TRUATA.

Pseudonymization as a Solution for International Data Transfers

Effective deployment of pseudonymization for the transfer of data

Your guide to pseudonymization for international data transfers

Overview

In light of the Schrems II ruling, the European Data Protection Board (EDPB) delivered a firm message that international data transfers must be brought into line with the decision of the Court of Justice of the European Union (CJEU); this came in its November 2020 release of Recommendations on measures that supplement transfer tools.

The EDPB's Recommendations outline the key steps that organizations need to take when conducting international data transfers in order to ensure compliance with the EU level of protection of personal data; that is to say, the threshold of "an essentially equivalent level of protection to that guaranteed in the EU" must be met. In the case where this threshold is not met, the data exporter must suspend or cease transfer activity immediately.

Data pseudonymization is one of the supplementary measures that the EDPB's Recommendations has highlighted as being legally effective. As a result, pseudonymization has come to the forefront of privacy and data protection discussions as a viable technical measure which, when deployed effectively, can enable organizations to continue conducting lawful transfers of personal data out of the EEA in certain cases, and also provide a number of commercial benefits.

An introduction to supplementary measures for international data transfers

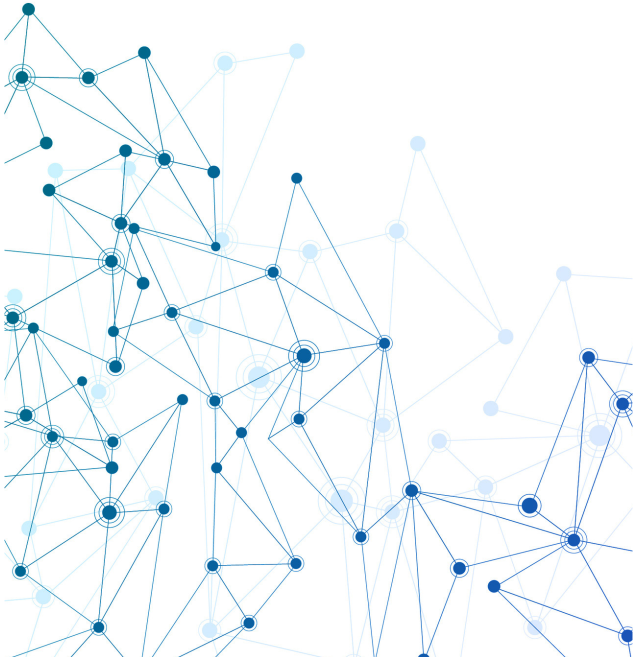
Our fast-evolving data economy has given rise to a number of intricate challenges surrounding the individual right to privacy and data protection; however, in recent years, the challenges pertaining to international data transfers, which support trillions of dollars in business every year, have become particularly complex.

The introduction of the GDPR intended to provide a level of protection to European citizens' personal data, irrespective of the location of that data. Through the release of its Recommendations on measures that supplement transfer tools, the EDPB has made it clear to all data exporters that it is seeking an active approach to compliance. Furthermore, it has reinforced that the GDPR principle of *accountability* applies to data transfers. The EDPB stated that it sees supplementary measures as being necessary to satisfy requirements of data protection by design and data protection by default, as laid out in Article 25 and Article 32 of the GDPR.

In these cases, it would seem that the EDPB views supplementary measures as a baseline requirement that make sense as part of an overall data transfer risk strategy.

Since international data flows are essential to organizations of all sizes and across all industries, business leaders across the globe are now turning to the EDPB's Recommendations to assess, as part of a Transfer Impact Assessment (TIA), what supplementary measures they may need to implement - above and beyond Standard Contractual Clauses (SCCs) - to see where the vulnerabilities are in these data flows.

The EDPB explains that the supplementary measures can either be *contractual, technical* or *organizational* in nature. However, while organizational and contractual measures may reinforce the safeguards that the SCCs provide, the EDPB notes that these measures by themselves *"do not meet all the conditions required to ensure a level of protection essentially equivalent to that guaranteed within the EU."*



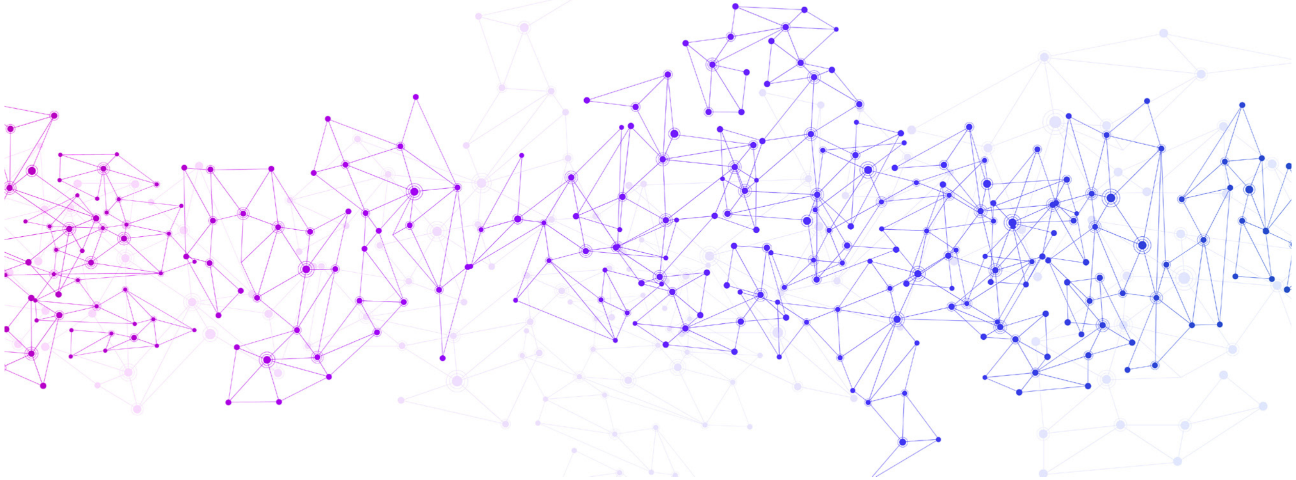
It should be noted that there will be situations where only technical measures will prove to be sufficient in overcoming the powers of access to data by authorities in third countries. One of these technical measures is that of **pseudonymization**, which is cited in Use Case 2.

Pseudonymization as a technical measure

The key criteria for implementing supplementary measures hinges on the identifiability of a data subject ("**Identifiability Criteria**"); therefore, the threshold for effective pseudonymization will also hinge on identifiability. To that effect, in order for pseudonymization to be used as an effective supplementary measure, it will be critically important that the data is sufficiently pseudonymized.

The EDPB refers to pseudonymization as going beyond merely addressing *names, addresses* or other plain identifiers; it must also address information such as *time of access*, *location* and other meta data relating to events within the data. That is to say that it will be imperative for organizations to address both direct and indirect identifiers in order to guarantee that the pseudonymized data cannot be attributed to an identified or identifiable natural person.

It will also be necessary for organizations to look at the potential for re-identification attacks and, by taking account of "all means reasonably likely to be used by a third party" to re-identify individuals, determine what information could end up in the public domain or be made available to a government agency, and what information - in that eventuality - could re-identify a data subject.



In order for organizations to have the confidence that they are taking the right steps and meeting their accountability obligations when relying on pseudonymization, it will be important for them to have the ability to measure, both objectively and quantitatively, the risk of re-identification in any given dataset.

This will enable the organization to judge if the data will achieve an *essentially equivalent level* of protection and can be lawfully transferred.

More specifically for those organizations who regularly transfer data to the U.S., there will be a need to ensure that any legal exposure of the data to s.702 FISA or EO12333 is rendered practically "impossible or ineffective".

For any organization, transitioning to a new era of working with international data transfers is no easy task; in fact, many may even be questioning if they can achieve any of this at all. However, by following some practical guidance from our legal and technical experts, it is possible to deploy effective pseudonymization and feel confident that you can rely on it to support continued international data transfers.

Getting started with pseudonymized data

The portmanteau "pseudonymization" is a blend of "pseudo" and "anonymization". The term has gained widespread recognition within the world of data, capturing that a pseudonymized dataset should not facilitate the identification of individuals directly from the data it contains. However, unlike anonymization, individuals may be re-identified if the pseudonymized dataset is combined with "additional information", such as a mapping table or an identifiable dataset that has overlapping columns.

The key difference with an "anonymized" dataset is that the re-identification of individuals is not possible by a third party, through "all means reasonably likely to be used", taking into consideration the technology that is available at the time of processing.

Pseudonymization seeks to make it very difficult, or impossible, to re-identify an individual from a dataset by transforming or processing either the entire dataset, the individual columns or the records so that the resulting dataset is qualitatively different than the original. In fact, pseudonymization is not one technique or transformation, but rather a process that can be implemented in a number of ways and with increasing sophistication, depending on the use case and legal or regulatory landscape at hand.

Each field or record in a dataset can be transformed or distanced from its original value using a variety of techniques, depending on the risk profile and desired purpose of the analytical use. The variety of available techniques that can be applied during the pseudonymization process means analytical utility can be preserved for particular use cases while reducing re-identification risks.

This makes pseudonymization a valid approach for organizations to continue supporting business processes that require international transfers of data while remaining compliant with the CJEU ruling. Pseudonymization can deliver a configurable level of protection to data so that it is equivalent to the level of protection that is provided under EU law.

Furthermore, the additional processing required to pseudonymize large datasets, which could possibly need processing multiple times to serve different uses, is likely to be far less costly than ceasing all transfers or making large-scale changes to an organization's data processing infrastructure, changes such as moving all data storage and processing to EEA jurisdictions.

Processing techniques

Any data processing technique that obscures or changes the original values of a column or an entire dataset can be used for the purposes of pseudonymization.

Examples of such techniques include:

- Tokenization
    - Dynamic tokens
- Noise addition
- Formal-preserving encryption (FPE)
- Pertubation
- Redaction
- Generalization
- Binning
- Masking
- Data synthesis

To illustrate an example, *Table 1* and *Table 2,* show the original and pseudonymized versions of an employee information table, respectively. Table 2 has had many of the re-identification risks removed, using a number of data processing techniques. The first two visible rows are indistinguishable from each other and, therefore, could not be linked back to the original records. It is important to note that Table 2 highlights one of the shortcomings of pseudonymization, detailed in the following section, which is that a pseudonymized table may retain re-identification risks.

The last row shown in the table could potentially re-identify the associated individual, even with all the transformations applied as shown, since they may be the only employee with a UK location; their higher salary may set them apart also. These residual risks, outlined below, must also be considered and potentially mitigated.

Name	Age	Occupation	Location	Salary
John Smith	33	Dental Assistant	24 Elm Tree Avenue, Foxrock, Dublin 18, Ireland	€35,500
Ravi Singh	35	Paediatric intern	3 Glen Drive, Drumcondra, Dublin 5, Ireland	€39,750
....	...	...	...	...
Mia Fischer	29	Kindergarten owner	26 Green Avenue, Croydon, London BR3 3BY, UK	€95,950

Table 1: *Original table containing precise values for employee name, age, occupation, location and salary.*

In Table 2, we can see that two of the visible records have been made virtually indistinguishable from each other and relative to their original values, while the remaining record is different for all fields.

This highlights both the power and potential pitfall of pseudonymizing an entire table for general use; it is indeed challenging to ensure all records are indistinguishable from each other, especially where there are many columns in the dataset. To address this, for particular use cases a dataset may be aggregated, or a subset of columns and records may be used, so that pseudonymization is more feasible. This can be done multiple times to support different uses of the dataset.

Name	Age	Occupation	Location	Salary
6yhu...i6zt	30-39	Healthcare	Dublin, Ireland	€38,675
ui76...o99l	30-39	Healthcare	Dublin, Ireland	€38,575
....	...	...	...	...
8izt...e31w	20-29	Early Education	London, UK	€93,950

Table 2: *Pseudonymized table with names tokenized, age binned to a range of 10 years, occupation and location generalized, and salary values perturbed.*

Pseudonymization

The stem, “pseudonym”, speaks to the core goal of this technical approach to protecting privacy. Similar to the purpose of an alias, pseudonymization involves replacing identifiable fields with a token or a suitable alternative representation that reduces, or removes, the ability to infer the original values of those fields.

How pseudonymization works for international transfers

When an EU-based organization wishes to conduct transfers of data to a non-EEA territory, the processing techniques previously noted can be used to complete a successful transfer in a manner that provides the equivalent level of protection to data subjects as that provided under European law. There is no single technique that works well in all cases; in fact, in most scenarios a number of techniques must be combined to achieve a desired outcome.

When considering which technologies can be used, and with what configurations, it will depend on the nature of the relationship between the exporter and importer companies and the purpose for which the transfer is being made. Broadly speaking, there are four categories of techniques that we will use to describe the use cases; these are outlined in Table 3.

Category of Technique	Description	Trūata Product(s)
Measurement	Objectively quantify different types of privacy risk within a dataset. Identify low-utility/high-risk quasi-identifiers to enable informed decisions around the use and protection of the data.	Trūata Calibrate
Transformation	Apply transformation techniques to reduce re-identification risk. This includes traditional and advanced pseudonymization techniques, such as tokenization, generalization and noise addition to achieve differential privacy or *k*-anonymity.	Trūata Calibrate
Synthesis	Generate datasets based on a learned model of an existing dataset, with variable privacy risk and analytical accuracy. Different techniques may reproduce quasi-identifiers from the original dataset, which will represent a re-identification risk.	Trūata Synthesize
Separation	Separate the analyst from the data so that row-level data cannot be accessed while facilitating the desired analytics. The interface can abstract analysis from the underlying data, mitigate privacy risks in outputs and enable combinations to achieve the desired analytical output.	Trūata Pioneer for AI and BI

Table 3: *Techniques that can be used to pseudonymize data for international transfer.*

The system depicted in Figure 1 produces a pseudonymized version of an input dataset that incorporates risk quantification, dynamic tokenization and transformation techniques to make it much more difficult to re-identify an individual from the dataset.

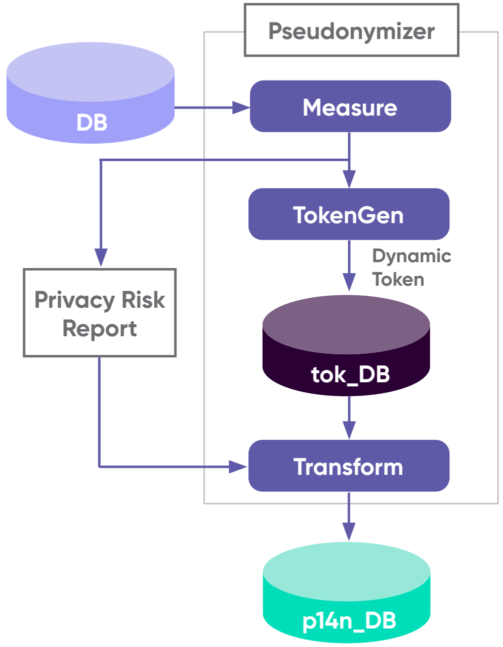


Figure 1: *Solution for pseudonymizing a dataset that incorporates a number of PETs, with dynamic tokenization to produce the pseudonymized "p14n_DB" dataset from the input "DB", driven by quantitative privacy risk scores.*

If longitudinal analysis of user-level metrics is required, a mapping table may be maintained to enable association of all tokens to each individual. This mapping table could allow re-identification to occur, so it should be kept separate from the pseudonymized table. An interface can be used to separate the analyst from the underlying data, joining the mapping and pseudonymized tables to facilitate longitudinal analysis, without the analyst accessing or reviewing the longitudinal data.

Figure 2 depicts how analytics that require data transfer between EU and non-EEA companies can be supported. Dynamic tokens are mapped to a single static token within the "token_DB" table that is hosted within the EEA. The pseudonymized "p14n_DB" table is transferred directly to the non-EEA importer and when user-level analytics are required, the "Join" module, for joining dynamic tokens together, is used. This interface joins the pseudonymized dataset with the token mapping table without allowing access by the analyst.

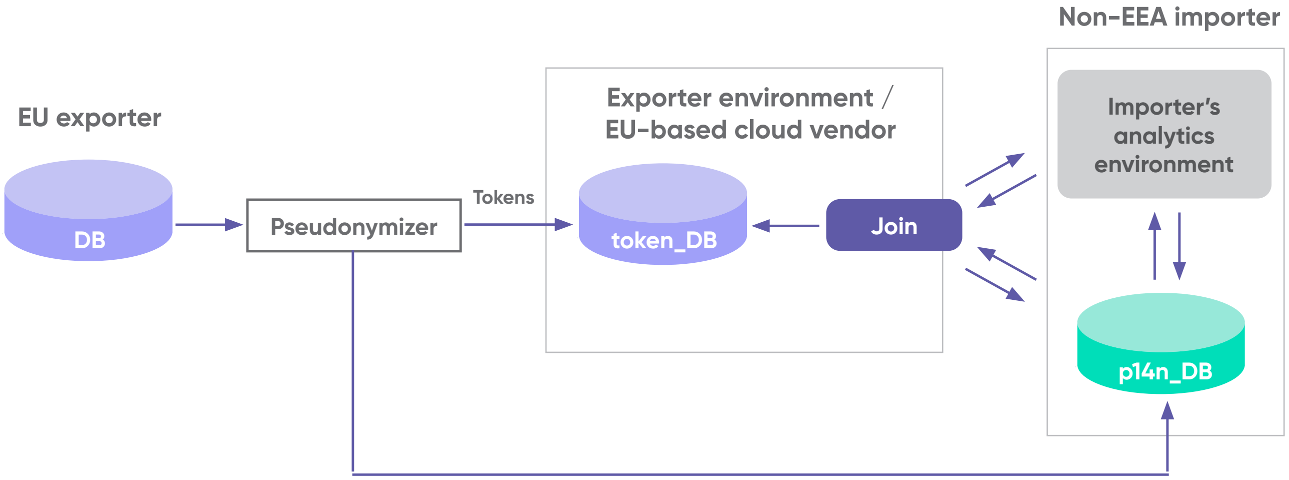

Non-EEA importer
EU exporter
Exporter environment / EU-based cloud vendor
Importer's analytics environment
DB
Pseudonymizer
Tokens
token_DB
Join
p14n_DB

Figure 2: *A mapping table to allow longitudinal analysis is maintained within the EU, while a software module to join dynamic tokens allows analytics workflows to be performed without providing access to row-level data.*

As shown in *Figure 3*, in some cases where effectively pseudonymizing the dataset for transfer cannot be achieved with the required level of protection, a synthetic version of the dataset, with known privacy and utility characteristics, may be transferred to the non-EEA entity for analysis. This enables business logic to be generated, which can then be applied to the original data residing within the EU.

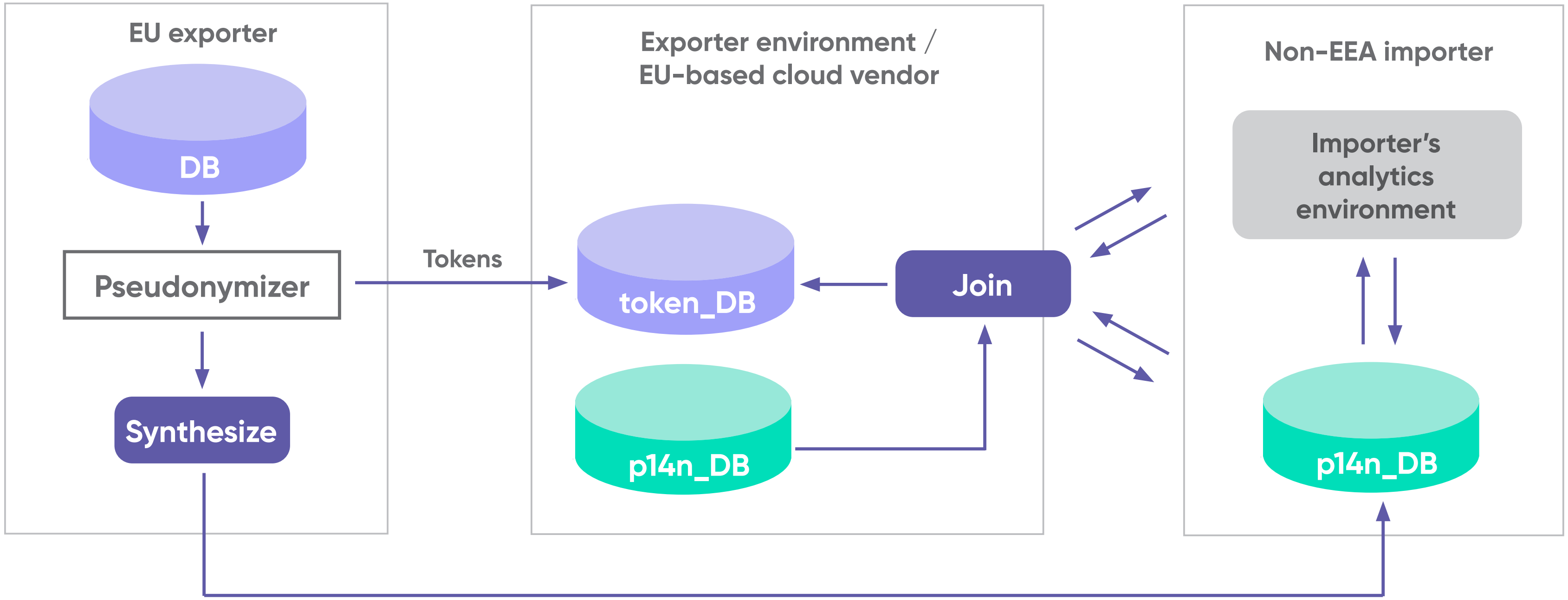


Figure 3: *Importer outside the EEA receives synthetic version of the dataset and uses a module to join dynamic tokens to produce final results.*

The benefits of pseudonymization

For most organizations, even the thought of having to suspend or cease analytical activities that involve transfers of data to non-EEA companies is simply not a viable option. Similarly, for many organizations, it is just not feasible to copy or move their entire data processing infrastructure to the EU. The good news, however, is that effective pseudonymization can, and will, enable businesses to continue these activities - even in a post-Schrems II world.

Pseudonymization has several benefits for organizations that rely on international transfers to support their business:

Confidence

Organizations can continue with critical data transfers to non-EEA countries by adopting privacy-enhancing technologies that enable them to harness data-driven insights without risking regulatory sanctions or compromising customer trust.

Control

Companies can decide exactly what data they wish to transfer based on the data risk profile. They will be able to make an effective risk-based decision on how it should be processed before transfer; they can also effectively balance data privacy and data utility.

Compliance

The transferred data has several layers of protection, so the data exporter can rest assured that the levels of protection required by EU law are replicated in the transferred data.

Customization

Companies can produce customized pseudonymization configurations and combine the technologies described in different ways; this will allow for the flexibility to support a greater number of data transfer scenarios.

Leveraging privacy-enhancing data solutions

The immense challenge that organizations face right now is to urgently figure out how to pivot and transition to a new way of sharing personal data on an international scale in order to continue to drive commercial growth and innovation.

This is combined with the need to maximize data utility while maintaining compliance with the ever-tightening European regulations surrounding data privacy and data protection. This is certainly a tall order for any organization, but not so much for those who leverage the right privacy-enhancing technologies and solutions to support the process.

Real-world scenarios

A **financial institution** may wish to transfer transaction data to a processor outside the EEA to build fraud detection models. These models can be trained and validated over pseudonymized data, then transferred back to the EU to be deployed against the original, identifiable data.

A **mobile network operator group** that has an analytics department located outside the EEA may receive pseudonymized call detail records (CDRs) from regional operating companies based in the EU. This will allow them to perform group-level analytics.

A **retailer** based in the EU can transfer pseudonymized data to a market intelligence analytics provider outside the EEA. The resulting market intelligence reports would be in aggregated form, allowing them to be used directly when transferred back to the EU.

A **marketing automation provider** based in the US may receive behavioral data that has been tokenized with a reversible algorithm from a customer. When configured triggers or events occur, the tokenized identifiers can be transferred back to the EU, re-identified and used to send the required outreach.

A purpose-built solution for international data transfers

At Trūata, we have already developed a purpose-built pseudonymization software solution that can be deployed in an organization's own data environment to effortlessly assist with international data transfers.

Trūata Calibrate

...a seamless pseudonymization solution that caters to business-specific needs.

Trūata Calibrate empowers organizations to seamlessly transform data to meet SafeUse levels. This ensures they can confidently leverage personal data to drive growth and innovation while complying with the highest global data protection regulations.

Trūata Calibrate is designed to operationalize privacy-compliant data flows, providing an easy-to-use solution for those wanting to activate data, innovate with data or conduct international data transfers. It is specially designed to solve everyday data privacy issues, such as navigating the complexities of global regulations, overcoming data flow inefficiencies, and providing an auditable trail of compliance.

The benefits:

- Use our patent-pending Fingerprint technology to automate the statistical analysis of datasets
- Access quantitative risk scores that consider data-centric and contextual signals
- Improve awareness and communications around privacy risk in datasets
- Receive recommendations for improving privacy
- Make informed decisions about data
- Seamlessly transform personal data into pseudonymized data
- Minimize the impact on utility for analytics
- Retain control over the level of data transformation
- Enable the automatic re-run of privacy risk assessments on newly created privacy-enhanced datasets
- Access reports that highlight the reduced risk
- Analyze numerical risk scores based on re-identification metrics to support the most stringent regulatory environments

References

Introduction to the hash function as a personal data pseudonymisation technique
https://edps.europa.eu/data-protection/our-work/publications/papers/introduction-hash-function-personal-data_en

De-identification Guidelines for Structured Data
https://www.ipc.on.ca/resource/de-identification-guidelines-for-structured-data/

Pseudonymisation techniques and best practices
https://www.enisa.europa.eu/publications/pseudonymisation-techniques-and-best-practices

Recommendations on shaping technology according to GDPR provisions: An overview on data pseudonymisation (2018)
https://www.enisa.europa.eu/publications/recommendations-on-shaping-technology-according-to-gdpr-provisions

Anonymisation and pseudonymisation
https://dataprotection.ie/en/guidance-landing/anonymisation-and-pseudonymisation

The EU Article 29 Working Party's Opinion on Anonymisation Techniques
https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

U.S. Secretary of Commerce, Wilbur Ross, Statement on Schrems II Ruling and the Importance of EU-U.S. Data Flows
https://www.commerce.gov/news/press-releases/2020/07/us-secretary-commerce-wilbur-ross-statement-schrems-ii-ruling-and

Recommendations 01/2020 on measures that supplement transfer tools to ensure compliance with the EU level of protection of personal data
https://edpb.europa.eu/sites/edpb/files/consultation/edpb_recommendations_202001_supplementarymeasurestransferstools_en.pdf

Get in touch

Arrange a free demo session today: book a demo

Speak to our team to learn more about pseudonymization solutions: info@truata.com

Follow Trūata for more: in

Trūata Headquarters: Silverstone House, 1st Floor, Ballymoss Road, Sandyford. Dublin D18 A7K7. Ireland | +353 1 566 8468 | truata.com | ©2021 Truata Limited ("Trūata") | All rights reserved. Trūata's marks and logos and all other proprietary identifiers used by Trūata herein ("Trūata Marks") are all trademarks and/or trade names of Trūata and/or its affiliates. Any other products, services, or company names referenced herein may be trademarks of Trūata or of other companies; if trademarks of other companies, there is no endorsement or affiliation, expressed or implied, claimed by Trūata. WP_PSEUDONYMIZATION_07_2021_012

14