EUCAIM

CANCER IMAGE EUROPE

Project title: European Federation for Cancer Images
Project acronym: EUCAIM
Grant Agreement: 101100633
Call identifier: DIGITAL-2022-CLOUD-AI-02

D3.5: [Report describing the EUCAIM’s anonymisation and pseudonymisation strategy]

Partner(s): KUL, UV
Author(s): Janos Meszaros, Ricard Martinez, Francisco R. Soriano, Alma Virtu, Isabelle Huys, Valia Kalokyri, Jose Munuera Mora
Date of delivery: 30\12\2024
Version: **1.5**

Contents

Executive summary

This deliverable presents the EUCAIM Project’s anonymization and pseudonymization strategy, ensuring compliance with the GDPR, EHDS, and other key EU regulations. It outlines technical and organizational measures to protect patient privacy while enabling responsible secondary use of health data for research and innovation.

Key features include secure processing environments, de-identification profiles, and tools like the EUCAIM Anonymizer Wizard Tool, designed to mitigate re-identification risks and enhance data security. Governance structures and risk assessment methodologies further reinforce accountability and compliance.

Acronyms and Abbreviations

AI Artificial Intelligence
CJEU Court of Justice of the European Union
DGA Data Governance Act
DoA Description of Actions
EDPB European Data Protection Board
EDPS European Data Protection Supervisor
EHDS European Health Data Space
EHDSR Proposal for a Regulation of the European Health Data Space
ENISA European Union Agency for Cybersecurity
EU European Union
GDPR General Data Protection Regulation
HLEG AI High-Level Expert Group on Artificial Intelligence
NIS Network and Information Security
WP Work Package
WP29 Article 29 Working Party

Disclaimer

The opinions stated in this report reflect the opinions of the authors and not the opinion of the European Commission.

All intellectual property rights are owned by the consortium of EUCAIM under terms stated in their Consortium Agreement and are protected by the applicable laws. Reproduction is not authorized without prior written agreement. The commercial use of any information contained in this document may require a license from the owner of the information.

I. Introduction

The EUCAIM Project proposes to deploy a pan-European digital federated infrastructure of FAIR cancer-related anonymized images from Real-World. The infrastructure is designed to preserve the data sovereignty of providers and provide a platform, including an Atlas of Cancer Images, for the development and benchmarking of AI tools towards Precision Medicine. EUCAIM will address the fragmentation of existing cancer image repositories by building on repositories of the AI4HI initiative, European Research Infrastructures and national/regional repositories and include clinical images, pathology, molecular and laboratory data.

This **deliverable outlines the anonymization and pseudonymization strategies implemented** by EUCAIM to safeguard privacy while maintaining the usability of data for research and innovation. Moreover, **Annex I: Legal requirements for anonymization** is a report from Dr. Ricard Martinez, which reflects the **"legal state of art" regarding anonymization**. In adherence to the General Data Protection Regulation (GDPR), the European Health Data Space (EHDS) proposal, and other regulatory frameworks, the report details the technical and organizational measures undertaken to balance data utility with robust privacy protections.
By addressing challenges such as re-identification risks and variability in data formats across the EU, **EUCAIM introduces tools and methodologies, including secure processing environment, de-identification profiles, and risk mitigation frameworks**. These strategies ensure that the project not only meets its compliance obligations but also aims to serve as a benchmark for ethical and **secure data management in the health research domain**.

II. Legal Background for technical and organizational measures to protect privacy

In the ever-evolving landscape of data-driven innovation, the responsible handling of personal data, especially within the health data domain, is paramount. **This chapter introduces the most important regulations** that contour our approach to safeguarding the rights and freedoms of data subjects. With a primary focus on the GDPR, the unique challenges posed by the European Health Data Space, and the emerging regulatory landscape marked by the AI Act. Nevertheless, variations in national regulations within the EU underscore the need for partners and their Data Protection Officers (DPOs) to stay informed and comply with these diverse measures.

1) General Data Protection Regulation (GDPR)

The General Data Protection Regulation (GDPR) is a comprehensive data protection and privacy regulation implemented by the European Union (EU) in 2018. It establishes **a framework for the collection, processing, and storage of personal data of EU citizens**, granting them greater control over their information. The GDPR imposes strict requirements on organizations handling such data, including the need for clear consent, the right to access and rectify personal information, and the obligation to implement robust security measures to safeguard data. Non-compliance with GDPR can result in significant fines, making it a crucial legal framework for protecting individuals' privacy rights.

Information security plays an instrumental role in guaranteeing fundamental rights. Therefore, the **GDPR defines it as one of its fundamental principles** in Article 5, as an essential objective, and as an obligation for controllers and processors. Under the 'risk-based approach', the controller must take appropriate measures in accordance with the state of the art.

GDPR Recital 78
Personal data should be processed in a manner that ensures **appropriate security and confidentiality** of the personal data, including for preventing unauthorised access to or use of personal data and the equipment used for the processing.[1]

GDPR Article 32
Taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the **controller and the processor shall implement appropriate technical and organisational measures** to ensure a level of security appropriate to the risk, including inter alia as appropriate:
(a) the pseudonymisation and encryption of personal data;
(b) the ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services;
(c) the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident;
(d) a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing.

[1] GDPR Recital 39

2. In assessing the appropriate level of security account shall be taken in particular of the risks that are presented by processing, in particular from accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to personal data transmitted, stored or otherwise processed.

The European Data Protection Board highlighted the important of data protection by design and provided a non-exhaustive list of the most important requirements in the **guidelines 4/2019 on Article 25 Data Protection by Design and by Default Version 2.0 Adopted on 20 October 2020**[2]

"85. Key design and default integrity and confidentiality elements may include:
· **Information security management system (ISMS)** – Have an operative means of managing policies and procedures for information security.
· **Risk analysis** – Assess the risks against the security of personal data by considering the impact on individuals' rights and counter identified risks. For use in risk assessment; develop and maintain a comprehensive, systematic and realistic "threat modelling" and an attack surface analysis of the designed software to reduce attack vectors and opportunities to exploit weak points and vulnerabilities.
· **Security by design** – Consider security requirements as early as possible in the system design and development and continuously integrate and perform relevant tests.
· **Maintenance** – Regular review and test software, hardware, systems and services, etc. to uncover vulnerabilities of the systems supporting the processing.
· **Access control management** – Only the authorized personnel should have access to the personal data necessary for their processing tasks, and the controller should differentiate between access privileges of authorized personnel.
o **Access limitation (agents)** – Shape the data processing in a way that a minimal number of people need access to personal data to perform their duties, and limit access accordingly.
o **Access limitation (content)** – In the context of each processing operation, limit access to only those attributes per data set that are needed to perform that operation. Moreover, limit access to data pertaining to those data subjects who are in the remit of the respective employee.
o **Access segregation** – Shape the data processing in a way that no individual needs comprehensive access to all data collected about a data subject, much less all personal data of a particular category of data subjects.
· **Secure transfers** – Transfers shall be secured against unauthorized and accidental access and changes.
· **Secure storage** – Data storage shall be secure from unauthorized access and changes. There should be procedures to assess the risk of centralized or decentralized storage, and what categories of personal data this applies to. Some data may need additional security measures than others or isolation from others.
· **Pseudonymization** – Personal data and back-ups/logs should be pseudonymized as a security measure to minimise risks of potential data breaches, for example using hashing or encryption.
· **Backups/logs** – Keep back-ups and logs to the extent necessary for information security, use audit trails and event monitoring as a routine security control. These shall be protected from unauthorised and accidental access and change and reviewed regularly and incidents should be handled promptly.
· **Disaster recovery/ business continuity** – Address information system disaster recovery and business continuity requirements to restore the availability of personal data following up major incidents.

[2] • EDPB, "Guidelines 4/2019 on Article 25 Data Protection by Design and by Default" Version 2.0. Adopted on 20 October 2020, <https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-42019-article- 25-data-protection-design-and_en>.

· **Protection according to risk** – All categories of personal data should be protected with measures adequate with respect to the risk of a security breach. Data presenting special risks should, when possible, be kept separated from the rest of the personal data.
· **Security incident response management** – Have in place routines, procedures and resources to detect, contain, handle, report and learn from data breaches.
· **Incident management** – Controller should have processes in place to handle breaches and incidents, in order to make the processing system more robust. This includes notification procedures, such as management of notification (to the supervisory authority) and information (to data subjects)

2) The European Health Data Space (EHDS)

Since EUCAIM aims to provide a **secure processing environment (SPE),** aligned with the European Health Data Space, it is crucial to align with the EDHS proposal.[3] The European data strategy of February 2020 announced the creation of data spaces in 10 strategic fields.[4] The first legislative proposal to emerge in a specific area was for a European Health Data Space (EHDS).[5] The EHDS' main objectives are three-fold, namely to: 1) **increase control for natural persons** over their electronic health data, 2) create a legal framework consisting of **trusted governance mechanisms** and a secure processing environment, and 3) contribute to a genuine **single market for digital health products and services**.[6]
The EHDS builds upon legislation such as the GDPR, the medical devices and cybersecurity legal frameworks,[7] the Data Act and AI Act. The EHDS establishes rules for the primary and secondary use of data. Primary use is defined as "the processing of personal electronic health data for the provision of health services",[8] whereas secondary use encompasses the use of electronic health data for broader needs, such as health research or policy making.[9]

EHDS Proposal Recital 54.
The health data access body or the data holder providing this service should remain at all time in **control of the access to the electronic health data** with access granted to the data users determined by the conditions of the issued data permit. Only non-personal electronic health data which do not contain any electronic health data should be extracted by the data users from such secure processing environment. Thus, it is an essential safeguard to preserve the rights and freedoms of natural persons in relation to the processing of their electronic health data for secondary use.

EHDS Proposal Article 50
Secure processing environment

[3] See Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space, available at https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52022PC0197
[4] https://digital-strategy.ec.europa.eu/en/policies/strategy-data
[5] Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space COM/2022/197 final.
[6] Explanatory memorandum accompanying the EHDS proposal and EHDS proposal Recital 1
[7] Regulation (EU) 2017/745 on medical devices, Regulation (EU) 2017/746 on in vitro diagnostic medical devices, Directive 2016/1148 on security of network and information systems
[8] EHDS proposal Article 2(2)(d)
[9] EHDS proposal Article 2(2)(e) and Chapter IV; Marcus et al. The European Health Data Space. Study requested by the ITRE committee, December 2022.

1. The health data access bodies shall provide access to electronic health data only through a secure processing environment, **with technical and organisational measures and security and interoperability** requirements. In particular, they shall take the following security measures:
(a) **restrict access** to the secure processing environment to authorised persons listed in the respective data permit;
(b) minimise the risk of the unauthorised reading, copying, modification or removal of electronic health data hosted in the secure processing environment through state-of-the-art technological means;
(c) limit the input of electronic health data and the inspection, modification or deletion of electronic health data hosted in the secure processing environment to a limited number of authorised identifiable individuals;
(d) ensure that data users have access only to the electronic health data covered by their data permit, by means of individual and unique user identities and confidential access modes only;
(e) keep **identifiable logs of access** to the secure processing environment for the period of time necessary to verify and audit all processing operations in that environment;
(f) ensure **compliance and monitor** the security measures referred to in this Article to mitigate potential security threats.

2. The health data access bodies shall ensure that electronic health data can be uploaded by data holders and can be accessed by the data user in a secure processing environment. The data users shall only be able to download non-personal electronic health data from the secure processing environment.
3. The health data access bodies shall ensure **regular audits** of the secure processing environments.
4. The **Commission shall, by means of implementing acts, provide for the technical, information security and interoperability requirements for the secure processing environments**. Those implementing acts shall be adopted in accordance with the advisory procedure referred to in Article 68(2).

3) Data Governance Act

The Data Governance Act (DGA) is seeking to establish a framework for the responsible and sustainable use of data within the European Union. It aims to **increase data availability by facilitating data sharing, promote trust and security in data handling**, and foster innovation and competitiveness. The DGA introduces guidelines for data intermediaries, implements a data access obligation, and proposes a data quality framework to achieve these objectives.[10]

DGA Article 2
(20) ‘**secure processing environment’** means the physical or virtual environment and organisational means to ensure compliance with Union law, such as Regulation (EU) 2016/679, in particular with regard to data subjects’ rights, intellectual property rights, and commercial and statistical confidentiality, integrity and accessibility, as well as with applicable national law, and to allow the entity providing the secure processing environment to determine and supervise all data processing actions, including the display, storage, download and export of data and the calculation of derivative data through computational algorithms.

[10] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act)

II. Measures implemented, and to be implemented

In this part, we will describe in detail which technical measures are already in place and those that are planned to be implemented during the development, testing, and application of the services in the EUCAIM Project. **This is not an exhaustive list of measures. The listed measures and additional ones need to be continuously reassessed according to the risks and state-of the art technology.** The status of current implementation or future ambition are indicated for each single measure:

1) Secure Processing Environment

EUCAIM provides a secure processing environment while providing flexible requirements to fit a wide range of use cases. Although security requirements will be further refined throughout the project, the basis will be the following:

- Users of the platform will be properly identified and will need to register to the platform to access the services of EUCAIM. **Only registered users** will be able to submit access data access request proposals.
- **Restricted access to data** will be made through either Virtual Research Environments, with on-site access to data, restricting downloading the data out of the platform and providing users with tools and resources to process the data; or blind distributed processing in which users will be able to run processing pipelines without being able to display the data.
- Actions performed in the platform will be registered in a non-repudiable form, and system administrators will be able to browse these **logs** when requested.
- The platform will go through a Privacy Assessment and Cyber Security Analysis.
- All communications will be performed under **secure protocols and encrypted channels**.
- Data in rest will be stored unencrypted, but access to the physical resources will be **restricted** to the system administrators and developers of the platform.
- **Applications will be audited** prior to be registered in the platform and will follow strict guidelines in terms of security and vulnerabilities.

These points are described in more detail in the following subsections.

2) Anonymization and pseudonymisation of personal data

I. The legal status of anonymization and pseudonymization

Pseudonymization (art. 4 (5) GDPR) is a data management and de-identification procedure that increases privacy by replacing the identifying fields of individuals by using one or more artificial identifiers, or pseudonyms, that cannot be linked directly to their corresponding nominative identities. Pseudonymized data is usually still regarded to be personal data and remains under the scope of the GDPR in the EU Member States.

As regards anonymization, Recital 26 of the GDPR states that the principles of data protection do not apply to anonymous. It should be ensured that anonymization is engineered appropriately in order to place the processing and storage of data declared as anonymous outside the scope of GDPR.

However, there is a recent relevant regulatory development concerning the April 2023 ruling of the **European General Court on the case Single Resolution Board vs the European Data Protection Supervisor.** With respect to this development, **potential implications for EUCAIM could be significant, depending on the outcomes.** This is because the case concerns the fundamental question of how the key concepts "personal data" and "anonymous data" should be interpreted under the GDPR. The ruling by the General Court favors the so-called "**contextual**" interpretation of the concepts, which is concerned with whether data can be related to an identifiable natural person by a particular party; if the party cannot link the data to an identifiable individual by employing means reasonably likely to be used in the context of re-identifiability, the data should be deemed anonymous from this party's perspective. This, however, differs from the **"absolute" interpretation** of the concepts, which holds that a piece of information must be considered personal data if it is possible (through means reasonably likely to be used) to attribute it to a natural person by any party. The absolute view on the concepts of personal data and anonymity is based on the literal interpretation of Recital 26 GDPR and has been embraced by most EU data protection authorities. Consequently, the ruling of the General Court has been challenged in the Court of Justice of the European Union with outcome pending.

The Proposal for a Regulation of the European Health Data Space (EHDS) [11] defines anonymisation as the preferential rule for data processing, admitting pseudonymisation as a secondary rule (when the goal of the processing cannot be reached with the use of anonymised data). This issue is clearly demonstrated in the additional regulation of the principles of data minimisation and purpose specification by this Regulation:

Article 44
Data minimisation and purpose limitation
(…)
3. Where the health data user has sufficiently demonstrated that the purpose of processing cannot be achieved with anonymised data in line with Article 46(1), point (c), the health data access bodies shall provide access to electronic health data in pseudonymised format. The information necessary to reverse the pseudonymisation shall be available only to the health data access body or a body that acts as trusted third party in accordance with national law.

The primary obligation and strategic decision for EUCAIM must therefore be to ensure the anonymisation of information. However, the question on this issue is controversial. To the extent that there is a general consensus that anonymisation is challenging in the field of health data.

The European Health Data Space Regulation aims to support the secondary use of pseudonymized electronic health data. Bellow, the summary of the paragraphs in the EHDS, describing the use of pseudonymized (or anonymized) data.
- Recital (37b): The **secondary use** of electronic health data must be based on **pseudonymised or anonymised** data to prevent identification of data subjects.
- Recital (43): **Health data access bodies** are encouraged to cooperate internationally to develop best practices and techniques for anonymisation. The **Commission** may provide procedures, requirements, and technical tools for **uniform anonymisation and pseudonymisation.**
- Recital (49): To protect privacy, non-personal electronic health data should be used when sufficient. If personal data is necessary, it **should be pseudonymised, with justification** required for its use. Anonymisation or pseudonymisation should occur early in the data processing chain, using advanced technologies and standards to prevent re-identification.
- Article 44: If the processing **purpose cannot be achieved with anonymised** data, **pseudonymised data can be provided**. Only the health data access body or a trusted third party should be able to reverse pseudonymisation.

[11] PROVISIONAL AGREEMENT RESULTING FROM INTERINSTITUTIONAL NEGOTIATIONS. Proposal for a regulation the European Parliament and the Council of the European Union on European Health Data Space (COM(2022)0197 – C9-0167/2022 – 2022/0140(COD)(COD))

- Article 45: **Data access applications must describe** whether data needs to be **pseudonymised or anonymised and justify** the need for pseudonymisation if applicable. They must also describe compliance with data protection laws.
- Article 46: Access to electronic health data is granted only **if pseudonymised data is necessary and justified.** The data permit will detail the conditions for data use, including the format and source of pseudonymised data.
- Article 60a: Personal electronic health data must be stored and processed within the EU **in secure environments** when performing pseudonymisation, anonymisation, or other personal data processing operations.

The Articles:
Recital (37b) …. The ***secondary use of electronic health data is based on pseudonymised or anonymised*** *data, in order to preclude the identification of the data subjects.*

Recital (43) In that regard, ***health data access bodies should*** *cooperate across borders to* ***develop and exchange best practices and techniques****. This includes rules for* ***anonymisation*** *of microdata sets. When relevant, the Commission should set out the procedures and requirements, and provide technical tools, for a unified procedure for anonymising and* ***pseudonymising*** *the electronic health data.*

Recital (49) Given the sensitivity of electronic health data, it is necessary to reduce risks on the privacy of natural persons by applying the data minimisation principle as set out in Article 5(1), point (c), of Regulation (EU) 2016/679. Therefore, non-personal electronic health data should be made available in all cases where this is sufficient. If the data user needs to use personal electronic health data, it should clearly indicate in its request the justification for the use of this type of data and the health data access body should assess the validity of that justification. The personal electronic health data should only be made available in pseudonymised format. Taking into account the specific purposes of the processing, data should be anonymised or pseudonymised as early as possible in the chain of making data available for secondary use. ***Pseudonymisation and anonymisation can be carried out by the health data access bodies or by the health data holders.*** *As data controllers,* ***health data access bodies and health data holders may delegate these tasks to data processors****. When providing access to an anonymised or pseudonymised dataset, a health data access body should use state-of-the-art anonymisation or pseudonymisation technology and standards, ensuring to the maximum extent possible that natural persons cannot be re-identified by health data users. Such technologies and standards for data anonymisation should be further developed.*

Section 3
Access to electronic health data for secondary use
Article 44
Data minimisation and purpose limitation
Where the health data user has sufficiently demonstrated that the purpose of ***processing cannot be achieved with anonymised data*** *in line with Article 46(1), point (c), the health data access bodies shall provide* ***access to electronic health data in pseudonymised format.*** *The information necessary to reverse the pseudonymisation shall be available only to the health data access body or a body that acts as trusted third party in accordance with national law.*

Article 45
Data access applications

*2 (c) a description whether electronic health data need to be made available in a **pseudonymised or anonymised format**, in case of pseudonymised format, a **justification** why the processing cannot be pursued using anonymised data;*
*4. Where the applicant seeks to **access the personal electronic health data in a pseudonymised format**, the following additional information shall be provided together with the data access application:*
(a) a description of how the processing would comply with applicable Union and national law on data protection and privacy, notably Regulation (EU) 2016/679 and, notably, Article 6(1) of Regulation (EU) 2016/679;

Article 46
Data permit
*1. The health data access bodies shall decide to **grant access** to electronic health data only when the following cumulative criteria are fulfilled:*
*(c) the processing complies with Article 6(1) Regulation (EU) 2016/679, in particular that in the case of **pseudonymized** data, there is sufficient justification that the purpose **cannot be achieved with anonymized data**;*
*6. When the health data access body issues a **data permit**, it shall set out the general conditions applicable to the health data user in the data permit. The data permit shall contain the following:*
(a) categories, specification and format of electronic health data accessed, covered by the data permit, including their sources and if the electronic health data will be accessed in a pseudonymised format in the secure processing environment;

Article 60a
Storage of personal electronic health data by health data access bodies and secure processing environments
*1. Health data access bodies, single data holders and the Union data access service shall store and process personal health electronic data in the European Union when performing **pseudonymisation**, anonymisation and any other personal data processing operations referred to in Articles 45 to 49, through secure processing environments within the meaning of article 50 and article 52(8) or through HealthData@EU. This requirement shall apply to any entity performing these tasks on their behalf.*

II. EUCAIM's approach to anonymization and pseudonymization aligned with the European Health Data Space

EUCAIM **will provide tools to perform the anonymization or pseudonymization** of the data for those providers which do not have their own method to do so. Regarding metadata in imaging studies, based on the strategies followed in the AI4HI projects, the EUCAIM tools will de-identify the data based on a de-identification profile for each cancer type. This profile specifies how to process the DICOM tags, removing those that contain Personal Identifiable Information such as the Institution Name and Address, and the Patient's Name. In addition, it will replace other tags with values of similar meaning that can still be of interest to the final user without compromising the identification of the patient.

Concerning clinical data, EUCAIM will define a set of variables to be provided by cancer type. However, the information provided will be analyzed and the variables with **personal information will be removed or replaced.**

Moreover, a **wizard tool will be developed under the scope of EUCAIM**. This tool will support the **identification of risks and propose ways to mitigate them**. In addition, it will raise awareness on weak points of each process, foster a secure-by-design anonymization planning and facilitate compliance with EUCAIM requirements and accountability obligations**.**

I. Anonymisation is a technical procedure subject to risk analysis methodologies.

As can be seen in D3.6 DMP, in the Working Package 5 guidelines, and in the preliminary report issued by Dr Martinez (WP3, Annex 1) EUCAIM deploys appropriate methodologies to achieve these tasks. These are consistent with the EHDSR objectives.

(37b) (….) This Regulation provides the necessary safeguards to mitigate certain risks involved in the realisation of those benefits. The secondary use of electronic health data is based on pseudonymised or anonymised data, in order to preclude the identification of the data subjects.

II. There is no legal certainty as to the conditions of anonymisation.

There are different approaches that confront the vision of the data protection authorities with that of the General Court of the European Union, which have been incorporated into the Project's risk management plans since Dr. Martínez's technical report.

In principle, the Court's jurisprudence accepts the EUCAIM solution as valid since:

a) The anonymisation is robust.

b) There are technical means that prevent the user from accessing the information and legal means that oblige the user to affirm a commitment not to re-identify.

The risk management objectives proposed by EHDS are thus achieved:

EHDSR (64) Certain categories of electronic health data can remain particularly sensitive even when they are in anonymised format and thus non-personal, as already specifically foreseen in the Data Governance Act. Even in situations of the use of state of the art anonymization techniques, there remains a residual risk that the capacity to re-identify could be or become available, beyond the means reasonably likely to be used. Such residual risk is present in relation to rare diseases (a life-threatening or chronically debilitating condition affecting not more than five in 10 thousand persons in the Union), where the limited numbers of cases reduce the possibility to fully aggregate the published data in order to preserve the privacy of natural persons while also maintaining an appropriate level of granularity in order to remain meaningful. It can affect different types of health data depending on the level of granularity and description of the characteristics of data subjects, the number of people affected or and for instance in cases of data included in electronic health records, disease registries, biobanks, person generated data etc. where the identification characteristics are broader and where, in combination with other information (e.g. in very small geographical areas) or through the technological evolution of methods which had not been available at the moment of anonymisation, can lead to the re-identification of the data subjects using means that are beyond those reasonably likely to be used. **The realisation of such risk of re-identification of natural persons would present a major concern and is likely to put the acceptance of the policy and rules on secondary use provided for in this Regulation at risk.** Furthermore, aggregation techniques are less tested for non-personal data containing for example trade secrets, as in the reporting on clinical trials and clinical investigations, and enforcement of breaches of trade secrets outside the Union is more difficult in the absence of a sufficient international protection standard. Therefore, for these types of health data, there remains a risk for re-identification after the anonymisation or aggregation, which could not be reasonably mitigated initially. This falls within the criteria indicated in Article 5(13) of Regulation (EU) 2022/868. These types of health data would thus fall within the empowerment set out in Article 5(13) of Regulation (EU) 2022/868 for transfer to third countries. The protective measures, proportional to the risk of re- identification, would need to take into account the specificities of different data categories or of different anonymization or aggregation techniques and will be detailed in the context of the Delegated Act under the empowerment set out in Article 5(13) of Regulation (EU) 2022/868.

III. The technological design of the platform supports the use of pseudonymised data.
In other words, as stated in Recital (37c) "the safeguards under Chapter IV will continue to apply, notably the ban against re- identification, including attempts, by data users". The approach is based on a model derived from Spanish legislation and tested in previous European projects such as BigMedylitics[12].This is based on:
(a) A technical and functional separation between the research team and those who carry out the pseudonymisation and keep the information that enables re-identification.
(b) Pseudonymised data shall only be accessible to data users when
- There is an express commitment to confidentiality and not to carry out any re-identification activity, and
- Specific security measures are adopted to prevent re-identification and access by unauthorised third parties.
(c) Re-identification **of data at source** may take place when, in the course of research using pseudonymised data, there is a real and concrete danger to the safety or health of a person or group of persons, or a serious threat to their rights, or it is necessary to ensure adequate health care[13].

IV. Rigorous legal support is provided.
It is based on the following measures:
a) A governance framework supervised by an Access Committee.
b) Legal guarantees are based on multiple levels. Definition of admissible and secure conditions of use in accordance with EU provisions and national law (data holders side) based on:
1. Data Sharing Agreement (Federated processing)
2. Data Transfer Agreement (Central Node Processing)
- Clear definition of the obligations of the entities with access to data and their individual users: Application framework and ethical and legal requirements in calls for proposals
3. Terms and conditions to be signed
4. Security obligations and commitment not to re-identify each user's first access to the platform.

V. EUCAIM is designed as a Secure Processing Environment.
This design allows secure processing of data with any legal basis, whether anonymised, pseudonymised or purely personal data. Data will not be transferred to data users; they can only process it in EUCAIM's secure processing environment.

This highly secure design will meet all the requirements of Article 50 of the EHDSR:
(a) restrict access to the secure processing environment to authorised natural persons listed in the respective data permit;
(b) minimise the risk of the unauthorised reading, copying, modification or removal of electronic health data hosted in the secure processing environment through state-of- the-art technical and organisational measures;
(c) limit the input of electronic health data and the inspection, modification or deletion of electronic health data hosted in the secure processing environment to a limited number of authorised identifiable individuals;

[12] See https://www.bigmedilytics.eu/blueprint/

[13] This type of soultion was awarded the Proactivity and Good Practices Award granted by the Spanish Data Protection Agency in 2021 to Fundación 29 de Febrero for the Project 'Healthdata29: Legal guide and repository to encourage the sharing of health data' based on the research of Dr. Martinez. https://www.uv.es/uvweb/uv-noticias/es/noticias/proteccion-datos-premia-proyecto-basado-investigacion-catedra-privacidad-transformacion-digital-microsoft-universitat-valencia-1285973304159/Novetat.html?id=1286172429159&plantilla=UV_Noticies/Page/TPGDetaillNews

(d) ensure that health data users have access only to the electronic health data covered by their data permit, by means of individual and unique user identities and confidential access modes only;
(e) keep identifiable logs of access to and activities in the secure processing environment for the period of time necessary to verify and audit all processing operations in that environment. Logs of access should be kept for not shorter than one year;
(f) ensure compliance and monitor the security measures referred to in this Article to mitigate potential security threats.

III. Technical background and description of the EUCAIM anonymization and pseudonymization strategy

1) Technical aspects of anonymization and pseudonimization

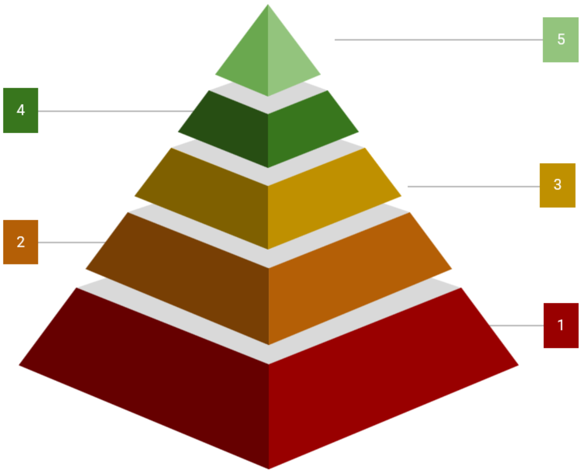
Anonymization and pseudonymization are essential practices in the protection of sensitive data, especially when processing healthcare data which is the case of EUCAIM. Both methods are used to safeguard patient privacy while enabling data to be used for research, analysis, and sharing within regulatory guidelines.

Anonymization refers to the irreversible process of removing or altering data elements that make it possible to identify an individual, making re-identification highly unlikely or practically impossible. This means stripping all direct and indirect identifiers, such as names, or any unique codes linked to a patient. In medical imaging this may include metadata embedded in the files, burnt-in text in images or image uniqueness aspects that may lead to identify an individual patient. Once anonymized, data is no longer considered personally identifiable, making it largely exempt from stringent data protection regulations.

On the other hand, pseudonymization involves replacing private identifiers with a pseudonym or key, which allows data to be re-identified with the proper key or additional information. This approach is reversible, unlike anonymization, and in EUCAIM it could be used in scenarios where some degree of linkage is necessary, such as longitudinal studies. Pseudonymization strikes a balance between data utility and patient privacy, allowing controlled data re-identification when necessary. However, it is important to highlight that in EUCAIM, the correspondence in the references pre and post pseudonymization will bu out of EUCAIM access in order to guarantee the privacy of the patient information.

While anonymization and pseudonymization serve similar purposes, the choice between the two often depends on the regulatory requirements, data utility needs, and sensitivity of the data in question. It is also important to mention that in the process of anonymization, as it is irreversible, clinical and imaging data should be anonymize simultaneously to be able to associate correctly both types of data to their correspondent patients and studies. On the other hand, using pseudonymization may reduce this risk, as the data holders will be able to track the data back to the original patient.

Figure 1: Anonymization of Medical Data



Properly Anonymized Data
Direct identifiers are removed.
Indirect identifiers are further processed.
Users can process data only in a Secure Processing Environment, making the combination with additional data impossible.
Therefore, re-identification is practically impossible.

Weakly anonymized
Direct identifiers are removed or altered.
Some indirect identifiers that remain, are further processed (e.g. suppression, generalisation) to minimize risks of re-identification even further in the context.

De-identified Data
Direct identifiers are removed or altered,
Some indirect identifiers may remain.
Potentially reversible if combined with additional data.

Pseudonymized Data
Identifiers are replaced with pseudonyms, enabling re-linking.
Reversible with a secure re-linking key.

Personal Identifiable Information (PII) - Raw Data
Fully identifiable, no anonymization.

I. Anonymization and pseudonymization in Imaging data

Medical imaging modalities that will be part of EUCAIM data like MRI (Magnetic Resonance Imaging), CT (Computed Tomography), and ultrasound generate vast amounts of data, primarily stored and shared using the DICOM (Digital Imaging and Communications in Medicine) standard. This data is essential for diagnostics, research, and training in medicine, making anonymization and pseudonymization practices critical for protecting patient privacy.

The DICOM standard comprises images, metadata, and headers that often contain identifiable information, such as patient names, dates of birth, and hospital details. Therefore, the anonymization or pseudonymization of DICOM files involves modifying both the image data (if applicable) and the metadata to ensure patient privacy is maintained.

Regarding the metadata present in the imaging files, each of the DICOM tag contains information that should be processed differently (https://www.dicomlibrary.com/dicom/dicom-tags/). For instance, some tags may contain information about the manufacturer of the machine used for the data acquisition which has very low risk to be used to re-identify patients and it could be very important information for research purposes, thus, this tag may be kept as it was in the original DICOM. On the other hand, other tags may contain much sensible information such as unique ids that should be removed or substituted. The definition of what process should be applied to each individual tag is what is called the anonymization or pseudonymization profile.

- Anonymization in DICOM: For reaching the proper anonymization of DICOM files, they must undergo a process of de-identification of metadata in the DICOM tags without keeping a link to the original information. For example, a patient's id may be substituted by a hashed code, dates could be shifted to a relative time frame, and institution names might be stripped out. DICOM headers typically contain structured tags, such as Patient Name (0010,0010) and Patient ID (0010,0020), that must be carefully redacted or modified. The data de-identification, together with the Secure Process Environment ensures the proper anonymization of the imaging metadata avoiding data to be traced back to individuals using the information they contain. This approach, while offering the highest level of privacy protection, may reduce the dataset's utility if de-identified fields are needed for analysis or research.
- Pseudonymization in DICOM: Unlike proper anonymization, pseudonymization of DICOM files preserves some linkage potential through the use of pseudonyms or keys in place of directly identifiable fields. This approach allows researchers to correlate data across studies while protecting sensitive patient information. Pseudonymized DICOM files often use unique

identifiers that map to patient records, which are securely stored in a separate database or system. Access to the mapping key is restricted to authorized personnel, providing a controlled mechanism for re-identification when necessary for legitimate purposes.

- Challenges Specific to Medical Imaging: Anonymizing or pseudonymizing DICOM data presents unique challenges, including the risk of inadvertently revealing patient identity through image features, such as unique anatomical structures, or through improperly redacted metadata. EUCAIM will provide the wizard tool which aims to help on detecting some of these potential risks, but it should be used with combination of manual review for effective anonymization or pseudonymization.

II. Anonymization and pseudonymization in clinical data

Clinical data includes electronic health records (EHRs), lab results, patient-reported outcomes, and more. Compared to medical imaging, this data often contains structured, text-based information that must be protected through robust anonymization or pseudonymization practices.

- Anonymization: In the case of proper anonymization, all direct identifiers such as names, sensitive dates or unique patient codes are stripped or masked entirely. Care must be taken to also anonymize indirect identifiers that could allow re-identification when combined with other data, such as rare disease conditions, dates of service, or demographic details. Effective anonymization processes often employ algorithms to generalize or perturb data, ensuring it cannot be linked back to individuals while retaining statistical utility.
- Pseudonymization: For clinical data, pseudonymization offers a middle ground by replacing identifiable data fields with codes or pseudonyms. This allows re-identification if needed, using a secure, controlled mapping mechanism. For example, a patient identifier might be replaced by a randomly generated number, while a secure key-management system ensures access to the original data is restricted to authorized personnel. This approach is particularly valuable in ongoing patient care, clinical trials, or cohort studies where longitudinal data linkage is essential.

In addition, in the context of healthcare data, a critical risk associated with anonymization and pseudonymization is re-identification at the dataset level. Even after direct identifiers are removed, unique combinations of quasi-identifiers—attributes like age, gender, diagnosis date, and dates—can lead to the identification of individual patients. This risk is heightened in datasets with rare conditions, unusual treatment paths, or small subpopulations, where unique data combinations may inadvertently single out individuals. Effective mitigation strategies involve techniques such as data generalization, suppression, and k-anonymity, which ensure that individual records are indistinguishable from a group of similar records within the dataset. This risk identification and minimization is also included as part of the wizard tool data processing pipeline.

2) Anonymization strategy in EUCAIM

I. De-identification profiles

The development of the EUCAIM de-identification profile followed a multi-phase approach. In the initial phase, a comprehensive comparison at the DICOM tag level was conducted across de-identification profiles from various AI4HI projects, including ProCAncer-I, CHAIMELEON, PRIMAGE, INCISIVE and EUCANIMAGE. This analysis also included default profiles from the widely used Clinical Trial Processor (CTP) application, alongside the de-identification guidelines recommended by NEMA, association that maintains the DICOM standard. The results of this extensive analysis, which were published in *Documenting the de-identification process of clinical and imaging data for AI for health imaging projects*[1], enabled to determine the kept and modified DICOM tags by all profiles, as well as the differences between the projects. As a result of this phase, a first draft of the EUCAIM profile was proposed, based on the least restrictive de-identification profile in order not to lose the usefulness of some metadata for image processing and analysis and to be able to make a more robust decision after further analysis.

However, this initial approach did not address the distinct variations in DICOM tags required by specific imaging modalities. To incorporate these modality-specific distinctions, a second analysis phase was conducted. During this phase, we compiled a comprehensive inventory of DICOM tags specific to each imaging modality, including magnetic resonance (MR), computed tomography (CT), computed radiography, nuclear medicine, ultrasound, positron emission tomography (PET), digital X-ray, and digital mammography. Tools such as the Innolitics DICOM Tag Browser and the NEMA DICOM browser were instrumental in obtaining exhaustive tag information for each modality. This focused analysis allowed for a more precise refinement of the EUCAIM de-identification profile, ensuring that it adequately accommodates the unique requirements and data sensitivities of each imaging type.

During this phase, four partner teams participated, each independently assigned to propose de-identification actions for a particular imaging modality. The actions proposed were based on standard de-identification procedures outlined in the NEMA DICOM guidelines, which include actions such as "Keep," "Clean," "UID," and multiple "Delete" actions, tailored by attribute type. Additionally, a new action labelled "Level" was introduced to allow for finer control over anonymization, offering variable restriction levels depending on data use requirements.

To ensure objective evaluations, each team proposed a de-identification action for each tag within its assigned modality by comparing it against the suggested NEMA basic profile and the preliminary EUCAIM draft, while avoiding any bias from the actions suggested for the same tags in other modalities. This approach followed a blacklisting methodology, meaning that each team identified tags to be removed if they contained sensitive information or posed a re-identification risk for patients. This independent assessment by modality allowed for a focused and secure approach to tagging, while avoiding cross-modality influence in determining sensitive data identifiers.

Once these proposals were compiled across modalities, it was initiated a review of tags where conflicting actions were proposed, aiming to reach a consensus on a final de-identification profile. For tags marked with the new "Level" action, multiple options are being considered to enhance granularity, with a default level recommendation to be included in the final profile.

Within the CHAIMELEON project, a re-identification challenge was conducted, led by HULAFE. The challenge involved researchers from AI4HI and EUCAIM attempting to re-identify DICOM studies that had been de-identified using the profiles developed by the CHAIMELEON and ProCancer-I projects. The results demonstrated that these profiles are robust, as no successful re-identification was achieved. Nonetheless, the challenge highlighted both the critical importance of thoroughly reviewing free-text fields, which could inadvertently contain identifiable information, and the necessity of maintaining compliance with the DICOM standard when modifying tags.

The insights gained from this challenge will contribute to enhancing the robustness of the EUCAIM de-identification profiles and ensuring compliance with the DICOM standard.

II. EUCAIM Anonymizer tool

EUCAIM offers a de-identification tool to support the secure integration of data, even the ones belonging to Tier 1 exhibiting the lowest level of compliance within EUCAIM and can be accepted with no additional requirements with the common data formats (EUCAIM’s hyper-ontology). Datasets belonging to Tier 2, Tier 3 and Tier 3* also adhere to the minimum requirements of data in Tier 1 and also fulfil additional criteria for homogeneity and security and are thus less susceptible to any possibility of personal information exposure.

The mandatory requirements for Tier 1 data focus on data de-identification to ensure that EUCAIM datasets do not contain personal information about patients and minimize the risks of patient re-identification. Imaging data must be in DICOM format. The Anonymizer tool, developed by FORTH, performs a de-identification of imaging data DICOM tags. It takes as input a folder with one or multiple patients/cases, and de-identifies the DICOM tags according to a predefined de-identification profile which specifies how to process each of them by taking into account both data usability and data security, as well as to preservation of the DICOM structure integrity. The Anonymizer tool imposes specific actions to each DICOM tag possibly containing protected information, such as masking, hashing, deleting or updating information depending on the specific DICOM tag type.

III. Wizard tool

When clinical data is compliant with the CDM, EUCAIM can offer tools such as the Wizard tool which is a new tool developed during the EUCAIM project based on an open source solution. This tool is designed to validate the de-identification applied to datasets, identify potential risks, and propose mitigation strategies. By providing these mechanisms and tools, EUCAIM aims to streamline the participation of Data Holders and ensure robust data anonymization practices across varying levels of compliance, enhancing the overall security and privacy of patient information within the ecosystem. Its goals are to support the identification of risks and propose ways to mitigate them, to raise awareness of the weak points of each process, to foster a secure-by-design anonymization planning, and to facilitate compliance with EUCAIM requirements and accountability obligations. Clinical data in Tier 1 is not compliant with the Common Data Model (CDM) thus the functionalities for the clinical data of the tool are deactivated.

The overall effort is based on maintaining data clinical value and data security which often pose contradicting requirements. The whole process needs to be adaptive to different use cases since the endpoints of future data contributions are not yet defined. Moreover, the workflow needs to consider the degree of homogeneity which is an important factor for data usability.

The aim of the Wizard, as discussed in the initial meetings among the EUCAIM participants, can be summarised in:

- Support the identification of risks - propose ways to mitigate them
- Raise awareness of the weak points of each process
- Foster a secure-by-design anonymization planning
- Facilitate compliance with EUCAIM requirements and accountability obligations

The Wizard functionalities are listed herein:

W1 Blacklisting per modality (Imaging data, cohort/cancer type agnostic) Starting from the different de-identification profiles of the five AI4HI projects, a working sheet has been built with the different views of each project regarding the handling of imaging metadata, i.e. DICOM header tags.

W2 Whitelisting per modality (Imaging data, cohort/cancer type agnostic) Whitelisting may precede blacklisting or may be performed as a separate task. The rationale for whitelisting-as opposed to blacklisting- is to keep only the necessary tags for further processing rather than removing the dangerous tags. The whitelisting was based on the defined tags per modality and levels of security that resulted from the multidisciplinary work that was described in the de-identification section among many EUCAIM partners.

W3 Risk assessment (clinical and imaging dataset)

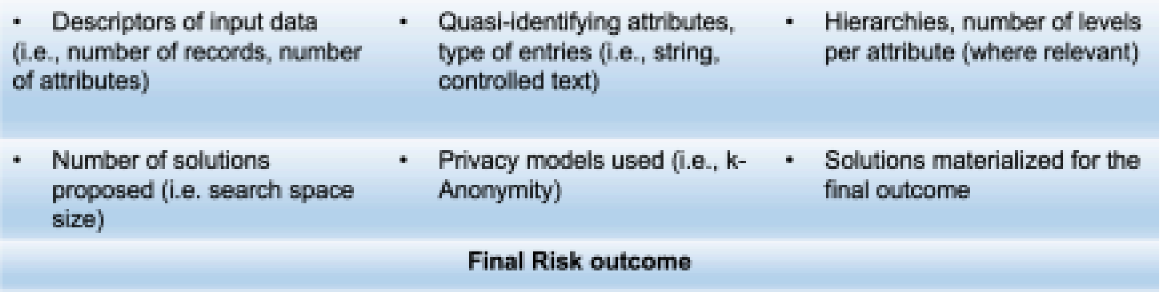
Starting from the fact that anonymization is not a binary concept, the Wizard aims to analyze, measure, and optimize the degree of anonymization. Since the question about when data is anonymous can hardly be answered, a feasible goal is to take measures against re-identification in order to minimize the risk of data being re-identified under means reasonably likely to be used in terms of time, cost, and available technology. The factors that have to be taken into account for increasing the security regarding a specific cohort are –but are not limited to– the size and variable distribution. However, the distributions may be subjected to changes in order to optimize the distributions, especially for underpopulated categories. That has to be performed with caution so as not to obscure the specific identity of each participant and thus compromise data value. This entails a double-faced attention to both data security and usability.

To calculate the total risk of a specific dataset, it is important to integrate the total amount of information that is provided for any individual. Thus, it is necessary to aggregate all available information in an appropriate format that can be effectively inspected. Inspection aims to detect missing information, overlapping or duplicated entries, as well as problematic entries or entry types.

W4 Risk Optimization - Iterative process

Re-identification risk, as risks derived from population uniqueness, can be estimated with different statistical methods. The Wizard tool aims to initiate an iterative process where possible re-identification scenarios are considered by the user and will be tailored according to both needs of security and data integrity. The solution that best fits, given the individual characteristics of each cohort, will be the final output of the iterative risk estimation process, requiring the feedback of the data provider to conclude the final plan. Firstly, for each quasi-identifier, a graph showing the initial distribution will be presented to identify the underpopulated categories. Moreover, combinations of attributes will be formed to analyze the degree to which different attribute combinations of variables separate the records from each other, which is an indication of the number of individuals -and attributes- at higher risk for re-identification. Population uniqueness in a sample are the individuals with a unique combination of attributes. It is considered that cohort uniques are also unique within the underlying population from which the data has been sampled. Different models are considered to return the estimates of population unique. Their results will be compared, as each model conveys different assumptions. During this process, some user-defined inputs will be required regarding the sample size, the prevalence of the specific pathology in the general population, etc. Each specific anonymization schema will be graded differently according to its security and usability under different sample sizes or representations in the cohort.

Many possible solutions will be the outcome of the user's interaction with the Wizard tool, where the user will inspect different ratings of data security and usability under proposed de-identification schemas. The user will be invited to decide on the best-tailored list of tags with specific accuracy and specific care taken for individuals at risk. The whole process aims to make the user aware of the danger of escalation, to exhaust the possibility of protecting data while ensuring usability and maintaining a reasonable balance among these two contradicting needs. Specific algorithms will provide qualitative or quantitative metrics for the data security and usability before and after the iterative Wizard-guided process. This information will be part of the dataset identity.



• Descriptors of input data (i.e., number of records, number of attributes)	• Quasi-identifying attributes, type of entries (i.e., string, controlled text)	• Hierarchies, number of levels per attribute (where relevant)
• Number of solutions proposed (i.e. search space size)	• Privacy models used (i.e., k-Anonymity)	• Solutions materialized for the final outcome
Final Risk outcome		

Figure X. Indicative information produced during the risk minimization process and presented in the final report (source: EUCAIM deliverable 5.4)

W5 Report

The final report will contain a descriptive list of proposed and completed actions regarding the optimization process of data security and data usability that was guided within the EUCAIM Wizard tool. The estimated level of data security as well as the level of data usability will be stated in metrics and units that will be defined by the Wizard team. The report will be issued after the whole cohort set is known as it will describe the level of data security for the specific population with the given distribution of each quasi-identifier. Indicatively, the number of individuals at relatively higher danger for re-identification will be stated, at different stages of the Wizard workflow, most importantly at the beginning and the end of the process. The individuals at risk for re-identification will be identified based on unique combinations of quasi-identifiers within the specific cohort, given the levels of tags and corresponding attribute intervals of the final de-identification plan.

Annex I: Legal requirements for anonymization

This report from Dr. Ricard Martinez reflects the “legal state of art” regarding anonymization.

Title:	**Document Version:**
Legal requirements for anonymization	v.1.1 updated

Project Number:	**Project Acronym:**	**Project Title:**
GA 101100633 DIGITAL-2022-CLOUD-AI-02	EUCAIM	European Federation for Cancer Images

Authors:	**Organisation:**	**WP**	**Contributing with WP:**
Dr. Ricard Martínez	UV	WP3	WP 5

Abstract:

This report reflects the “legal state of art” regarding anonymization. This not precludes technical decisions on the use of DICOM images anonymization tools or any other technique applied to personal data (or any other linking identifier) to be removed at the data sets.

The version 1.1 has been updated to with information coming from the Final Compromise Text of the Proposal for a Regulation on the European Health Data Space, published by the General Secretariat of the Council No. Cion doc.: 8571/22 ADD1-8.

Modified sections: Sections 1.2, 1.3,3.2,3.3

Keywords:

GDPR, Anonymisation WP5, T5.3.3 Data De-identification

REVISION HISTORY

REVISION	DATE	DESCRIPTION	AUTHOR (ORGANISATION)
V1.1	22/09/2024	LEGAL REQUIREMENTS FOR ANONYMIZATION	DR. RICARD MARTÍNEZ

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under Grant Agreement № 101100633.

Legal requirements for anonymization.

Summary

The anonymization objectives imposed on the platform can be achieved through so-called ‘de facto anonymization’. De facto anonymization consists of a technical deployment that includes a robust anonymization process complemented by a secure processing environment. The latter prevents both the improper downloading or manipulation of data and any access by third parties. It thus excludes the risk of re-identification by third parties and the risk associated with the use of malicious software since no processing tools other than those previously authorised by the platform are supported.
This technical method must be complemented by legal governance based on:

▪ A registration process on the platform, which implies an awareness of the rules of the platform.

▪The acceptance of a clear set of terms and conditions. This legal act does not correspond, strictu sensu, to the applicant for access to the data or to the user of the data, if they are different subjects, but to the legal representative of the entity in which they carry out their activities.

▪ The explicit acceptance by the user(s) of the information system of the system-specific security obligations and of the commitments not to re-identify. This should be done by taking a positive action such as ticking a box and accepting by receiving a confirmed message. The process must be digitally evidenced.

This method should be acceptable, adequate and legally admissible. Otherwise, not only EUCAIM, no data holder, data access agency or HealthData@EU will be able to offer properly anonymised datasets.

1. Legal framework.

1.1 General Data Protection Regulation Scenario

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (GDPR) does not define the concept of anonymization. Only Article 4 defines the concept of pseudonymisation in the following terms:

> (5) ‘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;

Pseudonymisation is conceived in the GDPR as a security measure. In any case, and as a starting point, data protection authorities consider pseudonymised data to be personal data. In the area of the use of personal data for research purposes, the GDPR states anonymization as a priority option. If we apply the GDPR's risk-based approach, everything leads to this solution. Clearly, from the point of view of the analysis of risks to rights arising from the processing (Articles 24 and 25 GDPR), it presents the lowest possible risk. It is also in line with the minimisation principle of Article 5 of the GDPR. Moreover, in the field of health research, anonymization is one of the additional safeguards required by Article 89 of the GDPR.

Moreover, the Proposal for a Regulation on the European Health Data Space (EHDS) expressly states the preference of anonymization as a strategy for the processing of data for research purposes. Pseudonymisation will be reserved for those cases, such as rare diseases (whatever the case may be) in which it is more complex to anonymise and always upon justification of the need to process pseudonymised data.

> (43) (...) In addition to the tasks necessary to ensure effective secondary use of health data, the health data access body should strive to expand the availability of additional health datasets, support the development of AI in health and promote the development of common standards. They should apply tested techniques that ensure electronic health data is processed in a manner that preserves the privacy of the information contained in the data for which secondary use is allowed, including techniques for pseudonymisation, anonymization, generalisation, suppression and randomisation of personal data. Health data access bodies can prepare datasets to the data user requirement linked to the issued data permit. This includes rules for anonymization of microdata sets.
>
> (...)
>
> (64) Certain categories of electronic health data can remain particularly sensitive even when they are in anonymised format and thus non-personal, as already specifically foreseen in the Data Governance Act. Even in situations of the use of state of the art anonymization techniques, there remains a residual risk that the capacity to re-identify could be or become available, beyond the means reasonably likely to be used. Such residual risk is present in relation to rare diseases (…)

However, it is essential to take into account the way in which the GDPR conceives anonymization. In this regard, our reference framework is defined by recital 26 of the GDPR:

> (26) (...) To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments. 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.

It should be noted that the requirements of the regulation have been interpreted differently by the Court of Justice of the European Union and by data protection authorities. The reference document for data protection authorities is the Opinion 05/2014 on anonymization techniques issued by the Article 29 Working Party. According to this document, the minimum requirements that an anonymization procedure should meet have two dimensions, the legal one and the dimension relating to the techniques applied by the organisation.

From a legal point of view, anonymization is linked with the original legal basis for processing at source and it is considered as a natural process for data storage and reuse. This entails the need to:

▪ Justify the existence of a basis that legitimises the anonymised reuse of data:

> - consent given unambiguously,
>
> - performance of a contract to which the data subject is party or for the implementation of pre-contractual measures taken at the request of the data subject, or
>
> - compliance with a legal obligation to which the controller is subject,
>
> - necessary to protect the vital interest of the data subject,

- necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller or in a third party to whom the data are disclosed, or

- satisfaction of the legitimate interest provided that the interest or fundamental rights and freedoms of the data subject are not overriding.

▪ Ensure conditions of coherence with the purpose:

- relationship between the purposes for which the personal data were collected and the purposes of further processing (purpose limitation);

- ensure that in the context in which personal data were collected, the reasonable expectations of data subjects as to their further use were met;

- be appropriate to the nature of the personal data and the impact of further processing on the data subjects;

▪ The controller must implement and demonstrate the safeguards adopted to ensure proper processing and to prevent any undue negative impact on data subjects.

Ensure the storage of data in an identifiable format during the legal time limits as a guarantee of access (CJEU case C-553/07):

- information on the recipients or categories of recipients to whom the data are disclosed,

- to the content of the information communicated.

It is therefore advisable to bear in mind that:

- The possibility of reversibility should be assessed in advance

- The technological means "are not only those of the controller" but rather "those available".

- Linking must be impossible both for the controller and for any third party.

- It is not just a matter of retrieving first name, surname, address, if there is the slightest possibility of potential identifiability by singling out, linkability or inference: **data protection law applies**.

- If the context (technology, relationship of data with other subsets of data...) involves risks: **data protection law applies**.

From the point of view of the anonymization process, the Article 29 Working Party, or the European Data Protection Board, establishes a procedure based on a risk analysis approach. The first step is to establish whether we face any of the three key risks:

▪ Singularisation: the possibility of extracting from a data set some records (or all records) that identify a person.

▪ Linkability: the ability to link at least two records of a single data subject or a group of data subjects, either in the same database or in two different databases.

- If the attacker can determine (e.g. by correlation analysis) that two records are assigned to the same group of persons but cannot single out the persons in this group, then the technique is resistant to singling out, but not to linkability.

▪ Inference: the possibility of deducing with significant probability the value of an attribute from the values of a set of other attributes. The second step would be to apply the set of anonymization techniques that eliminate this risk or reduce it as much as possible (see Annex).

Three key ideas should be retained from this document:

▪ Anonymisation is a processing operation in itself. Therefore, legality requirements must apply, including the conclusion of a processor contract when it is carried out by a third party.

▪ Pseudonymisation is not the same as anonymization.

▪ Anonymisation is conceived as irreversible or equivalent to erasure.

> It is therefore recommended that a variety of complementary techniques be used to achieve the goal of anonymization (see the Opinion 5/2014 categorisation).

On the other hand, the General Court of the European Union, in its recent judgment in Case T-557/20 (JUR v. SEPD), has confirmed the criterion applied in Breyer case (C-582/14). The opinion of the Working Party just referred to defines the risk of re-identification on the basis of the ability of any entity, of any third party worldwide to re-identify taking into account the current state of the art and its evolution in the near future. This approach makes anonymization an almost impossible objective. However, the Court of Justice establishes as a criterion that of the capacity of the data subject to carry out any re-identification action:

> 90. In so far as recital 16 of Regulation 2018/1725 refers to the means likely reasonably to be used by both the controller and by 'any other person', its wording suggests that, for information to be treated as 'personal data' within the meaning of Article 3(1) of Regulation 2018/1725, it is not required that all the information enabling the identification of the data subject must be in the hands of one person (see, by analogy, judgment of 19 October 2016, Breyer, C-582/14, EU:C:2016:779, paragraph 43).
>
> (...)
>
> 96. It is true that, as the EDPS maintains, in the light of paragraph 43 of the judgment of 19 October 2016, Breyer (C-582/14, EU:C:2016:779), cited in paragraph 90 above, the fact that the additional information necessary to identify the authors of the comments received during the consultation phase was held not by Deloitte, but by the SRB, does not appear such as to exclude a priori that the information transmitted to Deloitte constituted, for Deloitte, personal data.
>
> 97. However, it is also apparent from the judgment of 19 October 2016, Breyer (C-582/14, EU:C:2016:779), that, in order to determine whether the information transmitted to Deloitte constituted personal data, it is necessary to put oneself in Deloitte's position in order to determine whether the information transmitted to it relates to 'identifiable persons'.

The project will have to take into account both legal positions on two levels:

- The level of anonymization requirements must be rigorous, bearing in mind that each data provider, or data subject, and the project itself, face primarily the possibility of being investigated by a data protection authority. In the event that they do not change their approach, a possible sanction would cause serious reputational damage and force them to go to court to enforce the standard defined by the Court of Justice.

- The design of the platform should be aligned with the objective of excluding the presence in the information system of any unauthorised third party and to ensure conditions of use and traceability that exclude the ability of users to perform any re-identification actions.

The principle of data protection by design and by default requires that the technology and design of the platform contribute to the achievement of these objectives. And these requirements are projected not only on the central node (ATLAS dashboard, etc.), but also on the technical requirements that apply to each of the local federated nodes on which data processing is planned.

The Spanish Data Protection Agency has based itself on the model proposed by the Data Protection Authority of Singapore to define a workflow that can be useful for the project, if properly implemented.

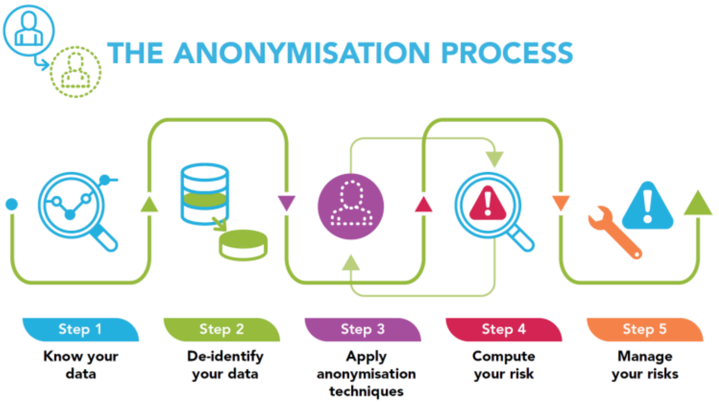


Figure 1. Steps for anonymization. Source: AEPD. Guide to basic anonymization. Prepared by the National Data Protection Authority of Singapore (PDPC - Personal Data Protection Commission Singapore

The dynamics of analysis proposed by this authority could be useful, both from the point of view of anonymization at source and of the risk approach on the platform.

1.2 Proposal for a Regulation on the European Health Data Space Scenario[14].

The EHDSR is in line with the GDPR's commitment to data re-use. European lawmakers still consider anonymization or pseudonymisation of real data to be essential.

> (37b) The secondary use of electronic health data can bring great societal benefits. The uptake of real-world data and real-world evidence, including patient-reported outcomes, for evidence-based regulatory and policy purposes as well as for research, health technology assessment and clinical objectives should be encouraged. Real-world data and real-world evidence have the potential to complement health data currently made available. To achieve this goal, it is important that data sets made available for secondary use by the present Regulation are as complete as possible. This Regulation provides the necessary safeguards to mitigate certain risks involved in the realisation of those benefits. The secondary use of electronic health data is based on pseudonymised or anonymised data, in order to preclude the identification of the data subjects.

From a risk management perspective, the possibility of re-identification Recital (64) identifies specific risks:

- **High identifiability in the case of rare diseases.**

> (64) (...) Even in situations of the use of state of the art anonymization techniques, there remains a residual risk that the capacity to re-identify could be or become available, beyond the means reasonably likely to be used. Such residual risk is present in relation to rare diseases (a life-threatening or chronically debilitating condition affecting not more than five in 10 thousand persons in the Union), where the limited numbers of case reduce the possibility to fully aggregate the published data in order to preserve the privacy of natural persons while also maintaining an appropriate level of granularity in order to remain meaningful.

- **Anonymisation is highly dependent on the relational structure of the dataset or its content.**

> (64) (...) It can affect different types of health data depending on the level of granularity and description of the characteristics of data subjects, the number of people affected or and for instance in cases of data included in electronic health records, disease registries, biobanks,

[14] Final Compromise Text of the Proposal for a Regulation on the European Health Data Space, published by the General Secretariat of the Council No. Cion doc.: 8571/22 ADD1-8. Available at https://www.consilium.europa.eu/media/70909/st07553-en24.pdf

> person generated data etc. where the identification characteristics are broader and where, in combination with other information (e.g. in very small geographical areas) (…)

- **It is technologically dependent**

> (64) (…) or through the technological evolution of methods which had not been available at the moment of anonymization, can lead to the re-identification of the data subjects using means that are beyond those reasonably likely to be used.

- **Lack of experience in safeguarding other rights**

> (64) (…) Furthermore, aggregation techniques are less tested for non-personal data containing for example trade secrets, as in the reporting on clinical trials and clinical investigations, and enforcement of breaches of trade secrets outside the Union is more difficult in the absence of a sufficient international protection standard.

The same recital highlights the profound social impact of a potential re-identification of health data that «would present a major concern and is likely to put the acceptance of the policy and rules on secondary use» provided for the Regulation at risk. That is the reason why it is necessary for the Commission to define, by means of a Delegated Act, criteria for the management of some of the specific risks in relation to international data transfers under the terms provided for in the Data Governance Act

> (64) (…) Furthermore, aggregation techniques are less tested for non-personal data containing for example trade secrets, as in the reporting on clinical trials and clinical investigations, and enforcement of breaches of trade secrets outside the Union is more difficult in the absence of a sufficient international protection standard. Therefore, for these types of health data, there remains a risk for re-identification after the anonymization or aggregation, which could not be reasonably mitigated initially. This falls within the criteria indicated in Article 5(13) of Regulation (EU) 2022/868. These types of health data would thus fall within the empowerment set out in Article 5(13) of Regulation (EU) 2022/868 for transfer to third countries. The protective measures, proportional to the risk of re-identification, would need to take into account the specificities of different data categories or of different anonymization or aggregation techniques and will be detailed in the context of the Delegated Act under the empowerment set out in Article 5(13) of Regulation (EU) 2022/868.

1.2.1 The focus is on the data user

The criteria for re-use of electronic health data present particular characteristics in the European Health Data Space Proposal. Firstly, it seems to be assumed that data holders will have difficulties in anonymising personal data. Thus, Articles 36, and specifically 37, of the EHDSR attribute to Health Data Access Bodies (HDABs), among other tasks, the task of anonymization or pseudonymisation. This is without prejudice to the fact that Recital (49) states that the task of anonymization may be performed by the data controller, by the HDAB and may be performed by a processor. The first article establishes it as a priority for the Member States, which must guarantee the exclusion of conflicts of interest in this area. The second defines it as a specific task of these bodies.

> Article 36
>
> Tasks of health data access bodies
>
> 1. Health data access bodies shall carry out the following tasks:
>
> (…)
>
> (d) process electronic health data referred to in Article 33 such as the receiving, combination, preparation and compiling of necessary requested data from health data holders, the pseudonymisation or anonymization of the data;

On the other hand, the regulatory framework also operates from the point of view of data access applicants for secondary use and on the conditions of access to the data themselves. Thus, the data minimisation principle of Article 44 EHDSR includes the core and backbone element of this matter:

> (…)
>
> 2. The health data access bodies shall provide electronic health data in an anonymised format, where the purpose of processing by the health data user can be achieved with such data, taking into account the information provided by the health data user.
>
> 3. Where the health data user has sufficiently demonstrated that the purpose of processing cannot be achieved with anonymised data in line with Article 46(1), point (c), the health data access bodies shall provide access to electronic health data in pseudonymised format. The information necessary to reverse the pseudonymisation shall be available only to the health data access body or a body that acts as trusted third party in accordance with national law.

The HDAB is under an obligation to check whether the intended purpose, the specified secondary use, together with the specific purpose of the processing, can be achieved with anonymised data. Pseudonymisation would be considered as a residual and conditional option. It is not sufficient that the purpose cannot be achieved with anonymised data; there must be proof of this on the part of the applicant. Nothing else follows from the provisions of Article 46 of the ECHR.

> Article 46 Data permit
>
> 1. The health data access bodies shall decide to grant access to electronic health data only when the following cumulative criteria are fulfilled:
>
> (a) the purpose described in the data access application matches one or more of the purposes listed in Article 34(1) of this Regulation;
>
> (b) the requested data is necessary, adequate and proportionate for the purpose or purposes described in the health data access application taking into account the provisions of data minimisation and purpose limitation in Article 44;
>
> **(c) the processing complies with Article 6(1) Regulation (EU) 2016/679, in particular that in the case of pseudonymized data, there is sufficient justification that the purpose cannot be achieved with anonymized data;**

This forces us to draw a preliminary conclusion. Firstly, whether it is the user with a data access permit who has the obligation to do the processing of anonymised data. Secondly, it is this user who is obliged to justify the use of pseudonymised data. It should be concluded that the aim of the system is that even this user should not have the possibility of re-identification of health data. Moreover, if the processing must necessarily take place in a secure processing environment (Article 50), the anonymization risk analysis must necessarily be carried out from the point of view of the capabilities of an authorised user.

1.3 De facto anonymization

We need to consider whether there is an appropriate strategy, or whether we should simply assume that anonymization is impossible, given the extremely strict position of data protection authorities in defining when personal data can be considered properly anonymised. If we consider the ever-expanding horizon of ever-increasing computing power, we can't come to any other conclusion. In addition, we have to take into account the possibility of using analytical methods with big data models or the cross-referencing of data from very different domains. Finally, in the medium term, we should think that all these capabilities, including the possibility of breaking encrypted data sets, will be feasible with quantum computing technologies. In this sense, the only conclusion that can be drawn is to consider the impossibility of anonymising large electronic health data sets.

We need to consider whether there is an appropriate strategy, or whether we should simply assume that anonymization is impossible, given the extremely strict position of data protection authorities in defining when personal data can be considered properly anonymised. If we consider the ever-expanding horizon of ever-increasing computing power, we can't come to any other conclusion. In addition, we have to take into account the possibility of using analytical methods with big data models or the cross-referencing of data from very different domains. Finally, in the medium term, we should think that all these capabilities, including the possibility of breaking encrypted data sets, will be feasible with quantum computing technologies. In this sense, the only conclusion that can be drawn is to consider the impossibility of anonymising large electronic health data sets.

However, the EHDSR's vision is very limited in its assessment of the risk of re-identification, as it only considers the most obvious cases. That is, rare diseases or the handling of genetic information, without taking into account the multiplier effect that can occur when handling large amounts of individual patient data, for example in studies of co-morbidity. From this point of view, we would be faced with the same risk situation as soon as we recombine these data with data from other sources, even if the use of e.g. K-anonymization-based analysis techniques[15] would allow us to eliminate those subjects potentially identifiable by inference or linkage.

Finally, if the determining criteria for the risk analysis oblige us to take into account, for example, the evolution of technology in the next ten years and/or the capabilities of any third party in the world, it is clear that we should consider the theoretical possibility of developing tools based on artificial intelligence that are capable of this re-identification.

If, in addition, the processing power of quantum computers lives up to its promises, the methods based on encryption will have to be upgraded without any delay. However, the aforementioned position of the General Court of the European Union opens up an alternative path. Perhaps this is the reason for the reference in recital (49) to the achievement of a 'possible' standard:

> (49) (…) When providing access to an anonymised or pseudonymised dataset, a health data access body should use state-of-the-art anonymization or pseudonymisation technology and standards, ensuring to the maximum extent possible that natural persons cannot be re-identified by health data users. Such technologies and standards for data anonymization should be further developed. Health data users should not attempt to re-identify natural persons from the dataset provided under this Regulation, subject to administrative fines and the enforcement measures laid down in this Regulation or possible criminal penalties, where the national laws foresee this

On the other hand, the positions of the Article 29 Working Party, now the European Data Protection Committee, and the General Court of the European Union are clearly divergent. If the criteria are not changed, Opinion 5/2014 will lead to a dead end[16]. In practice, irreversible anonymization, which is equivalent to erasure, is not feasible in the field of health, given the evolution of technology.

[15] Agencia Española de Protección de Datos (2021). ***K***-*ANONYMITY AS A PRIVACY MEASURE* (technical note). See at https://www.aepd.es/documento/nota-tecnica-kanonimidad-en.pdf

[16] Opinion 05/2014 on Anonymisation Techniques. Available at https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

However, there is another essential question from the point of view of secondary use, when it comes to the processing of massive amounts of data, the so-called big data, or the use of artificial intelligence. The inspiring principles of the AI Regulation call for a risk-based approach. For this, there are essential values such as accountability and the exclusion of bias. And when it specifically addresses data governance in high-risk systems, it emphasises that data governance requires a delicate balance between the principle of data minimisation and the variety, variability and diversity of data.

If the above statement is true, then 'traditional' anonymization techniques necessarily involve the reduction of variables, the exclusion of whole groups of subjects, or the use of techniques that may have certain collateral effects. There is nothing to prevent a properly anonymised data set from losing its scientific value or, even more painfully, from failing to support the risk analysis that is essential for the use of artificial intelligence. In other words, the anonymised dataset is fully acceptable from a GDPR perspective, but creates risks in areas such as the exclusion of bias or explainability. On the other hand, given the practical impossibility of obtaining the consent required in most jurisdictions, large-scale processing of pseudonymised data is currently impossible. Thus, unless the EHDSR is seen as legitimising the processing of large volumes of pseudonymised data without consent, the system would be unworkable.

Therefore, it seems necessary to adopt clearly and directly the concept of "de facto anonymization"[17]. The de facto anonymization methodology is compatible with a methodology based on risk based approach. This involves applying anonymization methodologies in successive phases following the steps recommended by the Spanish Data Protection Agency (see below) combined with the design of secure processing environments with security and traceability. In addition to anonymization techniques such as synthetic data, the software intermediation - such as differential privacy, secure multi-party computation or homomorphic encryption - could be considered as processes equivalent to anonymization or as legitimate uses of pseudonymised data. These are technologies with a high cost in terms of processing capabilities[18]. But at least they offer a framework of seemingly reliable solutions. Such strategies would be consistent with the proposal in recital (49).

[17] «De facto anonymization (sometimes also referred to as relative anonymization) describes de-identification operations by which so many identifiers are removed and further techniques (see 6.1) to reduce personal reference (e.g. randomization or generalization) are applied that re-identification with reasonable efforts in accordance with the current state of the art (see 3.3.4) is no longer possible and the personal reference is eliminated»

BDI (2021). *Anonymization of personal data. A cross-sector practical guide for industrial companies*. Available at https://issuu.com/bdi-berlin/docs/202103_handbook_bdi_anonymization-of-personal-data

[18] « **Anonymisation: Secure Multiparty Computing**

Secure Multiparty Computation150 or SMPC. This is a cryptographic protocol that, by means of Additive Secret Sharing, allows a secret data to be segmented into different parts, so that, when the information is shared, the original data cannot be revealed by any of the sources. In the protocol, the desired result is obtained without the need to reveal any sensitive data, and the result obtained does not suffer any type of deviation.

This strategy is useful in certain scenarios and requires technological assistance to implement it.

Anonymisation: Differential privacy

Differential privacy guarantees, by incorporating random noise to the original information, that in the result of the analysis process of the data to which this technique has been applied, there is no loss in the utility of the results obtained. It is based on the Law of Large Numbers, a statistical principle that states that when the sample size grows, the average values derived from it approach the real mean value of the information. Thus, the addition of random noise to all the data compensates for these effects and produces an 'essentially equivalent' value.

One example of use can be found in the US Census Bureau, which applies differential privacy to ensure the accuracy of its statistics and prevent personal information from being disclosed even through the statistics, and thus increase citizens' confidence in the security of the data they provide.

Anonymisation: Anonymisation-oriented documents

The technical deployment is completed by a legal apparatus based on three levels of interaction:

1.-The data holder provides the data by means of a data sharing and/or data transfer agreement.

To make this agreement viable, a set of formal and material guarantees are required from the entity:

- Ensuring that the data processing is legitimate and has been authorised by the data controller.
- Ensuring compliance with any applicable ethical requirements.
- Declare and catalogue restrictions on use in accordance with national law.
- Implement and accredit a reliable anonymization process.

In this regard, the platform from which data is processed should integrate a risk analysis. If there is a risk that the data are not correctly anonymised or that the infrastructure itself is used for this process, the legal support to the data holder necessarily implies the signing of a data processor contract.

2.-A legal-technical governance framework for data access must be established by the platform.

The achievement of this objective implies that, irrespective of the data access rights referred to in the EHDSR, the user with access to data must be subject to certain requirements:

Recital 9 of the DGA, in the case of re-use of data, states the need to develop data processing in which anonymisation is built into the concept of the data and in which data formats allow for efficient anonymisation 'by design': *'In order to facilitate the protection of personal data and confidential data and to speed up the process of making such data available for re-use under this Regulation, Member States should encourage public sector bodies to create and make available data in accordance with the principle of 'open by design and by default' referred to in Article 5(2) of Directive (EU) 2019/1024 and to promote the creation and the procurement of data in formats and structures that facilitate anonymisation in that regard.'*

Other techniques for safeguarding data protection

Without aiming to be exhaustive, there are other techniques used to safeguard data protection when sharing data. For example, homomorphic encryption, the recovery of private information, or the federated learning techniques in machine learning. The following is a brief overview of each of these techniques.

Homomorphic encryption is a privacy-by-default technique that is suitable for cases where a controller outsources a part of an activity to a processor, and wants to technically ensure that the processor will not access the data.

In a traditional scheme, the data controller transmits the information to the processor in encrypted form, to protect confidentiality during transit. Once the processor has received it, it is decrypted and processed. However, this scheme presents both legal and technical risks, so ideally, to minimise the risks, the processor should not have the possibility to decrypt the information, and all processing should be carried out on the data encrypted by the data controller. This would prevent a disloyal processor or a third party from accessing the data and using it for different purposes. One way to achieve this protection is through the so-called homomorphic encryption.

Homomorphic encryption therefore makes it possible to perform operations on encrypted data and obtain results, also encrypted, equivalent to the operations performed directly on the original information.

On the other hand, Private Information Retrieval (PIR) is a cryptographic technique that allows the user to retrieve an entry from a database without revealing to the data custodian the item that has been retrieved and unlink the information that could be inferred regarding who is performing the access. (...)

Lastly, we can also highlight federated learning techniques, both horizontal and vertical, for artificial intelligence applications based on Machine Learning. Federated learning techniques are a category of PET (Privacy-Enhancing Technology) that allow the development of machine learning systems without the need to communicate personal data between participants. These techniques can be both horizontal and vertical and are key in new scenarios for the improvement and development of society, such as Data Spaces».

Agencia Española de Protección de Datos (2023). *Approach to Data Spaces from GDPR Perspective*, pages 51-53. Available at https://www.aepd.es/documento/approach-to-data-spaces-from-gdpr-perspective.pdf

▪ A registration process on the platform, which implies an awareness of the rules of the platform.

▪The acceptance of a clear set of terms and conditions. This legal act does not correspond, strictu sensu, to the applicant for access to the data or to the user of the data, if they are different subjects, but to the legal representative of the entity in which they carry out their activities.

▪ The explicit acceptance by the user(s) of the information system of the system-specific security obligations and of the commitments not to re-identify. This should be done by taking a positive action such as ticking a box and accepting by receiving a confirmed message. The process must be digitally evidenced.

De facto anonymization therefore consists of a technical deployment that includes a robust anonymization process complemented by a secure processing environment. The latter prevents both the improper downloading or manipulation of data and any access by third parties. It thus excludes the risk of re-identification by third parties and the risk associated with the use of malicious software since no processing tools other than those previously authorised by the platform are supported.

This method should be acceptable, adequate and legally admissible. If this matter is not adequately resolved, we are faced with a particular situation. In most national laws there are no exceptions to consent for the processing of pseudonymised data. This means that during the transition to the EHDSR the processing of pseudonymised data will not be feasible. If consent is to be used, it is necessary to take into account the position of the European Data Protection Committee in the Guidelines 05/2020 on consent under Regulation 2016/679 and its consequences, i.e. the impossibility to collect large volumes of data on every occasion and for every data access request[19].

[19] See paragraphs 143 to 160. Available at https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-052020-consent-under-regulation-2016679_en

2. Legally recommended working scheme on anonymization.

The structure of the project makes it necessary to define two areas to be considered from the point of view of anonymization, the one that refers to the ATLAS or repository and the one that applies to processing in federated data spaces.

2.1 ATLAS

In this area, it should be noted that the data must be uploaded to the repository under the control of the project. In this phase the task scheme would be defined by the following steps:

Step 1: The data holder shall:

→ Provide legal certainty about the lawful use of the data

→ Be responsible for anonymization at source.

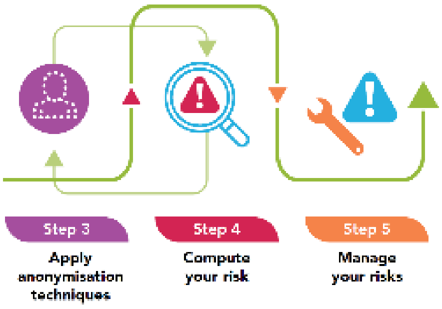
Step two: Risk analysis.

A risk analysis should be carried out when loading the datasets, which could preferably be integrated in the validation phase, or in any case in phases related to the curation or preparation of the data set.

Depending on the outcome of the risk analysis, two scenarios may arise:

→ The uploading of the data into the information system either because the risk does not exist or because the risk was managed with a second anonymization.

→ Proceed to a second processing in order to assure that the properly anonymization has been achieved.



Step 3: Risk management.

The information system shall integrate tools that guarantee the proper definition of user profiles and competences, ensure the traceability of their actions, and guarantee that users cannot perform actions aimed at re-identification.

It is clear that there are risks beyond the scope of the platform's development, such as the use of automatic screen capture tools. In the absence of methodologies to avoid such a risk, it would be advisable to adopt complementary strategies such as marking the images so that the copied images can be traced in the future. In any case, it is considered that the sum of the recommended measures, together with the binding commitments to be made by the users of the system, constitute a sufficient and adequate guarantee. Zero risk does not exist and the level of diligence can never reach the level of being able to prevent any malicious intent on the part of a user.

2.2 Procedures to be applied in the case of data analysis in federated areas.

The basic recommendations in this area are as follows.

▪ Step 1: Risk analysis

→ The data holder must guarantee at source the legal and technical conditions for anonymization of the data.

→ It is recommended to design a risk analysis task for the data set to be processed using federated methodologies

→ If there are risks of re-identification that cannot be corrected by project resources, a second re-identification will be recommended as a matter of course.

→ No data set shall be admitted or authorised without proper safeguards.

From a project sustainability point of view, lessons learned should be documented and training of experts in this task should be considered. It should be noted that the risk analysis processes of the re-identification conditions of the datasets provided by the data holders could be set up as a specific service of the platform.

Step 2: intermediation technology.

The development of federated data processing technologies should consider measures aimed at integrating an additional layer in the service of anonymization from the point of view of software intermediation. In this sense, state-of-the-art techniques related to the use of differential privacy or Multi-Party-Computation algorithms among other possible techniques applicable at the software layer have been proposed.

3. Additional considerations.

In addition, the following risks should be considered:

3.1 Combination of datasets

The combination of datasets from different sources requires the application of re-identification risk analysis methodologies and anonymization methodologies. It should be particularly taken into account that the interaction of the data contained in the project with other data spaces, be it in the field of health or in any other field, where it may generate re-identification risks.

The Proposal for a Regulation on the European Health Data Space makes it necessary to consider the possibility of processing using pseudonymised data. In this regard, the current Spanish law (Ad. Disp. 17 of LO 3/2018) offers criteria related to the conditions of the environment that are reproduced below.

> (d) the use of pseudonymised personal data for health and, in particular, biomedical research purposes is considered lawful.
>
> The use of pseudonymised personal data for public health and biomedical research purposes will require:
>
> 1. A technical and functional separation between the research team and those who carry out the pseudonymisation and keep the information that makes re-identification possible.
>
> 2. Pseudonymised data should only be accessible to the research team when:
>
> (i) There is an express commitment to confidentiality and not to engage in any re-identification activity.

> (ii) specific security measures are taken to prevent re-identification and access by unauthorised third parties.
>
> The re-identification of data at source may take place when, in the course of an investigation using pseudonymised data, it is established that there is a real and concrete danger to the safety or health of a person or group of persons, or a serious threat to their rights, or it is necessary to ensure adequate health care.
>
> (...)
>
> (f) Where, in accordance with Article 89 of Regulation (EU) 2016/679, processing is carried out for the purposes of public health research and, in particular, biomedical research, the following shall be carried out:
>
> 1. Carry out an impact assessment identifying the risks arising from the processing in the cases provided for in Article 35 of Regulation (EU) 2016/679 or those established by the supervisory authority. This assessment shall specifically include the re-identification risks linked to the anonymization or pseudonymisation of data.
>
> 2. Subject scientific research to quality standards and, where appropriate, to international guidelines on good clinical practice.
>
> 3. Adopt, where appropriate, measures aimed at ensuring that researchers do not access data that identifies the data subjects.
>
> 4. Designate a legal representative established in the European Union, in accordance with Article 74 of Regulation (EU) 536/2014, if the sponsor of a clinical trial is not established in
>
> the European Union. That legal representative may be the same as that provided for in Article 27(1) of Regulation (EU) 2016/679.
>
> g) The use of pseudonymised personal data for public health and, in particular, biomedical research purposes must be subject to the prior report of the research ethics committee provided for in the sectoral regulations.
>
> In the absence of the existence of the aforementioned Committee, the entity responsible for the investigation shall require a prior report from the data protection officer or, failing this, from an expert with the prior knowledge set out in Article 37(5) of Regulation (EU) 2016/679.

Our recommendation, without prejudice to the best criteria of developers, is to integrate by default this safeguards in any data processing, including anonymised data.

→ The use of any standardised tool for the development of the anonymization process must necessarily be documented including at least the following elements:

- Analysis of identified risks.
- Justification of the capabilities of the risk management tool.
- Basic definition of the functional characteristics of the tool.
- Procedure for the implementation and use of the tool.

→ The impact of the use of the platform for primary uses should be taken into account for risk management. The result of this risk analysis shall generate a report to be incorporated as evidence in the project documentation.

3.2 Updating of datasets.

The generation of health data is a dynamic process. In this sense, it is logical to consider that a dataset can, and even should, be updated at regular periods of time. In this scenario we refer exclusively to those cases in which the update occurs over the entire dataset regardless of whether the data of a specific patient is affected. This poses at least two risks that need to be considered:

▪ Each dataset must be subject to version control and a locked backup should be kept if not in use for a minimum period of time to be determined. In Spain the prescription period for civil liability is 5 years. The proposed Directive on AI liability obliges to apply the harmonised product liability framework and the state-specific civil liability framework. Therefore, the retention period of a dataset by means of a blocking of the dataset that proves that it was not tampered with should be programmed in an automated way taking into account the longer applicable period.

▪ It would be necessary to keep an active copy for those users who have a data access permission based on a previously defined specific dataset to which access was given We refer to those cases where a minimal alteration of the dataset could generate risks for areas such as reproducibility of the analysis, explainability or generate a bias. In these cases the marking, traceability and maintenance of the version of the dataset for which access was authorised and for the established period will be essential to maintain the integrity of the experiment and to avoid risks.

3.3 Updating datasets incorporating data that relate to a specific patient

This is one of the highest risk scenarios as it requires a patient identifier. It will be necessary to verify whether the encryption-based technologies or equivalent methodologies described in section 1.3 can guarantee the so-called de facto anonymization. In other words, the following conditions should be met:

1.-Anonymisation at source allows the generation of an encoding recognisable by the dataset that will integrate the data in such a way that they can be subject to update.

2.-This process is neither transparent nor accessible to EUCAIM nor to the user with a data access permit.

For this reason, the ideal method would be one in which the entire process takes place on the data holder's resources. Again, the preservation of older versions and the proper marking of each version would still be necessary.

If such a process is to be carried out by EUCAIM, it should be legally supported by a data processor contract. Technologically, the necessary safeguards should be adopted:

- Separate the integration environments from the storage environments of the updated dataset.

- Ensure that the linkage of a given patient in the final dataset made available for use with the original patient in the integrated dataset cannot be accessed.

- Ensure that an unauthorised user cannot reverse the process.

- Ensure absolute functional separation from the point of view of user profiles. If human intervention is required, the users responsible for the integration should under no circumstances be allowed to have permissions that would allow them to cross-reference data with the final dataset in use. If necessary for security and integrity reasons the comparison of datasets should preferably be articulated by automated means. The legal duties of security and confidentiality of the persons in charge of such processes should be very precisely and specifically defined.

Legal framework

- Council of Europe, Convention for the Protection of Individuals with Regard to the Automatic Processing of Individual Data, 28 January 1981, ETS 108 (**Convention 108**).

- Council of Europe, European Convention for the Protection of Human Rights and Fundamental Freedoms, as amended by Protocols Nos. 11 and 14, 4 November 1950, ETS 5 (**ECHR**).

- Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union OJ L 194 (**NIS Directive**).

- Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast) PE/28/2019/REV/1, OJ L 172 (**Open Data Directive**).

- Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148, PE/32/2022/REV/2, OJ L 333 (**NIS 2 Directive**).

- Directive 95/46/EC of the European Parliament and the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, OJ L 281, 23 November 1995 (**Data Protection Directive**).

- Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on adapting non- contractual civil liability rules to artificial intelligence (**AI Liability Directive**) COM/2022/496 final.

- Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on harmonised rules on fair access to and use of data (**Data Act**) COM/2022/68 final.

- Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (**Artificial Intelligence Act**) and amending certain Union legislative acts - General approach, <https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf>.

- Proposal for a Regulation Of The European Parliament And Of The Council on the European Health Data Space COM/2022/197 final (**EHDS**).

- Proposal for a Regulation Of The European Parliament And Of The Council Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts, COM/2021/206 final, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>.

- Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No. 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC 2017, OJ L 117/1 (**MDR**).

- Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU, OJ L 117 (**IVDR**).

- Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC, PE/31/2018/REV/1, OJ L 295.

- Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union, PE/53/2018/REV/1, OJ L 303 (**FFNPDR**).

- Regulation (EU) 2021/2282 of the European Parliament and of the Council of 15 December 2021 on health technology assessment and amending Directive 2011/24/EU (2021) OJ 458/

- Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (**Data Governance Act**) PE/85/2021/REV/1, OJ L 152.

- Regulation (EU) No 536/2014 of the European Parliament and of the Council of 16 April 2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC, OJ L 158 27.5.2014, p. 1 (**Clinical Trial Regulation**).

Official Documents

- Article 29 Working Party, “Annex to Letter from the WP29 to the European Commission”, DG CONNECT on mHealth, <https://ec.europa.eu/justice/article-29/documentation/other-document/files/2015/20150205_letter_art29wp_ec_health_data_after_plenary_annex_en.pdf>.

- Article 29 Working Party, “Opinion 01/2012 on the data protection reform proposals”, Adopted on 23 March 2021, <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2012/wp191_en.pdf>.

- Article 29 Working Party, „Opinion 05/2014 on Anonymisation Techniques”, Adopted on 10 April 2014, <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf>.

- Article 29 Working Party, “Guidelines on Data Protection Impact Assessment (DPIA) and Determining Whether Processing Is ‘Likely to Result in a High Risk’ for the Purposes of Regulation 2016/679” (2017)

- 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, <https://eur-lex.europa.eu/legal-

content/EN/TXT/?uri=COM:2019:250:FIN>.

- Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions - A European strategy for data COM/2020/66 final, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066>.

- Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions - Fostering a European approach to Artificial Intelligence COM/2021/205 final, <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=COM%3A2021%3A205%3AFIN>.

- Council of Europe, "Guidelines to respect, protect and fulfil the rights of the child in the digital environment", Recommendation CM/Rec(2018)7 of the Committee of Ministers, <https://rm.coe.int/guidelines-to-respect- protect-and-fulfil-the-rights-of-the-child-in-th/16808d881a>.

- EDPB, "Opinion 3/2019 concerning the Questions and Answers on the interplay between the Clinical Trials Regulation (CTR) and the General Data Protection regulation (GDPR) (art. 70.1.b))", <https://edpb.europa.eu/sites/default/files/files/file1/edpb_opinionctrq_a_final_en.pdf>.

- EDPB, "Recommendations 01/2020 on measures that supplement transfer tools to ensure compliance with the EU level of protection of personal data", Version 2.0, Adopted on 18 June 2021, <https://edpb.europa.eu/our-work-tools/our-documents/recommendations/recommendations-012020- measures-supplement-transfer_en>.

- EDPB, Guidelines 05/202 on consent under Regulation 2016/679, Version 1.1, Adopted on 4 May 2020, <https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202005_consent_en.pdf>.

- EDPB-EDPS (2022) Joint Opinion 02/2022 on the Proposal of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act), Adopted on 4 May 2022, <https://edpb.europa.eu/our-work-tools/our-documents/edpbedps-joint-opinion/edpb-edps-joint-opinion- 22022-proposal-european_en>.

- ENISA, "Cybersecurity and privacy in AI - Medical imaging diagnosis", 07.06.2023, <https://www.enisa.europa.eu/publications/cybersecurity-and-privacy-in-ai-medical-imaging-diagnosis>.
- ENISA, "Cybersecurity Challenges of Artificial Intelligence", 15.12.2020, <https://www.enisa.europa.eu/publications/artificial-intelligence-cybersecurity-challenges>.
- ENISA, "Data Pseudonymisation: Advanced Techniques and Use Cases", 28.01.2021,

<https://www.enisa.europa.eu/publications/data-pseudonymisation-advanced-techniques-and-use-cases>.

- ENISA, “Securing Machine Learning Algorithms”, 14.12.2021, <https://www.enisa.europa.eu/publications/securing-machine-learning-algorithms>.

- EDPB, “Guidelines 4/2019 on Article 25 Data Protection by Design and by Default” Version 2.0. Adopted on 20October 2020, <https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-42019-article- 25-data-protection-design-and_en>.

- European Commission, “Coordinated Plan on Artificial Intelligence”, <https://digital-strategy.ec.europa.eu/en/policies/plan-ai#:~:text=The%20key%20aims%20of%20the,AI%20policy%20to%20avoid%20fragmentation.&text=The%20Coordinated%20Plan%20on%20Artificial%20Intelligence%202021%20Review%20is%20the,global%20leadershi p%20in%20trustworthy%20AI.>.

- European Commission, “The EU’s Cybersecurity Strategy for the Digital Decade”, <https://digital- strategy.ec.europa.eu/en/library/eus-cybersecurity-strategy-digital-decade-0>.

- European Commission, ANNEXES to the Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions - Fostering a European approach to Artificial Intelligence COM(2021) 205 final, ANNEX, <https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review>.

- European Commission, Directorate-General for Research and Innovation, Eechoud, M., Study on the Open Data Directive, Data Governance and Data Act and their possible impact on research, Publications Office of the European Union, 2022, <https://data.europa.eu/doi/10.2777/71619>.

- European Parliament, “DRAFT Compromise Amendments on the Draft Report Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts”, Version 1.1, 16/05/2023, <https://www.europarl.europa.eu/resources/library/media/20230516RES90302/20230516RES90302.pdf>.

- High-Level Expert Group on Artificial Intelligence (HLEG), “Assessment List for Trustworthy AI (ALTAI) for self assessment”, 17 July 2022, <https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy- artificial-intelligence-altai-self-assessment>.

- High-Level Expert Group on Artificial Intelligence, “A Definition of AI: Main capabilities and disciplines”, 8 April 2019, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

- High-Level Expert Group on Artificial Intelligence, “Ethics Guidelines for

Trustworthy AI", 8 April 2019, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

- Joint Communication To The European Parliament And The Council, The EU's Cybersecurity Strategy for the Digital Decade JOIN/2020/18 final, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020JC0018>.

- Medical Device Coordination Group (MDCG) 2019-11 Guidance on Qualification and Classification of Softwarein Regulation (EU) 2017/745 – MDR and Regulation (EU) 2017/746 – IVDR, October 2019, <https://health.ec.europa.eu/system/files/2020-09/md_mdcg_2019_11_guidance_qualification_classification_software_en_0.pdf>.

- Medical Device Coordination Group (MDCG) 2019-16 Rev.1 Guidance on Cybersecurity for medical devices,December 2019, July 2020 rev.1, <https://ec.europa.eu/docsroom/documents/41863>.

- UNCRC, The United Nations Convention on the Rights of the Child, <https://www.unicef.org.uk/wp-content/uploads/2016/08/unicef-convention-rights-child-uncrc.pdf>.
- WHO, Ethics and governance of artificial intelligence for health, <https://www.who.int/publications/i/item/9789240029200>.

- WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects, 6 September 2022, <https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for- medical-research-involving-human-subjects/>.

- WMA Declaration of Taipei on Ethical Considerations regarding Health Databases and Biobanks, 4 June 2020, <https://www.wma.net/policies-post/wma-declaration-of-taipei-on-ethical-considerations-regarding-health- databases-and-biobanks/>.

- WMA, Declaration of Taipei – Research on Health Databases, Big Data and Biobanks, <https://www.wma.net/what-we-do/medical-ethics/declaration-of-taipei/>.

- ARTICLE 29 DATA PROTECTION WORKING PARTY, Opinion n. 5/2014 on Anonymisation Techniques, WP216 (10.04.2014), https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf, 0829/14/EN WP216

- Cyber Essentials by the UK National Security Centre, https://www.cyberessentials.ncsc.gov.uk/

- The CNIL`s Guides (2018) Guidance on the Security of Personal Data

- The ICO Guide on data security, https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/security/

- The ICO guidelines on the safe disposal of IT devices, https://ico.org.uk/media/for-organisations/documents/1570/it_asset_disposal_for_organisations.pdf

- UKRI Engineering and Physical Sciences Research Council, EPSRC Policy Framework on Research Data
- ICO (2012) Bring your own device (BYOD) guidance, https://ico.org.uk/media/for-organisations/documents/1563/ico_bring_your_own_device_byod_guidance.pdf
- ICO (2016) A practical guide to IT security – Ideal for the small business p. 13., https://ico.org.uk/media/for-organisations/documents/1575/it_security_practical_guide.pdf
- ICO draft guidance, Chapter 1: introduction to anonymisation - Chapter 2: How do we ensure anonymisation is effective? - Chapter 3: pseudonymisation - Draft anonymisation, pseudonymisation and privacy enhancing technologies guidance, February 2022, https://ico.org.uk/about-the-ico/ico-and-stakeholder-consultations/ico-call-for-views-anonymisation-pseudonymisation-and-privacy-enhancing-technologies-guidance/

Opinions of the European Data Protection Board (EDPB) Regarding Data Protection Impact Assessments (DPIA)

- Opinion 7/2020 on the draft list of the competent supervisory authority of France regarding the processing operations exempt from the requirement of a data protection impact assessment (Article 35(5) GDPR). Date: 22 April 2020. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). EDPB France.
- Opinion 13/2019 on the draft list of the competent supervisory authority of France regarding the processing operations exempt from the requirement of a data protection impact assessment (Article 35(5) GDPR). Date: 12 July 2019. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). EDPB France.
- Opinion 12/2019 on the draft list of the competent supervisory authority of Spain regarding the processing operations exempt from the requirement of a data protection impact assessment (Article 35(5) GDPR). Date: 12 July 2019. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). EDPB Spain.
- Opinion 11/2019 on the draft list of the competent supervisory authority of the Czech Republic regarding the processing operations exempt from the requirement of a data protection impact assessment (Article 35(5) GDPR). Date: 12 July 2019. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). EDPB Czech Republic.
- Opinion 10/2019 on the draft list of the competent supervisory authority of Cyprus regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35(4) GDPR). Date: 12 July 2019. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). EDPB Cyprus.
- Opinion 7/2019 on the draft list of the competent supervisory authority of Iceland regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 12 March 2019. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Iceland EDPB.
- Opinion 6/2019 on the draft list of the competent supervisory authority of Spain regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 12 March 2019. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). EDPB Spain.
- Opinion 2/2019 on the draft list of the competent supervisory authority of Norway regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 23 January 2019. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). EDPB Norway.

- Opinion 01/2019 on the draft list of the competent supervisory authority of the Principality of Liechtenstein regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 23 January 2019. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). EDPB Liechtenstein.
- Opinion 27/2018 on the draft list of the competent supervisory authority of Slovenia regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 4 December 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Slovenia.
- Opinion 26/2018 on the draft list of the competent supervisory authority of Luxembourg regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 4 December 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Luxembourg.
- Opinion 25/2018 on the draft list of the competent supervisory authority of Croatia regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 4 December 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Croatia.
- Opinion 24/2018 on the draft list of the competent supervisory authority of Denmark regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 4 December 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Denmark.
- Dictamen 25/2018 sobre el proyecto de lista de la autoridad de control competente de Croacia en relación con las operaciones de tratamiento que requieren una evaluación de impacto relativa a la protección de datos (artículo 35, apartado 4, del RGPD). Date: 4 December 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Croatia.
- Dictamen 24/2018 sobre el proyecto de lista de la autoridad de control competente de Dinamarca en relación con las operaciones de tratamiento que requieren una evaluación de impacto relativa a la protección de datos (artículo 35, apartado 4, del RGPD). Date: 4 December 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Denmark.
- Opinion 9/2018 on the draft list of the competent supervisory authority of France regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). France.
- Opinion 8/2018 on the draft list of the competent supervisory authority of Finland regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Finland.
- Opinion 7/2018 on the draft list of the competent supervisory authority of Greece regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Greece.
- Opinion 6/2018 on the draft list of the competent supervisory authority of Estonia regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Estonia.
- Opinion 5/2018 on the draft list of the competent supervisory authorities of Germany regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Germany.

- Opinion 4/2018 on the draft list of the competent supervisory authority of Czech Republic regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Czech Republic.
- Opinion 3/2018 on the draft list of the competent supervisory authority of Bulgaria regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Bulgaria.
- Opinion 22/2018 on the draft list of the competent supervisory authority of the United Kingdom regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). United Kingdom.
- Opinion 21/2018 on the draft list of the competent supervisory authority of Slovakia regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Slovakia.
- Opinion 20/2018 on the draft list of the competent supervisory authority of Sweden regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Sweden.
- Opinion 2/2018 on the draft list of the competent supervisory authority of Belgium regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Belgium.
- Opinion 19/2018 on the draft list of the competent supervisory authority of Romania regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Romania.
- Opinion 18/2018 on the draft list of the competent supervisory authority of Portugal regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Portugal.
- Opinion 17/2018 on the draft list of the competent supervisory authority of Poland regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Poland.
- Opinion 16/2018 on the draft list of the competent supervisory authority of the Netherlands regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Netherlands.
- Opinion 15/2018 on the draft list of the competent supervisory authority of Malta regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Malta.
- Opinion 14/2018 on the draft list of the competent supervisory authority of Latvia regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Latvia.
- Opinion 13/2018 on the draft list of the competent supervisory authority of Lithuania regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Lithuania.

- Opinion 12/2018 on the draft list of the competent supervisory authority of Italy regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Italy.
- Opinion 11/2018 on the draft list of the competent supervisory authority of Ireland regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Ireland.
- Opinion 10/2018 on the draft list of the competent supervisory authority of Hungary regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Hungary.
- Opinion 1/2018 on the draft list of the competent supervisory authority of Austria regarding the processing operations subject to the requirement of a data protection impact assessment (Article 35.4 GDPR). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Austria.
- Dictamen 22/2018 sobre el proyecto de lista de la autoridad de control competente del Reino Unido en lo que respecta a las operaciones de tratamiento que requieran una evaluación de impacto relativa a la protección de datos. Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). United Kingdom.
- Dictamen 21/2018 sobre el proyecto de lista de la autoridad de control competente de Eslovaquia en lo que respecta a las operaciones de tratamiento que requieran una evaluación de impacto relativa a la protección de datos. Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Slovakia.
- Dictamen 20/2018 sobre el proyecto de lista de la autoridad de control competente de Suecia en lo que respecta a las operaciones de tratamiento que requieran una evaluación de impacto relativa a la protección de datos (artículo 35, apartado 4, del RGPD). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Sweden.
- Dictamen 2/2018 sobre el proyecto de lista de la autoridad de control competente de Bélgica en lo que respecta a las operaciones de tratamiento que requieran una evaluación de impacto relativa a la protección de datos (artículo 35, apartado 4, del RGPD). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Belgium.
- Dictamen 19/2018 sobre el proyecto de lista de la autoridad de control competente de Rumanía en lo que respecta a las operaciones de tratamiento que requieran una evaluación de impacto relativa a la protección de datos (artículo 35, apartado 4, del RGPD). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Romania.
- Dictamen 18/2018 sobre el proyecto de lista de la autoridad de control competente de Portugal en lo que respecta a las operaciones de tratamiento que requieran una evaluación de impacto relativa a la protección de datos (artículo 35, apartado 4, del RGPD). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Portugal.
- Dictamen 17/2018 sobre el proyecto de lista de la autoridad de control competente de Polonia en lo que respecta a las operaciones de tratamiento que requieran una evaluación de impacto relativa a la protección de datos (artículo 35, apartado 4, del RGPD). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Poland.
- Dictamen 16/2018 sobre el proyecto de lista de la autoridad de control competente de los Países Bajos en lo que respecta a las operaciones de tratamiento que requieran una

evaluación de impacto relativa a la protección de datos. Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Netherlands.

- Dictamen 15/2018 sobre el proyecto de lista de la autoridad de control competente de Malta en lo que respecta a las operaciones de tratamiento que requieran una evaluación de impacto relativa a la protección de datos (artículo 35, apartado 4, del RGPD). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Malta.
- Dictamen 14/2018 sobre el proyecto de lista de la autoridad de control competente de Letonia en lo que respecta a las operaciones de tratamiento que requieran una evaluación de impacto relativa a la protección de datos (artículo 35, apartado 4, del RGPD). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Latvia.
- Dictamen 13/2018 sobre el proyecto de lista de la autoridad de control competente de Lituania en lo que respecta a las operaciones de tratamiento que requieran una evaluación de impacto relativa a la protección de datos (artículo 35, apartado 4, del RGPD). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Lithuania.
- Dictamen 12/2018 sobre el proyecto de lista de la autoridad de control competente de Italia en lo que respecta a las operaciones de tratamiento que requieran una evaluación de impacto relativa a la protección de datos. Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Italy.
- Dictamen 10/2018 sobre el proyecto de lista de la autoridad de control competente de Hungría en lo que respecta a las operaciones de tratamiento que requieran una evaluación de impacto relativa a la protección de datos (artículo 35, apartado 4, del RGPD). Date: 3 October 2018. Opinion of the Board (Art. 64). Data Protection Impact Assessment (DPIA). Hungary.