

D2 – Method for digital evidence anonymisation



The project Automatization of digital forensics and incident response (ADFIR) funded by the European Union – Next GenerationEU through the Recovery and Resilience Plan of the Slovak Republic under project No. č. 09I05-03-V02-00079.

CONTENTS

1	Project description	3
2	Introduction	3
3	Task KPB1.2.....	4
3.1	Data inventory and data classification	5
3.1.1	Personal v non-personal data classification	5
3.1.2	Other types protected information	8
3.2	Compliance with legal conditions for personal data anonymisation.....	9
3.3	Selection of anonymisation methods.....	10
3.3.1	Randomization	11
3.3.2	Generalization	12
3.4	Output control	14
3.5	Usefulness of anonymized digital evidence for future analysis	14
4	Process overview.....	16

1 Project description

The project **Automatization of digital forensics and incident response** (hereinafter referred to as “**ADFIR**”) is funded by the **European Union – Next GenerationEU through the Recovery and Resilience Plan of the Slovak Republic** under project No. 09-I05-03-V02-00079. This project addresses one of the key challenges in cybersecurity and information security – how to process the massive volume of digital evidence generated during cybersecurity incidents or forensic investigations. Currently, this process is highly demanding in terms of human resources and time. Therefore, automation using machine learning methods can significantly **improve the quality of digital forensic analysis** and reduce the time required to perform it. Overall, this enables security teams to respond more effectively to cyber threats. Main benefits of this project are:

- **Accelerated Resolution of Cybersecurity Incidents.** The project ADFIR introduces automated approaches to collecting, processing, and analyzing digital traces. As a result, security teams can identify the causes of incidents more quickly and adopt effective measures to address them.
- **Reduced Workload for Forensic Analysts.** Routine and time-consuming tasks involved in processing digital traces will be replaced by automated methods. This will allow analysts to focus on more complex cases and strategic decision-making.
- **Higher Quality and Consistency of Outputs.** The use of unified methodologies and tools ensures that the processed digital traces will be more accurate, consistent, and easily verifiable. This significantly reduces the risk of errors caused by human factors.
- **Potential Use in Criminal Proceedings.** The project outputs will be developed in compliance with legal requirements and standards, allowing the digital traces to be accepted as relevant evidence for investigations and court proceedings.

2 Introduction

The objective of the work package KPB1 is to analyse and **respond to legal issues that relate to the proposed project**. This work package defines minimal standards and legal guarantees necessary to ensure the application of the personal data protection legal framework to the processing of digital evidence and the admissibility of acquired outputs of the automatized forensic analysis as evidence in criminal court proceedings, while considering the ethical, legal, legitimacy and proportionality aspects of the machine learning extraction of digital evidence. The features of the digital evidence (integrity, authenticity, admissibility and relevance) will also be examined.

Since the project implements a hybrid approach combining data from simulated attacks in a controlled environment with anonymised artifacts from real practice in the implementation phase, provision of a corresponding legal analysis of data anonymisation is necessary. The output provided is a proposal for digital evidence anonymisation method, identifying the

necessary steps to be taken to ensure legal compliance. The aim is to provide a toolkit that allows for the differentiation between personal and non-personal data and further, if personal data is identified, to ensure their processing and anonymization in accordance with the applicable regulation.

3 Task KPB1.2

Task KPB1.2 – Analysis of the conditions for digital evidence anonymization and inclusion of data relating to security incidents into dataset. Consideration of legal requirements for dataset creation and conditions for the inclusion of data into the dataset (UPJS). In this task, we examine the applicable regulation on personal data protection and how it may be applied in relation to digital evidence examination and processing within the scope of the project.

Digital evidence can encompass a wide range of different information of probative value stored or transmitted in digital form.¹ Specific information included in digital evidence may be awarded legal protection under different regulations. An example of such information includes, e. g. data that may be classified as **personal data** under the applicable personal data protection regulation contained primarily in the 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**) (hereinafter only as 'GDPR') that has general application, is binding in its entirety and is directly applicable in all Member States. In the national context the corresponding regulation can be found in the Act No. 18/2018 Coll. on personal data protection as amended.

If digital evidence is to be examined and processed for purposes different from the criminal proceedings for which it was originally obtained by competent authorities or other subjects², it must adhere to the applicable regulation. In the following chapters, we outline the necessary steps to be taken in order to allow further use of digital evidence through its anonymisation. In summary, these include:

¹ Scientific Working Group on Digital Evidence (SWGDE). *Digital Evidence: Standards and Principles*. In *Forensic Science Communications*. 2000, No. 2. Available: <https://archives.fbi.gov/archives/about-us/lab/forensic-sciencecommunications/fsc/april2000/swgde.htm#Introduction>

² As Dražová points out, within the scope of their competences, various entities secure data as potential digital evidence outside criminal proceedings, including national or government units for dealing with cyber security incidents, public authorities for civil or administrative proceedings, digital investigation teams operating in the commercial sphere, and the like. Dražová, P. *Electronic evidence in criminal proceedings*. Dissertation thesis, 2021. Available:

<https://opac.crzp.sk/?fn=detailBiblioFormChildA52T&sid=4BD9478A3A7038F60D223DAC8853&seo=CRZP-detail-kniha>

1. data inventory and data classification as personal or non-personal data,
2. ensuring compliance with legal conditions for personal data anonymisation,
3. selection of appropriate data anonymisation methods,
4. implementation of selected data anonymisation methods,
5. output control.

3.1 Data inventory and data classification

The first necessary step is the **creation of a data inventory**. In practice, this includes the determination of all information being processed and creating a list of such information, e. g. in a form of a table.

The following step is the examination of this information with the objective to determine its legal nature, specifically whether such information possesses specific legal protection. The created table of information should provide the conclusions reached and the categories of information determined. In most instances, the categories of information will include the distinction between **personal and non-personal data**, but other categories may also be used (e.g. sensitive data, personally identifiable information, confidential information, trade secrets), depending on the applicable regulation.

Table 1 Example of data inventory and classification

INFORMATION	CLASSIFICATION OF INFORMATION
name	personal data
email address	personal data / confidential information
file name	personal data
IP address	personal data
time stamp	personal data
list of customers	trade secret

3.1.1 Personal v non-personal data classification

Pursuant to Article 4 (1) of GDPR, **personal data** covers „any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.“

Article 9 of GDPR also differentiates the **special categories of personal data** that include personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation. Processing of special categories of personal data

is, in principle, **prohibited**, unless one of the legal basis for the processing of such data defined in Article 9 (2) GDPR is applicable.

The classification of information as personal data is key for the determination of the applicability of personal data protection regulation in relation to the information in question. If specific information cannot be categorised as personal data, its processing will not be subject to conditions and restrictions defined in the applicable personal data protection regulation.

In determining whether information can be classified as personal data, **all means reasonably likely to be used** by the controller (the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data) or other person must be considered to reach a definitive conclusion on the nature of information in question. Data can be identifiable, e.g., when it contains directly identifying information (name, video or audio recordings, email address, location data, etc., it allows to single out an individual (in the presence of unique data points or behavioural patterns that can apply to only one person), if it is possible to infer information about an individual on the basis of the dataset created or to link records that relate to an individual (through combination of attributes in the dataset or linking of multiple datasets), or if de-identification is possible (pseudonymization). In general, if information is collected from natural persons, it can be assumed that it will present personal data.

Within the personal data protection framework, the information that does not fall under the wide definition of personal data is considered as **non-personal data** that includes:

- a) data which originally did not relate to an identified or identifiable natural person, such as data on weather conditions generated by sensors installed on wind turbines or data on maintenance needs for industrial machines, and
- b) data which were initially personal data but were later made anonymous (**anonymous data**). Properly anonymised data cannot be attributed to a specific person, not even by use of additional data and are therefore non-personal data.³

Pursuant to Recital 26 of GDPR, *„the principles of data protection should therefore not apply to **anonymous information**, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, **including for statistical or research purposes**“*⁴.

³ Communication from the Commission to the European parliament and the Council. Guidance on the Regulation on a framework for the free flow of non-personal data in the European Union. COM(2019) 250 final.

⁴ Currently, GDPR does not contain a legal definition of the term ‘scientific research’. A non-binding definition can be found in Recital 159 GDPR: *„scientific research purposes should be interpreted in a broad manner including for example technological development and demonstration, fundamental research, applied research and*

This provision provides a **conceptual definition of anonymisation**. It requires that data subjected to anonymisation are **stripped of sufficient elements such that the data subject can no longer be identified**. An important factor is that the processing must be **irreversible**. The regulation does not clarify how such a deidentification process should or could be performed but focuses on **the outcome that must be achieved** that is the **impossibility of data subject identification**.

Anonymisation is also defined in **international standards** such as the ISO 29100 one – being the *“Process by which personally identifiable information (PII) is irreversibly altered in such a way that a PII principal can no longer be identified directly or indirectly, either by the PII controller alone or in collaboration with any other party”* (ISO 29100:2011). Irreversibility of the alteration undergone by personal data to enable direct or indirect identification is the key also for ISO.

The assessment of whether data is properly anonymised and therefore whether in fact the data protection regulation is inapplicable **depends on specific and unique circumstances of each individual case**. Several examples of re-identification of datasets that were supposedly anonymised have showed that such an evaluation may be demanding.⁵ The creation of a truly anonymous dataset from a rich set of personal data, whilst retaining as much of the underlying information as required for the task, is not a simple proposition. For example, a dataset considered to be anonymous may be combined with another dataset in such a way that one or more individuals can be identified.⁶ Inherent to the anonymisation process is **the risk of anonymisation failure of reversal**; the severity of this risk and the likelihood of its occurrence should be considered.

To establish whether an individual is identifiable, one has to look on **all means reasonably likely to be used** by a controller or by another person to identify an individual directly or indirectly. If non-personal data can be related to an individual **in any way**, causing them to be either directly or indirectly identifiable, the data must be considered as personal data.

privately funded research“. Moreover, according to the Explanatory Report to the Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, **“statistical purposes”** refers to the elaboration of statistical surveys or the production of statistical, aggregated results; statistics aim at analysing and characterising mass or collective phenomena in a considered population. Moreover, processing of data for **“scientific research purposes”** aims at providing researchers with information contributing to an understanding of phenomena in varied scientific fields (epidemiology, psychology, economics, sociology, linguistics, political science, criminology, etc.) with a view to establishing permanent principles, laws of behaviour or patterns of causality which transcend all the individuals to whom they apply.

⁵ See The study on future data flows conducted for the European Parliament’s ITRE Committee by Blackman, C., Forge, S.: Data Flows — Future Scenarios: In-Depth Analysis for the ITRE Committee, 2017, p. 22, Box 2. Available online

at: [http://www.europarl.europa.eu/RegData/etudes/IDAN/2017/607362/IPOL_IDA\(2017\)607362_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/IDAN/2017/607362/IPOL_IDA(2017)607362_EN.pdf)

⁶ ARTICLE 29 DATA PROTECTION WORKING PARTY. Opinion 05/2014 on Anonymisation Techniques. Adopted on 10 April 2014. 0829/14/EN WP216.

In this regard, it is relevant to note the term pseudonymisation. Pursuant to Article 4 (5) of GDPR, pseudonymisation means *“the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.”* Pseudonymised data, which could be attributed to a natural person by the use of additional information, is to be considered **information on an identifiable natural person**, and is therefore personal.⁷ In simplified terms, pseudonymised data is **still personal data** and is subject to the applicable personal data protection regulation.

3.1.2 Other types protected information

Personal data are not the only type of information that is protected by the applicable regulation. Special protection is also awarded to:

1. **confidential information** that, in general, includes any information determined as confidential by the relevant subjects, primarily contractual parties within the scope of their contractual agreements;
2. **trade secrets** that protect information which meets all of the following requirements:
 - a) it is secret in the sense that it is not, as a body or in the precise configuration and assembly of its components, generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question;
 - b) it has commercial value because it is secret
 - c) it has been subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep it secret.⁸

Information may also be protected within the scope of intellectual property regulation, e. g. non-registered trademarks. The applicability of such specific legal protection must be considered on a case-by-case basis and reflected in the data inventory and data classification process.

⁷ European Data Protection Board. Guidelines 01/2025 on Pseudonymisation. Adopted on 16 January 2025. P. 3.

⁸ Article 2 (1) of the Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure. *OJ L 157, 15.6.2016, pp. 1–18.*

3.2 Compliance with legal conditions for personal data anonymisation

If it is determined that data under examination should be anonymised, the following necessary step is to ensure that the anonymisation process is realised in accordance with the applicable personal data protection regulation.

Anonymisation of personal data presents a form of personal data processing. Pursuant to Article 4 (2) of GDPR, 'processing' means *"any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction."* Any processing of personal data is subjected to **principles relating to processing of personal data** specified in Article 5 GDPR.

One of the most important principles of personal data processing is the **lawfulness of processing**. Personal data processing is lawful only if and to the extent that at least one the legal basis for processing provided in Article 6 (1) of GDPR is fulfilled. Specifically, this provision defines six legal bases for personal data processing that include:

- a) **data subject consent** to the processing of their personal data for one or more specific purposes,
- b) necessity of personal data processing for the **performance of a contract** to which the data subject is party or for actions required prior to entering into a contract by the data subject at their request,
- c) **compliance with a legal obligation** to which the controller is subject,
- d) **protection of vital interests** of the data subject or of another natural person,
- e) necessity of personal data processing for the **performance of a task carried out in the public interest** or in the exercise of official authority vested in the controller,
- f) necessity of personal data processing for the purposes of the **legitimate interests** pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.

The processing of personal data must also be **fair and transparent**. Data subjects should be aware about the processing of their personal data and the extent of such processing. Data subjects should be made aware of risks, rules, safeguards and rights in relation to the processing of personal data and how to exercise their rights in relation to such processing.⁹

⁹ Recital 39 GDPR.

The principle of purpose limitation requires that personal data are collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, **scientific** or historical **research purposes** or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes. The specific purposes for which the personal data are processed should be explicit and legitimate and determined **at the time of personal data collection**.¹⁰ Personal data should be processed only if the purpose of the processing could not reasonably be fulfilled by other means.¹¹

The principle of data minimisation requires that personal data are adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed.

The accuracy principle ensures that personal data being processed are accurate and, where necessary, kept up to date. Moreover, every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay.

The storage limitation principle specifies that personal data must be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed. Personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject. To ensure the compliance with this principle, time limits should be established by the controller for erasure or for periodic review.¹²

The integrity and confidentiality principle stipulates that personal data must be processed in a manner that ensures appropriate security of personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures.

The subject responsible for the compliance with these principles is the controller. Controller must be able to **demonstrate the compliance** in this regard (**principle of accountability**).

3.3 Selection of anonymisation methods

The next step is to determine which anonymisation methods will be implemented. Data anonymisation should always be performed using **the latest state-of-the-art anonymisation**

¹⁰ Recital 39 GDPR.

¹¹ Ibidem.

¹² Ibidem.

techniques, reflecting the developments in technology and data analytics that may enable the conversion of anonymised data into personal data. In this chapter, we describe the different personal data anonymization methods relevant in the examined context of digital evidence anonymization.

Personal data can be anonymized through different anonymization methods or techniques. The controller must consider the main risks associated with anonymization process, namely:

1. **singling-out** - possibility to isolate some or all records which identify an individual in the dataset;
2. **linkability** - the ability to link, at least, two records concerning the same data subject or a group of data subjects (either in the same database or in two different databases);
3. **inference** - the possibility to deduce, with significant probability, the value of an attribute from the values of a set of other attribute.¹³

3.3.1 Randomization

Randomization is a family of techniques that **alters the veracity of the data** to remove the strong link between the data and the individual. If the data are sufficiently uncertain then they can no longer be referred to a specific individual,¹⁴ and therefore cannot be classified as personal data, the processing of which is subject to the applicable regulation.

Randomization by itself will not reduce the singularity of each record as each record will still be derived from a single data subject but may protect against inference attacks/risks and can be combined with generalization techniques to provide stronger privacy guarantees. **Additional techniques may be required** to ensure that a record cannot identify a single individual.¹⁵

3.3.1.1 Noise addition

Noise addition consists of modifying attributes in the dataset such that they are less accurate whilst retaining the overall distribution. It alters attributes, introducing randomness to the dataset, ensuring that the data examined cannot be linked to individuals, while simultaneously preserving aggregate patterns relevant for future research. This method will often need to be combined with other anonymisation techniques such as the removal of obvious attributes and quasi-identifiers. The level of noise should depend on the necessity of the level of information required and the impact on individuals' privacy as a result of disclosure of the protected attributes.¹⁶

¹³ ARTICLE 29 DATA PROTECTION WORKING PARTY. Opinion 05/2014 on Anonymisation Techniques. Adopted on 10 April 2014. 0829/14/EN WP216.

¹⁴ Ibidem

¹⁵ Ibidem

¹⁶ Ibidem

3.3.1.2 Permutation

Permutation rearranges the values of attributes in a table so that some of them are artificially linked to different data subjects, thus breaking direct links between individual and their personal data. To illustrate, in a table including names of employees and their salaries, this method would shuffle the corresponding salaries across all individuals without altering the salary amounts. Such swapping will ensure that range and distribution of values will remain the same but correlations between values and individuals will not.¹⁷

3.3.1.3 Differential privacy

Differential privacy adopts a different approach: while, in fact, noise insertion comes into play beforehand when dataset is supposed to be released, differential privacy can be used when the data controller generates anonymised views of a dataset whilst retaining a copy of the original data.¹⁸ Differential privacy tells the data controller how much noise he needs to add, and in which form, to get the necessary privacy guarantees.¹⁹

3.3.2 Generalization

A second family of anonymisation techniques consists of generalizing, or diluting, the attributes of data subjects by modifying the respective scale or order of magnitude.²⁰

3.3.2.1 Aggregation and K-anonymity

Aggregation and K-anonymity techniques aim to prevent a data subject from being singled out by grouping them with, at least, k other individuals. To achieve this, the attribute values are generalized to an extent such that each individual shares the same value.²¹

¹⁷ ARTICLE 29 DATA PROTECTION WORKING PARTY. Opinion 05/2014 on Anonymisation Techniques. Adopted on 10 April 2014. 0829/14/EN WP216.

¹⁸ Ibidem

¹⁹ Cf. Ed Felten (2012) Protecting privacy by adding noise. URL: <https://techatftc.wordpress.com/2012/06/21/protecting-privacy-by-adding-noise/>.

²⁰ ARTICLE 29 DATA PROTECTION WORKING PARTY. Opinion 05/2014 on Anonymisation Techniques. Adopted on 10 April 2014. 0829/14/EN WP216.

²¹ Ibidem

3.3.2.2 L-diversity

L-diversity extends k-anonymity to ensure that deterministic inference attacks are no longer possible by making sure that in each equivalence class every attribute has at least l different values. One basic goal to achieve is to limit the occurrence of equivalence classes with poor attribute variability, so that an attacker with background knowledge on a specific data subject is always left with a significant uncertainty. It is useful to protect data against inference attacks when the values of attributes are well distributed. It cannot prevent the leakage of information if the attributes within a partition are unevenly distributed or belong to a small range of values or semantic meanings. In the end, l -diversity is subject to probabilistic inference attacks.²²

3.3.2.3 T-closeness

T-closeness is a refinement of l -diversity, in that it aims to create equivalent classes that resemble the initial distribution of attributes in the table. This technique is useful when it is important to keep the data as close as possible to the original one; to that end, a further constraint is placed on the equivalence class, namely that not only at least l different values should exist within each equivalence class, but also that each value is represented as many times as necessary to mirror the initial distribution of each attribute.²³

The selection of specific data anonymisation methods or their combination to be implemented in practice must be determined by the specific controller, reflecting:

- the nature of data in question and its sensitivity - different data categories require different levels of protection,
- the scope of the dataset,
- the intended purpose for future data use and the desired accuracy of the outcome sought,
- the risk of re-identification, e. g. the availability of external datasets that could be linked or other supplementing information,
- technical feasibility, scalability and maintainability,
- contextual integrity, and other relevant factors.

Given the broad scope of possible datasets and anonymisation techniques, as well as future developments in (re-)identification techniques, any assessments will require a case-by-case analysis.

²² ARTICLE 29 DATA PROTECTION WORKING PARTY. Opinion 05/2014 on Anonymisation Techniques. Adopted on 10 April 2014. 0829/14/EN WP216.

²³ Ibidem

3.4 Output control

Following the selection and implementation of data anonymisation techniques, the last necessary step is to ensure the achievement of the desired outcome, namely the successfulness of data anonymisation. Therefore, the output created through the application of selected data anonymisation techniques must be checked.

Following the recent developments in the evaluation of the anonymisation successfulness, the European Data Protection Board proposed three criteria to be considered, namely (1) No Record Isolation, (2) No Linkage and (3) No Inference. **The No Record Isolation criterion** is met if the data does not contain a unique combination of attribute values that relate to a single individual. **The No Linkage criterion** is met if the data does not contain an individual's record which could be linked to another record which (a) also relates (with certainty or high likelihood) to that same individual, and (b) comes from a different dataset. **The No Inference criterion** is met if no specific and meaningful inference can be drawn from the given data. Inferences may be based on the given data and/or on additional information, including information about the anonymisation process. In particular:

- An inference is specific if the inferred information relates to a single identified or identifiable individual (i.e. it is personal data); and
- An inference is both specific and meaningful if the processing of the inferred information is also liable to have an effect on the data subject's rights and interests, relies on the given data and could not be obtained from general knowledge or from data about the population at large.²⁴

Where if all three of the criteria are met, the information may be regarded as anonymous.

3.5 Usefulness of anonymized digital evidence for future analysis

Anonymisation of digital evidence can preserve its usefulness for future automated analysis if it is designed around structure preservation rather than blunt deletion. Instead of stripping or randomly masking identifying fields, different techniques that retain the statistical patterns, relationships, and formats that algorithms rely on should be used. Maintaining a consistent mapping scheme—where the same real-world entity is always replaced with the same anonymised token—ensures that pattern recognition, link analysis, and machine learning models can still detect correlations across records without ever exposing the original identities. It is also important to document the anonymisation methodology and retain metadata about data types, units, and relationships, so that automated tools can correctly

²⁴ European Data Protection Board. Guidelines 02/2026 on Anonymisation. Adopted on 07 July 2026 for public consultation.

interpret the transformed dataset. Finally, anonymisation should be validated against the specific analytical techniques intended for future use (e.g., clustering, anomaly identification, association rules, etc.) to confirm that essential features and variances in the data are not inadvertently flattened or destroyed, striking a balance between privacy protection and analytical utility.

4 Process overview



METHOD FOR DIGITAL EVIDENCE ANONYMISATION



DATA INVENTORY

- determination of all information being processed
- creation of a list of such information, e. g. in a form of a table



DATA CLASSIFICATION

- examination of information processed and determination of its legal nature, specifically whether such information possesses specific legal protection
- the table of information should provide the conclusions reached and the categories of information determined

PERSONAL V. NON-PERSONAL DATA

- Article 4 (1) of GDPR - **personal data** covers „any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.“
- **non-personal data:**
 - a. data which originally did not relate to an identified or identifiable natural person, and
 - b. data which were initially personal data but were later made anonymous (anonymous data)



LEGAL COMPLIANCE

- if it is determined that data under examination should be anonymised, this process must adhere to the applicable regulation
- anonymisation of personal data presents a form of personal data processing
- any processing of personal data is subjected to principles relating to processing of personal data specified in Article 5 GDPR



SELECTION OF ANONYMISATION METHODS

- data anonymisation should always be performed using **the latest state-of-the-art anonymisation techniques**, reflecting the developments in technology and data analytics that may enable the conversion of anonymised data into personal data



OUTPUT CONTROL

- the successfullness of data anonymisation must be checked

