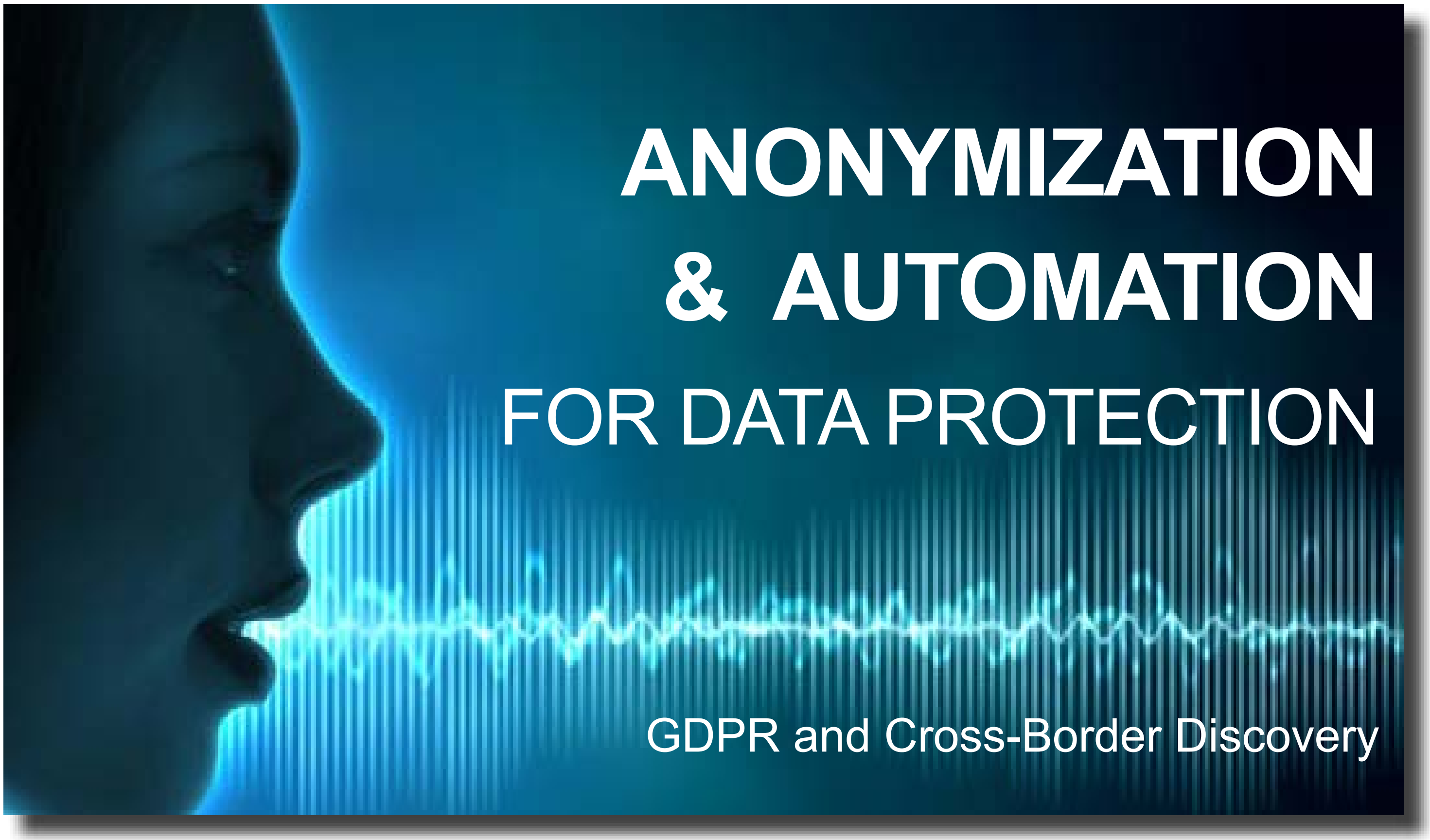

ANONYMIZATION & AUTOMATION

FOR DATA PROTECTION

GDPR and Cross-Border Discovery

With GDPR's presence being felt everywhere; data privacy is turning out to be an extremely difficult challenge for organizations plagued with data silos. Companies are on the lookout for new ways to process and utilize their personal data without having to violate the rules. Nevertheless, this is again challenging as GDPR narrows the ways through which personal data can be gathered and processed.

At times like these, anonymization, automation, and pseudonymization prove to be helpful in reducing the risks as well as in assisting the 'data processors' fulfil the data compliance regulations. Data anonymization is a must-have tool in every data scientist's 'privacy toolbox'. Not only does it protect private data, but it also does a great job in preserving the utility of the data to varying degrees; while, automation cut redaction costs.

GDPR, the abbreviated form of General Data Protection Regulations is introduced by the European Union to govern and regulate the processing of personal data relating to individuals in the EU by an organization, individual or company. U.S. companies that have a web presence and those which market their products and services over the web should also comply with GDPR.

Often times, U.S. companies without a physical presence in any EU countries collect data over the web. If so, the data collected is to be protected under the GDPR rules. Information such as health-related data, data regarding an individual's race, ethnic origin, religion, political opinions, and philosophical benefits, biometric data, and trade-union membership qualify as personal data and are considered sensitive.

In case of a breach, the EU regulator or 'supervising authority' should be notified within 72 hours. If a company fails to report a breach within the given timeframe, they are subject to a first-tier penalty of 2% of global revenue, or €10 million – whichever is higher.

Pseudonymization and Anonymization

The process of pseudonymization includes a specific something called artificial identifiers, or pseudonyms. These are tools which replace identifying fields in a data record in order to boost privacy. Often times there is just a single pseudonym per replaced field. However, a single pseudonym can work equally well for multiple replaced fields as well.

To put it simply, pseudonymization alters the records and replace personal information with alternate information. For instance, in a data record, every instance of 'Susan Brown" would be changed to a pseudonym 'Kathy Williams'.

According to Article 3 in GDPR, pseudonymization is “the processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information”.

Anonymization, unlike pseudonymization, removes or redacts the entire personal information from a data record. This makes it more difficult to identify a particular individual from a stored data collection.

According to Article 4(5) under GDPR Recital 26, the complete removal or redaction of names is also considered pseudonymization in case the organization still has the ability to link back to the pre-altered documents with the original information intact.

To perform anonymization that aligns with the GDPR’s requirements, every piece of personal data needs to be anonymized. **These mainly include information stated in the table.**

There are a variety of five techniques that could be employed to anonymize data, and each choice depends on the degree of risk related to the data. Below listed are a few methods available for anonymizing data.

SCRAMBLING refers to obfuscation or mixing of alphabets or letter, and often times, the process can be reversed. For instance, Matthew becomes Twmehta.

PERSONALIZEDANONYMIZATION, users can employ their own anonymization techniques. The custom anonymization is carried out using scripts or applications.

DIRECTORYREPLACEMENTmeans modifying the names of individuals incorporated within a data record, all the while maintaining uniformity between values.

DATA CATEGORIES

The SYNTRAN Anonymizer™ is programmed to anonymize data that has been categorized by the GDPR.

Standard Personal Data	Special Category Personal Data	Corporate Data
Name, Last Name	Gender	Company name
Postal Addresses	Physical Characteristics	Employee name
Location	Sexual Orientation	Product Name
Telephone Numbers	Political Affiliation	Department Name
Identification #’s	Religion	Financial Amounts
Credit Card Info	Race/Ethnicity	Job Title
Email Address	Personal Health Info (PHI)	Trade Secrets
IP Address	Country of Origin	Patent Info
Unique Device Identifier	Trade Union Membership	

MASKING gives data scientists the ability to identify data without actually manipulating identities. This is achieved by hiding part of the data with random characters or other data.

Blurring does an approximation of the data values and renders their meaning outdated thus making it impossible to identify individuals.

Though the techniques listed above are effective in anonymizing data, it is commonly applied to datasets containing personal data. Thus, you will have to get the consent to use the techniques.

SYSTRAN ANONYMIZER is analytics software that allows data anonymization. This is how we approach data anonymization:

- SYSTRAN Anonymizer allows mass automated anonymization and pseudonymization of electronic text-based documents with a high degree of accuracy thus cutting down the cost and amount of time required.
- Document containing personal information is sanitized for future use. This minimizes the risks associated with confidentiality and breach.
- Reduced redaction costs by employing automation.

Storing your personal data in a safe environment is of utmost importance and anonymization is definitely the best method to ensure the safety of the collected data. Moreover, the extra safety measures let you make use of the data in ways that non-anonymized data could not be used due to legal issues.

SYSTRAN Anonymizer allows you to add additional security measures to the data records, no matter how large your files are. What’s more, our tool does it all within the shortest timeframe and price point possible.

If you feel like learning more, feel free to contact us anytime.

Anonymization and Pseudonymization cannot guarantee GDPR compliance. However, these techniques can help lessen the risk of violations before personal data needs to be transferred outside of the European Union. Please note that SYSTRAN does not give legal advice and you should always consult with knowledgeable legal counsel about all the steps required for legal compliance under the GDPR or other regulations or statutes.

Anonymizer@systrangroup.com