EUCAN
IMAGE

A European Cancer Image Platform Linked to Biological and Health Data for Next-Generation Artificial Intelligence and Precision Medicine in Oncology

Deliverable D1.1: Legal and ethical interoperability study/retrospective analysis

Reference	**D1.1_ EucanImage_UPV**
Lead Beneficiary	University of the Basque Country (UPV)
Author(s)	Recuero, Mikel Nicolás, Pilar Durst, Ludovica Goisauf, Melanie Mayrhofer, Michaela Th. Schlünder, Irene Zaccagnini, Davide
Dissemination level	Public
Type	Report
Official Delivery Date	September 30th, 2021
Date of validation of the WP leader	September 29th, 2021
Date of validation by the Project Coordinator	September 29th, 2021
Project Coordinator Signature	

EuCanImage is funded by the European Union's H2020 Framework
Under Grant Agreement 952103

Version log

Issue Date	Version	Involved	Comments
18/01/2021	V1.1	Nicolás, Pilar Recuero, Mikel	1st draft
18/02/2021	V1.2	Durst, Ludovica Goisauf, Melanie Mayrhofer, Michaela Schlünder, Irene Zaccagnini, Davide	First common review and comments on the 1st draft
18/03/2021	V2.1	Nicolás, Pilar Recuero, Mikel	2nd draft
18/05/2021	V2.2	Durst, Ludovica Goisauf, Melanie Mayrhofer, Michaela Schlünder, Irene Zaccagnini, Davide	Second common review and comments on the 2nd draft
18/06/2021	V3.1	Nicolás, Pilar Recuero, Mikel	3rd draft
18/07/2021	V3.2	Durst, Ludovica Goisauf, Melanie Mayrhofer, Michaela Schlünder, Irene Zaccagnini, Davide	Third common review and comments on the 3rd draft
30/08/2021	V4	Nicolás, Pilar Recuero, Mikel	4th draft and first submission
13/09/2021	V4.1	Durst, Ludovica Goisauf, Melanie Mayrhofer, Michaela Schlünder, Irene Zaccagnini, Davide	Fourth common review and comments on the 4th draft
14/09/2021	V.4.2	Nicolás, Pilar Recuero, Mikel	Final amendments and comments
23/09/2021	V5	Nicolás, Pilar Recuero, Mikel	Last submitted version

Acronyms

Acronym	Name
CDO	Chief Data Officer
CJEU	Court of Justice of the European Union
DPA	Data Processing Agreement
DPO	Data Protection Officer
DTA	Data Transfer Agreement
EDPB	European Data Protection Board
EEA	European Economic Area
EU	European Union
GDPR	General Data Protection Regulation
PACS	Picture Archiving and Communication Systems
SCC	Standard Contractual Clauses
SMPC	Secure Multi-Party Computation
TCIA	The Cancer Imaging Archive
WP29	Article 29 Working Party

Table of Contents

Executive summary

Content overview

This report comprises the Deliverable 1.1 entitled "Legal and ethical interoperability study and retrospective analysis", as an essential part of the Task 1.1. "Overall assessment of legal and ethical constraints for data sharing in oncology imaging", both from WP1.

Objectives

The Deliverable covers a retrospective analysis and a preliminary general assessment of several issues such as the legislation in force in Europe as a whole (leveraging ongoing experiences in other projects such as EUCANCan, EUCAN-Connect or euCanSHare) in terms of data protection and anonymisation; roles and responsibilities of EuCanImage's users and stakeholders, legal differences and requirements within cancer data types (imaging and non-imaging data) as well as a comparative view across relevant jurisdictions (e.g. the European Union and the United States) for interoperability purposes.

Methodology

This document has been preceded by a review of documentation from the consortium members e.g., Data Processing Agreements, Data Transfer Agreements, Ethics Reviews, etc., together with a fluid and permanent communication with the different Work Packages of EuCanImage. Likewise, meetings and discussions are held weekly amongst the different partners within various Working Groups e.g., ELSI, platform, data, or AI.

Furthermore, this report is closely connected to other Tasks, not only in the Work Package 1 (i.e., Tasks 1.2, 1.3, 1.4 and 1.5), but also concerning other Work Packages (i.e., Tasks 3.1, 3.2, 3.3, 3.4, 3.5 and 4.1).

Ultimately, it further combines the background ELSI expertise of the WP1 members with the insights from the various meetings and discussions; followed by a literature, normative and jurisprudential review; leading to a comprehensive assessment to the present or foreseeable needs of EuCanImage.

Structure

The document is structured in three main parts:

i. Legal and ethical retrospective analysis: questionnaire and conclusions.
ii. Overall assessment of legal and ethical constraints for data sharing within EuCanImage.
iii. Legal interoperability study between relevant jurisdictions.

1 Introduction

Deliverable 1.1 aims to carry out a legal and ethical interoperability and retrospective analysis with a preliminary overall assessment for data sharing in oncology imaging. This Deliverable is included in Task 1.1. "Overall assessment of legal and ethical constraints for data sharing in oncology imaging", both from WP1.

Thus, Deliverable 1.1 should be understood as the foundations of the EuCanImage platform from a legal and ethical perspective, even though this report will not initially address all the issues related to its final design, which will depend on many factors that may *a priori* be difficult to shape unilaterally. It should be conceived, however, as an assessment of the initial needs of the whole consortium and as a sort of state-of-the-art which will eventually be integrated in the final privacy by design and by default report (Deliverable 1.4).

The initial steps of EuCanImage must be guided by a thorough assessment that allows to fully understanding the needs of the project itself and of the different partners and, at the same time, to conciliate them with a scrupulous legal and ethical compliance.

Therefore, this report will contribute to the achievement of the final goal of EuCanImage, the creation of a GDPR and ethical compliant integrated platform for large-scale cancer imaging, including well-validated AI solutions, as well as specific objectives such as:

- **Objective 3:** Build a multi-centre and multi-scale AI development platform for cancer imaging.
- **Objective 5:** Develop the legal framework, as well as innovative solutions, that will enable responsible data sharing and enhanced Open Science within EuCanImage and the cancer research community.

At the very same time, as a part of Task 1.1, this report is intended as a retrospective analysis and a preliminary general assessment of several issues such as the legislation in force in Europe as a whole (leveraging ongoing experiences in other projects such as EUCANCan, EUCAN-Connect or euCanSHare) in terms of data protection and anonymisation; roles and responsibilities of EuCanImage's users and stakeholders, legal differences and requirements within cancer data types (imaging and non-imaging data) as well as a comparative view across relevant jurisdictions (e.g. the European Union and the United States) for interoperability purposes.

To conclude, this document is clearly structured in three main parts:

i. **Legal and ethical retrospective analysis: questionnaire and conclusions.** As an initial part of Task 1, it was decided to set out a brief questionnaire to find out about various issues relating internal policies, procedures, tools or approaches of the institutions and centres that will be sharing data within the EuCanImage platform. This was conceived as a key step to implement most of the work in WP1, including forthcoming standardised policies and contracts and the global design of the data protection layer for the platform. Hence, this first approach will set out the different understandings and responses from the partners who will be sharing data and draw up several conclusions for the crucial interoperability between partners and jurisdictions.

ii. **Overall assessment of legal and ethical constraints for data sharing within EuCanImage.** This assessment is based on the main conclusions drawn from the previous questionnaire, meetings and contacts held with the different partners, Working Groups and Work Packages. The point is to achieve the right balance between the objectives of the project, the initial needs of the partners and the strict compliance with legal and ethical requirements. Hence, it is intended to analyse in a more detailed fashion the whole data flow throughout the project and the structure of the platform and catalogue. Accordingly, appropriate data sharing scenarios will be designed from a legal and ethical standpoint, ensuring meticulous respect for the privacy and data protection of individuals.

Ultimately, the aim is to implement, from the very design of the data flow (e.g., Data Management Plan) and the platform (e.g., platform architecture), the principles of privacy by design and by default, among others, which will be further developed and underpinned in the forthcoming Deliverables of this Work Package up to the subsequent privacy by design report (Deliverable 1.4.)

iii. **Legal interoperability study between relevant jurisdictions.** Finally, one last major hurdle to overcome is the divergence between EU Member States' laws and other different jurisdictions. Therefore, the final objective must be aligned with achieving interoperability between the diverse jurisdictions and countries involved, developing a comprehensive and interoperable legal governance framework with a robust privacy-by-design-based approach. In addition, the engagement of The Cancer Imaging Archive (hereinafter referred to as "TCIA"), based in the United States, raises numerous questions due to the regulatory divergences between jurisdictions. Consequently, it is also intended to analyse the feasibility of carrying out international data transfers to that territory, although this is initially not among the objectives of the present project but of an eventual further scalability.

2 Legal and ethical retrospective analysis: questionnaire and conclusions

As mentioned above, Task 1 is aimed at providing an overall assessment of legal and ethical constraints for data sharing in oncology imaging. This work shall therefore commence with an initial assessment and consultation phase which, prior to the development of the platform and the subsequent data flow, allows ascertaining the initial status and position of each of the participants, as a sort of state-of-the-art. A thorough understanding of the initial needs of stakeholders will *ab initio* contribute to the design of a data sharing framework in line with ethical and legal requirements and, specifically, with the General Data Protection Regulation[1] (hereinafter, "GDPR"). In addition to the GDPR, other regulations may apply throughout the development and subsequent operation of the platform, such as the Regulation on free flow of non-personal data[2], national or Member States' law, or even the forthcoming Data Governance Act[3].

Thus, regular communication and multidisciplinary teamwork between the different partners are inherent to this Task, which will contribute to bridge the gap between regulatory and ethical issues and the day-to-day operative of EuCanImage. For this reason, the present report has been preceded by a thorough review of institutional documentation from clinical partners and research centres including, when necessary, Ethics Committees' approvals.

Likewise, meetings and conversations are held almost every week amongst the different Working Groups involved in the development of EuCanImage, so that there is fluid and permanent communication between the Work Package 1 and the rest of Work Packages involved (e.g., clinical, anonymisation, data, AI, platform, etc.).

On top of the above, it was decided to set out a brief questionnaire to check what is the baseline of each partner in terms of data protection and ethical requirements implementation, what data is to be shared (source, storage, transfers, disclosure), what role they are assuming in the consortium (and/or in the forthcoming platform) and what initial conception each partner has, among others, on relevant issues such as pseudonymisation and anonymisation.

To this end, a series of questions have been formulated, presented in three different categories:

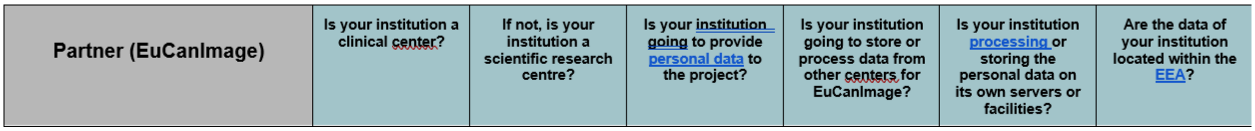
[1] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

[2] 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.

[3] Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act).

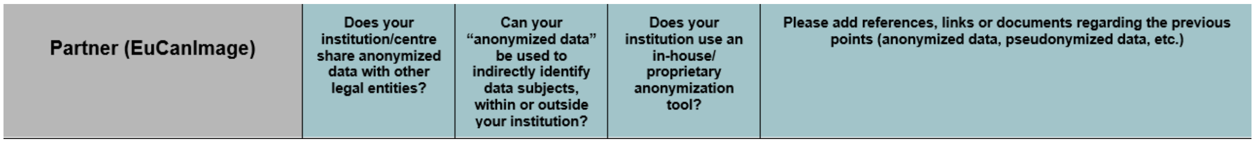
i. Data to be shared.
ii. Data source, data storage, and transfers.
iii. Data identifiability and anonymisation.

Lastly, it should be emphasised that this first section of the report only refers to *ex ante* data processing operations carried out internally by institutions, centres, and partners before sharing such data with EuCanImage or the platform itself. Issues concerning *ex post* data sharing (both internally and externally) will be further explored in the subsequent sections of this report and the forthcoming deliverables.



Partner (EuCanImage)	Is your institution a clinical center?	If not, is your institution a scientific research centre?	Is your institution going to provide personal data to the project?	Is your institution going to store or process data from other centers for EuCanImage?	Is your institution processing or storing the personal data on its own servers or facilities?	Are the data of your institution located within the EEA?

Figure 1. *Summary of the overall checklist sent to the partners and consortium members.*



Partner (EuCanImage)	Does your institution/centre share anonymized data with other legal entities?	Can your "anonymized data" be used to indirectly identify data subjects, within or outside your institution?	Does your institution use an in-house/ proprietary anonymization tool?	Please add references, links or documents regarding the previous points (anonymized data, pseudonymized data, etc.)

Figure 2. *Summary of the checklist on data identifiability sent to the partners and consortium members.*

2.1 Data to be shared

The first issue raised in the questionnaire sent to consortium members concerns the very nature of the data to be shared within EuCanImage. It is intended to determine, *ab initio*, with which type of data (both quantitative and qualitative) the processing operations prior to the constitution of the platform are to be carried out. These, together with issues relating to data identifiability, are two crucial points for the application of data protection rules and for assessing the initial risks entailed by the forthcoming processing operations.

Thus, the personal data (or potentially personal) to be initially shared within EuCanImage have been identified as corresponding to the following categories:

- **Imaging data.** In particular, cancer imaging data including raw imaging data (DICOM format) and annotated images in which the tumour is segmented. It is intended to populate initially the platform with imaging data relating to three types of cancer: colorectal, liver and breast. However, at a larger stage, it is planned to populate the platform with a wide range of cancer types (lung, brain, prostate, bone, among others). All imaging data and collections will be stored and managed at Euro-Bioimaging Archive.
- **Non-imaging data.** The imaging data itself will also be associated to other non-imaging data such as cancer outcomes, tumour gene expression or immunohistochemistry. In addition, a further key objective of the project is for

the platform to include other datasets relating to non-imaging data such as biological data, health records and -omics, in order to link the cancer images in the EuCanImage archive to repositories where corresponding non-image information will be hosted, curated and managed. Furthermore, the use of metadata for cancer health data is foreseen, which will include information such as donor, primary diagnosis, follow-up, specimen, treatment (chemotherapy, radiotherapy, hormone therapy, targeted therapy, and surgery), environmental exposures, family, comorbidity, and biomarkers). All non-imaging data, including genomic, biomarker and clinical information such as diagnoses, and outcomes will be stored and managed at European Genome-Phenome Archive (EGA).

Ultimately, it is foreseen that by the end of the project more than +20.000 unique human subjects will be linked to the platform. Therefore, it can be clearly noted that this is a vast and varied amount of data both from a quantitative (>20000) and qualitative (data concerning health, genetic data, biometric data) perspective. This entails several legal and ethical implications, notably, if the information relates to an identified or identifiable natural person.

Hence, as far as the GDPR is concerned, the first step is to determine whether the information (both imaging and non-imaging) is likely to be of a personal nature and, thus, fall under the umbrella of the Regulation. This issue is by no means easy to address, and most probably cannot be fully answered in this first report, even though it will be further explored in the section 2.3 of this report.

By contrast, it does seem clearer that the data to be used are highly sensitive and therefore may fall into what the GDPR designates as "special categories of personal data". Consequently, if the case-by-case analysis revealed that personal information is concerned, the data would most likely fall within these special categories. This will entail, *per se*, several legal implications such as that in addition to the legal basis for processing stated on the Article 6(1) of the GDPR, a separate condition for processing under Article 9(1) of the GDPR must be in place.

In fact, the data referred to above can easily fall into the following three legal categories, all of which relate to special or 'sensitive' data:

- **Data concerning health.** The GDPR defines them as "personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status"[4]. For instance, the health records and/or other data (imaging and/or non-imaging) that fall within the broad notion of data concerning health.
- **Genetic data.** The GDPR describes them as "personal data relating to the inherited or acquired genetic characteristics of a natural person which give unique information about the physiology or the health of that natural person

[4] Article 4(15) GDPR.

and which result, in particular, from an analysis of a biological sample from the natural person in question"[5]. For example, certain genomic data.

- **Biometric data.** The GDPR provides the following definition: "personal data resulting from specific technical processing relating to the physical, physio-logical or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data"[6]. Cancer imaging data could also fall into this definition (but not limited to).

Consequently, from a **qualitative** standpoint, due to their particular sensitivity, the data initially provided by the clinical partners must be subject (at/from source) to all applicable legal and ethical requirements, measures and safeguards including, *inter alia*, correspondent legal basis for the processing, conditions for the processing of special categories of data, compliance with data protection principles, subject's rights, etc.

In addition, it is paramount that these legal bases and conditions for the processing are compatible with the purpose of the processing to be carried out within EuCanImage and the forthcoming platform. Hence, where the consent constitutes the legal basis for the processing (Articles 6(1)(a) + 9(2)(a) of the GDPR), this must be sufficiently broad[7] to cover the further purposes of the processing to be carried out within EuCanImage. The same applies where other legal bases are relied upon, such as the public interest, legal obligation, or legitimate interests, combined with the correspondent derogation for the processing of special categories of data (e.g., Article 9(2)(j) of the GDPR). Therefore, this means that not all data providers need to rely on the same legal basis, as the data will predominantly be of clinical origin and the consortium either as a whole or the platform will be involved in the initial stages of such data collection.

Nevertheless, once the data have been collected and processed by the data providers, it will be necessary to rely on new legal grounds and conditions for processing them with secondary and/or compatible purposes. The following figure constitutes a graphic representation of the different alternatives that EuCanImage and its data providers may rely on for the (secondary or compatible) processing of special categories of data for scientific research purposes:

[5] Article 4(13) GDPR.

[6] Article 4(14) GDPR.

[7] Article 5(1)(b) *in fine* + Article 89(1) + Recital 33 of the GDPR.

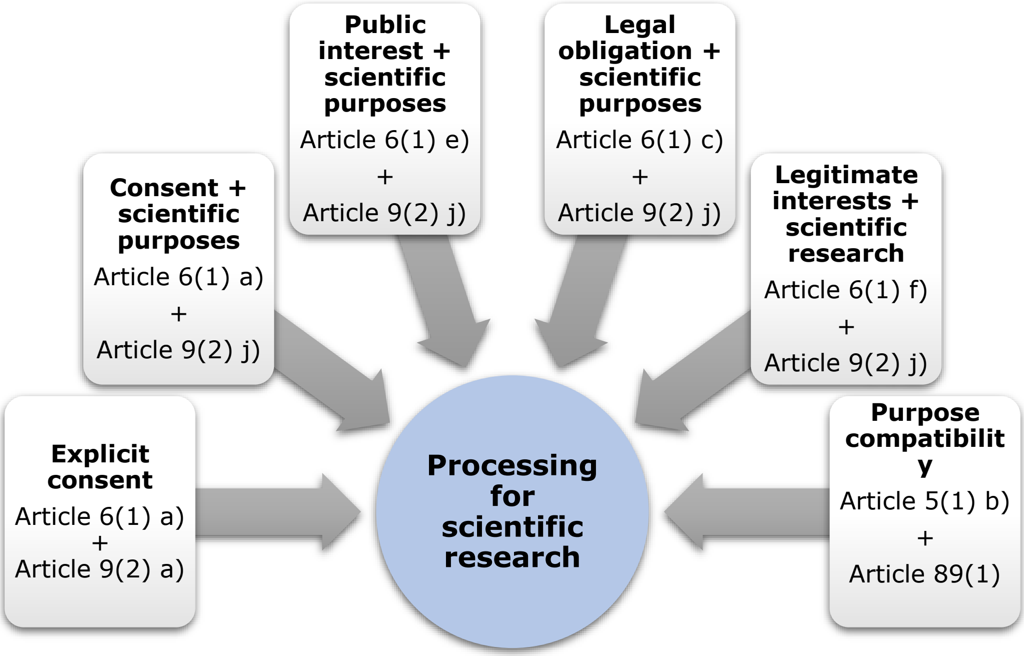


Figure 3. *Alternatives for processing special categories of data within EuCanImage.*

On the other hand, from a **quantitative** perspective, as mentioned above, a vast amount of data will be processed throughout the development of the platform. This, together with the special sensitivity of the data, renders it highly advisable for the controllers to carry out an initial risk analysis which, in the event that the processing operations result in a high risk to the rights of data subjects and in accordance with the legal criteria in force, may entail the obligation of the controllers to carry out a Privacy Impact Assessment (PIA), among other requirements.

2.2 Data source, storage, and transfers

The submitted questionnaire was followed by a series of questions aimed at understanding the origin (source) of the data and the nature of the entities initially providing the data or datasets. This point is relevant to identify which country the data or datasets originate from (inside or outside the EU) and the very nature of the source institution or entity (public or private research centres or clinical institutions).

Therefore, the questionnaire has revealed that initially EuCanImage will be populated with cancer imaging and non-imaging data from a total of 5 clinical partners or institutions plus the TCIA's data (5+1), which will be integrated in a federated manner, as can be seen in the figure below:

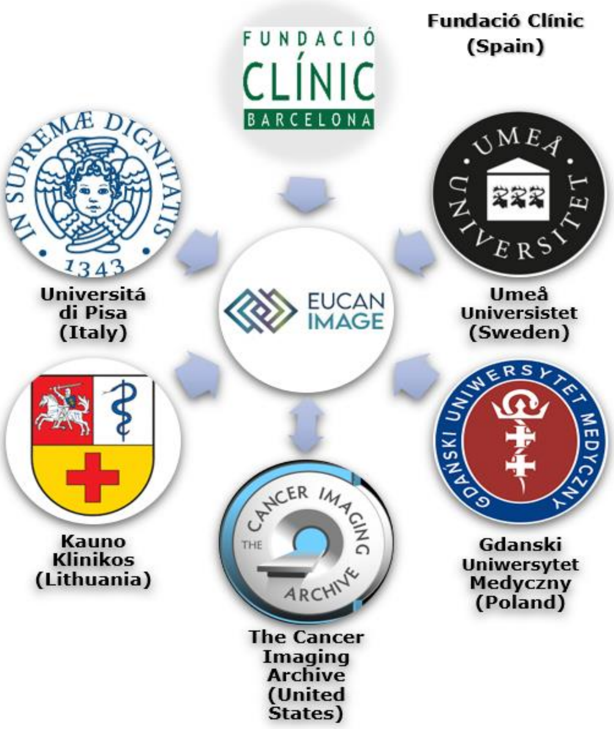


Figure 4. *Initial data providers of imaging and non-imaging for EuCanImage.*

From the above, it can be inferred that the data providers are either clinical centres, research institutions or universities. This may be relevant, for example, when relying on a certain legal basis or condition for the processing of personal data. Additionally, another relevant finding is that data providers have admitted that, initially, the data they are going to provide to the platform are of a personal nature and/or have not previously undergone an effective anonymisation process, as they may still be indirectly identifiable at least by the source institution. Finally, the TCIA will not directly provide data to the platform but integrated in a federated fashion. This means users of EuCanImage platform will also receive information about available similar datasets at TCIA, but not directly hosted or managed by EuCanImage.

Secondly, further questions were asked to understand the role that each partner is deemed to play in the data flow, given that EuCanImage is a consortium of up to twenty partners, not all of whom will be processing personal data, and some of whom will not even have access to them. In this regard, following the roles established by the GDPR, three essential legal roles or positions can be identified within the set of processing operations:

- **Data controllership (including data controllers and/or joint controllers).** The term data controller means "the natural or legal person, public authority, agency or other body which determines the purposes and means of the processing of personal data”[8]. Likewise, where two or more controllers jointly determine the purposes and means of processing, they shall be joint controllers[9].

This implies that data providers (clinical partners, research centres or universities) will operate as independent data controllers and thus will be responsible for the legitimate collection of data following national and European legal requirements, including ethical declarations and other requirements.

Therefore, at least in the first stage of the project, there will be five independent data controllers:

- Fundació Clínic Recerca Biomèdica (FCBR), Spain.
- Università di Pisa (UNIPI), Italy.
- Kauno Klinikos (KAUNO), Lithuania.
- Umeå Universitet (UMU), Sweden.
- Gdański Uniwersytet Medyczny (GUMED), Poland.

The EuCanImage platform-catalogue (managed by Euro-Bioimaging through the Erasmus Medical Centre Rotterdam) will not assume the role of a data controller as it will act as a mere intermediary and thus as a data processor or sub-processor (still under discussion). Nevertheless, as the former Article 29 Working Party (hereinafter, ‘WP29’) points out, regardless of the role or qualification given to a party by contract, if they decide how the data are to be processed (to a greater or lesser extent but only in an essential way) or determine the purposes or means of such processing, they may be qualified as data controllers[10]. More recently, this has also been confirmed by its successor, the European Data Protection Board (hereinafter, ‘EDPB’)[11].

Therefore, the qualification and categorisation of the EuCanImage platform and catalogue as mere ‘facilitator’ or ‘exchange platform’ will depend on the final scheme finally adopted within the data flow and on the set of processing operations to be carried out on behalf of the data providers.

Similarly, the case law of the Court of Justice of the European Union[12] (hereinafter referred to as "CJEU") has led to an expansive interpretation of

[8] Article 4(7) GDPR.

[9] Article 26(1) GDPR.

[10] Article 29 Data Protection Working Party. (2010). *Opinion 1/2010 on the concepts of “controller” and “processor”*. Pages 18-19.

[11] European Data Protection Board. (2020). *Guidelines 07/2020 on the concepts of controller and processor in the GDPR*. Paragraph 27.

[12] See, among others: Judgment of 29 July 2019, *Fashion ID v. Facebook Ireland Ltd*. (C-40/17); Judgment of 10 July 2018 Tietosuojavaltuutettu v. Jehovan todistajat (C-25/17);

the concept of joint controllership and to the need to assess the influence or concurrence of the entities at different stages and to different degrees of the processing operations. In other words, only a minimum level of involvement in determining the means and purposes of processing is needed to acquire controller or joint controller status, which means that in the present context, many participants or stakeholders may end up falling into this broad concept of controllership. Hence the crucial importance of clearly delimiting the responsibilities of each entity and limiting the processing operations to be carried out by either the platform or the catalogue to only those strictly designated by the data providers.

- **Data processors (including sub-processors).** Defined in the GDPR as the "natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller"[13]. In turn, it is lawfully conceivable for the data processor to engage another processor for carrying out specific processing activities on behalf of the controller[14].

  The application of these concepts to EuCanImage implies that, at least, the partners and entities de-identifying, curating, storing and annotating the data on behalf of the data providers will, unless otherwise evidenced, hold the role of data processors (or sub-processors). This may be the case, of the following partners:

  - Euro-Bioimaging (Erasmus Medical Centre Rotterdam). Management and storage of imaging data collection through XNAT platform.
  - European Genome-phenome Archive (Centre for Genomic Regulation). Management and storage of non-imaging data.
  - Barcelona Supercomputing Centre (ELIXIR). EuCanImage integrated web portal.
  - Collective Minds Radiology AB (CMRAD). Image annotation.
  - OncoRadiomics, SRL. AI research and development.
  - Siemens Healthineers. AI research and development.
  - Universitat de Barcelona. AI research and development.
  - Universiteit Maastricht. AI research and development.

and Judgment of 5 June 2018, *Wirtschaftsakademie Schleswig v. Facebook Ireland Ltd* (C-210/16).

[13] Article 4(8) GDPR.

[14] Article 28(4) GDPR.

In addition, these data processors would also include sub-processors[15] e.g., Amazon Web Services.

As can be observed, the data flow becomes complex and can lead to a chain of sub-contracting resulting in a blurring of responsibilities and legal roles. Therefore, it becomes essential to regulate properly the controller-processor and processor-sub-processor relationships in a transparent manner. Moreover, the data providers (controllers) will have to hold the pivotal role and liability towards the rights of the data subjects. Consequently, in addition to the corresponding contracts and agreements (Data Processing Agreements, Sub-processing Agreements) the controller shall maintain an absolute authority and accountability over the processing operations delegated to one or more processors with the legal limitation on the latter's ability to engage one or more sub-processors without controller's specific or general authorisation.

- **Authorised users or persons.** Persons who, under the direct authority of the controller or processor, are authorised to process personal data[16]. This may include employees of the data controller or processor and/or other users or subjects who do not exactly fit the definition of either controller or processor but who are authorised to carry out one or more data processing operations. In EuCanImage, an authorised user would be the healthcare professionals annotating the images on behalf of the controllers or the processors' staff with access to the data.

- **Third parties.** Means a "natural or legal person, public authority, agency or body other than the data subject, controller, processor and persons who, under the direct authority of the controller or processor, are authorised to process personal data"[17]. What is relevant for this role is to bear in mind that the fact that a subject or an entity is considered a third party with regard to one set of processing operations does not preclude that they may be considered later a controller or processor. For example, a third party receiving personal data may take on the role of controller or processor of such data. In EuCanImage, a third party would be the scientist or AI developer accessing the platform and/or requesting a particular dataset.

[15] Article 28(4) of the GDPR: «Where a processor engages another processor for carrying put specific processing activities on behalf of the controller».

[16] See Article 4(10) *sensu contrario*.

[17] See Article 4(10).

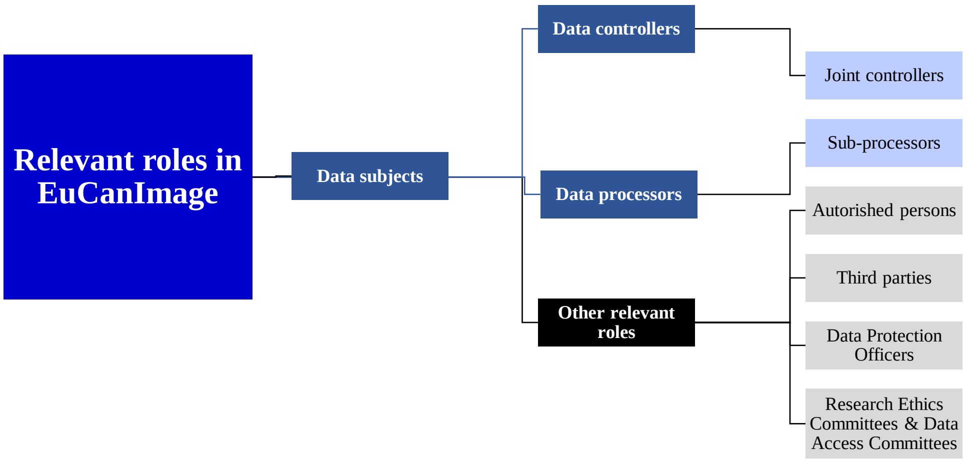


Figure 5. *Overview of relevant data protection roles for EuCanImage*.

Ultimately, a series of questions were asked to identify potential transfers of personal data to third countries outside the EU and associated risks. In this regard, all data providers have explicitly declared that the data is being (or it is going to be) hosted or stored on their own servers or facilities (or under their direct control and management by means of a sub-processor) and that all of them are located within the European Union. Nevertheless, it is essential to confirm that the data will not leave the European Union in any case, either to a controller or to a processor (or sub-processor) located in a third country. This is particularly relevant in the case of data flows to and with the TCIA, as it is located in the United States. At this stage, all data controllers and data processors will be located in the European Economic Area (hereinafter 'EEA') and no data is going to be transferred to the US. However, international data flows can take place in the opposite direction: from the US to EU. In any case, section 4.2 of this report will address issues concerning cross-border and international data flows, with special emphasis on the transfers of data to third countries (TCIA in the United States) to pave the way for possible future scalability.

Nevertheless, it should be stressed that the preliminary conclusions and legal assessment given in this first section mainly refer to the initial stage of the project and the platform. The respective partners and participants may see their roles and their corresponding ethical and legal responsibilities alter significantly when their purposes and duties are explicitly defined and when a definitive data sharing scheme has been agreed upon, as roles and responsibilities may vary significantly from a centralised to a federated or hybrid data sharing model.

2.3 Data identifiability and anonymisation

Finally, several questions have been asked related to the identifiability of the information and the eventual internal de-identification procedures and/or anonymisation techniques of the data providers. This preliminary approach is essential because, in order to fall within the scope of the GDPR, the processing operation must relate or involve personal data, namely information relating to an identified or identifiable natural person[18]. The concept of personal data is, therefore, crucial, as the GDPR will not apply to non-personal data or to anonymous or anonymised information but it will apply to pseudonymised data[19] [20].

In fact, the GDPR defines pseudonymisation as follows:

> *"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".*

Nonetheless, this distinction is not straightforward in theoretical nor practical terms and even more so when sensitive and particularly identifying (or even "individualising") data are concerned, as is the case of biometric and genetic data. However, the aim of this report is not to offer an academic complex approach to the concept of anonymisation or pseudonymisation, but rather to analyse the existing needs and facts and to offer a "by design" preliminary assessment to complete the initial ethical and legal study carried out within this deliverable.

Hence, it must be concluded that the original (raw) data obtained from clinical partners and research centres (data providers) are of a personal nature and cannot be considered as anonymised under the terms of the GDPR[21] and the current state-of-the-art (including, among others, the CJEU[22], the European Data Protection Board[23] and/or the Article 29 Data Protection Working Party[24]). This initial

[18] Article 4(1) of the GDPR.

[19] Recital 26 of the GDPR.

[20] European Data Protection Supervisor and Agencia Española de Protección de Datos. (2021). *10 misunderstandings related to anonymisation*. Page 3.

[21] Identifiability test outlined in Recital 26 of the GDPR.

[22] Court of Justice of the European Union (CJEU). Judgment of 19 October 2016, *Patrick Breyer v. Bundesrepublik Deutschland*, Case C-582/14 (hereinafter "Breyer").

[23] European Data Protection Board. (2021). *EDPB Document on response to the request from the European Commission for clarifications on the consistent application of the GDPR, focusing on health research*.

[24] See, among others: Article 29 Data Protection Working Party. (2007). *Opinion 4/2007 on the concept of personal data*. Adopted on 20th June 2007; Article 29 Data Protection

understanding is further strengthened by the responses received in the questionnaire and by discussions with the data providers.

As a result, there is *ab initio* no consensus on the use of effective and combined de-identification tools or anonymisation techniques and some institutions have stated that they still maintain links to the original data subject and/or that the indirect identification of the data subject remains feasible. It can therefore be concluded that datasets initially shared by clinical partners are of a personal nature and that the techniques or procedures internally applied by them constitute thereof a mere technical measure or safeguard under the GDPR.

Nevertheless, to overcome this situation and to mitigate risks, a consensus has been reached among the EuCanImage partners to apply the set of de-identification techniques of POSDA tool, with the help of the TCIA experience in its implementation. This de-identification process will be carried out locally before any dataset is transferred by the data controller (clinical site), under its strict control. It will be carried out by the controller itself or by a data processor on its behalf (e.g., CMRAD). In any event, any de-identification process of such data will constitute *per se* a processing operation subject to the GDPR and must rely, *inter alia*, in an adequate and justified legal basis[25]. For this very reason, such processing must be conducted in a manner that complies with the GDPR and its principles, namely data minimisation, purpose limitation (exemptions may apply to processing with scientific research purposes) lawfulness or conditions for the processing of special categories of data, inter alia.

This report is not prejudging the outcome of the application of such a technique and whether it is effective in achieving anonymisation, partly because there is currently no clear legal nor jurisprudential position on the matter. It is merely worth noting that, given the current state-of-the-art, it is questionable whether there can be an effective anonymisation in terms of the GDPR when the project seeks to link the cancer images in the EuCanImage archive (e.g., EuroBioImaging and local repositories) to repositories where corresponding non-image information will be hosted, curated and managed; resulting in a large number of linked and enriched databases or repositories that increase the risk of re-identification. In addition, the European Data Protection Board (hereinafter referred to as "EDPB") has recently stated, when asked about anonymisation of genetic and health data in scientific research, that:

Working Party. (2014). *Opinion 05/2014 on Anonymisation Techniques.* Adopted on 10 April 2014.

[25] European Data Protection Board. (2021). *EDPB Document on response to the request from the European Commission for clarifications on the consistent application of the GDPR, focusing on health research*. page 11.

> *"It should be taken into account that anonymisation of personal data can be difficult to achieve (and upheld) due also to ongoing advancements in available technological means, and progress made in the field of re-identification. For this reason, the anonymisation of personal data <u>should be approached with caution in the context of scientific research</u>* (…). *The EDPB points out that the possibility to anonymise genetic data <u>remains an unresolved issue</u>. (…) However, in the interests of protecting the rights and freedoms of individual data subjects, <u>it is strongly advised that such genetic data is treated as personal data</u> and that the processing thereof is conducted with the implementation of appropriate technical and organisational measures to ensure compliance with the Regulation"*[26]*.*

To conclude, this is an issue to be assessed on a case-by-case basis and in accordance with the measures and safeguards adopted and, therefore, will require further discussion and detail beyond the scope of this report. Consequently, it cannot be ascertained whether the data resulting from the de-identification process carried out by means of the POSDA suite will legally qualify as anonymised information or merely as pseudonymised data.

[26] Ibid. paragraphs 50 and 51.

3 Overall assessment of legal and ethical constraints for data sharing within EuCanImage

The preliminary conclusions drawn from the previous section constitute an essential building block for the final design and understanding of EuCanImage's data scheme and the set of processing operations to be carried out. As mentioned above, this point aims to analyse, in a more detailed fashion, the whole data flow and processing operations to be carried out throughout the project and the structure of the forthcoming platform and catalogue.

Therefore, this third section will further examine the set of data processing operations and activities, the subjects and entities involved and how all of them are interconnected to constitute a whole data flow or scheme that, despite its complexity, is already built on strong ethical and legal principles and premises. Subsequently, EuCanImage's participants, users and other stakeholders will be classified in terms of legal roles along with their liabilities and responsibilities.

Firstly, the very nature of the data to be processed must be considered personal, as it cannot yet be decisively concluded whether POSDA's de-identification procedure and techniques can produce truly anonymised information from the GDPR's standpoint and the current state-of-the-art.

Secondly, the storage and management system chosen by the data providers or clinical sites will be also taken into account, as EuCanImage has adopted a hybrid data management system supporting (1) a centralised approach based on the Euro-Bioimaging and EGA repositories, and (2) a de-centralised approach that will allow clinical sites to retain their data locally.

In addition, it is worth bearing in mind that this report has been preceded by a multitude of meetings and conversations held with the different partners involved in EuCanImage and in its various stages, such as the Data Management Plan, data model discussions, platform architecture and design, de-identification techniques and procedures, ethical requirements, etc. Therefore, the privacy by design and GDPR-compliant approach has been emphasised and underpinned from the very outset of the project and from the platform's design, architecture, and data flows.

The next lines and sections describe the different stages and data processing operations identified in EuCanImage as of the drafting of this report. Nevertheless, it should be noted that, as the project remains at an early phase, certain processes, functions, architectures, and dynamics are still being explored and developed. As a result, these reports should be conceived as dynamic documents and responsive to practical needs, while retaining their legal virtuality and accuracy.

3.1 Processing operations and stages within EuCanImage

EuCanImage's complexity stems largely from the need to balance and harmonise in a legally compliant fashion all the processing operations carried out by the different consortium members, ensuring accountability and transparency, and avoiding excessive, unnecessary, or inaccurate data processing. It is therefore paramount to carry out a detailed and individualised assessment of the processing operations and stages within EuCanImage, which needs to result in a comprehensive report of privacy by design (Deliverable 1.4).

This section aims to provide a preliminary overview of EuCanImage's data flow which will allow to identify the subjects and entities involved in the processing operations and their legal roles (together with their responsibilities and liabilities) and to reconcile them with the ethical and legal requirements.

From a technical standpoint, the following sets of operations will be carried out in EuCanImage:

- **De-identification.** This process consists in "removing or substituting all patient identifiers such as name, address, hospital identification number, from the patient data"[27]. Such a de-identification process will be carried out for imaging data (DICOM format) by means of the POSDA tool of the TCIA. However, at this stage of the project, it is still being explored what de-identification techniques or procedures will be applied to the non-imaging data. In any case, as already mentioned in this report, from a legal standpoint, de-identification is not equivalent to anonymisation, as the latter will only take place when the "information does not relate to an identified or identifiable natural person"[28]. For anonymisation to be effective it must be irreversible or, at least, unlikely that a data subject will be re-identified given the context of the processing, the circumstances of the individual case and the state of the technology. This implies that if the (source) data controller retains the raw data, or any key or other information that may be used to reverse the anonymisation and to identify the data subject, the information will still be personal data and therefore the GDPR will apply[29]. Therefore, removing direct identifiers is not enough and it will often be necessary to adopt additional measures and safeguards to prevent data subject's identification.
- **Curation**. It can be defined as "the entirety of procedures and actions after data gathering that refer to data management, creation, modification, verification,

[27] Diaz, O.; Kushibar, K.; Osuala, R.; Linardos, A.; *et. al.* (2021) "Data preparation for artificial intelligence in medical imaging: A comprehensive guide to open-access platforms and tools". *Physica Medica*. 83:2021. P. 26.

[28] Recital 26 of the GDPR.

[29] See among others: Article 29 Data Protection Working Party. (2014). *Opinion 05/2014 on Anonymisation Techniques*. p.5.; Data Protection Commission. (2019). *Guidance on Anonymisation and Pseudonymisation*. p. 6.

extraction, integration, standardisation, conversion, maintenance, quality assurance, integrity, validation, traceability and reproducibility"[30]. Hence, it constitutes a broad set of processing operations that may or not simultaneously involve different entities (i.e., roles), tools and applications.

- **Storage**. It can either be carried out centrally on an external server (e.g., through a data processor engaged by the data controller), or in a federated form by the data controllers themselves in their own infrastructures and servers. These set of stored data that are made available are often referred to as 'repository'. Similarly, in the case of imaging data storage, they are often referred to as 'Picture Archiving and Communication Systems' (PACS).
- **Annotation.** It is the process of "labelling the images with essential information e.g., spatial location and classification, in the same DICOM file or in a separate text report"[31]. In the case of EuCanImage, this annotation process will be carried out by means of a data processor (Collective Minds) that will provide the CMRAD platform so that experts can collaboratively annotate medical images on the cloud across sites. At this point, it will be essential to establish restricted or controlled access procedures and to ensure the secrecy and confidentiality (e.g., via duty of secrecy or even NDAs).
- **AI research and development.** EuCanImage is designed at its core to promote research reproducibility and data reuse for AI research and development. Therefore, the final stage comprises a number of other operations such as AI training and AI validation carried out by third parties (AI developers) with temporary access to the data or with the possibility to transfer the data to them pursuant pertinent agreements.

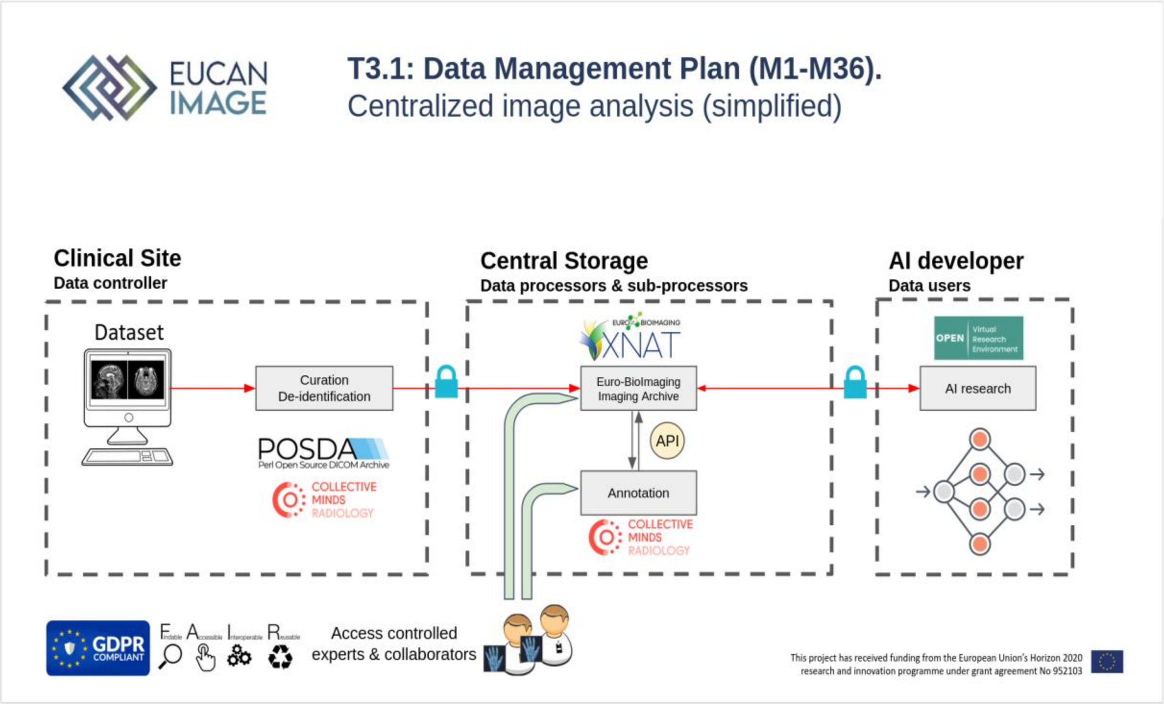
The above-mentioned operations may constitute different modalities of personal data processing operations insofar they relate to an identified or identifiable natural person. Thus, in addition to determining whether they refer to identified or identifiable information, it will also be necessary to examine which parties or entities carry them out and, where appropriate, on behalf of which controllers. In any case, these are not the only processing operations to be carried out, but the main ones, which will require further development in subsequent deliverables.

From this pure technical standpoint of the data management and analysis, as mentioned above, the consortium has adopted a hybrid system in which the data controllers themselves can decide between both a centralised and decentralised model. From this perspective, two scenarios need to be developed:

[30] Diaz, O.; Kushibar, K.; Osuala, R.; Linardos, A.; *et. al.* (2021) "Data preparation for artificial intelligence in medical imaging: A comprehensive guide to open-access platforms and tools". *Physica Medica*. 83:2021. Page 27.

[31] Ibid. page 30.

- **Centralised**. Data will be hosted centrally. On the one hand, imaging data will be stored in XNAT instances of Euro-BioImaging (The Netherlands). On the other hand, non-imaging data (genomics and phenotype data) will be stored in the European Genome-phenome Archive (EGA) and should accomplish the requirements specified in the EGA instructions and policies (e.g., informed consents, de-identification, etc.) and the access to the datasets will be governed by *ad hoc* Data Access Committees (DACs).



T3.1: Data Management Plan (M1-M36).
Centralized image analysis (simplified)

EUCAN IMAGE
Clinical Site
Data controller
Dataset
Curation
De-identification
POSDA
Perl Open Source DICOM Archive
COLLECTIVE MINDS RADIOLOGY
Central Storage
Data processors & sub-processors
EURO-BIOIMAGING XNAT
Euro-BioImaging Imaging Archive
API
Annotation
COLLECTIVE MINDS RADIOLOGY
AI developer
Data users
OPEN Virtual Research Environment
AI research
GDPR COMPLIANT
Findable Accessible Interoperable Reusable
Access controlled experts & collaborators

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 952103

Figure 6. *Simplified data flow of centralised management and storage system for imaging data in EuCanImage.*

- **Federated**. Data will not leave the hospitals or clinical centres because it will be locally stored. Therefore, all described operations (e.g., de-identification, curation, annotation, etc.) will be carried out locally and the AI training and research will be carried out in a distributed manner, by means of a distributed learning infrastructure. When it comes to imaging data, EuCanImage will offer support for setting up local instances of XNAT.

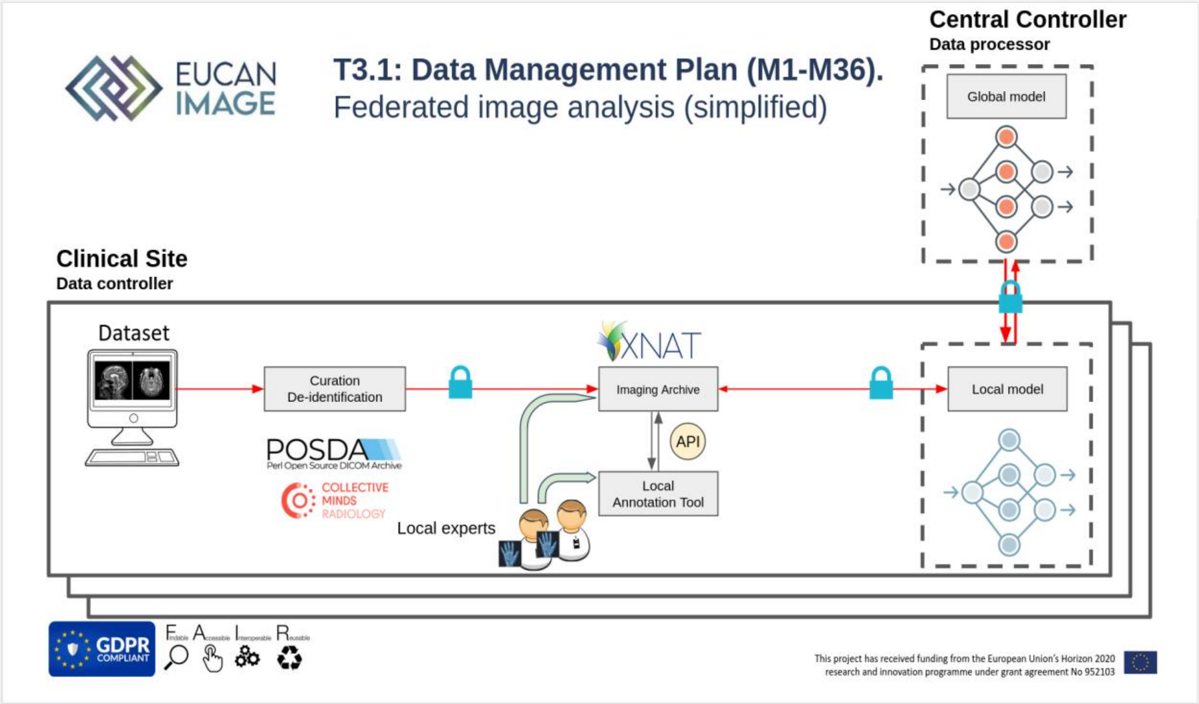

EUCAN IMAGE
T3.1: Data Management Plan (M1-M36).
Federated image analysis (simplified)
Central Controller
Data processor
Global model
Clinical Site
Data controller
Dataset
Curation
De-identification
XNAT
Imaging Archive
API
Local Annotation Tool
Local model
POSDA
Perl Open Source DICOM Archive
COLLECTIVE MINDS RADIOLOGY
Local experts
GDPR COMPLIANT
Findable Accessible Interoperable Reusable
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 952103

Figure 7. *Simplified data flow of federated management and storage system for imaging data in EuCanImage.*

Nevertheless, above two purely technical schemes or scenarios are not sufficient to address all relevant ethical and legal aspects or issues, such as the nature of the information to be processed at each stage (i.e., personal, or non-personal), the roles of the parties involved, the contracts or agreements to be signed, the technical and organisational measures and/or the additional safeguards to be adopted. Consequently, from an ELSI perspective, a more detailed assessment must be carried out, considering the legal and ethical issues and the possible implications of each technical decision.

3.2 Ethical and legal overview of EuCanImage data processing and sharing stages

The purpose of this section is to provide a comprehensive overview of, on the one hand, the different stages or phases and sets of processing operations identified in EuCanImage and, on the other hand, the legal and ethical implications and requirements for each proposed scenario.

Overall, following the findings that have preceded this report (e.g., meetings, questionnaires, checklists, and other discussions), it has been decided to reduce all possible data processing and sharing models for EuCanImage in three stages. Each stage presents its own particularities in terms of data protection and should therefore be separately addressed:

i. **Stage 1. Data source, minimisation, and de-identification.**

This first phase starts with the data collection by the data providers, i.e., the clinical partners initially integrated in EuCanImage. All data providers are located inside the European Union (Spain, Italy, Lithuania, Sweden and Poland). The collection of the data is carried out independently of EuCanImage, as the data sets in question are already available in each centre or hospital and of a clinical nature (primary uses or purposes). Therefore, EuCanImage does not involve prospective or *ad hoc* data collection, but re-use of already existing data for secondary purposes (scientific research).

As a result, data providers will be responsible and liable for the legitimate collection of the data following European and national legal requirements (e.g., legal basis, derogations for processing, compatibility of the secondary purposes, additional safeguards of article 89.1, DPIAs, licenses, etc.) and for complying with ethical requirements (REC's approval). Generally, neither EuCanImage nor WP1 will assess compliance with all ethical and legal requirements of the data collection, as this a responsibility of the data providers themselves and their legal and ethical experts and/or Data Protection Officers. However, this does not prevent legal and ethical support and guidance being given to clinical centres throughout this process, and even DTAs and other different agreements and documents have been reviewed and drafted.

Furthermore, an early data minimisation procedure will be carried out in this stage to enhance the privacy by design approach. Clinical partners and AI researchers are actively consulting with WP1 experts to apply data minimisation principle in the definition of clinical requirements and instructions (Task 2.1) with the selection of data strictly necessary. In addition to minimisation, other processing operations, referred to above as 'data curation', will be carried out in this first phase.

Lastly, in this first phase the de-identification process will take place using the POSDA tool and leveraging the TCIA's experience in a GDPR-compliant fashion. Besides the application of POSDA, other technical and organisational measures will be required and put in place to ensure that re-identification is not possible (e.g., disclaimers, no re-identification agreements or commitments, etc.).

Ultimately, data providers will integrate the EuCanImage data flow as (5) independent data controllers. At the same time, one or more data processors (or sub-processors) may be engaged to assist the controllers in performing certain non-essential technical or specialised tasks (e.g., de-identification, conversion, validation, etc.).

ii. Stage 2. Storage, linkage, annotation and EuCanImage catalogue.

The second phase, in contrast to the first one, consist of four substages which may be different depending on whether a federated or centralised system has been chosen:

- **Data storage.**
  - **Centralised.** On the one hand, in the case of centralised management and storage, this process starts with the coordinated submission by data providers of the previously curated and de-identified data to the Euro-BioImaging (imaging data) and EGA (non-imaging data) repositories. Hence, it will be necessary to sign the correspondent Data Processing Agreement (hereinafter, 'DPA') with the entity concerned in each case (Erasmus Medical Centre or Centre for Genomic Regulation). Furthermore, for each repository, specific requirements and procedures imposed by each entity for submitting datasets must be comply with by the data providers.
  - **Federated**. On the other hand, in the case of a federated system, the data will not leave the origin institution and the XNAT instance necessary for the storage and management of the imaging data will be installed and deployed in the data controllers' own infrastructures and servers (locally). Therefore, no data will be either transferred or processed by a data processor but by the controller itself through its authorised staff (authorised persons or users).

- **Data linkage.** Curation tools will assign the same identifier ('EuCanImage ID') to the images as well as to related genomics and phenotypic data. The hash function will be developed outside the data providers. This would result in a double codification carried out by different institutions.

  In terms of data protection, this constitutes a paramount point, as it is relevant to ascertain whether this EuCanImage ID may be accessed in any manner by third parties or users. Furthermore, it must be thoroughly assessed whether the metadata in the forthcoming EuCanImage Catalogue may be considered as personal data due to its potential identifiability by linkage or other likely reasonable means. This issue is currently under discussion and is being very carefully approached to ensure that the metadata in the catalogue does not pose any risk to the data subjects.

- **<u>Data annotation (imaging).</u>**
  - **Centralised.** CMRAD "Collective Minds" platform will be used (SaaS) with the purpose of enabling collaboration between radiologists during annotation exercises within EuCanImage. Accordingly, each of the controllers will have to sign a DPA with Collective Minds Radiology AB, which, in turn, has Amazon Web Services (AWS) as a sub-processor.
  - **Federated.** Imaging annotations will be made by means of local annotation tools and without granting access to third parties other than data controllers' authorised personnel.

- **<u>EuCanImage catalogue.</u>** Data controllers (providers) will share metadata to create a common EuCanImage catalogue managed and hosted by EMC (Euro-BioImaging) for which the corresponding DPA will also be required if such metadata may directly or indirectly relate to a natural person. The EuCanImage catalogue will enable platform users to identify the cancer imaging datasets of interest (by means of a Data Browser or Cohort Browser) and to locate the data source or provider. The data catalogue will be linked to the EGA's repository for phenotype and genomics information. This metadata will comprise:
  - Contact information of the institute and owner of the data collection.
  - General information regarding the setting within which the data collections where collected.
  - Information on the type of imaging data that is available. e.g., imaging modality, sequences, type of imaging processing, type of imaging annotations.

  This catalogue will be open for consultation and access and it will not involve data transfers but only display of relevant metadata.

  Nevertheless, to fully ensure that the data catalogue is aligned, *inter alia*, with the GDPR and relevant ethical requirements, it must be assured that these metadata may not directly or indirectly identify any individual and that the user or third party accessing them commits itself not to re-identify the subjects and not to process the data for any purpose other than the correspondent scientific research.

iii. Stage 3. Access to the EuCanImage catalogue by third parties or users (AI developers, researchers)

The last phase of the data flow involves the access by third parties or users to the catalogue to search datasets that may be interesting for their research. This step will be carried out by means of a Data Browser or Cohort Browser that will provide public metadata, as well as the necessary information and procedures about the protocol to access each restricted dataset.

A phased process is envisioned for user seeking data relevant to their activities, implementing three different access modalities:

- Preliminary access. Users who are still uncertain as to what data, specifically, they need for their activities can access certain public open data and metadata that have been previously anonymised. Such data sets are limited in scope, modalities and size, do not contain identifiers or quasi-identifiers and offer only rough and limited views of the underlying information. Synthetic copies of original data may also be used in this modality.
- Access to pseudonymised data. This process may follow the one above, after clarification of data requirement and identification of a specific data set to be accessed or be the first one when such requirements are already clear to the user. As above, the transaction initiates on the EuCanImage data catalogue relying, in the back-end on the Data Access sub-portal for imaging data, and EGA for genomic data. A direct relationship will be established between the user and the correspondent data controller (clinical site, provider) by means of a controller-controller agreement.
- Federated Learning. A third data access modality will be provided to develop and test AI models and more in general to execute queries on the distributed data infrastructure via the secure multi-party computation (SMPC) system. This modality will be utilised primarily by users with very clear data and analytics requirements and tasks. SMPC systems guarantee the highest level of anonymity thanks to the fact that queries only return analytical results and no records or fields or values from the underlying data. In this view, the type of data protection layer to be implemented will be similar to the one for anonymous data modality.

The different components of EuCanImage infrastructure, including data catalogue, data access sub-portal or AI computational platform will be made available through an integrated portal. This infrastructure will be fully virtualised to be installed at data repositories (central or federated) to minimise data transfers and risks.

3.3 Scenarios for data processing and sharing in EuCanImage

Lastly, a schematic overview of the possible scenarios for data processing and sharing in EuCanImage must be envisaged. The aim is not to examine all the processing operations to be carried out, which has already been undertaken above, but rather to design the data schema in its current initial conception. It is therefore intended that the next deliverables and tasks, both of this Work Package and others, can take this schema and data flow as a reference to understand their legal requirements, mainly in terms of data protection, their role and position.

This is particularly relevant for the forthcoming Deliverable 1.3 ("Ad hoc appointment letters, data sharing agreement, cross-border transfer clauses and data processing specific instructions written for signature or adhesion"), as a first approach to the set of legal texts, contracts, agreements, and other tools to be developed, as well as for the future development of the platform's policy and data governance framework.

Thus, in accordance with what has been anticipated in the preceding lines, two main scenarios of data processing and sharing can be distinguished, which are being or will be further developed and implemented in the course of EuCanImage:

i. **Scenario 1.- Data processing and sharing based on data providers' decisions and instructions**. As mentioned, EuCanImage has adopted a hybrid model and structure, allowing data providers to choose between a centralised or federated architecture and design. Although this does not necessarily pose direct legal or ethical implications, the data schema does vary slightly, and a subdivision will therefore be preferable:

    a. Centralised. The data will not be stored directly on the data providers' servers, instances or facilities but in external consolidated data repositories that act as data processors. The data providers do not lose their legal controller role, as they retain full control over their data, although non-essential operations may be carried out on their behalf.

    b. Federated. In contrast to the centralised approach, here all data management and storage is carried out directly by the data providers, either on their own servers and facilities or through data processors or sub-processors external to EuCanImage. The main implication is that the data flow becomes legally simpler and clearer, but technically more complex and costly for the data providers, as they will have to deploy the EuCanImage infrastructure in their own centres or institutions.

ii. **Scenario 2.- Distributed learning and SMPC.** This model is currently under exploration by EuCanImage members and stakeholders and will require further development in subsequent tasks and deliverables.

3.3.1 Scenario 1.- Data processing and sharing based on data providers' decisions and instructions

i. **Nature and legal status of the data throughout the processing and sharing stages.**

   a. Personal data. In the first stage, the subjects will remain identified or identifiable for the data providers.

   b. Personal data (pseudonymised). In the second stage, the data will undergone a de-identification process. This process will also implement a double coding both by the data provider and by the EuCanImage platform (with the EuCanImage ID).

   c. Non-personal (anonymous or anonymised). In the third stage, the EuCanImage catalogue will only contain metadata of a non-personal nature.

ii. **Legal roles of the participants.** The data providers will always determine the means and purposes of the processing operations, including an eventual data sharing or access to/by third users. As a result, they will be the data controllers within the EuCanImage both in a centralised and in a federated model. However, if in the centralised model the platform determines, to a greater or lesser extent, any essential means or purposes of the processing (e.g., decides about the access or disclosure of data to/by third users) this will imply that the legal entity behind the platform will be also considered a controller (or even joint controller). Lastly, third users accessing the platform or requesting a dataset will be considered as a third party and, eventually, a controller if they access the data for their own scientific research purposes.

iii. **Specific gaps and risks.**

   a. Unlawful data processing by third parties (platform users).
   b. Security and privacy breaches.
   c. Re-identification of the data subjects.
   d. Long-term data storage.
   e. Relevant changes on the data processing context.

iv. **Measures, safeguards, guarantees and other mechanisms envisaged for mitigating the gaps and risks.**

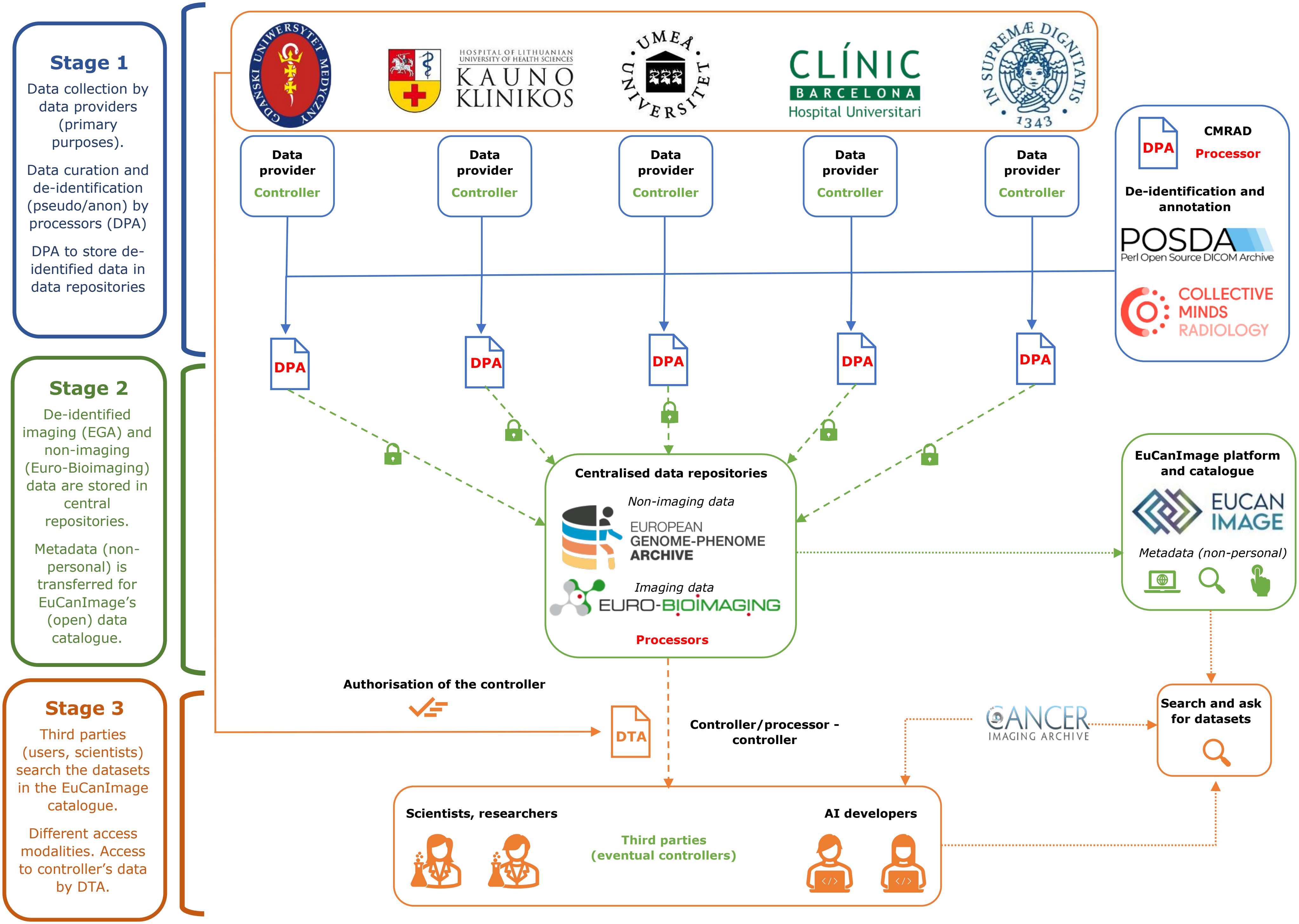
   a. Commitments and other legally binding acts ensuring that the data have been collected in accordance with data providers' national legislation and, in any case, in accordance with the GDPR and ethical requirements.

   b. Privacy by design report for the project and the platform (Deliverable 1.4). Compliance with data protection principles and requirements provided in this Deliverable will be taken into account to ensure that the project will follow a Privacy-by-design approach (T1.3) and will serve as basis to develop EuCanImage Privacy-by-design assessment document (D1.4) (see point 3.4). Technical requirements and specifications foreseen in the project will also be informed accordingly.

c. Technical measures such as the implementation of the POSDA de-identification suite.
d. Double-hash or codification. Data subjects will be double-hashed or coded by both the data providers and by the EuCanImage platform (EuCanImage ID).
e. All the data repositories and other data processors will be located within the European Economic Area.
f. No personal data will be contained within the EuCanImage catalogue. It will only incorporate minimum metadata to search and apply for the available datasets by third users.
g. A contractual, policy and governance framework will be developed and implemented (Task 1.2) to govern the relations between controllers, processors and third parties, as well as their correspondent limitations and liabilities.
h. Third users will have to apply for each dataset and sign the respective legal contract or agreement directly with the data provider (controller) or with the data processor on behalf of the data provider (controller) after the latter's authorisation.
i. Long-term sustainability. The project must ensure the sustainability of results over time, which will require medium to long-term data storage. This implies that EuCanImage platform will remain operative after the end of the project. This issue is already under discussion and legal and ethical implications will be thoroughly considered.
j. Constant monitoring and dynamic approach. Possible changes in the context of the processing operations (e.g., emerging re-identification risks) will be correspondingly monetarised and taken into account. Ultimately, the policy framework will be dynamic, and it will implement further mechanisms and tools for future updates.

Below we provide two schematic representations of the scenario described above with its different stages, participants and maim legal, ethical and contractual requirements, slightly distinguishing between a federated and a centralised model.

Figure 8. Representation of Scenario 1 with a centralised model or architecture



Stage 1

Data collection by data providers (primary purposes).

Data curation and de-identification (pseudo/anon) by processors (DPA)

DPA to store de-identified data in data repositories

Stage 2

De-identified imaging (EGA) and non-imaging (Euro-Bioimaging) data are stored in central repositories.

Metadata (non-personal) is transferred for EuCanImage's (open) data catalogue.

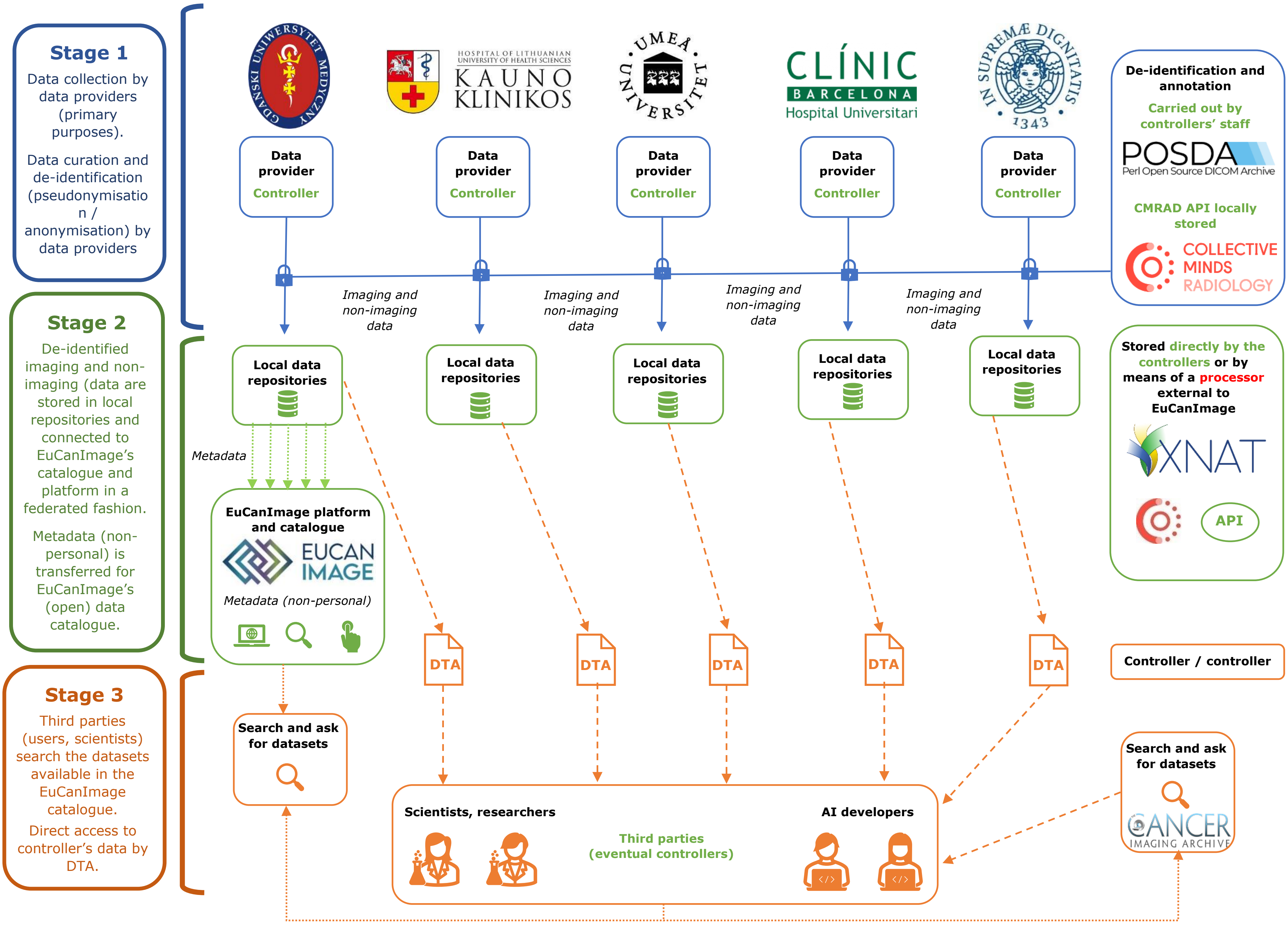
Stage 3

Third parties (users, scientists) search the datasets in the EuCanImage catalogue.

Different access modalities. Access to controller's data by DTA.

Data provider Controller
Data provider Controller
Data provider Controller
Data provider Controller
Data provider Controller
CMRAD
DPA
Processor
De-identification and annotation
POSDA Perl Open Source DICOM Archive
COLLECTIVE MINDS RADIOLOGY
DPA
DPA
DPA
DPA
DPA
Centralised data repositories
Non-imaging data
EUROPEAN GENOME-PHENOME ARCHIVE
Imaging data
EURO-BIOIMAGING
Processors
EuCanImage platform and catalogue
EUCAN IMAGE
Metadata (non-personal)
Authorisation of the controller
DTA
Controller/processor - controller
CANCER IMAGING ARCHIVE
Search and ask for datasets
Scientists, researchers
Third parties (eventual controllers)
AI developers

Figure 9. Representation of Scenario 1 with a federated model or architecture



Stage 1

Data collection by data providers (primary purposes).

Data curation and de-identification (pseudonymisation / anonymisation) by data providers

Stage 2

De-identified imaging and non-imaging (data are stored in local repositories and connected to EuCanImage's catalogue and platform in a federated fashion.

Metadata (non-personal) is transferred for EuCanImage's (open) data catalogue.

Stage 3

Third parties (users, scientists) search the datasets available in the EuCanImage catalogue.

Direct access to controller's data by DTA.

Data provider
Controller
Data provider
Controller
Data provider
Controller
Data provider
Controller
Data provider
Controller
De-identification and annotation
Carried out by controllers' staff
POSDA
Perl Open Source DICOM Archive
CMRAD API locally stored
COLLECTIVE MINDS RADIOLOGY
Imaging and non-imaging data
Imaging and non-imaging data
Imaging and non-imaging data
Imaging and non-imaging data
Local data repositories
Local data repositories
Local data repositories
Local data repositories
Local data repositories
Stored directly by the controllers or by means of a processor external to EuCanImage
XNAT
API
Metadata
EuCanImage platform and catalogue
EUCAN IMAGE
Metadata (non-personal)
DTA
DTA
DTA
DTA
DTA
Controller / controller
Search and ask for datasets
Search and ask for datasets
CANCER IMAGING ARCHIVE
Scientists, researchers
AI developers
Third parties (eventual controllers)

3.3.2 Scenario 2.- Federated learning and SMPC

Multi-party computation systems allow the distribution of encrypted and partitioned queries over a federated set of databases thus allowing third parties to retrieve the results that do not contain identifiable information. The Federated Learning infrastructure developed within EuCanImage will utilise such principle to connect, on a voluntary basis, data providers and data centres in the project.

In this scenario, a data scientist (third party or user) will connect to the central server, located and managed by our partner Universitat de Barcelona and distribute queries to train an AI-model in a cross-silo fashion. The individual nodes in each institution will interact with the AI-model, which will be trained on their data, but the data itself will never be transferred away from the data provider. The nodes involved in the process will request access to the model, which means that the owner of the model, i.e. a data scientist, will not be able to request data, thus offering an extra layer of security and privacy protection for the individuals.

Users (data scientists) will also be allowed to perform qualitative analysis on the data, in two modalities:

i. To aggregate queries that do not share, process, or disclose personal data, but rather return population statistics such as the average or the variance of a given variable in the cohort.

ii. To transform raw data in order to prevent analysis that could potentially leak data concerning an identified or identifiable person, such as scatter plots. To achieve this each individual datum will have noise added to it, so that the corresponding data subject (patient) has plausible deniability of his or her information. This concept is also known as 'differential privacy'.

This scenario in currently under development and exploration by the consortium members. In later stages, privacy-preserving and GPDR-compliant analysis will be further carried out, including possible data analysis automation. Data providers will be exhaustively and transparently informed of these modalities and of the risks related to the disclosure of potentially personal and sensitive information. In addition, as part of the policy and contractual framework developed in Task 1.2., guidelines in the use of this infrastructure and scenario will be further drafted and documented

3.4 Data protection by design and by default and next steps

A core pillar for the protection of the rights and freedoms of individuals with regard to the processing of their personal data is the adoption of technical and organisational measures to ensure that the requirements of the GDPR are met. This is what is commonly referred to as by the Regulation as "accountability". In order to be able to demonstrate compliance with the GDPR, the controller should adopt internal policies and implement measures that meet in particular the principles of data protection by design and by default[32].

- **Data protection by design (Article 25(1) of the GDPR).** "Taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing, the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in a effective manner and to integrate the necessary safeguards into the processing (…)".
- **Data protection by default (Article 25(2) of the GDPR).** "The controller shall implement appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed. That obligation applies to the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility".

Accordingly, when developing and designing applications, services or products which are totally or partially based on the processing of personal data, controllers and processors shall take into account the principles of data protection by design and by default from the minute one. Such measures could consist, *inter alia*, of minimising the processing of personal data, pseudonymisation, encryption, transparency measures, the ability to ensure a continued confidentiality, integrity, availability and resilience of systems and processing operations and a process of regular verification, evaluation and assessment of the effectiveness of the adopted measures.

EuCanImage constitutes a clear example of compliance with these principles, measures and guarantees from the early conception of the project, the design and development of its applications, functionalities, platform and systems and the final implementation of a secure and controlled multi-layered framework for data

[32] Recital 78 of the GDPR.

processing. As such, EuCanImage will implement the following actions, measures, tools, and safeguards to ensure and demonstrate compliance with GDPR principles[33]:

a. **Lawfulness, fairness and transparency.** Each data provider (controller) will ensure and control the fulfilment of this requirements and principles and must rely on an appropriate legal ground and, if necessary, on derogations for the processing of special categories of data. This will include safeguards pursuant Article 89(1) of the GDPR, such as pseudonymisation and local Research Ethics Committees' reviews. The imaging data (Task 3.2), genomics and phenotypic data (Task 3.3) integrated with the catalogue will enforce data access procedures, in accordance with GDPR principles via business rules encoded in the platform permissioning layer (Task 3.5). Through the catalogue, each data provider (controller) will ensure the right of transparency and information stated on Articles 13 and 14 of the GDPR. This will be also appropriately implemented as binding clauses in the correspondent DTAs and DPAs and in the platform policy documents (Task 1.2).

b. **Purpose limitation.** Data processed within EuCanImage will be used solely and exclusively for scientific research purposes. No further uses or purposes will be allowed without explicit consent as indicated in the DTAs, DPAs, disclaimers and other documents.

c. **Data minimisation.** Clinical and AI researchers are working together with WP1 team to apply this principle in the definition of clinical requirements (Task 2.1) with the selection of data only strictly necessary for the final purposes of AI and clinical research. Information will also be minimised on the data catalogue, which will be iteratively reviewed during the design phase (Task 3.4). This is intended to implement data protection by design and by default principles, ensuring that only strictly necessary data will be processed at the different stages of the data flow.

d. **Pseudonymisation.** Potential for direct or indirect linkage or re-identification will be critically assessed and addressed at multiple stages. As mentioned in section 2.3, data providers have, prior to the very commencement of the project, carried out different pseudonymisation procedures. In addition, to ensure consistency and standardisation in the processes and techniques applied, a new de-identification process will be carried out using TCIA's POSDA tool, before any data set is transferred or shared with any EuCanImage partner. Metadata made public through the EuCanImage catalogue must have been previously analysed to assess any risk of re-identification. Furthermore, measures and guarantees to avoid re-identification will be implemented e.g. disclaimers, non-reidentification commitments, etc.

e. **Accuracy.** This principle will be ensured and enforced with the extensive data annotation and curation efforts in Task 4.3 using the CMRAD infrastructure

[33] This is merely a preliminary *numerus apertus* list. The whole details and measures finally adopted will be developed (Task 1.3) in the corresponding report (Deliverable 1.4).

and platform. Data curation procedures and tools will check for consistency, integrity and accuracy in imaging and non-imaging data across sites.

f. **A Chief Data Officer (hereinafter, 'CDO') has been appointed in the person of the WP1 leader.** The Data Protection Officers (hereinafter, 'DPO') of all participants in data processing operations are identified and the WP1 team is in close liaison and communication with them. Therefore, EuCanImage CDO will assume the role of coordinating project-level decisions.

The implementation of aforementioned principles, measures and guarantees is being carried out in a close liaison and collaboration with the partners and entities in charge of each stage of development and design of EuCanImage. Moreover, with regard to WP1's specific work, three levels or steps of implementation are foreseen: i) developing the policy framework for EuCanImage (Tasks 1.1 & 1.2); ii) creating Data Transfer and Data Processing agreements to govern data transactions; and iii) translate legal and ethical requirements into technical ones and then implement these in EuCanImage models in a privacy by design fashion (Task 1.4).

4 Legal interoperability study between jurisdictions

As mentioned in the preceding sections, at this stage, all data controllers and data processors will be located inside the European Union and therefore no data will be transferred to third countries or international organisations. Consequently, it should be stressed that no international transfers or international access to European patients' or data subjects' data is foreseen.

Nevertheless, the opposite scenario is envisaged with the integration of the US-based TCIA, i.e., data of US citizens may be transferred or access to or from the EU. Furthermore, declining the eventual possibility of scaling up the EuCanImage model and platform in the future to an international level would be a major pitfall. Therefore, this section aims to provide a preliminary legal analysis outlining the applicable framework for cross-border or international data transfers and/or accesses.

On the one hand, accessing and processing data at a cross-border level within the European Union itself entails certain challenges and issues. Despite the advent of the GDPR, processing operations involving subjects and entities from several EU Member States must deal with slight differences between the national regulatory frameworks, or even with regional or sectoral fragmentation within the same Member State. Here, EuCanImage is not an exception as our consortium involves hospitals, institutions, universities, research centres and other entities from different member States such as The Netherlands, Italy, Spain, Lithuania, Poland, Sweden, Belgium, Germany and Austria, among others.

On the other hand, it is undeniable that scientific research and, in particular, cancer research, is closely connected to an international level and community. Indeed, this fact has been emphasised by a multitude of international institutions and organisations (e.g. the UNESCO[34]) and even by the European GDPR itself[35]. Thus, it cannot be overlooked that the future scalability of EuCanImage remains an open issue, although it is not currently envisaged either in this report or the data flow designed for the duration of the H2020's EuCanImage project.

Both cases will be further explored in the following lines in order to assess the state-of-the-art currently relevant for EuCanImage and for an eventual future scalability of the platform, the system and its tools.

[34] United Nations Educational, Scientific and Cultural Organization, UNESCO (2017). *Recommendation on Science and Scientific Researchers*. Records of the General Conference, 39th session, Paris, 30 October-14 November 2017. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000260889.page=116

[35] See Recitals 6 and 101 of the GDPR, among others.

4.1 Legal interoperability for the implementation of EuCanImage models between and within EU Member States: issues related to cross-border processing operations

Under the GDPR, cross-border processing is defined as (a) the processing of personal data which takes place in the context of the activities of establishments in more than one Member State of a controller or processor in the Union where the controller or processor is established in more than one Member State or (b) the processing of personal data which takes place in the context of the activities of a single establishment of a controller or processor in the Union but which substantially affects or is likely to substantially affect data subjects in more than one Member State[36].

The presence and the exponential growth of cross-border data processing operations are inextricably tied to the very existence of the European Single Market, and it would therefore seem pointless for the European legislator to limit or hinder its flourishing. In fact, the adoption of the GDPR already anticipated the intention of the European legislator to harmonise and establish a common data protection framework applicable to all EU Member States[37].

In the same vein, it does not seem at all that the GDPR aims to limit data flows or operations between several Member States for scientific or health research purposes. On the contrary, the GDPR's regime for scientific research purposes holds a prominent position and benefits from softened requirements and rules in many parts of the text[38].

In light of the above, accessing, sharing and processing personal data between or within several EU Member States should not pose any further issues or discussions than the application of the common precepts of the GDPR and the corresponding agreements or contracts in each case, depending on the role or position of the entities or institutions receiving or transmitting the data. However, these harmonisation efforts and ambitions of the GDPR have only been partially achieved concerning scientific research, as some issues remain unresolved[39], and others require or allow

[36] Article 4(23) of the GDPR.

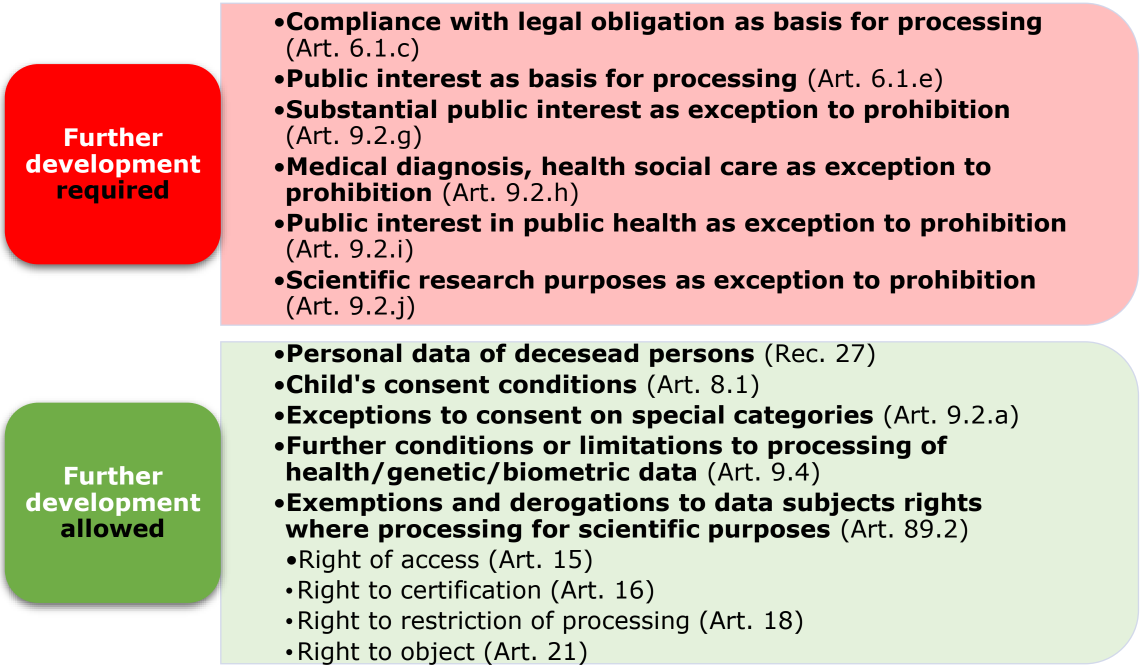
[37] In fact, Article 1(3) of the GDPR literally states, "the free movement of personal data within the Union shall be neither restricted nor prohibited for reasons connected with the protection of natural persons with regard to the processing of personal data". This approach also becomes evident in Recitals 5 to 8, among other provisions.

[38] See, *inter alia*: purpose limitation exemption (Article 5(1)(b) *in fine*), storage limitation exemption (Article 5(1)(e)), processing of special categories of data (Article 9(2)(j)), exemptions to the provision of information to the data subject (Article 14(5)(b), exceptions to the exercise of certain rights of the data subjects such as the right to be forgotten (Article 17(3)(d) or the right to object (Article 21(6)), etc.

[39] These include, among others, the anonymisation of genetic data, safeguards relating to processing for scientific research purposes, the effect of Member States' law with possible

further development by EU Member States' law. Such differences arise to a greater extent with regard to scientific research with genetic data and data concerning health (e.g., Articles 9(2)(j), 9(4), 89(2) of the GDPR, among others). For instance, processing of special categories of data pursuant Article 9(2)(j) necessary for scientific research purposes may be differently implemented and developed by the Member States, requiring varying safeguards and measures. Moreover, in some Member States, secondary uses of personal data for scientific research purposes are subject to an authorisation by a data permit authority[40].

Thus, Member States' law becomes a core element within the processing of personal data for scientific research purposes, as the GDPR repeatedly refers to Member States' law or allows them to introduce further additional provisions or conditions[41]. The following is a schematic representation outlining all aspects and areas relevant to EuCanImage aims and purposes where the GDPR either allows or requires Member States to introduce further provisions through national law:



Further development required

- **Compliance with legal obligation as basis for processing** (Art. 6.1.c)
- **Public interest as basis for processing** (Art. 6.1.e)
- **Substantial public interest as exception to prohibition** (Art. 9.2.g)
- **Medical diagnosis, health social care as exception to prohibition** (Art. 9.2.h)
- **Public interest in public health as exception to prohibition** (Art. 9.2.i)
- **Scientific research purposes as exception to prohibition** (Art. 9.2.j)

Further development allowed

- **Personal data of decesead persons** (Rec. 27)
- **Child's consent conditions** (Art. 8.1)
- **Exceptions to consent on special categories** (Art. 9.2.a)
- **Further conditions or limitations to processing of health/genetic/biometric data** (Art. 9.4)
- **Exemptions and derogations to data subjects rights where processing for scientific purposes** (Art. 89.2)
  - Right of access (Art. 15)
  - Right to certification (Art. 16)
  - Right to restriction of processing (Art. 18)
  - Right to object (Art. 21)

Figure 10. Provisions in the GDPR that require/allow further development through Member States' law.

restrictions or limitations to certain rights of data subjects or on the conditions for the processing of genetic data and data concerning health (Article 9(4) of the GDPR).

[40] An example of this divergences may be Finland, which by legal mandate of the Act on Secondary Use of Health and Social Data (552/2019) has created the public entity "Findata" (Social and Health Data Permit Authority). However, many other Member States do not even legally foresee the establishment of this type of authorities.

[41] The most glaring paradigm of this issue is the Article 9(4) of the GDPR, which literally states: "member States may maintain or introduce further conditions, including limitations, with regard to the processing of genetic data, biometric data or data concerning health."

As a result, EuCanImage partners and legal teams will have to cope with a complex regulatory framework that comprises not only the GDPR but also relevant Member States' rules, in addition to traditional divergent ethical requirements and procedures of the involved hospitals, centres and institutions. The existence of relevant collections and datasets in countries such as Spain, Italy, The Netherlands, Sweden, Poland, or Lithuania implies that each controller must bear in mind their national or internal law when it comes to the provisions where Member States can implement further exceptions and conditions to those of the GDPR.

The general and specific requirements to be met by each entity, institution or subject have already been detailed in the precedent sections of this report. In any case, it is not pertinent to elaborate now on all the contracts, agreements, measures, guarantees and safeguards to be adopted as this is the object of forthcoming Deliverables (1.3 and 1.4) and Tasks (1.2 and 1.3).

In addition, it is not intended to delve into those divergences since, apart from being beyond the scope of this report; the Commission has recently published a thorough report that echoes all the differences between Member States on scientific research and health and genetic data[42]. Nevertheless, the following summarises the applicable regulatory and legal framework per Member State that may be relevant for the development of EuCanImage:

Spain	Italy	Netherlands	Sweden	Poland	Lithuania
Organic Law 3/2018 on Protection of Personal Data and the Guarantee of Digital Rights	Legislative Decree 101/2018 that amends the "Data Protection Code" and implements the GDPR	GDPR Implementation Act of 16 May 2018.	Data Protection Act (2018:218) & Swedish Data Protection Regulation (2018:219)	Personal Data Protection Act of 10 May 2018 & Act on Amendments to Sectorial Acts of 21 February 2019	Law on Legal Protection of Personal Data
Law 14/2007 on Biomedical Research	Law 29/2019 on National Network of Cancer Registers and Surveillance Systems.	Law of 26 February 1998 on medical-scientific research with humans.	Act 543/1998 on Health Data Registers	Act on Higher Education (regulating data processing for scientific purposes)	Law on Ethics of Biomedical Research
Law 41/2002 on patient autonomy and rights		Law of 20 November 2003 on Statistics	Act 351/2006 on Genetic Integrity	Act on Patient Rights and Patient Ombudsman	

Figure 11. *Comparison between legal frameworks applicable to each Member State relevant to EuCanImage.*

[42] European Commission. (2021). *Assessment of the EU Member States' rules on health data in the light of GDPR*. Available at: https://ec.europa.eu/health/sites/health/files/ehealth/docs/ms_rules_health-data_en.pdf

Conclusively, the need to integrate requirements and further conditions stated on Member States' national laws implies that each EuCanImage data provider (controller) must have collected and processed the data in accordance not only with the GDPR, but also with the applicable national or internal regulations (i.e., state law, regional, federal, provincial, or local). Reconciling and balancing these legal differences at a supra-state (EU) level may be complex, in many cases, not even possible, as the rules govern, and regulate some areas in contradictory terms.

4.2 Towards a future scalability of EuCanImage: issues related to data transfers to third countries and international organisations

As anticipated, EuCanImage does not currently envisage data processing operations concerning European patients or data subjects with institutions or entities outside the European Economic Area (hereinafter, 'EEA'). In fact, all the involved parties in the data flow *i.e.*, controllers, processors or sub-processors are located within the EU, apart from the TCIA, although it will neither access nor process data concerning European citizens, an issue that will in any case be further discussed below.

Nonetheless, excluding potential future collaborations between internationally acknowledged institutions such as the TCIA would represent aprioristic hindrances to scientific progress and technological development in our area. Thus, data flows to and from countries outside the Union and international organisations are paramount for the international cooperation[43].

The main issue lies in the fact that the GDPR introduces a particular framework applicable to what is referred to as "transfers of data from third countries and international organisations" that is restrictive in order to ensure the rights of the data subjects and that is being strengthened by the latest case law stemming from the CJEU. As will be noted, the latter has been particularly controversial regarding the transfers of data to the US, to the extent that, in the current scenario, it has been decided not to carry out any of them due to the legal implications that may barely be reconcilable between both jurisdictions.

[43] This is the literal wording of recital 101 of the GDPR.

4.2.1 The rules for the transfers of data to third countries and international organisations in the GDPR

The GDPR contains, as mentioned above, a particular framework that applies to transfers of personal data to third countries and/or international organisations. This regime is laid down pursuant Articles 44 to 50 of the Regulation and it applies in a general fashion, *i.e.* it does not contain specific provisions on scientific research or for the processing of genetic data or data concerning health[44].

However, what the GDPR does not expressly provide is a definition of this concept, not even in Article 4. Therefore, it may be defined as the processing operation whereby a controller or a processor within the EEA ('data exporter') transfers or gives access to personal data to a controller or processor outside the EEA ('data importer') or to an international organisation. This means that the access or transfer of data concerning European individuals to a third country (such as the US) may be considered, for the purposes of the Regulation, a transfer of data to a third country or simply, an 'international data transfer'.

The point is that the GDPR prohibits international transfers as a rule, as it stipulates that they shall only take place if, subject to other provisions of the Regulation, the conditions laid down pursuant Chapter V are complied with by the controller and processor, including for onward transfers[45]. In other words, prior to carrying out any international transfer, the importer and the exporter must comply with the provisions of the GDPR (*e.g.,* principles, legal basis, processing of special categories of data, DPOs, DPIAs, etc.) and, in addition, they must meet the specific provisions of Chapter V *i.e.* the need to rely on one or more paths or tools such as adequacy decisions, appropriate safeguards or derogations.

On top of the above, data importers and exporters will not only have to demonstrate the compliance with the general provisions of GDPR and the specific ones contained in Chapter V, but also ensure that any processors or controllers to whom the importer will transfer the data also comply with them[46]. The overall purpose of this regime is to ensure that the level of protection of natural persons guaranteed by the GDPR and the European laws is not undermined when their data is being transferred to other jurisdictions. This concept of "adequate level of protection" is not straightforward and has been widely discussed throughout the case law of the CJEU. The word "adequate" does not mean that the third country is required to ensure a level of protection "identical" to that guaranteed in the EU. It must be understood as requiring the third country in fact to ensure, by reason of its domestic law or its international commitments, a level of protection of fundamental

[44] Notwithstanding the reference to EU or Member States' legislation in Article 49(4).

[45] Article 44 of the GDPR "General principle for transfers".

[46] Article 29 Data Protection Working Party, WP29. (2017). *Working Document on Adequacy Referential. Adopted on 28 November 2017*, WP254. Endorsed by the EDPB on 25 May 2018. P. 6·

rights and freedoms that is essentially equivalent to that guaranteed in the EU[47], even though the means to which that third country has recourse may differ from those employed within the EU[48].

Nevertheless, meeting the threshold set by the CJEU is by no means simple, and even actually can be concluded, as will be noted below, that third countries such the US do not ensure an adequate level of protection and may therefore require the adoption of supplementary measures by the controller in order to ensure compliance with that level of protection[49].

4.2.2 Transferring data to third countries and international organisations: adequacy decisions, appropriate safeguards and derogations for specific situations

As mentioned, to carry out an international transfer of data it is crucial to rely on, at least, one of the mechanisms or pathways stated on Chapter V of the GDPR. These provisions classify such tools in three categories that may be relied upon to carry out transfers of personal data according to the following order of priority:

i. **Transfers on the basis of an adequacy decision (Article 45 of the GDPR**). As a first step, the transfer of data may be carried out where it is covered by an adequacy decision. An "adequacy decision" is an implementing act by the European Commission[50] determining that a third country, a territory or one or more specified sectors within them ensures an adequate level of protection[51].

   The benefits of relying on adequacy decisions for carrying out transfers of data are obvious. The adequacy decision will have EU-wide effect[52] and no specific authorisation will be required[53]. Nevertheless, the case law of the CJEU has questioned its reliability in ensuring the rights and freedoms of European citizens and has led to significant legal and practical uncertainty. In fact, as will be seen in the next section, there is currently no adequacy decision in force legitimising transfers between the EU and the US as a result of the annulment first, of the Safe Harbour decision, and more recently, of the Privacy Shield, both by the CJEU.

[47] CJEU Judgment of 6 October 2015, *Maximillian Schrems v Data Protection Commissioner* (C-362/14, *Schrems I*). Parag. 73.

[48] Ibid. parag. 74.

[49] Ibid. parag. 133.

[50] Article 45(3) of the GDPR.

[51] Article 45(1) of the GDPR.

[52] Recital 103 of the GDPR.

[53] Article 45(1) in fine of the GDPR.

ii. **Transfers subject to appropriate safeguards (Article 46 of the GDPR).** In the absence of a valid adequacy decision data may be transferred to a third country or an international organisation if the controller or processor provides appropriate safeguards, and on condition that enforceable rights and effective legal remedies for data subjects are available. In other words, if there is no adequacy decision, the controller or processor may rely on one or more of the tools or mechanisms set out in Article 46(2) and (3) to provide what is referred to as "appropriate safeguards".

In this case, there is no need to observe any particular order and thus it may be feasible to rely on one or more tools pursuant Article 46 according to the context of the transfer, the risks or the purpose of the processing, among others. Nevertheless, the provision lays down two sets of appropriate safeguards according to whether or not they require any specific authorisation from a supervisory authority:

- **Appropriate safeguards that do not require specific authorisation from a supervisory authority (Article 46(2) of the GDPR).**
    - Legally binding and enforceable instruments between public authorities or bodies.
    - Binding Corporate Rules (BCRs).
    - Standard Contractual Clauses (SCCs).
    - Codes of conduct.
    - Certification mechanisms.
- **Appropriate safeguards that require authorisation from the competent supervisory authority (Article 46(3) of the GDPR).**
    - Contractual clauses authorised by a supervisory authority.
    - Provisions to be inserted into administrative arrangements between public authorities or bodies which include enforceable and effective data subject rights.

Although it may seem that many pathways are available, in fact, only few of them may be valid and suitable for carrying out transfers of data in the context of EuCanImage *i.e.,* Standard Contractual Clauses, contractual clauses authorised by a supervisory authority, codes of conduct and certification mechanisms. Moreover, the latter two are not yet developed and therefore underexploited. Definitively, this restricts the available pathways and furthermore taking into account the limitations and requirements set by the CJEU in Schrems II with regard to the SCCs, which shall also be extended to the rest of Article 46 tools and mechanisms.

iii. **Derogations for specific situations (Article 49 of the GDPR).** Finally, where the transfer is covered neither by an adequacy decision, nor by an appropriate safeguard mechanism, it shall only be carried out if it is covered by any of the exceptional derogations or situations set out in Article 49 of the GDPR. However, derogations under Article 49 "are exemptions from the general principle and due to this fact, they must be interpreted restrictively so that the exception does not become the rule[54]. Moreover, amongst all the envisaged derogations, only one may be feasible for transferring data to third countries in the context of EuCanImage. This would be the explicit context of the data subject laid down in Article 49(1)(a)[55].

The following section contains a brief study of the possible tools and pathways for international transfers of data in the context of EuCanImage, with special emphasis on the US and the arising controversies and challenges.

4.2.3 Tackling post-Schrems II international data flows to the United States: an interoperability gap for EuCanImage

Whereas the actual context of data transfers between the EU and third countries constitutes a critical point, the situation worsens when the importer is in the US. In fact, it has been the dispute about transatlantic data flows that has given rise root to the vast case law of the CJEU culminating, *inter alia*, in the annulment of two adequacy decisions in force with the US.

The first decision, under the name of "Safe Harbour"[56], was adopted while the former Data Protection Directive was still in force and declared that the US guaranteed an adequate level of protection. In this regard, it should be recalled that both the former Directive and the GDPR establish that transfers of data must be carried out only where a third country or international organisation ensure an "adequate level of protection". This must be understood as requiring the third country in fact to ensure, by reason of its domestic law or its international commitments, a level of protection of fundamental rights and freedoms that is essentially equivalent to that guaranteed in the EU[57], even though the means to which that third country has recourse may differ from those employed within the EU[58]. In order to ensure that level of protection, the Safe Harbour established a self-certification system whereby

[54] "the protection of the fundamental right to respect for private life at EU level requires that derogations from and limitations on the protection of personal data should apply only in so far as it is strictly necessary" See CJEU case law, among others, in: Judgment 18 December 2008, *Satakunnan Markkinapörssi v. Satamedia* (C 73/07, *Satamedia*), parag. 56; Judgment 8 April 2014, *Digital Rights Ireland Ltd v Minister for Communications* (Joined Cases C-293/12 and C-594/12 , D*igital Rights*) parag. 52.

[55] When "the data subject has explicitly consented to the proposed transfer, after having been informed of the possible risks of such transfer due to the absence of an adequacy decision and appropriate safeguards".

[56] 2000/520/EC: Commission Decision of 26 July 2000 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequacy of the protection provided by the safe harbour privacy principles and related frequently asked questions issued by the US Department of Commerce.

[57] CJEU Judgment of 6 October 2015, *Maximillian Schrems v Data Protection Commissioner* (C-362/14, *Schrems I*). Parag. 73. *OJ L 215, 25.8.2000, p. 7–47.*

[58] Ibid. parag. 74.

companies wishing to join voluntarily, under the control of the Federal Trade Commission, had to comply with a series of principles such as information, access and adequacy of onward transfers. However, the Decision also contained an important general derogation allowing limitations of Safe Harbour principles "to the extent necessary to meet national security, public interest, or law enforcement requirements".

This system was disrupted by the advent to the CJEU of the case referred to as "Schrems I"[59] which, inter alia, questioned the adequacy of international data transfers to the US on the grounds that the US did not ensure an adequate protection of European citizens' rights against surveillance, security, and intelligence activities from US national agencies. The CJEU argued that the general derogation allowing interferences, founded on national security and public interest, with the fundamental rights of persons whose personal data is transferred to the US, conflicts and proves incompatible with the European data protection legislation[60]. In particular, legislation permitting the public authorities to have access on a generalised basis to the content of electronic communications compromises the essence of the fundamental right to respect for private life[61]. Lastly, legislation not providing for any possibility for an individual to pursue legal remedies in order to have access to personal data relating to him, or to obtain the rectification or erasure of such data, does not respect the essence of the fundamental right to effective judicial protection[62].

Consequently, the CJEU concluded that the adequacy decision in force between the EU and the US *i.e.*, the Safe Harbour system was invalid and that US security laws directly contravened fundamental rights and data protection principles enshrined in European treaties and laws[63]. For the interoperability study carried out in this section, what is relevant here is that this first pronouncement of the CJEU strongly questioned the US national security laws, thus posing a structural problem of incompatibility of both 'privacy' and 'data protection' models and legal frameworks that is currently hardly reconcilable.

Following the annulment of the Safe Harbour decision, both parties negotiated a new framework for international data transfers that crystallised in the Decision of 12 July 2016 referred to as "Privacy Shield"[64]. As a result of the urgency with which the new agreement was negotiated, it was already obvious that the essential or structural changes highlighted by the CJEU had not been implemented and were not

[59] CJEU Judgment of 6 October 2015, *Maximillian Schrems v Data Protection Commissioner* (C-362/14, *Schrems I*).

[60] Ibid. parag. 87.

[61] Ibid. parag. 94.

[62] Ibid. parag. 95.

[63] Ibid. parag. 98.

[64] Commission Implementing Decision (EU) 2016/1250 of 12 July 2016 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequacy of the protection provided by the EU-U.S. Privacy Shield. C/2016/4176.

intended to be made in the medium or long term. Despite this, the new decision added slight amendments such as more detailed provisions on access and use of personal data by US public authorities[65] and the establishment of a new Ombudsperson Mechanism to guarantee "independent oversight and individual redress"[66].

However, US laws authorising mass surveillance programs on personal data relating European individuals remained in force (i.e., FISA Section 702, Executive Order 12333). Therefore, "Schrems II" case was promptly lodged to the CJEU, leading to a foreseeable new annulment of the existing US-EU adequacy decision, the Privacy Shield[67]. The reasons for overturning the new decision were similar to those given in Schrems I:

- **US law allows public authorities broad and disproportionate access to and processing of European citizens' personal data for national security purposes.** Surveillance programs based on Section 702 of the FISA or on the Executive Order 12333 do not establish any limitations to the powers of national security agencies and authorities[68] nor confer rights that are enforceable against US authorities in the courts[69] . Therefore, US law cannot ensure a level of protection essentially equivalent to that guaranteed in the EU.
- **European individuals whose personal data have been or will be transferred to the US are not guaranteed effective judicial protection.** The Ombudsperson Mechanism introduced by the US authorities does not constitute an independent body insofar as it is appointed by the Secretary of State[70] and has no powers to adopt binding decisions on intelligence services or programmes[71].

Notwithstanding the annulment of the decision, the (critical) point stemming from Schrems II for the interoperability and US-EU relations was that the judgment also examined the mechanisms for providing adequate safeguards pursuant Article 46 of the GDPR and, in particular, the Standard Contractual Clauses (hereinafter, 'SCC').

Consequently, EuCanImage is today placed in a critical context for interoperability purposes between jurisdictions and, specifically, in terms of international transfers of data to the US. There is no valid adequacy decision in force thus any transfer will have to be carried out on the basis of appropriate safeguards

[65] Ibid. parag. 64 et seq.

[66] Ibid. parag. 117 et seq.

[67] CJEU Judgment of 16 July 2020, *Data Protection Commissioner v Facebook Ireland Ltd, Maximillian Schrems*, (C-311/18, *Schrems II*).

[68] Ibid. 179.

[69] Ibid. 182.

[70] Ibid. 194 and 196.

[71] Ibid. 196.

pursuant Article 46 et. seq. of the GDPR and, ultimately, on the basis of derogations envisaged in Article 49 of the Regulation.

With regard to the instruments and mechanisms laid down in Article 46, only a few would be feasible in practice for scientific research purposes *i.e.,* Codes of Conduct, certification mechanisms and/or SCCs. The first two tools have not yet been developed nor implemented for international transfers and it is not realistic to promote them at project level due to the time and costs involved. Therefore, *a priori*, the most suitable option for EuCanImage will be the adoption of SCCs.

The SCCs are contractual arrangements or stipulations that include clauses relating to data protection, that regulate the relations between the data importer and the data exporter and that envisage technical and organisational measures, and mechanisms to ensure the rights and freedoms of data subjects, among other provisions. Although often referred to simply as SCCs, under the umbrella of Article 46(2) of the GDPR there are two different types of standard data protection clauses: **i)** those adopted by the Commission that do not require specific approval[72] (Article 46(2)(c) of the GDPR) and **ii)** those adopted by a supervisory authority and approved by the Commission (Article 46(2)(d) of the GDPR).

It could therefore be concluded that EuCanImage could rely on SCCs to carry out any future transfer of data to the US. Nevertheless, as mentioned above, Schrems II judgement had far-reaching consequences as the SCCs are held to the same threshold of "essential equivalence" as adequacy decisions. In other words, data subjects whose personal data are transferred to the US or other third country pursuant to SCCs should be afforded a level of protection essentially equivalent to that guaranteed within the EU[73]. In so far as those standard data protection clauses cannot, having regard to their very nature, provide guarantees beyond a contractual obligation to ensure compliance with the level of protection required under EU law, they may require, depending on the prevailing position in a particular third country, the adoption of supplementary measures by the controller in order to ensure compliance with that level of protection[74] .

This is exactly where the main issue arises for the adoption of SCCs within EuCanImage with institutions or entities located in the US. At the present time, the US and EU laws concerning privacy and data protection of individuals, respectively, constitute different models that are not interoperable, and which will not be interoperable in the medium nor long term. In fact, the European Parliament has

[72] Recently amended and adapted to the requirements of Schrems II by means of the Commission Implementing Decision (EU) 2021/914 of 4 June 2021 on standard contractual clauses for the transfer of personal data to third countries pursuant to Regulation (EU) 2016/679 of the European Parliament and of the Council. C/2021/3972. OJ L 199, 7.6.2021, p. 31–61.

[73] CJEU Judgment of 16 July 2020, (C-311/18, *Schrems II*). Parag. 96.

[74] Ibid. parag. 133.

recently conducted a study entitled "Exchanges of Personal Data After the Schrems II Judgment"[75] which analyses the present and future compatibility between both jurisdictions for carrying out international transfers and its conclusions have not been entirely optimistic:

- **No US federal or state privacy law is likely to provide "essentially equivalent" protection compared to the European GDPR in the present time nor in the foreseeable future**[76]. In fact, "there are serious and in practice insurmountable US constitutional and institutional as well and practical/political obstacles to the adoption of such laws"[77].
- **The FTC does not constitute an effective supervisory authority in the lines of the EU authorities.** The FTC Act would have to be expanded or amended to give it the power to seek penalties for violations of GDPR requirements and to cooperate with EU authorities and the EDPB[78].
- **The US Congress should significantly strengthen the rights of action of individuals, including non-US persons[79].**
- **The US should be urged to reform its federal surveillance legislation.** Unless such reform is carried out, no new adequacy decision with US can be approved without further annulment[80].

All of the above simply reaffirms that, according to the current state-of-the-art, the US does not ensure an adequate level of protection and that the adoption of SCCs by EuCanImage to legitimise transfers to that jurisdiction will not be sufficient to satisfy the threshold set by the CJEU in Schrems II. Consequently, data exporters and importers are obliged to adopt supplementary measures to those offered by the clauses or, if the data exporter established in the EU is not able to take these appropriate supplementary measures, suspend or end the transfer of data to the third country concerned[81]. However, neither the GDPR nor the CJEU specified what is referred to as "additional safeguards" or "supplementary measures". The only

[75] European Parliament. (2021). *Exchanges of Personal Data After the Schrems II Judgement*. Policy Department for Citizens Rights and Constitutional Affairs. Available at: https://www.europarl.europa.eu/RegData/etudes/STUD/2021/694678/IPOL_STU(2021)694678_EN.pdf

[76] Ibid. page 9.

[77] *Idem*.

[78] Ibid. page 9 and 10.

[79] Ibid. page 10.

[80] Ibid. page 11.

[81] CJEU Judgment of 16 July 2020, (C-311/18, *Schrems II*). Parag. 134 and 135.

document is the Recommendation issued by the EDPB on measures that supplement transfer tools to ensure compliance with the EU level of protection of personal data[82].

Accordingly, the text contains a comprehensive (and complex) procedure to be applied in the case of international transfers of data to a country that does not ensure an adequate level of protection and examples of supplementary measures to be adopted that may have a contractual, technical, or organisational nature (e.g., anonymisation, encryption, contracts, agreements and other binding instruments with penalties for non-compliance, etc.).

In short, any transfer of personal data within EuCanImage to a third country such as the US, where there is neither an adequacy decision in force nor an essentially equivalent level of protection, must be carried out via the adoption of SCCs supplemented by additional measures such as, *inter alia*, anonymisation, encryption or agreements and other binding instruments for the parties.

[82] European Data Protection Board (2021). *Recommendations 01/2020 on measures that supplement transfer tools to ensure compliance with EU level of protection of personal data*. Adopted on 18 June 2021.

5 Final considerations

I. This report comprises the Deliverable 1.1 entitled "Legal and ethical interoperability study and retrospective analysis", as an essential outcome from Task 1.1. "Overall assessment of legal and ethical constraints for data sharing in oncology imaging". Therefore, against the background of the objectives in WP1, this deliverable represents the first stage in the implementation of a comprehensive ethical and legal framework for the development of an AI-supported decision tools in oncologic imaging.

By means of this document the requirements and efforts pursuant Task 1.1 have been successfully completed by the consortium and the Work Package 1 team. It has been carried out a thorough retrospective analysis and a preliminary general assessment of several issues such as the legislation in force in Europe as a whole (leveraging ongoing experiences in other projects such as EUCANCan, EUCAN-Connect or euCanSHare) in terms of data protection and anonymisation; roles and responsibilities of EuCanImage's users and stakeholders, legal differences and requirements within cancer data types (imaging and non-imaging data) as well as a comparative view across relevant jurisdictions (e.g. the European Union and the United States) for interoperability purposes.

II. Furthermore, this deliverable is closely connected to other tasks and deliverables. In fact, it constitutes an essential building block for the forthcoming work in Work Package 1 *i.e.,* Task 1.2 ("Policy and contractual framework for governing transactions of cancer imaging data") and its correspondent Deliverable 1.3 ("Ad hoc appointment letters, data sharing agreement, cross-border transfer clauses and data processing specific instructions written for signature or adhesion"), as it accurately anticipates most of the aspects concerning EuCanImage's policy, governance and contractual frameworks. In addition, this report contains many aspects regarding the privacy-by-design approach, to be developed pursuant Task 1.3 ("Privacy-by-design review and requirements analysis") and its Deliverable 1.4 ("Privacy-by-design assessment document"). Lastly, next steps will also include a study of societal resp. socio-ethical aspects from the perspective of ethics of AI and general medical ethics, as well as stakeholders and experts as part of Task 1.4 ("Ethical and social implications of AI-based cancer imaging solutions"). Beside the legal framework, considerations regarding systematic biases in training of AI as well as medical black boxes are vital for the development and utilisation of trustworthy AI. Altogether, the findings will build the basis for the formulation of guidelines by means of Deliverable 1.5 ("Guidelines of ethics and social implications of AI in oncologic imaging").

From the initial conception of the project, a multidisciplinary teamwork and continuous communication has been established between all the members of EuCanImage and the different Work Packages, by means of, *inter alia*, a variety of internal Working Groups. This reflects the impact that other tasks and deliverables from other Work Packages had on this document. In the same vein, Task 1.1 and

Deliverable 1.1 are relevant for other tasks such as Task 3.1 ("Data Management Plan"), Task 3.2 ("Imaging data repository"), Task 3.3 ("Linkage of imaging with omics and phenotypic data types"), Task 3.4 ("Implementation of EuCanImage's catalogue"), Task 3.5 ("Data access management") and Task 4.1 ("Suite for GDPR-compliant data anonymisation and transfer"), among others.

III. EuCanImage is being developed in a rapidly evolving context, overlapping with important projects and initiatives such as the European Health Data Space (EHDS). This poses several uncertainties, such as the determination of an ethical and legal framework that combines the compliance of the GDPR requirements, Member States' law and ethical standards with the scientific progress. Most of the arising controversies are shared by many other projects and initiatives such as the EHDS itself or the 1+MG initiative. Additionally, these common gaps and hindrances are being discussed within the Artificial Intelligence for Health Imaging ELSI Group (AI4HI ELSI).

IV. Throughout our Task 1.1, these issues have been thoroughly identified and explored concerning the particular needs and considering the context of the processing operations to be carried out. Accordingly, EuCanImage proposes a platform with the objective of making databases "findable" for the scientific community, as the starting point that will allow data providers and controller to take control of the decisions about their datasets.

V. The legal roles and corresponding obligations of the participants in EuCanImage have been identified and classified. Based on this, in parallel with the project's Data Management Plan, the contractual framework governing the relationships between all the partners and participants have been determined and will be further specified in forthcoming deliverables. However, it is crucial to emphasise the pivotal role that the data providers and their legal teams play, most notably, through the first stages of the data processing operations.

VI. In terms of legal interoperability, it has been stressed that EuCanImage's partners and legal teams will have to cope with a complex regulatory framework comprising not only the GDPR but also relevant Member States' laws, in addition to divergent ethical requirements and procedures of the involved data providers. This leads to the need to integrate requirements and further conditions by each data provider, that must have collected and processed the data in accordance not only with the Regulation but also with the applicable national legislation. At project level, to contribute as far as possible to the interoperability, a common contractual and governance framework will be developed, including the correspondent policies.

Finally, it has been further analysed the eventual international scalability of EuCanImage even though the project does not currently envisage data processing operations with organisations or individuals outside the European Economic Area. Excluding potential future collaborations between internationally acknowledged institutions such as the TCIA in the US would represent aprioristic hindrances to scientific progress and technological development in our area. EuCanImage has to cope with a critical context of legal interoperability between jurisdictions and, specifically, in terms of international transfers of data to the US. There is no valid adequacy decision in force thus any transfer will have to be carried out on the basis of appropriate safeguards pursuant Article 46 et. seq. of the GDPR. It could therefore be concluded that EuCanImage could rely on SCCs to carry out any future transfer of data to the US, supplemented by additional measures such as, *inter alia*, anonymisation, encryption or agreements and other binding instruments for the parties, following the case-law from the CJEU.

6 References

6.1 Legislation and normative instruments

Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act). COM/2020/767 final.

Commission Implementing Decision (EU) 2021/914 of 4 June 2021 on standard contractual clauses for the transfer of personal data to third countries pursuant to Regulation (EU) 2016/679 of the European Parliament and of the Council. C/2021/3972. OJ L 199, 7.6.2021

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.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

Commission Implementing Decision (EU) 2016/1250 of 12 July 2016 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequacy of the protection provided by the EU-U.S. Privacy Shield. C/2016/4176.

2000/520/EC: Commission Decision of 26 July 2000 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequacy of the protection provided by the safe harbour privacy principles and related frequently asked questions issued by the US Department of Commerce.

Directive 95/46/EC of the European Parliament and of 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, 23/11/1995 P. 0031 – 0050.

6.2 Jurisprudence and case law.

Court of Justice of the European Union (CJEU). Judgment (Grand Chamber) of 16 July 2020, *Data Protection Commissioner v. Facebook Ireland Ltd, Maximillian Schrems*. (C-311/18).

Court of Justice of the European Union (CJEU). Judgment (Second Chamber) of 29 July 2019, *Fashion ID v. Facebook Ireland Ltd*. (C-40/17).

Court of Justice of the European Union (CJEU). Judgment (Grand Chamber) of 10 July 2018, *Tietosuojavaltuutettu v. Jehovan todistajat*. (C-25/17).

Court of Justice of the European Union (CJEU). Judgment (Grand Chamber) of 5 June 2018, *Wirtschaftsakademie Schleswig v. Facebook Ireland Ltd*. (C-210/16).

Court of Justice of the European Union (CJEU). Judgment of 19 October 2016, *Patrick Breyer v. Bundesrepublik Deutschland*. (C-582/14).

Court of Justice of the European Union (CJEU). Judgment of 6 October 2015, *Maximillian Schrems v Data Protection Commissioner*. (C-362/14).

Court of Justice of the European Union (CJEU). Judgment (Grand Chamber) of 8 April 2014, *Digital Rights Ireland Ltd v Minister for Communications* (Joined Cases C-293/12 and C-594/12).

Court of Justice of the European Union (CJEU). Judgment (Grand Chamber) of 18 December 2008, *Satakunnan Markkinapörssi v. Satamedia* (C-73/07).

6.3 Reports, opinions, guidelines and other non-binding instruments

Article 29 Data Protection Working Party (2017). *Working Document on Adequacy Referential*. Adopted on 28 November 2017, WP254. Endorsed by the EDPB on 25 May 2018. Available at: https://ec.europa.eu/newsroom/article29/redirection/document/57550

Article 29 Data Protection Working Party. (2014). *Opinion 05/2014 on Anonymisation Techniques*. Adopted on 10 April 2014. Available at: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

Article 29 Data Protection Working Party. (2010). *Opinion 1/2010 on the concepts of "controller" and "processor"*. Available at: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2010/wp169_en.pdf

Article 29 Data Protection Working Party. (2007). *Opinion 4/2007 on the concept of personal data*. Adopted on 20th June 2007. Available at: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf

Data Protection Commission. (2019). *Guidance on Anonymisation and Pseudonymisation*. Available at: https://www.dataprotection.ie/sites/default/files/uploads/2019-06/190614%20Anonymisation%20and%20Pseudonymisation.pdf

European Commission. (2021). *Assessment of the EU Member States' rules on health data in the light of GDPR*. Available at: https://ec.europa.eu/health/sites/health/files/ehealth/docs/ms_rules_health-data_en.pdf

European Data Protection Board (2021). *Recommendations 01/2020 on measures that supplement transfer tools to ensure compliance with EU level of protection of personal data*. Adopted on 18 June 2021. Available at: https://edpb.europa.eu/system/files/2021-

06/edpb_recommendations_202001vo.2.0_supplementarymeasurestransferstools_en.pdf

European Data Protection Board. (2021). *EDPB Document on response to the request from the European Commission for clarifications on the consistent application of the GDPR, focusing on health research*. Available at: https://edpb.europa.eu/sites/default/files/files/file1/edpb_replyec_questionnaireresearch_final.pdf

European Data Protection Board. (2020). *Guidelines 07/2020 on the concepts of controller and processor in the GDPR*. Available at: https://edpb.europa.eu/sites/default/files/consultation/edpb_guidelines_202007_controllerprocessor_en.pdf

European Data Protection Supervisor and Agencia Española de Protección de Datos. (2021). *10 misunderstandings related to anonymisation*. Available at: https://edps.europa.eu/system/files/2021-04/21-04-27_aepd-edps_anonymisation_en_5.pdf

European Parliament. (2021). *Exchanges of Personal Data After the Schrems II Judgement*. Policy Department for Citizens Rights and Constitutional Affairs. Available at: https://www.europarl.europa.eu/RegData/etudes/STUD/2021/694678/IPOL_STU(2021)694678_EN.pdf

United Nations Educational, Scientific and Cultural Organization, UNESCO (2017). Recommendation on Science and Scientific Researchers. *Records of the General Conference*, 39th session, Paris, 30 October-14 November 2017. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000260889.page=116

6.4 Literature

Diaz, O.; Kushibar, K.; Osuala, R.; Linardos, A.; Garrucho, L.; Igual, L.; Radeva, P.; Prior, F.; Gkontra, P.; Lekadir, K. (2021) “Data preparation for artificial intelligence in medical imaging: A comprehensive guide to open-access platforms and tools”. *Physica Medica*. 83:2021. pp. 25-37.

6.5 Figures