SPOTLIGHT
Sponsored by: Anonos

As cloud adoption grows and organizational cloud maturity increases, enterprises focus more on privacy and security for data in use. Statutory Pseudonymization is emerging as a privacy-enhancing technology that effectively supports many enterprise data use cases.

Data Without the Drama — How to Process Data in the Cloud While Being Fully Compliant

January 2023

Written by: Ralf Helkenberg, Research Manager, European Privacy and Data Security; and Carla Arend, Associate Vice President, Cloud Research, Europe

Introduction

Organizations are adapting to a digital-first world in which cloud data platforms play a key role. For instance, they are embracing edge computing and hybrid multi-cloud architectures and shifting their compute strategies from centralized on-premises data centers to distributed data infrastructure models. In addition, organizations are using powerful data analytics, increasingly infused with artificial intelligence (AI) and machine learning (ML), to drive valuable customer insight, smarter decision-making, and competitive advantage.

Today, businesses must look beyond themselves for new ideas. We're seeing the emergence of industry platforms and ecosystems for business collaboration to gain access to different capabilities and new sources of data to build new business models or develop new products and services. The European-initiated GAIA-X project is an example of a digital ecosystem in which data and services are to be available, collated and shared in an environment of trust.

Underpinning this digital transformation is digital trust – from trust in the data and technology to trust in the business partner or supplier. Data privacy and security are foundational to building digital trust and provide the glue for digital transformation. However, organizations anticipate a wide range of external risks to potentially affect their digital transformation and technology investment plans. Their foremost risk concern is business exposure to cyber incidents and data breaches, whether from ransomware attacks, technical cloud misconfigurations or via the supply chain. From the CEO perspective, the twin challenges of cybersecurity and data sovereignty are seen as the foremost risk factors impacting their businesses during the next two years, according to IDC research (see Figure 1).

AT A GLANCE

WHAT'S IMPORTANT

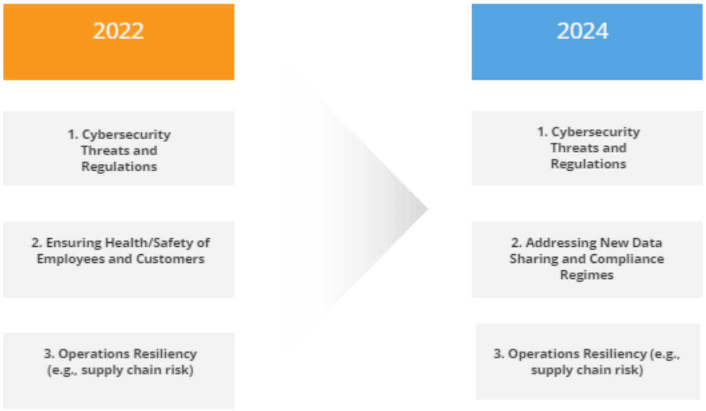
Cleartext processing in the cloud is the new frontier for privacy compliance. Statutory Pseudonymization is a technical safeguard to protect data in use.

KEY TAKEAWAYS

Statutory Pseudonymization has many business use cases:

- Secure computation in the cloud
- Data sovereignty and Schrems II-compliant processing
- Privacy-compliant data analytics and collaboration
- Unlocked datasets to expand trusted AI/ML capabilities

Figure 1: ***CEO Sentiment: Top 3 Risks with Greatest Impact on Business Today and in 2 Years***


2022
1. Cybersecurity Threats and Regulations
2. Ensuring Health/Safety of Employees and Customers
3. Operations Resiliency (e.g., supply chain risk)
2024
1. Cybersecurity Threats and Regulations
2. Addressing New Data Sharing and Compliance Regimes
3. Operations Resiliency (e.g., supply chain risk)

Source: IDC WW CEO Sentiment Survey, 2022

Meanwhile, privacy regulation is tightening, adding additional complexities to turning data into business insight. More data protection regulations, either modeled closely on the General Data Protection Regulation (GDPR) or with strong standards for protection, are emerging across the globe. Countries are taking more sovereignty measures to control the infrastructure and data generated in their jurisdictions, from rules on data residency to conditions on transborder data flows. The growing extraterritorial application of data governance laws subject organizations to a growing tension between allowing digital innovation to accelerate and ensuring data and IT infrastructures comply with regulations and guidelines. The data sovereignty implications extend to the cloud environment given that organizations are increasingly moving their services and data to platforms managed by international cloud providers.

Using the public cloud is connected to the level of trust in the cloud service provider. Privacy and security concerns are inhibitors to moving regulatory sensitive data to the cloud, particularly in Europe. Security concerns center around the perceived increased security risk of storing data with a third-party provider, the lack of visibility into what data is within cloud applications, the extent to which they have control of who can access sensitive data — and, lastly, the extent to which government and law enforcement can access and request customer data. For example, the reach of the U.S. CLOUD Act has inhibited public authorities in some European countries from using the cloud.

Cloud data can only be safeguarded if security and compliance features are well understood and properly configured from the outset. The industry-wide shared responsibility model sets out security responsibilities across all types of cloud platforms, providing guidance to securely leverage the benefits of the cloud. Typically, the cloud service provider is responsible for the security of the cloud (the infrastructure), and the cloud customer is responsible for security in the cloud (data and resource configuration).

Encryption and access controls in the cloud can solve security and compliance requirements for data at rest and data in transit. But problems arise when you want to process or collaborate with the data because it must be de-encrypted and is therefore transformed into a state that is unprotected. Statutory Pseudonymization, however, closes privacy and security gaps by protecting data in use, helping organizations realize their cloud and data strategies in a regulatory-compliant manner.

Definitions

Following are key terms and their definitions:

- **Data at rest:** Data that is in storage, including but not limited to archived data, databases, files stored on hard drives, USB thumb drives, backup files, and files stored off site on cloud storage platforms
- **Data in transit:** Data that is moving across networks between computer systems, applications, or locations
- **Data in use:** Data that is being updated, processed, erased, accessed, or shared by a system
- **Anonymization:** Personal data transformed in such a way that it is no longer possible to link back to a specific individual
- **Traditional pseudonymization:** Replaces directly identifiable data value with a token or masking that reduces or removes the ability to infer the original values of those fields
- **Statutory Pseudonymization:** As defined in the GDPR and similar legislation, both direct and indirect identifiers are replaced through a combination of de-identification techniques and dynamically changing pseudonyms or codes, which are kept separately and protected by technical and organizational measures. The re-identification of individuals is not possible without the use of additional information – or "keys" – held separately and securely by the data controller or their designee. More importantly, the process can be reversed only by using the keys when reidentification of the source data is authorized.

Although anonymization and Statutory Pseudonymization may appear similar at first, they perform different functions in data protection law, such as the GDPR. The difference rests on whether the source data can be re-identified. Data that has been irreversibly anonymized ceases to be personal data and does not require compliance with data protection law.

However, uncertainties exist as to whether such procedures can provide a sufficient degree of anonymity. Studies have shown that even within independent anonymized datasets, identifying individuals is not that difficult when the databases are combined. This means knowing whether anonymization has been achieved is rarely a black-and-white proposition and a challenging assessment to make. Another downside to anonymization is that it decreases data utility. To preserve levels of utility, traditional anonymization techniques restrict data processing to enclaves or silos to mitigate the risks of reidentification.

Statutory Pseudonymization as a Technical Safeguard to Protect Data in Use

The attack surface that organizations need to protect has expanded due to the shift to hybrid work, greater reliance on cloud services, accelerated digitalization and sprawling data infrastructures across a multitude of applications, devices, and locations. The threat landscape is evolving, and cyberattacks are growing in volume, variety, complexity, and precision and so are data breaches. In a 2022 IDC European security survey, 54% of respondents had experienced an increase in the volume of cyberattacks during the prior 12 months.

The business cost and reputational impact from cyber incidents requires organizations to rethink their data protection strategies as well as upscale data privacy and security infrastructure for cloud computing environments. Protecting data while in use, though, is challenging because applications have generally required clear text — not encrypted — to compute. Traditional technical trust controls such as access management do not provide the privacy and security assurance to share and collaborate with sensitive datasets across IT environments.

Furthermore, organizations are trying to obtain more value from their data to improve their products and services. For example, more chief data officers and data analytical roles are being created to drive such data-enabled transitions. However, data privacy and security has become a flashpoint in the drive to achieve digital transformation. Concerns about potential privacy violations and the prioritization of locking data down on premises through standard security measures have mistakenly led many organizations to forego the benefits of data computation in the cloud and multi-party data sharing and collaboration. But organizations don't have to make an either-or choice between data utility and regulatory privacy and security. They can have both through Statutory Pseudonymization.

While not a new privacy-preserving technique, Statutory Pseudonymization has been redefined by and gained special prominence through the GDPR. According to the regulation, the ability to Statutorily Pseudonymize data has several advantages:

- Organizations can lawfully repurpose data beyond the original stated reason for its collection.
- Statutory Pseudonymization is an important safeguard when processing data for scientific, historical or statistical purposes.
- It decouples privacy and accuracy, enabling data protection by design and by default without compromising data accuracy.
- It constitutes a data security risk reduction technique that can reduce the danger of identification or harm from a data breach.
- It constitutes a security measure that can help organizations meet their data security obligations.

Business Benefits of Statutory Pseudonymization

The following are some of Statutory Pseudonymization's primary business benefits:

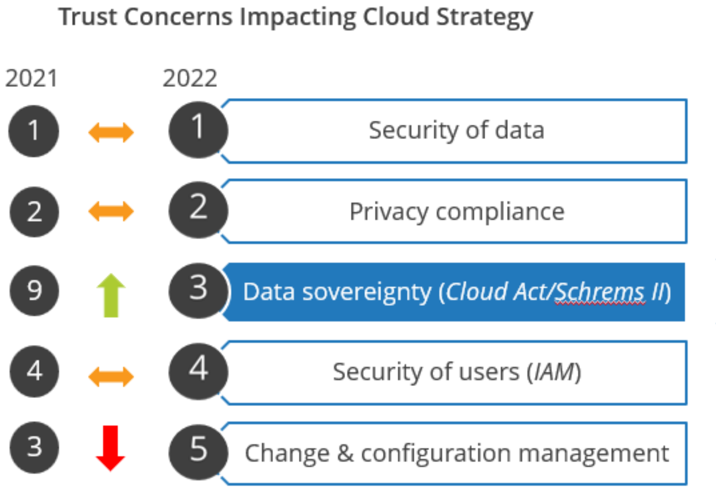
Secure Computation in the Cloud

Cloud adoption is expanding. Organizations are modernizing their on-premises IT estates with a hybrid and multi-cloud architecture for greater agility, flexibility, and cost and operational efficiencies. Organizations don't like the idea of losing control of their data, a concern that extends to multi-tenant cloud environments. There is the perceived increased security risk of storing data with a third-party provider, the lack of visibility into what data is within cloud applications, and the extent to which they have control over who can access sensitive data. By implementing Statutory Pseudonymization, organizations can deploy and process sensitive workloads in the cloud without compromising data privacy and compliance, and with the security assurance that the cloud provider has no access to or control of the data.

Data Sovereignty and Schrems II-Compliant Processing

When dealing with on-premises infrastructures, data sovereignty is clear cut. However, storing and processing data in the cloud is more complex. Since the 2020 Schrems II ruling that invalidated the EU-U.S. Privacy Shield, organizations have been grappling with whether they can legally and safely transfer personal data under the GDPR outside the EU. IDC research shows data sovereignty becoming a leading trust and compliance concern for organizations when shaping their cloud strategies. Ranked nineth in IDC's 2021 European security survey, data sovereignty jumped six spots in 2022, coming in third behind privacy compliance and data security (see Figure 2).

Figure 2: ***Trust Concerns Impacting Cloud Strategy***


Trust Concerns Impacting Cloud Strategy
2021
2022
1 ↔ 1 Security of data
2 ↔ 2 Privacy compliance
9 ↑ 3 Data sovereignty (*Cloud Act/Schrems II*)
4 ↔ 4 Security of users (*IAM*)
3 ↓ 5 Change & configuration management

Source: IDC European Security Survey, 2022

For most organizations, having to suspend or cease analytical activities that involve transfers of data to non-EU companies is simply not a viable option. Similarly, it is not feasible to copy or move their entire data processing infrastructure to the EU. Statutory Pseudonymization has come to the forefront of privacy and data protection discussions as a viable technical measure that, when deployed effectively, can enable organizations to continue conducting lawful transfers of personal data outside the EU.

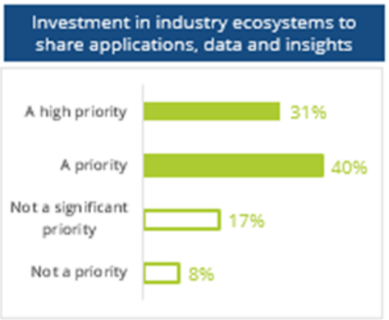
The European Data Protection Board (EDPB) 2021 Schrems II guidance sets out the steps that organizations should follow to legitimize the transfers of personal data to third countries. It includes where required, adopting additional contractual, organizational, and technical protective measures to safeguard against overreaching government surveillance. Statutory Pseudonymization is one of the supplementary technical measures that the EDPB's recommendations highlight as being legally effective.

Privacy-Compliant Data Analytics and Collaboration

Encouraged as a data protection best practice, data-at-rest encryption has become the default modus operandi for securing sensitive data. Consequently, datasets from such highly regulated industries as financial services and healthcare often sit siloed, unable to be shared or combined with third parties for analytics.

Yet encryption is primarily a security measure for making data unintelligible against unauthorized users. In environments where data is constantly moving between different parties and combined with other data, encryption (though providing effective data protection) is an inhibitor to creating valuable business insights. But organizations are becoming more aware of the importance of working together to analyze their collective data, with the assumption that the whole is greater than the parts. In an IDC survey, investment in industry ecosystems to share and combine applications, data, and insight is a priority for 71% of organizations (see Figure 3).

Figure 3: ***Sharing Applications, Data, and Insights Is an Investment Priority***



Source: IDC, 2023

Unlock Datasets to Expand Trusted AI/ML Capabilities

Machine-learning algorithms are rapidly increasing their demand for more computation and larger datasets. IDC research shows that few organizations have enough internal data to achieve this on their own. Traditional ML approaches require centralized data collection and model training. Regulatory restrictions around data privacy, latency problems and high transfer costs are inhibitors to bringing larger datasets together, which in turn limits the training and fine-tuning of algorithms. Statutory Pseudonymization enables multiparty AI/ML models to be trained and processed in the cloud on a joint dataset from multiple organizations, without revealing the constituent datasets that would compromise data privacy.

Privacy-Enhancing Technologies

Privacy-enhancing technologies (PETs) are emerging technologies that embody data protection principles to achieve specific privacy or data protection functionality. They help implement data protection by design and default principles effectively and integrate necessary safeguards for data processing. There are various PET types, and each technology has particular data protection capabilities and use limitations (see Table 1). Only Statutory Pseudonymization provides the combined data safeguards — data confidentiality and de-identification — that enable protected multi-cloud processing, compliant internal data transfers, and confidential AI/ML analytics without compromising data utility and accuracy.

Table 1: ***Different Types of PETs***

PETs	**Description**	**Weakness**
Trusted Execution Environment (Confidential Computing)	Preserves data confidentiality by performing computation on encrypted data in a hardware-based Trusted Execution Environment.	Many necessitate code changes to existing applications Security risk via side-channel attacks.
Homomorphic Encryption	Provides strong security and confidentiality by enabling computations on encrypted data without first decrypting it.	Requires significant computational resource. Supports a limited number of computational operations.
Secure Multi-Party Computation	A protocol that allows computation or analysis on combined data without the	Does not protect the data output. High communications overhead to support data exchanges between nodes.

	different parties revealing their own private input.	
Differential Privacy	Random injection of noise to alter the data in the dataset	Difficult to tailor the optimal trade-off level between data privacy and utility.
Synthetic Data	Artificial data generated by data synthesis algorithms, which replicate patterns and the statistical properties of real data	Synthetic data may not represent outliers present in the dataset.

Source: IDC, 2023

Considering the Anonos Data Embassy Software Platform

Anonos is a global provider of data privacy and security technology, with more than a decade of experience in developing privacy-preserving software that helps organizations turn regulated and sensitive data into a competitive advantage. With 26 granted international patents, the company's platform leverages GDPR-compliant Statutory Pseudonymization to make it possible to legally analyze, combine and use data across untrusted environments, turning it into a business asset without violating privacy, security, or regulatory restrictions.

Anonos' Data Embassy software centers around Variant Twins: non-identifiable yet 100% accurate variations of the source data required for specific use cases. The patented system uses a combination of state-of-the-art privacy-enhancing and de-identification techniques, including Statutory Pseudonymization, to replace directly and indirectly identifying personal data, such as a person's name and date of birth, with unique de-identifiers that prevent attribution of the data to a specific person without permission. Protection rules are centrally customized, configured and enforced, and they can be adjusted as required with downstream changes automatically applied.

Variant Twins provide unique feature advantages compared to more traditional solutions in the following ways:

- Universal data protection – Variant Twins can travel anywhere because granular data privacy and security controls are embedded within them, so they remain protected while in use regardless of their location.
- Specific scope of use – Variant Twins are designed to provide only the minimal level of identifying information necessary for a particular purpose, preventing processing or analysis beyond that specific use case.
- 100% accuracy – Variant Twins have the same mathematical properties as equivalent source data, ensuring no degradation in accuracy or value, so they deliver the same results as processing cleartext but without the risks of unauthorized re-identification.
- Dynamic pseudonymization – Protections in each transformed Variant Twin vary by time, purpose, place and use to prevent multiple datasets from being recombined to re-identify a data subject, defeating linkage and inference attacks.
- Controlled relinking – Only when authorized, can a Variant Twin be relinked to other Variant Twins or to source data, which is held separately and under the exclusive control of the designated party, so the resulting insights can be applied.

- Auditability – The re-linkability of Variant Twins, combined with audit logs of all activities within Data Embassy, ensure full capture of data lineage and protections applied.

Anonos' product strategy aligns with addressing the key privacy challenges of multi-party data sharing and analytics, particularly in cloud environments. The main data protection use cases for Data Embassy are the following:

- Breach- and ransomware-resistant data processing: Reduced surface area for external attacks or internal data misuse by obscuring identifying elements of personal data.
- Surveillance-proof data processing: Lawful international data transfers and processing to protect the identity of EU data subjects.
- External and internal data sharing: Breakdown of barriers that limit expanded access to third-party expertise in analytics, AI, ML, and new data monetization opportunities delivered via public cloud infrastructure.
- Data supply chain defensibility: Insulation of parties in the data supply chain from non-compliance liability

Challenges

Anonos also faces some market challenges:

- More education is needed on the issue of clear text processing in the cloud and Statuary Pseudonymization as one potential solution to the issue. IDC believes many organizations have not yet fully recognized the benefits of Statutory Pseudonymization as a security and data protection by design mechanism for personal data processing in the cloud. Cloud and security professionals are not yet fully attuned to its potential, with their focus having concentrated on traditional technical security controls.
- Cloud providers are working on solutions for confidential computing, which are addressing the topic of cleartext processing from a different angle.

Anonos will need to invest in further market education to broaden the general understanding of the topic of cleartext processing in the cloud, and partner with the large cloud providers to position Statuary Pseudonymization as a complementary technology solution.

Conclusion

Cloud adoption continues to grow and as organizations' cloud maturity increases, they are focusing on privacy and security for data in use. Statutory Pseudonymization is emerging as one PET that effectively supports many enterprise data use cases, so IDC believes the market for it will continue to grow. To the extent that Anonos can address the challenges outlined in this paper, the company has a significant opportunity for success.

About the Analysts

	Carla Arend, *Associate Vice President, Cloud Research* Carla Arend is an AVP with the European research team and heads up IDC's European cloud research. Arend provides industry clients with key insight into market dynamics, vendor activities, and end-user adoption trends in the European cloud market. As part of her research, she covers topics such as how European organizations are adopting cloud, how cloud drivers and inhibitors are evolving, cloud management, cloud security, data management in the cloud, IoT and cloud, AI and cloud, Devops and cloud, as well as GDPR impact on cloud and cloud code of conduct.
	Ralf Helkenberg*, Research Manager, European Privacy and Data Security* As research manager for the European Security group, Ralf Helkenberg provides insight and analysis on the European privacy and data security markets. His research covers the evolving regulatory landscape, and the market dynamics and technology trends within privacy management, de-identification, data discovery, encryption, key management, and data loss prevention.

MESSAGE FROM THE SPONSOR

More About Anonos

Anonos® is a global innovator in data privacy and security, providing the only enterprise software that protects data in use with total accuracy. The globally patented and award-winning Data Embassy® platform uses a combination of state-of-the-art privacy-enhancing technologies, including synthetic data, Statutory Pseudonymization, and other de-identification capabilities, to create Variant Twins®. These are non-identifiable yet 100% accurate variations of source data engineered for specific use cases to meet desired business outcomes. Because granular data privacy and security controls are centrally customized, configured and technologically enforced, Variant Twins can travel safely anywhere – across departments, outside the enterprise, or around the globe. Accuracy is 100% guaranteed because Variant Twins have the same mathematical value as the equivalent source data on which they're based, and their cleartext utility leads to faster enterprise insight and ROI. From testing though production, organizations can expand and expedite data-driven initiatives without compromising privacy, security, accuracy or speed.

To learn more, schedule a briefing at www.anonos.com.

IDC Custom Solutions

The content in this paper was adapted from existing IDC research published on www.idc.com.

IDC Research, Inc.
140 Kendrick Street
Building B
Needham, MA 02494, USA
T 508.872.8200
F 508.935.4015
Twitter @IDC
idc-insights-community.com
www.idc.com

This publication was produced by IDC Custom Solutions. The opinion, analysis, and research results presented herein are drawn from more detailed research and analysis independently conducted and published by IDC, unless specific vendor sponsorship is noted. IDC Custom Solutions makes IDC content available in a wide range of formats for distribution by various companies. A license to distribute IDC content does not imply endorsement of or opinion about the licensee.

External Publication of IDC Information and Data — Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Vice President or Country Manager. A draft of the proposed document should accompany any such request. IDC reserves the right to deny approval of external usage for any reason.