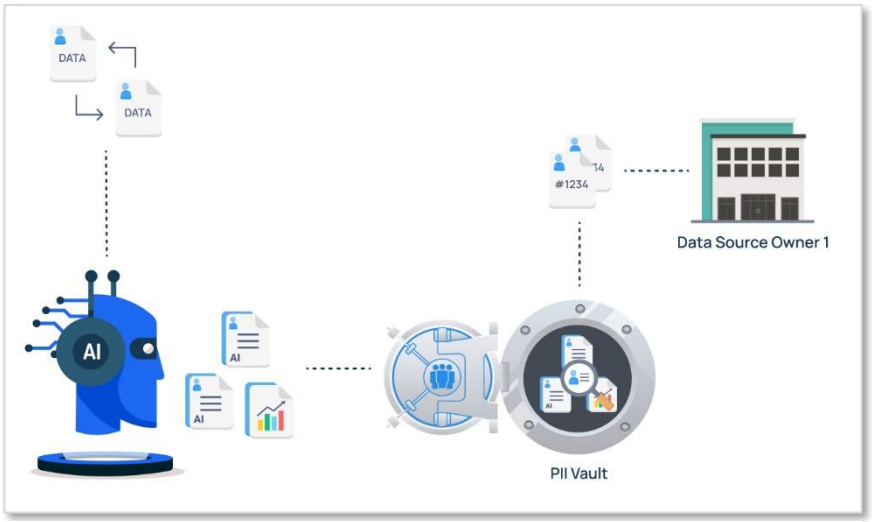
These matching procedures are used to link the PII profile from one source to matching profiles from the other sources. These matching procedures are completely configurable by the data processor and provide a means by which they can attribute a level of confidence to each match.

The match results the data processor receives from PII Vault is a table which includes a pair of Poly-Ids. These are the same values the processor received from the data sources. The pre- defined confidence level, and other details to make the links fully useable, also accompany each pair of Poly-Ids.

With the pseudonymous data from the medical clinics, the school district, and the matching table, the data processor can now link all the data from all sources at the individual level without ever having received any actual identifiers. At this point, the processor has everything it needs to execute the analysis they have been tasked to perform.

Step 4– Anonymous Results Delivery

Being able to safely work on pseudonymous data, from multiple sources and combined at the individual level, is a new and powerful capability that the PII Vault provides. However, being able to return the results of data science operations on that data can also be extremely meaningful. This is accomplished with the Anonymous Result Delivery capability of PII Vault.



PII Vault receives individual result, reidentifies and sends back to Data Owner

Anonymous Result Delivery works in a very simple and secure flow. The data

SUMMARY

PII Vault minimizes PII exposure by:

1. Providing a unique and pseudonymous value so data can be fully pseudonymized before it is shared with a data processor.
2. Enabling a data processor to link pseudonymous records at the individual level without ever receiving identifying details.
3. Optionally, re-identifying an individual result from the data processor back to one or more of the sources who provided the initial data.

processor provides an individualized result back to PII Vault. PII Vault cannot read this result. PII Vault's makes the result available for delivery to the intended data source. Because the data source does not store any Poly-Ids, PII Vault re-identifies this result so the data source can apply the results to their own records. PII Vault accomplishes this task by replacing the Poly-Id with a unique value provided by the data source, for example the individual's account number.

How PII Vault Can Be Used as an EDPB Supplemental Measure

There are many ways to implement PII Vault. Adhering to the best practices is critical to ensuring the proper level of protection is provided within the process of sharing of identified data. To illustrate how this could work, consider the following scenario: A company in France is in a data sharing agreement with a company in the US. Their arrangement allows all customer activity to be combined and analyzed so the most relevant marketing offers can be provided to these customers.

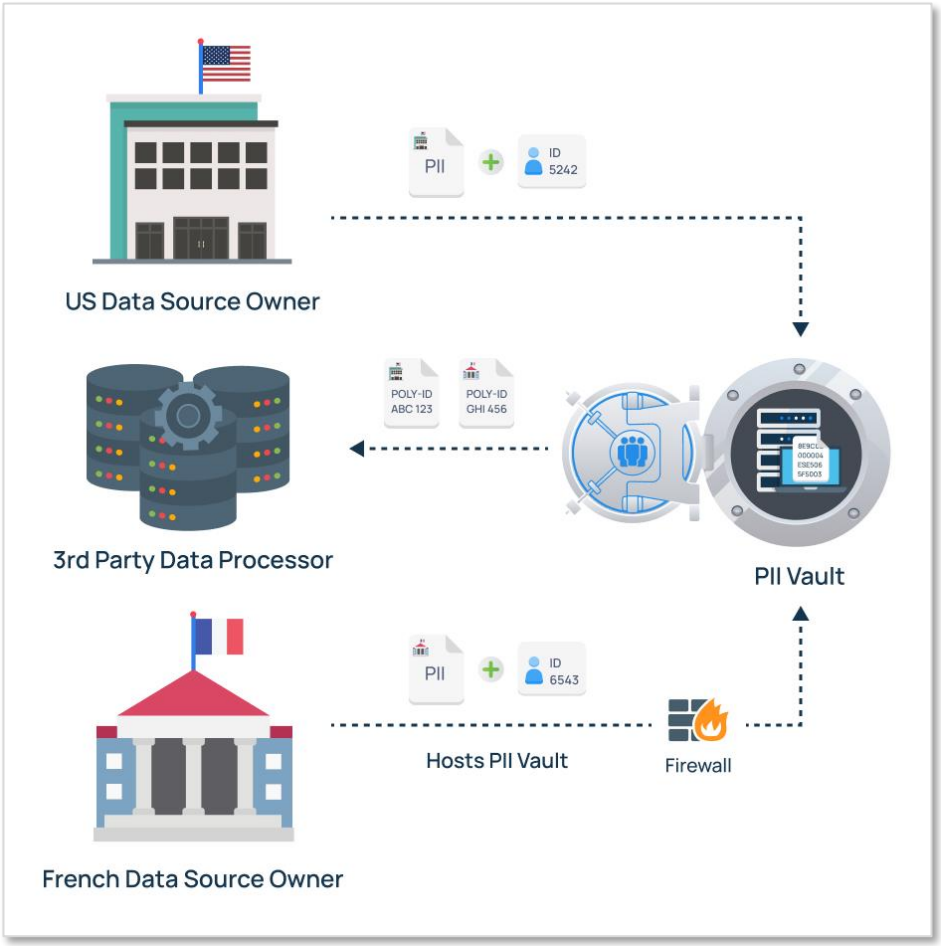
Step 1– Define the Participants

In this case, the best approach would be to have three participants:

1. The French company
2. The American company
3. A third-party data processor (an organization who has no access to the source data from either of the other participants)

Step 2– Setup the Environment

Due to the French rules being more stringent than the American rules, it would make sense to host PII Vault in France, most likely behind the firewall of the French company. From a technical standpoint, it does not matter where PII Vault resides because no entity, not even Anonomatic, has direct access to the data inside PII Vault.



Three PII Vault accounts created; PII Vault hosted behind firewall by French company

There would be three

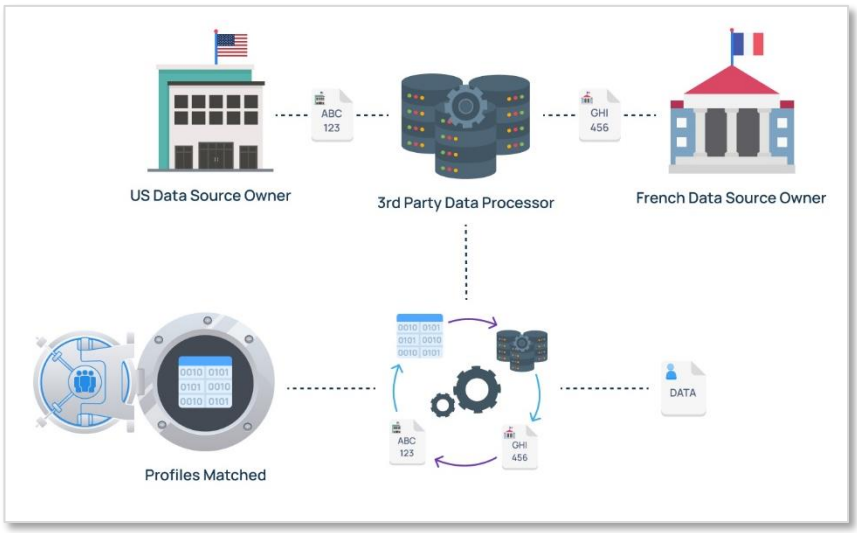
accounts created in PII Vault. One would be for the French company, one for the American and one for the data processor.

Step 3– Provide the Data Processor with Pseudonymous Data

The French and American companies would use PII Vault to pseudonymize their data independently of each other. Included in the PII is that company's unique and identifying account number for the individual customer. The companies would take the resulting Poly-Ids and send those along with only pseudonymous fact data to the data processor.

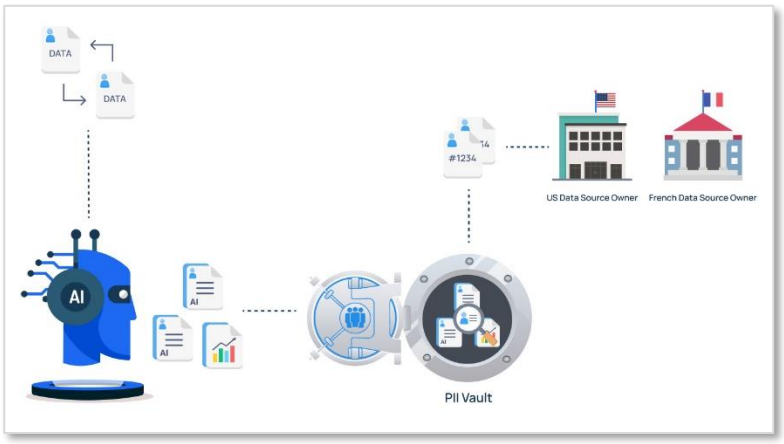
Step 4: Process the Pseudonymous Data

The data processor receives the pseudonymized fact data from both companies then obtains the permissioned matching Poly-Ids from PII Vault. They use this matching table to combine the pseudonymous data from the French and American companies which they then process. They never received any identifiers and have no means to directly re-identify any of the data. Once their processing is complete, they may have generated the pseudonymous, individualized results.



Data processor uses PII Vault matching table to combine pseudonymous data from French and American companies

Step 5: Return the Pseudonymous, Individualized Offer



The data processor returns a completely pseudonymous, but individualized, result to the appropriate source of the data through PII Vault. Each individualized result would have the unique Poly-Id from the source of the data. PII Vault would replace the Poly-Id with the account number of the individual as provided by the source company and return that result back to the source.

Summary

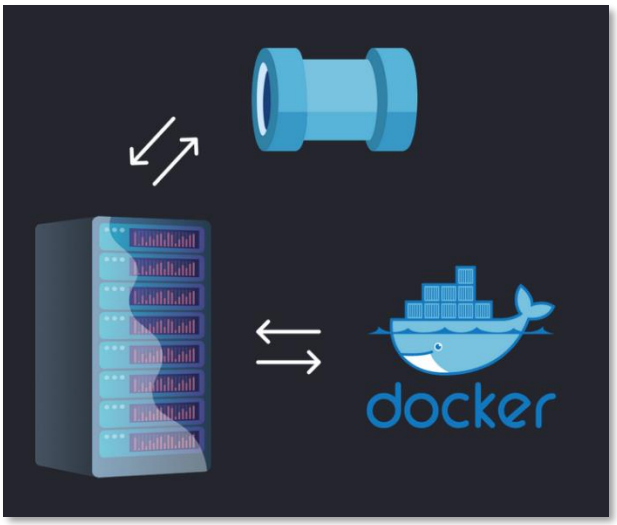
In this process both the French and the American companies have not had to share identified data outside of their own firewalls. The Data Processor has not had to receive any identified data. Still each party can share and receive the data they need but without risking exposure of any identifying information.

How PII Vault Keeps PII Safe

PII Vault provides unique and game-changing capabilities, but none of these capabilities matter if the data within PII Vault was not secure. Accordingly, Anonomatic has implemented an unprecedented combination of security measures to protect its data.

PII Vault Implementation

PII Vault is not a vendor-hosted service. It is instead delivered as a self-contained, black-box, Docker container. Docker is a leading solution in the field of virtual machines. Your IT people most likely love it. With PII Vault being deployed as a Docker container, an IT staff can not only install it anywhere within an IT infrastructure, but it can also be setup and working within minutes of being downloaded. There are essentially only two commands a knowledgeable IT technician needs to run to deploy PII Vault within either a private, public or hybrid cloud. PII Vault can even be installed completely on-prem, so your PII data never leaves your firewall. It also means that all the internal security measures and access controls can be applied to PII Vault. As a reminder, PII Vault does NOT have a user interface. In addition, there are no reports to run against PII Vault and there is absolutely no monetization of the data in PII Vault. PII Vault is also referred to as a "black-box" container because there are no access methods provided for anyone, including Anonomatic, to view or edit the data in PII Vault's data repository other than through the API.



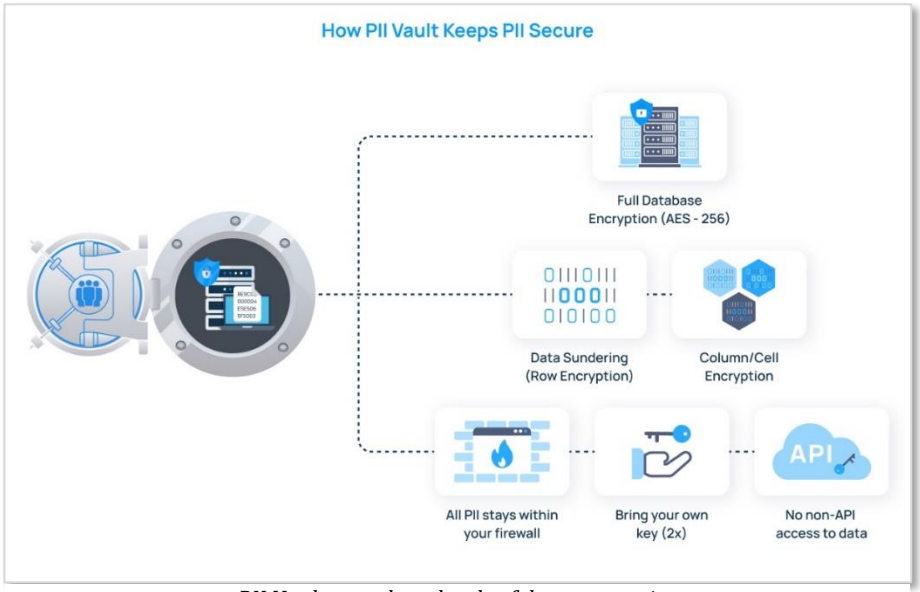
Data Separation and Poly-Anonymization

These two security measures are so effective they deserve repeating. With the actual identifiers being physically and logically separated from the underlying fact data it identifies, the PII itself carries a fraction of the risk it does when the actual identifiers are paired with that fact data. When the PII is further protected by Poly-Anonymization, which prevents the direct linking of values between data sources, the protection becomes exponentially stronger.

Additionally, Poly-Anonymization is not only used for externally shared data, but it is

also used internally. PII Vault uses Poly-Anonymization to protect individual data profile records, so there is no direct link between the identity of the source of the data and the data within PII Vault provided by that source.

Encryption Methods



PII Vault uses three levels of data encryption

PII Vault uses three levels of data encryption. The first is that the entire database is protected by industry standard 256-bit AES encryption. The second level is a custom-developed row-level data protection technology called Data Sundering. Data Sundering can be considered similar in functionality with column-level encryption but without the performance impact and other limitations. With Data Sundering any hacker getting through the first level of encryption will be unable to ever decipher the values of any row of data without also having the necessary multi-part keys. These keys are never stored in the database. Of course, PII Vault customers can bring their own key. Finally, the third level encodes individual values with an accompanying third set of keys.

As a final level of security, those organizations who do not need to maintain continuity of Poly-Ids over a period of time may purge all PII stored in the PII Vault at any time.

Conclusion

Maintaining data security and complying with international data privacy regulations is a constantly moving target. Large, complex, and expensive enterprise systems which

concentrate on protecting a single organization's data are totally inadequate for the world of data sharing. PII Vault is a revolutionary new solution which strips many of the compliance regulations away from common tasks because many of those regulations do not apply to fully anonymized data.

It is this capability, specifically the ability:

1. For an EU-based organization to share critical data outside of the EU.
2. Have data merged at the individual level with data from other organizations either inside or outside the EU.
3. To have that data processed and the results returned.

These capabilities allow any organization to share, receive, or process data without the risks, costs and PII compliance overhead normally involved in such activities. This makes PII Vault a uniquely well-suited solution to assist organizations in complying with the Supplemental Measures recommendations of the European Data Protection Board.

[1] Anonomatic, Inc. is a software company based in the United States. Nothing in this paper should be taken as legal advice. This paper reflects our point of view of how, by utilizing PII Vault and following best practices, organizations within the EU can comply with the recommendations of the EDPB and other applicable supervisory authorities.

[2] GUIDs are 128-bit values with 10^38 possible values