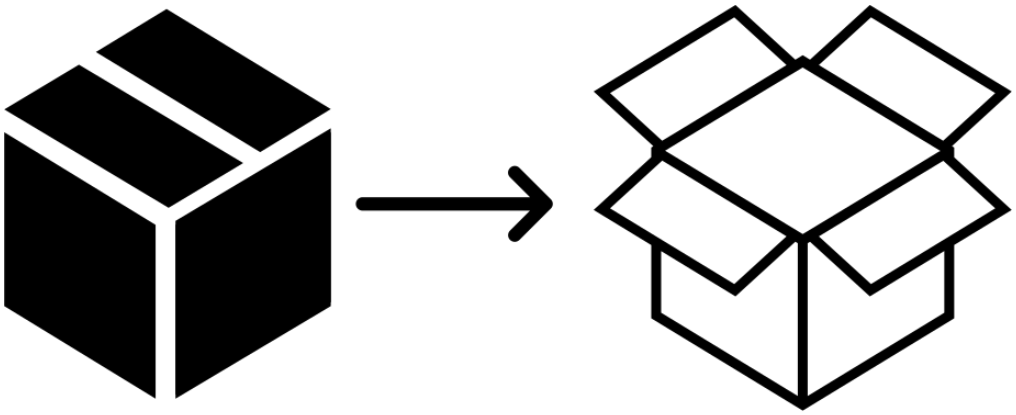
chist-era



XAIface

Measuring and Improving Explainability for AI-based Face Recognition

Annex 1 Legal Aspects (Image Dataset)

Deliverable number: Part of D3.2

Version: 2.0

EURECOM Sophia Antipolis

JOANNEUM RESEARCH DIGITAL

universität wien

it

EPFL

Acronym of the project: XAIface
Title of the project: Measuring and Improving Explainability for AI-based Face Recognition.
Grant: CHIST-ERA-19-XAI-011
Web site of the project: https://xaiface.eurecom.fr/

Short abstract

Since the consortium relies on already existing data sets, legal modalities of the usage of such data sets need to be addressed. The purpose of the document is to provide guidance and aid in providing an adequate level of data protection. Therefore, this deliverable will outline the requirements, which must be considered to comply with the current data protection regime.

Table of Contents

Definitions **4**

1. Introduction **6**
1.1. Scope of this document 6
1.1.1. Choosing of the dataset(s) and Data Protection 6
1.1.2. Goal of this deliverable 7

2. Processing existing data sets in accordance with the General Data Protection Regulation (GDPR) **8**
2.2. Applicable Law – Territorial Scope of the GDPR 12
2.2.1. Territorial Application of the GDPR 13
2.2.2. Opening Clause: Assessment of national Law 15
2.2.3. Austria: Exceptions established by case law 18
2.2.4. Interpretation 19
2.3. Role allocation in Data Protection Law 20
2.3.1. The different “roles” in the GDPR 20
2.2.5. Controller 21
2.2.6. Joint Controllership 22
2.2.7. Relevant Case Law of the European Court of Justice 23
2.2.8. Applying above considerations on joint research projects 26
2.4 Transfer of personal data to third countries 30
2.4.1. Defining the term “transfer of personal data to third countries” 31
2.4.2. Publishing results online 35
2.4.3. Regulatory framework of Chapter V GDPR 36

3. Lawfulness of processing in a scientific context under the GDPR **38**
3.1. Lawfulness of processing personal data 38
3.2. Specificities in the context of scientific research 42
3.3. Processing of Data(sets) which have been obtained from a Third Party 45
3.4. Processing Data that have been obtained unlawfully 45
3.4.1. Consequences of processing unlawfully obtained Data 48

4. Database Protection & Licenses **48**
4.1.1. International Law 49
4.1.2. EU Law – Directive 96/9/EC 50
4.1.2.1. Material Scope 50
4.1.2.2. Territorial Scope 51
4.1.2.3. Rights 51
4.1.2.4. Licensing 51

5. Results **52**
Checklist – Choosing from existing face image data sets 53

Definitions

Biometric Data means personal data resulting from specific technical processing relating to the physical, physiological 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.[1]

Controller is a person, who alone or jointly with others decide on the means and purposes of the data processing and can be seen as the main addressee of the GDPR.

Data subjects are natural persons, whose data will be processed.

Data Protection Directive 1995 was repealed through the GDPR and was in force until the entry of the GDPR on 25th of May in 2018.

General Data Protection Regulation (GDPR): The General Data Protection Regulation is a European legal act, which lays down European-wide harmonized provisions regarding the processing of personal data and is directly applicable in all EU-member states.[2]

Personal data is any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.[3]

Processing means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.[4]

Processor is a person who is acting on behalf of the controller(s) within a data processing activity and does not decide over the means and purposes of the processing.

[1] Art 4 (1) (14) GDPR.
[2] Regulation 2016/679/EU 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).
[3] Art 4 (1) (1) GDPR.
[4] Art 4 (1) (2) GDPR.

1. Introduction

1.1. Scope of this document

This document addresses legal issues regarding data protection law. A special point of focus are the legal criteria according to which the dataset(s) should be chosen.

1.1.1. Choosing of the dataset(s) and Data Protection

In the first iterative discussions on the possible legal implications, various issues that arise with processing of existing data sets within the project were discerned. These issues touch the following topics and provisions:

1. Scope of application (GDPR)
   - Material Scope (Art 2 GDPR)
   - Territorial scope
     - Which law is applicable?
       - GDPR (for partners situated in EU territory)
       - Swiss law (EPFL)
       - National law (based on opening clauses)
     - What if a Swiss controller processes data from (exclusively) third countries?
     - Does a single national law regarding the processing of personal data for purposes of scientific research apply (e.g. in Austria FOG [“Federal Research Organization Act"]), or do separate laws for each controller apply?
2. Role allocation
   a. Demarcation of joint and sole controller;
      i. Requirements;
   b. Joint controller: third-party state and Member State actor?
3. Lawfulness of processing in scientific research;
4. Is the original unlawful acquisition of personal data for the database relevant for current legal assessment of the processing for this research project?
   If yes, how does this factor influence the assessment of the lawfulness of the processing?

1.1.2. Goal of this deliverable

In this working document, the legal issues regarding the processing activities within this research project will be evaluated and described. A key problem will be choosing an existing data set without prior knowledge of the concrete circumstances of the data collection.

The aim of this deliverable is to provide guidelines on how to choose data sets in a GDPR compliant manner by laying down the decisive requirements, which must be considered to process personal data in accordance with the GDPR.

This document has been structured according to the legal questions established in chapter 1.2.1.

At the end of this document, a checklist will be formulated and provided. The checklist should give guidance on how to decide which data sets should primarily be used. The provided checklist should make the assessment more accessible and should not be seen as a comprehensive decision tree, which covers all aspects of data protection law.

2. Processing existing data sets in accordance with the General Data Protection Regulation (GDPR)

2.1. Material Scope of the GDPR

The General Data Protection Regulation (GDPR) is a European legal act, which is in force since 25th of May 2018 and has repealed the Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data.[5]

By changing the instrument from European directive to European regulation, the GDPR became directly applicable in all member states. Hence, no legal implementation into national law is necessary anymore and therefore it guarantees a Europe-wide harmonised level of protection.[6] With the entry into force of the regulation, it became part of the national legal systems of the member states and enjoys priority of application.

However, the GDPR contains so-called “opening clauses”, which allow the national legislator to enact additional provisions, which specify the GDPR. Hence, to assess a case from a data protection perspective national and European law is pertinent. This juxtaposition of national and European law inevitably leads to a legal fragmentation in the field of data protection, which again causes legal uncertainty for those who are affected by the law or who should apply the provisions for compliance. However, as mentioned above, only with the national concretisations and specifications the abstract provisions of the GDPR will get more enforceable. Consequently, even though a harmonised regulation exists on a European level, it is inescapable to also apply the national legislation parallel to the GDPR. European data protection law is therefore characterised by a co-regulation by the European Union and the member states.[7] Where national law is conflicting with the European data protection framework, the GDPR enjoys priority of application in all areas.[8]

Besides the juxtaposition of European and national law, national legislation of the member states can collide, where it is not finally clarified how to solve conflicting national legislation of the EU-Member States. This issue will be explained in detail in the next chapters as it also concerns processing activities for scientific purposes.

---

[5] 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) OJ L 119, 4.5.2016, p. 1–88.
[6] Art 288 TFEU.
[7] *Roßnagel* Alexander, Gesetzgebung im Rahmen der Datenschutz-Grundverordnung, Aufgaben und Spielräume des deutschen Gesetzgebers?, Datenschutz und Datensicherheit DUD, 2017/5, 277 (278).
[8] *EuGH*, Urt. v. 15.7.1964, 6/64, EU:C:1964:66, Slg. 1964, 1253, 1269 – Costa/E. N. E. L.

In any case, the applicability of the GDPR must first be assessed. Every legal act has a material, territorial and temporal scope, which determines when and where the provisions of the legal act are applicable.

For which processing activities is the GDPR applicable?

Article 2 of the GDPR defines the material scope of the regulation.[9] The material scope of the GDPR is comprehensive. The GDPR is applicable in general, **for all processing activities of personal data**, which are wholly or partly processed by automated means.[10] It is also applicable for non-automated processing like filing systems, provided the personal data is structured in a way that "*it is accessible to specific criteria, whether centralised, decentralised or dispersed on a functional or geographical basis*".[11] If the material scope is not met, the GDPR is not applicable.

There are also **some exemptions** laid down by the framework, where the GDPR is not applicable, even though personal data is processed:

"*This Regulation does not apply to the processing of personal data:*

(a) In the course of an activity which falls outside the scope of Union law;

(b) by the Member States when carrying out activities which fall within the scope of Chapter 2 of Title V of the TEU;

(c) by a natural person in the course of a purely personal or household activity;

(d) by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security."

Additionally, the regulation also does not apply to processing of personal data of deceased persons or legal persons.[12] However, these exemptions are not relevant within the context of XAIface and do not need to be considered in detail at this point.

Therefore, the main point to decide if a processing activity falls within the material scope of the GDPR is **the notion of "personal data".** But when can data be considered "personal" in the sense of the GDPR?

The GDPR defines personal data in Article 4 (1) (1) GDPR as "*any information relating to an identified or identifiable natural person ('data subject')".* Hence, the understanding of personal data is quite broad, when is a person **identified or at least identifiable**. In the wording of the GDPR there is no further explanation of when a natural person can be

[9] Territorial scope see: Chapter 2.2.
[10] Art 2 (1) GDPR.
[11] Art 4 (1) (6) GDPR.
[12] In these cases, it might be possible that national law exists and is pertinent.

considered identified, but there are indications of what **identifiable** means. According to definition in Article 4 GDPR

"an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;".

Furthermore, Recital 26 of the GDPR specifies the meaning of "personal data" and determines that all means, which are reasonably likely to be used, must be taken into account. "*To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments.*"

Therefore, the GDPR does not follow a comprehensive and absolute understanding of personal data. In order to decide whether personal data will be processed, it must not only be considered if the data alone has potential to be seen as "personal data". Rather, other circumstances of the processing must be taken into account, especially the identity of the controller and the means he has at his disposal, which are likely to be used and what his intention – the purpose of the processing – is. Furthermore, not only the situation of the controller needs to be evaluated. In addition, **the means of third parties** (see Recital 26, "*[…] or by another person*") need to be considered. This understanding of personal data has already been confirmed by consistent case law of the Court of Justice (ECJ), which also follows a very broad approach of personal data.[13] The assessment if personal data in the sense of the GDPR is processed has to be evaluated from a holistic point of view, where all the circumstances of the processing situation are considered.

Processing of personal data within XAIface?

According to the technical partners within the project, several processing activities will take place. Especially, when it comes to face recognition technology (in the following „FRT"), the processing of face images is unavoidable.

Even though not every processing of face images can be seen as the processing of personal data, it is obvious, especially, when it comes to FRT, personal data must be seen as given in this context. The fundamental nature of a biometric recognition system is the verification or identification of individuals. Also, within the project the technical partners must necessarily process face images in a way that individuals – natural person – will be identified; only then the results of the project can be validated properly, and substantiated conclusion can be made.

Processing of special category of personal data? Biometric data?

[13] Breyer-Case: ECJ 19. 10. 2016, C-582/14, Breyer, ECLI:EU:C:2016:779.

The data protection regime also presumes that some data – information – needs a higher level of protection than others, therefore additional requirements have been laid down for these types of processing activities. The GDPR differentiates between “general” personal data and special categories of personal data. This is due to the fact that some data is more sensitive (therefore this data is also referred to as “sensitive data”) than others and thus a higher level of protection is justified to protect the rights and freedoms of the data subjects in a comprehensive way.

Due to the system of the GDPR, it must also be evaluated if processing of special categories of personal data takes place.

To guarantee the aforementioned higher level of protection the GDPR follows a risk-based approach, which can be derived among others from Article 9:

“*Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited*.”

Hence, Article 9 (1) GDPR constitutes a general prohibition for processing of special categories of personal data. Several exemptions to the prohibition exist, for example in Article 9 (2) a to j, which allow the processing of sensitive data under certain circumstances.[14] However, many authors claim that the exemptions in Article 9 (2) cannot be seen as a legal base for data processing. The lawfulness criteria of the processing of sensitive data would only be fulfilled, if a legal base in Article 6 and exemption in Article 9 cumulatively exist.

Other provisions, which are depending on the risk of the processing, are Article 24, 25 or 35 of the GDPR. Additionally, the fact that a controller will process special categories of personal data influences the scaling of the measures which must be implemented. In conclusion this means, if a controller is also processing sensitive data additional requirements for the lawfulness of the processing (especially laid down in Art 9 (2) a to j) must be met. Furthermore, the risk, which arises from the fact that sensitive data is processed, must be taken into account in the assessment of the implemented measures. The riskier the processing, the more data protection requirements a controller must fulfil to act in compliance with the GDPR.

Article 9 of the GDPR also refers explicitly to biometric data. The repealed Data protection Directive from 1995 did not contain provisions on biometric or genetic data. The category of biometric data was introduced to data protection law with the GDPR in 2018 and first defined in Article 4 (14):

[14] There is also an ongoing litigation in literature if Art 6 and Art 9 of the GDPR must be applied cumulatively or not if a controller wants to process special category of personal data.

"*biometric data means personal data resulting from specific technical processing relating to the physical, physiological 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*".

First, the data needs to be a result of a specific technical processing activity, which means that face recognition done by a human being e.g. does not fall under the definition of biometric data from a data protection point of view. A technical system which aims to identify or verify an individual by its biometric characteristics is indispensable. Whether a natural person (e.g. an image comparison expert) makes the final decision is irrelevant. Even though the definition of biometric data mentions facial images explicitly, not every image of a human face can automatically be considered biometric data as defined in the GDPR. Many authors claim that biometric material is generated only, when the raw data gets gentrified.

Furthermore, the data must be derived from a physical, physiological, or behavioural characteristic of an individual. A biometric characteristic in general must fulfil several requirements to function as a biometric identifier. Examples for a biometric identifier can be the face, the Retina, the DNA, the gait or voice of an individual, the tooth print et cetera.

Finally yet importantly, according to the GDPR biometric proceeding allows and confirms the uniquely identification of a natural person. No further elaboration can be found in the *verba legalia* on the difference between a data subject being identified or uniquely identified. Some authors claim unique identification can be assumed, if the characteristic of the data subject is unique.[15] This must be countered by the fact that every biometric feature must have a certain degree of uniqueness, because only then it can serve as a biometric identifier.

Biometric data will be processed and must be considered in the legal assessment, especially with regard to the lawfulness of the processing and implementing of technical and organisational measures.

2.2. Applicable Law – Territorial Scope of the GDPR

In addition to the material scope of the GDPR, the territorial scope must also be examined, especially since this project will be conducted by a multinational consortium of researchers. The consortium consists of institutions from EU-Member States such as France, Portugal, and Austria. Furthermore, researchers from Switzerland, which is not part of the European Union, will participate. All activities conducted within the project must adhere to pertinent data protection laws.

At first, suitable databases for AI-training and validation have been identified. The second step will be the selection of one or more databases, which are best suited for the needs of the project. One of the requirements is compliance with the GDPR (EU) and/or other pertinent data protection laws. Hence, at first the applicable law must be identified.

[15] *Hödl* in Knyrim, DatKomm Art 4 DSGVO 147 (1.12.2018, rdb.at)

In essence, five distinct processing activities can be discerned:

1. Data collection for the database
2. Transfer (download) of the database
3. Training of the AI
4. Validation of the AI
5. Publishing of the results

The first activity will not be conducted by members of the consortium. Instead, the members will rely on already available and open accessible databases provided by other researchers. It should also be noted that the following passage only examines the application of data protection law (as opposed to copyright laws p.ex).

2.2.1. Territorial Application of the GDPR

The territorial scope of the GDPR is defined as following in Article 3:

1. "*This Regulation applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union,* ***regardless of whether the processing takes place in the Union or not.***

2. *This Regulation applies to the processing of personal data of data subjects who are in the Union by a controller or processor not established in the Union, where the processing activities are related to:*

   (a) the offering of goods or services, irrespective of whether a payment of the data subject is required, to such data subjects in the Union; or
   (b) the monitoring of their behaviour as far as their behaviour takes place within the Union.

3. *This Regulation applies to the processing of personal data by a controller not established in the Union, but in a place where Member State law applies by virtue of public international law*."[16]

The regulation follows three principles to determine its scope:

- Territorial principle
- Establishment principle

[16] Art 3 GDPR.

- Market location principle[17];

According to the regulation, every processing activity within the context of a task of a controller or processor, who is established within the European Union, is subject to the regulation. Whether the processing takes place in the European Union or not is irrelevant.

Hence, processing conducted by controllers or processors established in France, Portugal or Austria falls within the scope of the provisions.

A controller or processor is established in the Union, if activities are effectively exercised through stable arrangements within the territory.[18] In case of research institutions, the criteria are undoubtedly satisfied, especially since the threshold is very low.[19] Whether some servers or technical equipment are located elsewhere is irrelevant for the application.[20]

It should also be noted that the GDPR applies, if the criteria are satisfied irrespective of the nationality or residence of the data subjects.[21] Therefore, the fact that some of the databases mainly consist of data from data subjects from states such as China, is not relevant for the determination of the applicability.

Research institutions, which are not established in the territory of the European Union, however, only fall within the territorial scope, if a sufficient connection can be made according to Art 3 (2) or (3) GDPR.

Art 3 (2) is an expression of the market location principle. Processing conducted by controllers or processors established outside the territory of the European Union, such as Switzerland, is only subject to the GDPR, insofar as the controller or processor offers goods or services to data subjects in the Union or monitors behaviour, which takes place within the Union.

The second case primarily concerns profiling and tracking activities[22], which will not be part of the research project.

The first case requires a controller or processor to offer goods or services within the territory of the European Union. Whether or not payment is involved is irrelevant. The aim is to broaden the scope of protection.[23] In this case, the research institutions do not offer physical products. According to the services-directive[24] a "*'service' means any self-employed economic activity, normally provided for remuneration, as referred to in Article 50 of the Treaty*". Due to the wording of Art 3 (2) 2 GDPR, the criterion of remuneration is of no

[17] *Leissler/Wolfbauer* in *Knyrim,* DatKomm Art 3 DSGVO Rz 6 (1.3.2021, rdb.at)
[18] Rec. 22 GDPR.
[19] EDPB, Guidelines 3/2018 on the territorial scope of the GDPR (Article 3) 6.
[20] *Leissler/Wolfbauer* in *Knyrim*, DatKomm Art 3 DSGVO Rz 7 (1.3.2021, rdb.at).
[21] *Leissler/Wolfbauer* in *Knyrim*, DatKomm Art 3 DSGVO Rz 10 (1.3.2021, rdb.at).
[22] *Leissler/Wolfbauer* in *Knyrim*, DatKomm Art 3 DSGVO Rz 20 (1.3.2021, rdb.at).
[23] Rec. 23 GDPR.
[24] EU-RL 2006/123/EG vom 12. 12. 2006.

particular relevance.[25] Article 50 now refers to Article 57 of the Treaty on Functioning of the European Union (TFEU[26]). According to article 57 TFEU, services shall in particular include activities of industrial or commercial character, activities of craftsmen and activities of the professions. The enumeration is not exhaustive. According to *Budischowsky,* the definition in the TFEU only excludes non-profit activities and activities subject to special regulations.[27]

The only research institute located outside of the territory of the European Union is EPFL. EPFL is a university and therefore a non-profit organisation. Hence, it could be argued that research activities conducted by the EPFL cannot be considered an economic activity. Contrary to this opinion, however, the opinion of the EDPB specifically includes an example of a fictitious Swiss university offering a master's degree or summer courses.[28] In this example, the EDPB makes the application of the GDPR solely dependent on whether or not these courses are specifically advertised to German and Austrian universities. Thus, the argument could be made that the distinction must be made more granularly based on the particular 'service' or task. In another example the EDPB also implies that there would have to be a link between the offer of a service and the processing activities in question.[29] This link could be seen in the wording of Art 3 par 2 GDPR ("such data subjects").

Since research activities conducted by a university are not economic activities and there would be no link between the offering of goods and services and the processing in the project, the GDPR does not apply to processing conducted by EPFL based on Art 3 par 2 GDPR.

Art 3 (3) GDPR extends the scope of the GDPR to organisations established in other countries, which are subject to member state law by virtue of public international law. Such organisations include for example consular posts and diplomatic missions.[30] Therefore, this provision is not applicable.

Processing conducted by controllers or processors established in an EU member state falls within the scope of the GDPR. This includes EURECOM (France), Joanneum Research (Austria), Instituto de Telecomunicações (Portugal) and University of Vienna (Austria), insofar as these institutions process personal data. Processing conducted by the EPFL is not subject to the GDPR.

2.2.2. Opening Clause: Assessment of national Law

Insofar as the GDPR does not apply or only applies partially, research institutions must adhere to national data protection law. This could be the case in two scenarios:

[25] *Leissler/Wolfbauer* in *Knyrim*, DatKomm Art 3 DSGVO Rz 18 (1.3.2021, rdb.at) refering to *Klar* in *Kühling/Buchner*, DS-GVO Art 3 Rz 73.
[26] EU, Consolidated version of the Treaty on the Functioning of the European Union, 13 December 2007, 2008/C 115/01.
[27] *Budischowsky* in *Jaeger/Stöger*, EUV/AEUV Art 57 AEUV RZ 10 (1.10.2018, rdb.at).
[28] EDPB, Guidelines 3/2018 on the territorial scope of the GDPR (Article 3) 19.
[29] EDPB, Guidelines 3/2018 on the territorial scope of the GDPR (Article 3) 16.
[30] Rec. 25 GDPR.

a) Scenario A: The GDPR does not apply due to its territorial scope.
b) Scenario B: The GDPR does apply, but the processing activities are subject to an opening clause.

This section will focus mainly on scenario B. A relevant opening clause can be found in article 89 of the GDPR with regards to processing for scientific purposes. Member state law may provide derogations from the GDPR for such purposes. As a result of this clause, some states, such as Austria, passed a law on data protection and scientific research. Hence, the question arises, which of these national laws apply. Such problems are usually subject to national collision law. In the case of the Austrian legal system, such law does not exist anymore.[31] The GDPR does not provide any provisions on which material law is applicable, as opposed to Art 4 of the former data protection directive. In commentary literature, two answers to the problem can be found:

a) Since there is no specific provision in the GDPR, regulating the applicable law falls within the competence of the member states. In general, this means that national courts or public bodies will apply the "lex fori" rule.[32] As a consequence, those bodies will generally apply their respective national law. However, there are usually many exceptions to the general rule, especially in case of damages. In such a case, for example according to the IPRG (Austrian international private law act)[33], the law in effect at the location of the infringing act applies. Of course this solution can only apply insofar as there would be an infringing act and insofar as the problem is a subject of private law. This, however, is not the case. Instead the question at hand is one of compliance and hence a matter of public law.

b) According to some authors[34], it must be concluded that there is a gap in the law and an analogy must be created according to the articles on the competence of the lead supervisory authority.[35] The lead supervisory authority will then apply national law. This idea is meritorious if not simply for its simplicity and will therefore be elaborated below.

Authors who generally accept the theory that the applicable national law is to be determined through the GDPR, consequently refuse to accept the competence of the national legislator to determine the applicable law. They argue, it must be concluded, that there is an unplanned gap in the law, that must be filled by means of analogy[36] to the provisions of competence of the lead supervisory authority of the GDPR. The procedure to determine the supervisory authority is complex and depends on factors such as the role allocation and connection of the individual controllers and processors. The general idea behind the

[31] *Leissler/Wolfbauer* in *Knyrim*, DatKomm Art 3 DSGVO Rz 29ff (1.3.2021, rdb.at).
[32] *Leupold/Schrems* in *Knyrim*, DatKomm Art 79 DSGVO (1.6.2021, rdb.at).
[33] *Leissler/Wolfbauer* in *Knyrim*, DatKomm Art 3 DSGVO Rz 33 (1.3.2021, rdb.at).
[34] *Zavadil* in *Knyrim*, DatKomm Art 56 DSGVO Rz 26 (1.3.2021, rdb.at) refering to *Feiler/Forgó*, EU-DSGVO Art 92 Anm 5.
[35] Art 56 GDPR.
[36] *Feiler/Forgó*, EU-DSGVO Art 56 Anm 12.

provision is to implement the well-established “One-Stop-Shop”-principle.[37] The principle of the lead supervisory authority, however, is new:

> “*1. Without prejudice to Article 55, the supervisory authority of the main establishment or of the single establishment of the controller or processor shall be competent to act as lead supervisory authority for the cross-border processing carried out by that controller or processor in accordance with the procedure provided in Article 60.*
>
> “2. *By derogation from paragraph 1, each supervisory authority shall be competent to handle a complaint lodged with it or a possible infringement of this Regulation, if the subject matter relates only to an establishment in its Member State or substantially affects data subjects only in its Member State.*
>
> 5. *In the cases referred to in paragraph 2 of this Article, the supervisory authority shall inform the lead supervisory authority without delay on that matter. Within a period of three weeks after being informed the lead supervisory authority shall decide whether or not it will handle the case in accordance with the procedure provided in Article 60, taking into account whether or not there is an establishment of the controller or processor in the Member State of which the supervisory authority informed it.*
> 6. *Where the lead supervisory authority decides to handle the case, the procedure provided in Article 60 shall apply. The supervisory authority which informed the lead supervisory authority may submit to the lead supervisory authority a draft for a decision. The lead supervisory authority shall take utmost account of that draft when preparing the draft decision referred to in Article 60(3).*
> 7. *Where the lead supervisory authority decides not to handle the case, the supervisory authority which informed the lead supervisory authority shall handle it according to Articles 61 and 62.*
> 8. *The lead supervisory authority shall be the sole interlocutor of the controller or processor for the cross-border processing carried out by that controller or processor*.”

The lead supervisory authority is usually determined by the location of the main establishment of the controller or processor.[38] In the case of joint controllers, no such establishment can be determined and there are no further specific provisions. Hence, it is recommended that joint controllers determine a «main establishment» of one controller that would have the necessary permissions in such a case.[39] In case the controllers should not be found to be joint controllers, the applicable law must be determined by each institution respectively. The applicable law would be (in accordance with Art 56 GDPR) the law of the main establishment of the respective controller, provided that no closer link can be drawn to any other Member State. (Art 56 par 1 & 2 GDPR). Furthermore, it should be noted that collision law may still exist in other states.

[37] *Zavadil* in Knyrim, DatKomm Art 56 DSGVO (1.3.2021, rdb.at) Rz 1.
[38] Exception in Art 56 par 1 GDPR.
[39] *Zavadil* in *Knyrim*, DatKomm Art 56 DSGVO Rz 16 (1.3.2021, rdb.at).

2.2.3. Austria: Exceptions established by case law

In 2020 the Austrian data protection authority had to render a decision on the request of a research institute.[40] According to Austrian law, specifically § 7 of the data protection act, the data protection authority may grant authorisation for the processing of personal data for scientific research purposes at the request of the person responsible, if

a) the consent of the data subject cannot be obtained, because the data is not accessible or would require a disproportionate effort,
b) the processing is in the public interest, and
c) the professional qualification of the data controller is credibly demonstrated.

The research institute in question tried to develop algorithms in the area of autonomous driving and described their project as follows:

The objective of the project was to verify the interpretation of images through said algorithms. The challenge was to analyse and categorise images sufficiently. The algorithms needed to be capable of distinguishing streets, humans, traffic signs, bikers etc. In order to obtain this objective, pictures should be pre-classified by humans and later used for validation. For validation, the researchers needed a large number of images. Hence, the researchers created data manually by filming in public areas from the viewpoint of a driver. As a next step, the researchers planned to repeat the data collection in the territory of other European Member States. The cars used for the data collection were specifically marked with stickers with a link that led to a website (in German and English language), which provided information on the data processing purposes. Any planned routes were also announced on the website. The processing would necessarily include pictures of humans. Even though it was necessary that the humans were clearly visible, the researchers had no use for their respective identity.

Data collection and classification was a lengthy process, but the data could be used as reference material for many other algorithms. Therefore, the research institute planned to provide the data to other research institutes working in the same area, insofar as the interests of depicted persons did not outweigh the scientific interest of the research institutes. Research institutes in third countries would only receive the data on the basis of standard contractual clauses.

The research institute provided a lengthy risk assessment and argued that consent cannot be obtained due to disproportionate effort. The researchers argued that the processing was necessary to realise autonomous driving and to avoid accidents and would therefore be in the public interest. The research institute noted that processing would only be conducted by certain, trained staff members, who were bound to secrecy and offered further guarantees.

The DPA elaborated that images of persons are personal data, but do not fall within the special categories of personal data. The collection and evaluation of images for scientific

[40] Austrian DPA, 21.01.2020, 2020-0.013.649 (DSB-D202.235).

purposes falls within § 7 of the data protection act. The DPA agreed that obtaining consent would not be feasible and the public interest as well as the qualification was sufficiently demonstrated. The DPA allowed the researchers to create footage in public spaces in Austria.

However, the DPA did not grant authorisation to obtain footage in other Member States. The authority elaborated on the opening clause in Art 89 GDPR. In general, the DPA argued, the Austrian legislator was competent to and did in fact create specific provisions on the processing for scientific purposes in § 7 of the data protection act. However, the DPA argued that since other Member States also made use of the opening clause, other national laws are also relevant. **The DPA ascertained that national laws of other Member States also apply as "leges speciales".** Furthermore, the DPA noted that each data protection authority only has jurisdiction within the respective territory of their Member State. Therefore, the request was rejected.

2.2.4. Interpretation

The decision of the Austrian data protection authority highlights a significant gap in the full harmonisation of data protection law in the EU. The "One-stop-shop"-principle seemingly does not apply in cases where chapter IX of the GDPR is applicable.[41] As a result, research institutes in different territories must apply different national data protection laws, which of course can lead to fragmentation and also foster legal uncertainty.

[41] *Zavadil* in *Knyrim*, DatKomm Art 56 DSGVO (Stand 1.3.2021, rdb.at).

2.3. Role allocation in Data Protection Law

The “role allocation” plays a crucial part in the application of the GDPR since this determines who shall be responsible for the compliance with the obligations laid down in the regulation. The Controller is the primary point of contact for the data subjects for exercising their rights, which make the determination of “for whom are obligations mandatory” to a key element of the functioning of data protection law. This section therefore shall provide a brief overview of the role concept in the GDPR and show their demarcation.

2.3.1. The different “roles” in the GDPR

Following the *verba legalia,* the GDPR generally distinguishes between

- controller[42],
- joint controller[43]
- processor[44] and
- third party.

The **controller** is the main addressee of most of the provisions and in general decides about the means and purposes of the processing activity; simply put, the “why” and “how” of the processing. In the GDPR the controller is defined as: “*the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law*”[45] What this decision-making criteria mean in detail will be shown shortly in this section.

In addition to what has already been said about controllership and its importance in the GDPR, these above-mentioned criteria are also applicable for the institute of **joint controllership**, but with one main difference: there are two or more actors who are deciding jointly about these key elements (means and purposes) of the processing.

In contrast to the role of the controller, the **processor** acts on behalf of the controller and processes personal data only for the purposes of the controller. The processor is defined according to the GDPR as “*a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller*.”[46] The main difference between controller and processor is the actual influence on the data processing itself. The processor is bound by the instructions of the controller and only acts on behalf and for the purposes of the controller.[47]

---

[42] Art 24 GDPR ff.
[43] Art 26 GDPR.
[44] Art 28 GDPR.
[45] Art 4 (1) Z7 GDPR.
[46] Art 4 (1) Z8 GDPR.
[47] *European Data Protection Board*, Guidelines 07/2020 on the concepts of controller and processor in the GDPR adopted on 02. September 2020, p.24.

Another condition for being considered as a processor is the fact that there is also another actor (the controller) involved in the data processing.[48] So, the processor must be seen as a separate entity in relation to the controller to which the controller delegates part or all of the processing.[49]

A processing activity can also involve multiple processors since the controller can engage another processor or the processor itself can engage another party, but the latter only is justified under the GDPR with the authorisation of the controller.[50]

The **Third Party** is defined according to the GDPR as a "*as a natural or legal person, public authority, agency or body other than the data subject, the controller, the processor and person who, under the direct authority of the controller or processor, are authorised to process personal data*."[51]

Of course, in any collaboration between multiple parties it can be difficult to assess who determines (and to what extend) the means and purposes of the processing. This section aims to provide guidelines on how to apply the concepts laid down in the GDPR and the existing case law as well as in the literature, to an international research project such as XAIface.

2.2.5. Controller

How to determine a controller in practice is the first question that needs to be answered to be able to understand the concept of the role allocation within the GDPR. As mentioned above, the controller is the actor who decides about the key elements of the processing. According to the Guidelines of the EDPB this control can be stemmed from a legal provision or the factual influence on the processing.[52]

Furthermore, the controllership must always be understood as a **functional concept** when the control is stemmed from the factual influence on the processing.[53] To determine whether or not an entity should be qualified as a controller, it needs to be determined if that entity actually exerts a decisive influence on the purposes and means of the processing rather than only, for example, a (contractual) designation as "controller".[54]

---

[48] *European Data Protection Board*, Guidelines 07/2020 on the concepts of controller and processor in the GDPR adopted on 02. September 2020, p.24.
[49] *European Data Protection Board*, Guidelines 07/2020 on the concepts of controller and processor in the GDPR adopted on 02. September 2020, p.24.
[50] *European Data Protection Board*, Guidelines 07/2020 on the concepts of controller and processor in the GDPR adopted on 02. September 2020, p.24.
[51] Art 4 (1) Z10 GDPR.
[52] *European Data Protection Board*, Guidelines 07/2020 on the concepts of controller and processor in the GDPR adopted on 02. September 2020, p.10.
[53] *Wyrobek* in Knyrim, DatKomm Art 26 DSGVO 14 (1.6.2021, rdb.at).
[54] *European Data Protection Board*, Guidelines 07/2020 on the concepts of controller and processor in the GDPR adopted on 02. September 2020, p.11.

Of course, the designation can be seen as an indicator for controllership in many cases. Nevertheless, the factual influence on the processing is crucial for the allocation as controller.

Furthermore, it is not mandatory that the controller himself processes the data by participating operationally, which makes the assessment exceedingly complicated in practice. Especially for outsiders or those who are affected by the data processing (data subjects), it is often not clear which actor has the factual influence on the means and purposes of the processing due the fact that controllership does not depend on who is actually processing the data. In absence of a legal provision, only the factual influence on the key elements is relevant for the role allocation.

So, the determination of the **means and the purposes** of the processing of personal data is the essential part of the role allocation because the party who determines these means and purposes is to be considered the controller of the processing activity at hand.

2.2.6. Joint Controllership

If the controller of a processing activity is determined, you must determine further, whether other entities that take part in that processing activity should be considered “joint controllers” (together with the initially determined controller).[55]

To be qualified as joint controllers, it is necessary that two (or more) actors decide jointly about the means and purposes of the data processing. Hence, the element of a pluralistic decision-control is indispensable for the joint controllership. Therefore, the European Data Protection Board (EDPB) elaborated in its guidelines that it is not a decisive factor if one actor just has a minimal influence on the means and purpose. The crucial factor is the joint determination of both key elements: the how and why of the processing.[56]

Only if both (or more) actors determine the means and the purpose (it is not possible that actor A determines the means and actor B the purpose), joint controllership can be assumed. If this is not the case the other actor could qualify as a processor or a (sole/separate) controller. The difference between sole or joint controller seems vague, since there are no solid differentiating factors. The crucial element for distinction is the collaborative cooperation, whereby **collaborative cannot** be understood as **equative**.

Therefore, it is necessary to evaluate all processing activities separately since there can be a different outcome for the different processing activities.

[55] Concrete examples are given in the discussion of the relevant case law, below.
[56] *Wyrobek* in Knyrim, DatKomm Art 26 DSGVO 20 (1.6.2021, rdb.at); *Martini* in Paal/Pauly, DS-GVO BDSG Art 26 DSGVO 19.

2.2.7. Relevant Case Law of the European Court of Justice

It is therefore hardly surprising that the courts have already had to deal with the distribution of roles on several occasions and that there is case law further developing the notion of joint controllership. It can already be said at the outset that the ECJ takes a very broad approach on the institute of **joint controllership**.

An important decision regarding the determination of the controller is the judgment of the European Court of Justice (ECJ) in the case C-131/12 "**Google Spain and Google".**

This decision concerned a Spanish national resident, who complained that when an internet user entered his name in the search engine "Google Search", the user would obtain links to two pages of La Vanguardia's newspaper on which an announcement mentioning Mr. Costeja González's name appeared for a real-estate auction connected with attachment proceedings for the recovery of social security debts.

Google Spain and Google Inc. argued that the activity of search engines could not be regarded as processing of the data which appear on third parties' web pages, given that search engines process all the information available on the internet without effecting a selection between personal data and other information. They further argued that even if that activity would be classified as 'data processing', the operator of a search engine cannot be regarded as a 'controller' in respect to that processing since it has no knowledge of those data and does not exercise control over the data.[57]

However, the ECJ found that, in exploring the internet automatically, constantly and systematically in search of the information which is published there, the operator of a search engine 'collects' such data which it subsequently 'retrieves', 'records' and 'organises' within the framework of its indexing programmes, 'stores' on its servers and, as the case may be, 'discloses' and 'makes available' to its users in the form of lists of search results.[58]

The ECJ also clearly distinguishes between the processing of personal data carried out in the context of the activity of a search engine from that carried out by publishers of websites, (consisting in loading those data on an internet page).[59]

So, while the ECJ ruled that the concept of "controller" should – in view of its objective (which is to ensure, through a broad definition of the concept of 'controller', effective and complete protection of data subjects[60]) be interpreted broadly, the ECJ also strictly differentiated between the different processing activities.[61]

[57] ECJ 13 May 2014, C-131/12, *Google Spain and Google,* § 22.
[58] ECJ 13 May 2014, C-131/12, *Google Spain and Google,* § 28.
[59] ECJ 13 May 2014, C-131/12, *Google Spain and Google,* § 35.
[60] ECJ 13 May 2014, C-131/12, *Google Spain and Google,* § 34.
[61] This strongly indicates that their respective processing activities are not part of a joint controllership of website administrators and search engine operators.

In the case C-210/16 (**Wirtschaftsakademie Schleswig-Holstein**) the ECJ qualified the user of a Facebook fanpage as joint controller with Facebook with regard to certain processing activities (including the placing of "cookies" on visitors of the fanpage).

The ECJ had to determine to which extent an administrator of a fan page hosted on Facebook contributes in the context of that fan page decides, jointly with Facebook, the purposes and means of processing the personal data of the visitors to the fan page. The ECJ considered that there was a contract between the fanpage administrator and Facebook[62] and that the intention for placing of cookies and the processing of personal data was primarily to enable Facebook to improve its system of advertising transmitted via its network but as well to enable the fan page administrator to obtain statistics produced by Facebook from the visits to the page for the purposes of managing the promotion of its activity.

But while the EDPB considers the fact that the fanpage user makes use of this service in its recently adopted guidelines on concepts of controller and processor, the ECJ stresses the fact, that "*the mere fact of making use of a social network such as Facebook does not make a Facebook user a controller jointly responsible for the processing of personal data by that network*".[63]

The main reason for the ECJ to consider the fanpage administrator in this specific case jointly responsible is, that the administrator of a fan page hosted on Facebook, by creating such a page, gives Facebook the opportunity to place cookies on the computer or other device of a person visiting its fan page, whether or not that person has a Facebook account.[64] This means that the fanpage user increases the reach of Facebook by establishing a fanpage and promoting not only his or her services but also the network itself.

The court also points out that the fanpage administrator can define certain parameters (e.g. the target audience) which have an influence on the processing activities and also on the statistics provided by facebook to the user.[65]

Additionally, the ECJ stated that it is not required for the fanpage administrator to have access to the processed data to be seen as joint controller with facebook.[66]

The ECJ concludes, that the existence of joint responsibility does not necessarily imply equal responsibility of the various operators involved in the processing of personal data and that those operators may be involved at different stages of that processing of personal data

[62] ECJ 05 June 2018, C-210/16, *Wirtschaftsakademie Schleswig-Holstein*, § 32
[63] ECJ 05 June 2018, C-210/16, *Wirtschaftsakademie Schleswig-Holstein*, § 35
[64] Ibid; see also § 41 („It must be emphasised, moreover, that fan pages hosted on Facebook can also be visited by persons who are not Facebook users and so do not have a user account on that social network. […]"
[65] ECJ 05 June 2018, C-210/16, *Wirtschaftsakademie Schleswig-Holstein*, § 36.
[66] ECJ 05 June 2018, C-210/16, *Wirtschaftsakademie Schleswig-Holstein*, § 38 (see also EDPB, Guidelines 07/2020 on the concepts of controller and processor in the GDPR v2.1, 07 July 2021, 21 [§ 66]).

and to different degrees, so that the level of responsibility of each of them must be assessed with regard to all the relevant circumstances of the particular case.[67]

This last aspect is further elaborated on in the case C-40/17 (**Fashion ID**). This case concerns the social plugin ("Like"-button) that a website operator (Fashion ID) implemented on its website.

The ECJ referred to its ruling in the previously mentioned cases and held that the objective of that provision [on the concept of "controller"[68]] is to ensure, through a broad definition of the concept of 'controller', effective and complete protection of data subjects.[69]

Refering to the definition of processing of personal data[70] the court points out that the **processing** of personal data **may consist of one or a number of operations**, each of which relates to one of the different stages that the processing of personal data may involve.[71] The ECJ concludes (with reference to the Advocate General) that a natural or legal person may be a controller, jointly with others **only in respect to operations** involving the processing of personal data **for which it determines jointly the purposes and means**.[72] This means that one cannot be considered to be a controller for operations that precede or are subsequent in the overall chain of processing for which that person does not determine either the purposes or the means.[73]

Fashion ID embedded on its website a social plugin causing the browser of a visitor to that website to request content from the provider of that plugin and, to that end, to transmit to that provider the personal data of the visitor.

The ECJ concludes that Fashion ID can be considered a controller in respect to the collection and disclosure by transmission of personal data (but not to the preceding and subsequent processing activities by Facebook) since Fashion ID exerts a decisive influence on the collection and transmission of the personal data of visitors to that website to the provider of that plugin, Facebook Ireland, which would not have occurred without that plugin.[74]

In the case C-25/17 (**Jehovan todistajat**) the ECJ elaborated on joint controllership between a religious community and its members in respect of notes taken by its members who engage in door-to-door preaching (i.e. notes about the people visited by the members of the community).

---

[67] ECJ 05 June 2018, C-210/16, *Wirtschaftsakademie Schleswig-Holstein*, § 43
[68] Art 4(7) GDPR.
[69] ECJ 29 July 2019, C-40/17, *Fashion ID*, §§ 64, 70.
[70] The ECJ refers to Art 2(b) Directive 95/46/EG (Data Protection Directive) which is quite similar to the one in Art 4(2) GDPR.
[71] ECJ 29 July 2019, C-40/17, *Fashion ID*, § 72
[72] ECJ 29 July 2019, C-40/17, *Fashion ID*, § 74
[73] Ibid.
[74] ECJ 29 July 2019, C-40/17, *Fashion ID*, § 78, 85.

The ECJ considered the religious community and its members as joint controllers, since it encourages its members who engage in preaching to carry out data processing in the context of their preaching activity.[75]

The religious community has relevant influence on the processing activity by organising, coordinating, and encouraging the preaching activities of its members intended to spread its faith and therefore participates, jointly with its members who engage in preaching, in determining the purposes and means of processing of personal data of the persons contacted.[76]

2.2.8. Applying above considerations on joint research projects

The role allocation in a scientific research project requires further elaboration. In every research project that includes various project partners within a consortium, different partners can potentially take one of the roles described above regarding processing activities concerning personal data.

Any partner that determines the purposes and means of processing personal data is to be considered a controller according to Art 4 (7) GDPR, even if data are processed in a scientific context. If these purposes and means are not determined solely by one of the partners of a scientific consortium, it is possible, that all partners who determine these purposes jointly with each other, can be considered joint controllers.

This section determines at which point two entities are to be qualified as joint controllers, that in many ways share responsibilities under the GDPR and further discusses if this is relevant for this research project.

There will always be an intertwined relationship between the partners, seeing as every joint research project has an overall goal and therefore there is at least to some degree a joint purpose of the association. However, this does not necessarily lead to “joint controllership” in the sense of the GDPR by all the partners and for any processing of personal data, as will be shown below.

The preceding section shows, that the ECJ has recently applied a very broad understanding of jointly determining the purposes and means of the processing (as joint controllers). Before the recent case law by the ECJ, starting with the case “Wirtschaftsakademie Schleswig-Holstein”, the literature was quite reluctant to apply the concept of joint controllership and rather considered controller-processor relationships or separate controllership for the respective processing activities instead. Only after this landmark

[75] ECJ 10 July 2018, C-25/17, *Jehovan todistajat*, § 72.
[76] ECJ 10 July 2018, C-25/17, *Jehovan todistajat,* § 73, 75.

decision every joint project required a specific evaluation on whether or not this case law on joint controllership applies to processing conducted in that project.

In this regard, clarification on how to apply these complex role concepts was required, since some concepts (especially that of the "joint controllership") have suddenly become highly relevant in light of the recent decisions of the ECJ. The Guidelines of the EDPB on the concepts of controller and processor in the GDPR (07/2020) provide various clarifications on the different concepts. Two of the many examples provided in the Guidelines focus on a research project:

> *"Several research institutes decide to participate in a specific **joint research project** and to use to that end the existing platform of one of the institutes involved in the project. **Each institute feeds personal data** it holds **into the platform** for the purpose of the joint research and uses the data provided by others through the platform for carrying out the research. **In this case**, **all institutes qualify as joint controllers** for the personal data processing that is done by storing and disclosing information from this platform since they have decided together the purpose of the processing and the means to be used (the existing platform). Each of the institutes however is **a separate controller for any other processing** that may be **carried out outside the platform** for their respective purposes."*[77]

This example concerns the joint collection of data by various research institutes. If all partners of a research institute collect and share data via a common platform with each other for a joint purpose, all of them are joint controllers. The EDPB highlights, however, that there are various processing activities carried outside of this shared platform, for which there is no joint controllership.

> "*A health care provider (the investigator) and a university (the sponsor) decide to launch together a **clinical trial with the same purpose**. They collaborate together to the **drafting of the study protocol** (i.e. purpose, methodology/design of the study, data to be collected, subject exclusion/inclusion criteria, database reuse (where relevant) etc.). They may be considered as joint controllers, for this clinical trial as they jointly determine and agree on the same purpose and the essential means of the processing. The collection of personal data from the medical record of the patient for the purpose of research is to be distinguished from the storage and use of the same data for the purpose of patient care, for which the health care provider remains the controller.*
>
> *In the event that the **investigator does not participate to the drafting of the protocol** (he just accepts the protocol already elaborated by the sponsor), and the protocol is only designed by the sponsor, the investigator **should be considered as a processor** and the sponsor as the controller for this clinical trial.*"

[77] EDPB, Guidelines 07/2020 on the concepts of controller and processor in the GDPR v2.1, 07 July 2021, 22 [highlighted by the authors]

This example provided by the EDPB concerns a specific clinical trial. The study protocol is very specific and – as described in the example – clearly defines the means of the processing.

Another example provided by the Guidelines concerns the distinction between joint controllership and a controller-processor relationship:

> *"Company ABC, the developer of a blood pressure monitoring app and Company XYZ, a provider of apps for medical professionals, both wish to examine how blood pressure changes can help predict certain diseases. The companies decide to set up a joint project and reach out to Hospital DEF to become involved as well.*
>
> *The personal data that will be processed in this project consists of personal data which Company ABC, Hospital DEF and Company XYZ are separately processing as individual controllers. The decision to process this data to assess blood pressure changes is taken jointly by the three actors. Company ABC, Hospital DEF and Company XYZ have jointly determined the purposes of processing. Company XYZ takes the initiative to propose the essential means of processing. Both Company ABC and the Hospital DEF accept these essential means after they as well were involved in developing some of the features of the app so that the results can be sufficiently used by them. The three organizations thus agree on having a common purpose for the processing which is the assessment of how blood pressure changes can help predict certain diseases. Once the research is completed, Company ABC, Hospital DEF and Company XYZ may benefit from the assessment by using its results in their own activities. For all these reasons, they qualify as joint controllers for this specific joint processing.*
>
> *If Company XYZ had been simply asked by the others to perform this assessment without having any purpose of their own and merely been processing data on behalf of the others, Company XYZ would qualify as a processor even if it was entrusted with the determination of the non-essential means."*[78]

In this last example, the EDPB highlights the benefits for the parties involved and the joint decision taken by the companies and the hospital, which, at least in the opinion of the EDPB makes them joint controllers rather than one of them being the processor of the others.

All of these examples are not directly applicable to the use case of the processing of image data sets. The use case in the current project would be, that the consortium defines an at least generally shared purpose with specific goals, that the partners of the consortium want to achieve. Each partner then uses one image data set – chosen by each partner on their own to be the basis of the research of various factors of explainable facial recognition technology – separately. The results are only discussed in the consortium.

[78] EDPB, Guidelines 07/2020 on the concepts of controller and processor in the GDPR v2.1, 07 July 2021, 23.

It appears, that the overarching purpose and goal of the project does, however, not allow to determine the purpose of all of the separate processing activities conducted by the different partners for their specific research focus in the project in the sense of Art 4 (7) GDPR. Rather it appears that all of these processing activities are – similar to the first example mentioned above and the processing carried out outside the shared platform – carried out as separate controllers.

There are several other considerations, which lead to this conclusion:

- Firstly, it can be stated that it is likely that not all partners of a consortium are included in the determination of the means and purposes of the data processing, which take place in the project, because every partner brings different expertise into the project. The cumulation of this expertise serves the aim of the project.

- The processing activity can be divided in several processing steps, which must be assessed separately.

- It can also be possible that no personal data need to be processed from some partners that goes beyond the employment of their researchers at all – even though it isn't required that a joint controller participates operationally to be referred to as a controller, it is extremely unlikely that these partners will have any influence on the key elements of the processing activities.[79]

We can see that the assessment of sole or joint controllership is a multi-layered task, which needs to be evaluated by following a holistic approach. For a valid elaboration it is necessary to determine the actual individual processing activities taking place. As we only want to evaluate the legal aspects of the use of the determined data sets in this deliverable, we will restrict the subject matter to the processing activities that will be conducted in this context.

First, the data sets will not be shared on a joint platform and second, every partner must decide by themselves which data they are processing and for which purpose. Furthermore, each institute must decide by themselves and with their own Data Protection Officer (DPO), which data sets they will use within the project. No joint elaboration will be taking place on this subject. It will be recommended not to share the data sets on a joint platform or in any other way with the other partners to guarantee a separate controllership.

Although all the partners are involved in the same project and decided about overall purpose of this project, it can be possible to separate the influence on the processing activities itself. Every partner has his tasks according to the proposal of the project. These tasks must be fulfilled in a certain time. For the fulfilment of the tasks every partner is free in his method of operation and can decide how the aim of the task will be achieved.

Also, the decision about the essential elements of the means shall be conducted by the partners solely. Essential elements of the means can be, which data will be processed, how long it will be processed, who has access to the data set, et cetera. In all processing

[79] *Wyrobek* in Knyrim, DatKomm Art 26 DSGVO 20 (1.6.2021, rdb.at).

activities the partners must be seen and treated as sole controllers for their processing of personal data. To be in accordance with the GDPR also the other provisions must be considered to legitimate process personal data.

Regardless of the explanations above there is another crucial fact not considered at this point of the role allocation and hence, the decision who is responsible for the compliance with the obligation laid down by the GDPR. Due to the international association in different areas of the scientific field, it can sometimes occur, that it is unclear which legal provisions are applicable for whom. The consortium of the current project, for example, not only includes partners from the EU, but also one partner from Switzerland, where the GDPR in this particular case – due the circumstance of EPFL being a university, which is not offering any goods or services et cetera – is not applicable.

If the processing activities can be separated clearly, the legal assessment is relatively clear. Switzerland does not fall under the scope of the GDPR, as pointed out above, and hence, the provisions of the GDPR are not applicable for the processing of the Swiss partner, EPFL. So, to decide which data sets EPFL can use, the Federal Swiss data protection law is pertinent, and the selection must be assessed under the Swiss provisions. The Swiss Data Protection is quite similar to European standards, but there are several differences compared to the European framework, which must be considered.

So, if we determine no joint controllership at all for the processing activities, the problem seems relatively obsolete, because third party-actors must act in accordance with their national data protection law.

2.4 Transfer of personal data to third countries

Since the project partner EPFL is located in Switzerland (no Member State of the European Union and therefore a *"third country"*), the question arises whether transmissions of personal data to the Swiss partner constitute "transfers of personal data to a third country".

Chapter V of the GDPR (Art 44 – 50) regulates the transfer of personal data to third countries and international organisations. Its regulatory measures are based on the assumption that third countries in general do not grant the same level of data protection as the GDPR does.[80] Art 44 GDPR states that:

> *"Any **transfer** of personal data […] to a third country or to an international organisation shall take place only if, **subject to the other provisions of this Regulation, the conditions laid down in this Chapter are complied with by the controller and processor**, including for **onward transfers** of personal data from the third country or an international organisation to another third country or to another international organisation. All provisions in this Chapter shall be applied in order to*

[80] *Jahnel/Pallwein-Prettner*, Datenschutzrecht[3] (2021) 91.

ensure that the level of protection of natural persons guaranteed by this Regulation is not undermined.'

Whenever Chapter V of the GDPR applies, the exporter, either a controller or a processor, has to ensure compliance with the regulatory measures of Chapter V.

In order to determine the scope of Chapter V GDPR, it is of great importance to first define the term "*transfer"* of personal data to third countries.

2.4.1. Defining the term "transfer of personal data to third countries"

The term "*transfer*" is neither explicitly defined in Chapter V GDPR, nor in Article 4 GDPR ("Definitions"). The term *"cross-border processing"* defined in Article 4 (23) GDPR might at first glance seem relevant for the interpretation of the term *"data transfer"*. Cross-border processing, however, solely addresses situations where a controller or processor has multiple establishments **inside** one or more Member States. Therefore, transfer of personal data to third countries is not being addressed.

When defining the term *"data transfer"* it is helpful to examine various translations of the term, as different versions of the GDPR can lead to different results in the interpretation of the term in question:

The German version of Art 44 GDPR uses the term *"Daten**übermittlung**"* for data transfer. Under the German translation that same term is also being used as an example of processing operations covered by the definition of processing under Art 4 (2) GDPR ("***Übermittlung** durch Offenlegung**"***). Similarly, the Hungarian version of the GDPR uses the term *"(adat)**továbbítás**"* both for data transfer (Art 44 GDPR) and for an example of processing operations in Art 4 (2) GDPR (*"közlés **továbbítás**"*).

By equating the term "*transfer"* ("*Übermittlung*") that is being used in Art 44 GDPR with the term "*disclosure by transmission"* ("*Übermittlung durch Offenlegung*"), as the German and the Hungarian version of the Regulation arguably do, disclosure of personal data by transmission to a processor in a third country does not fall under Chapter V of the GDPR: A processor is no third party but a recipient according to the GDPR.[81] Therefore, personal data cannot be disclosed to a processor by "transmission".

You can not assume, that the exclusion of any data being disclosed to processors in third countries was the intention of the legislator, since the main aim of Chapter V GDPR is to *"ensure that the level of protection of natural persons guaranteed by this Regulation is not undermined"*.[82]

[81] *Pauly* in Paal/Pauly, DS-GVO BDSG³ (2021) Art 44 GDPR 3.

[82] Art 44 GDPR.

Other versions of the GDPR, such as the French version and the English version, explicitly differentiate between the term *"transfer"* (in French: *"transferts"*) that is being used in Article 44 GDPR, and the term "disclosure by *transmission"* (in French: *"transmission"*) which is mentioned as an example of data processing in Art 4 (2) GDPR.

The distinction between *"transmission",* Art 4 (2) GDPR, and *"transfer",* Art 44 GDPR, that can be found in multiple versions of the regulation, makes it clear that the legislator did not intend to equate those two terms and therefore did not want to exclude the transfer of data to processors located in third countries from Chapter V GDPR.

This leads to the first conclusion that even the transfer of data to a processor in a third country falls under the term data transfer as used in Art 44 GDPR. This essentially means that any transfer of personal data from a controller, joint controller or processor to another controller, joint controller or processor in a third country constitutes "a transfer of personal data to a third country". The question of proper role allocation is hence of little importance for questions regarding the scope of Chapter V GDPR.

According to the prevailing opinion in literature, **any processing operation that transfers personal data out of EU territory or that makes personal data accessible from outside of the EU** falls under the term *"data transfer to a third country"* as defined in Art 44 GDPR.[83]

This very broad interpretation of the term is criticized due to the very broad scope it leads to, especially with regards to the internet. The European Court of Justice rightfully stated that if the term "transfer":

> *"were interpreted to mean that there is 'transfer [of data] to a third country' every time that personal data are loaded onto an internet page, that transfer would necessarily be a transfer to all the third countries where there are the technical means needed to access the internet. Thus, if the Commission found […] that even one third country did not ensure adequate protection, the Member State would be obliged to prevent any personal data being placed on the internet."*[84]

A too broad interpretation of the term "transfer" would make Chapter V a *"regime of general application"*[85] with regards to the internet. Therefore, the European Court of Justice used to hold a differing opinion regarding the interpretation of data transfer. In its *Case C-101/01 Bodil Lindquist* judgement from 2003, the ECJ held that there is no transfer of personal data to a third country where personal data is being uploaded onto an internet page, despite thereby making those data accessible to anyone connecting to the internet, including people

[83] *Ehmann/Selmayr*, *DS-GVO: Datenschutz-Grundverordnung²* (2018) Art 44 GDPR 7; *Knyrim* in Knyrim, DatKomm (1.10.2018, rdb.at) Art 44 GDPR 19.

[84] ECJ C-101/01, *Lindqvist*, ECLI:EU:C:2003:596 para 68.

[85] ECJ C-101/01, *Lindqvist*, ECLI:EU:C:2003:596 para 69.

in third countries.[86] The ECJ suggested that **a data transfer to a third country should be an active act, and not simply making personal data passively accessible**.[87] This interpretation of the term greatly narrows down the meaning of *'data transfer'* and thereby the scope of Chapter V GDPR.

It is unsure, whether the ECJ still upholds this narrower interpretation of the term today: In its *Case C-362/14 Schrems I* judgement, the ECJ argued, that any transfer of personal data from a member state to a third country constitutes a processing of personal data and did not differentiate between active acts of transferring data and passively making personal data accessible.[88] The ECJ has since also started to put much more emphasis on the negative impact internet search engines[89] and the publication of data on websites in general[90] can have on fundamental rights such as Art 7 and Art 8 of the EU Charter. It is therefore possible that the ECJ would today revise its earlier case law and decide differently in a case similar to *Case C-101/01 Bodil Lindquist* by interpreting the term data transfer broader.[91]
With its Guidelines 05/2021, the *EDPB* recently tried to clarify the term "data transfer" in order to create legal certainty and came up with the following three cumulative criteria that qualify a processing as a transfer:

1. A controller or a processor is subject to the GDPR for the given processing.
2. This controller or processor (exporter) discloses by transmission or otherwise makes personal data, subject to this processing, available to another controller, joint controller or processor. (importer)
3. The importer is in a third country or is an international organization, irrespective of whether this importer is subject to the GDPR in respect to the given processing in accordance with Article 3.[92]

The guidelines of the EDPB are not binding but have already been cited in recent rulings of National Data Protection Authorities[93] and will most likely also affect the interpretation of the term "data transfer" by the ECJ in future cases. The rather short guidelines did not answer

[86] ECJ C-101/01, *Lindqvist*, ECLI:EU:C:2003:596 para 71.

[87] *Kuner/Docksey/Bygrave*, Commentary on the EU general data protection regulation (GDPR). A commentary (2019) 763.

[88] ECJ C-362/14, *Schrems,* ECLI:EU:C:2015:650 para 45; *Ehmann et al*, DS-GVO Art 44 GDPR Rn 8.

[89] ECJ C-131/12, *Google Spain and Google,* ECLI:EU:C:2014:317 para 38.

[90] ECJ C-92/09 and C-93/09, *Volker und Markus Schecke and Eifert* ECLI:EU:C:2010:662 para 46.

[91] *Ehmann/Selmayr*, *DS-GVO: Datenschutz-Grundverordnung²* (2018) Art 44 GDPR 8; *Kuner/Docksey/Bygrave*, Commentary on the EU general data protection regulation (GDPR). A commentary 763.

[92] *EDPB*, Guidelines 05/2021 on the Interplay between the application of Article 3 and the provisions on international transfers as per Chapter V of the GDPR (2021) 4.

[93] *DSB (Austrian Data Protection Authority),* 2021-0.586.257 (D155.027) 31.

the question whether passively uploading data constitutes a transfer or not. Therefore, this deliverable will on the one hand rely on the cumulative criteria created by the EDPB, when defining the term "*data transfer*". On the other hand, the deliverable will also take into account the ECJ's differentiation between passively uploading data (no transfer) and active acts of transfers (transfer).

Criterion 1 of the EDPB Guidelines requires, that the processing meets the requirements of Art 3 GDPR.[94] In XAIFace, any processing conducted by the controllers established in France, Portugal or Austria falls within the scope of the provisions of the GDPR (See Section 2.2.1.).

Criterion 2 requires the exporter to make the data available to the importer, who can be another controller, a joint controller or a processor. This criterion is not fulfilled where data are disclosed directly and on their own initiative by the data subjects, since no controller or processor is making data available.[95]

Criterion 3 requires the importer to be in a third country. As already mentioned, the Swiss partner EPFL is such an importer as it is located in a third country.

Therefore, according to the EDPB Guidelines, whenever a partner of XAIface who is subject to the GDPR (EURECOM, Joanneum Research, Instituto de Telecomunicações, University of Vienna) decides to make personal data accessible to the Swiss partner EPFL, this act of processing constitutes a "transfer".

This would for example be the case, if one partner decided to directly transmit data to the Swiss partner EPFL. It would also constitute a transfer of personal data, if a partner shared data sets via a joint platform with the Swiss partner. (This deliverable, however, recommends **not** to share the data sets on a joint platform or in any other way with the other partners).

Whenever the Swiss partner shares personal data with another partner of the project, such an act of processing does not constitute an international data transfer, since EPFL is not subject to the GDPR (Criterion 1).

Neither does it constitute an international data transfer, if a partner of a third country gains access to data sets by a provider, as long as the provider is not subject to the GDPR. (If the provider of data sets is subject to the GDPR, it constitutes a transfer of personal data to third countries and the provider of the data sets (the exporter) must ensure compliance with Chapter V of the GDPR [the importer].)

In addition to the transfer of personal data to third countries, Chapter V also covers any "*onward transfer*" following the initial transfer of data from an exporter to an importer. The

[94] *EDPB*, Guidelines 05/2021 on the Interplay between the application of Article 3 and the provisions on international transfers as per Chapter V of the GDPR (2021) 5.

[95] *EDPB*, Guidelines 05/2021 on the Interplay between the application of Article 3 and the provisions on international transfers as per Chapter V of the GDPR (2021) 5.

term "*onward transfer*" must be interpreted broadly, covering not only onward transfers to other third countries (as Art 44 GDPR suggests), but also transfers inside the initial third country.[96]

According to the wording of Art 44 GDPR, not only Chapter V of the GDPR but the whole Regulation applies to any onward transfer (*"subject to the other provisions of this Regulation"*). However, this does not mean that the GDPR fully applies to any upcoming processing following the initial transfer (the territorial scope of the GDPR is defined in Art 3, not in Art 44 GDPR). The scope of the GDPR is solely extended also to cover any act of onward transferring of personal data following the initial data transfer:

Hence, whenever the Swiss partner decides to transfer personal data - that was initially made accessible to him by a project partner that is subject to the GDPR - inside Switzerland or to another third country, the GDPR still applies to that onward transfer. Further processing that occurs following the onward transfer itself does not fall under the extended scope of the GDPR.

2.4.2. Publishing results online

It is rather questionable, whether the publication of results on the internet also constitutes a transfer of personal data to third countries. All three criteria of the EDPS' Guidelines would be fulfilled: Partners of XAIFace that are subjects to the GDPR, make personal data accessible to importers located in third countries. Such a broad scope of Chapter V GDPR could, however, not have been the intention of the European Legislator:

First, at the time the Directive 95/46 and with it the first provisions regarding international data transfer were drawn up, due to the state of development of the internet back then, it should not be presumed that the legislator intended the term transfer to cover any loading of personal data onto an internet page.[97]

Secondly, as already stated before, a too broad scope would make Chapter V a *"regime of general application"*[98] with regards to the internet:

Any upload of personal data on internet pages by an exporter would constitute a transfer of personal data. Since Chapter V GDPR essentially obliges the exporter of personal data to ensure an essentially equivalent level of data protection compared to the EU standard by complying with Chapter V's regulatory framework, an exporter uploading data on an internet page would be obliged to ensure an adequate level of data protection in any third country, where people have the technical means to access the data online. This would include most

[96] *Pauly* in Paal/Pauly, DS-GVO BDSG³ (2021) Art 44 GDPR 13.

[97] ECJ C-101/01, *Lindqvist*, ECLI:EU:C:2003:596 para 68.

[98] ECJ C-101/01, *Lindqvist*, ECLI:EU:C:2003:596 para 69.

third countries worldwide, making any exporter face an arguably unfulfillable task. Such a broad interpretation of the scope of Chapter V would mean, that the vast majority of uploads of personal data onto the internet happening right now violates the GDPR, hence no exporter can ensure an adequate level of data protection in any third country, where people have access to the internet.

Therefore, the most convincing arguments speak for an exclusion of any act that "simply makes data passively accessibly by uploading them onto the internet" from the scope of Chapter V GDPR. By following the ECJ's opinion on the interpretation of the term "transfer" and its differentiation between active acts of transfers and simply making data passively available, this deliverable concludes that the publication of results online does not fall under the scope of Chapter V GDPR.

2.4.3. Regulatory framework of Chapter V GDPR

Whenever partners of XAIface who are subject to the GDPR (EURECOM, Joanneum Research, Instituto de Telecomunicações, University of Vienna) decide to make personal data accessible to the Swiss partner EPFL, they have to ensure compliance with the regulatory measures of Chapter V GDPR.

In regulating international data transfer, the GDPR follows a two-step approach: [99]

First, international transfer of personal data has to comply with all other relevant provisions of the GDPR, before personal data may be transferred outside of the EU.[100] Therefore, any international transfer of personal data has to be for example in compliance with the principles of data processing as lied out in Art 5 GDPR, has to be lawful according to Art 6 GDPR, Art 9 GDPR and Art 10 GDPR, etc (just as any other processing of personal data inside the European Union).

Secondly, any transfer of personal data to third countries has to happen in accordance with Chapter V of the GDPR (*two-step approach*[101]). Chapter V contains a complete list of methods to transfer personal data to third countries and introduces a three-tiered structure[102], consisting of the adequacy decision, appropriate safeguards, and derogations - with the adequacy decision on top and the derogations at the bottom: Only if no adequacy

[99] *Kuner/Docksey/Bygrave*, Commentary on the EU general data protection regulation (GDPR). A commentary 757.

[100] ibid.

[101] Ibid.

[102] *Kuner/Docksey/Bygrave*, Commentary on the EU general data protection regulation (GDPR). A commentary 764.

decision has been reached, appropriate safeguards should be used, and only if appropriate safeguards have not been implemented, derogations should be relied upon.[103]

Since 26th of July 2000, a Commission Decision on the adequate protection of personal data provided in Switzerland is in place,[104] ensuring an essentially equivalent level of data protection compared to that guaranteed by the European Union.

As long as this adequacy decision remains valid, any transfer of personal data by a European project partner to the EPFL that relies upon this adequacy decision and complies with any other provisions of the GDPR (two-step approach) occurs in compliance with the GDPR.

[103] *Kuner/Docksey/Bygrave*, Commentary on the EU general data protection regulation (GDPR). A commentary 765.

[104] 2000/518/EC: Commission Decision of 26 July 2000 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequate protection of personal data provided in Switzerland (notified under document number C(2000) 2304) (Text with EEA relevance.) OJ L 215.

3. Lawfulness of processing in a scientific context under the GDPR

The GDPR recognises the importance of scientific research. However it neither exempts processing activities for scientific research from its scope,[105] nor are processing activities for scientific research purposes generally to be considered lawful under the GDPR.[106] Instead the GDPR allows member states, within certain limits, to establish additional specific rules for the processing of personal data for research purposes e.g. exemption of data subject rights obligations et cetera, which should facilitate the research work and drive innovation.

Nonetheless, for the processing activities in the context with scientific research, a legal base for the processing is pertinent and the other requirements laid down by the GDPR have to be considered. This chapter includes a brief outline on the concept of lawful processing under the GDPR and continues with the demonstration of the specificities of the processing for scientific research purposes. Finally, it will be examined, whether and under which conditions and wrongfully acquired data may be further processed.

3.1. Lawfulness of processing personal data

According to Article 2(1) GDPR the Regulation applies to the processing of personal data fully or partially by automated means and to the processing other than by automated means of personal data which form part of a filing system or are intended to form part of a filing system. The notion of personal data must be understood in a broad sense and follows a relative understanding, which concludes the comprehensive evaluation of certain processing conditions. Cases can occur, where personal data will be processed, but due to the exemptions laid down in Article 2 (2) GDPR, the GDPR is not applicable e.g., personal and households' activities – although this exemption cannot be interpreted too extensively.

In cases, where no exemptions apply, all the obligations, which are laid down in the regulation are usually mandatory and must be – if applicable – considered in a certain processing situation by the controller.[107] The main addressee of most of the obligations is the controller, who decides about why and how the processing takes place. The obligations, which the controller has adhere to, are not laid down in one catalogue of obligations. They must be understood as the holistic concept of the GDPR and can be derived from various – not always concise – provisions within the entire legal act.

But the GDPR also recognises processing activities, which will be treated in a privileged way by the data protection framework e.g. processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes. This can especially be derived from Chapter IX GDPR "Provision relating to specific processing situations". In this processing situations national law must be considered if an opening clause applies.

[105] Such an exemption from the material scope exists only for the processing of personal data "*by a natural person in the course of a purely personal or household activity*" (see Art 2(2)(c) GDPR).
[106] Compare to Art 6 and 9 GDPR.
[107] There might be some exemption if there is an opening clause laid down and a national additional legislation was enacted e.g. Art 89 GDPR or Art 9 (4) GDPR.

Due the system of the GDPR the obligations, which the controller must adhere to, can be derived from several provisions.

The general obligations of a controller include e.g.:

- Principles of Processing laid down in Article 5 GDPR
- Lawfulness of the Processing according to Article 6 GDPR and following
- Guarantee of Data Subject Rights laid down in Chapter III
- General Obligations of the Controller and Processor laid down in Chapter IV;

Every processing activity in the sense of Art 2 (1) GDPR must be conducted in a manner consistent with the "principles" defined in chapter II GDPR. This includes the adherence to the principles under Art 5 GDPR as well as the general requirements of the lawfulness of processing personal data (Art 6 GDPR) and – if applicable – the specific requirements for special categories of personal data (Art 9 and 10 GDPR). In this specific context special consideration will also be given to the specific provision in Art 6 (4) GDPR which concerns "further processing" of personal data (i.e. for purposes other than those for which the personal data have been collected).

Art 5 GDPR contains a multiplicity of principles, which demonstrate the general paradigm of European data protection law. Although, the principles were formulated in a programmatic way, they are more than mere symbolic affirmations.[108] Their obligatory character must evidently be considered, especially due the amount of the administrative fines. Any infringement of the principles can be fined up to 20 million or up to 4% of the total worldwide annual turnover according to Art 83 GDPR.[109] The substantive content of Art 5 GDPR will additionally be underlined by the circumstance that member states can enact exemptions or derogations from the principles especially in certain processing situations like scientific purposes.[110]

The principles according Art 5 GDPR are the following:

- Lawfulness, fairness, and transparency
- Purpose limitation
- Data minimisation
- Accuracy
- Storage limitation
- Integrity and confidentiality
- Accountability

Art 5 (2) GDPR can be divided in two obligations. Firstly, the controller must process in accordance with the above-mentioned principle and secondly the controller must also demonstrate compliance with Art 5 (1) GDPR. Hence, the controller is accountable for the

[108] *Martini* in Paal/Pauly, DS-GVO BDSG Art 5 DSGVO 2; otherwise *Hötzendorfer/Tschohl/Kastelitz* in Knyrim, DatKomm Art 5 DSGVO (Stand 7.5.2020, rdb.at) 11.
[109] Art 83 (5) lit a GDPR.
[110] *Martini* in Paal/Pauly, DS-GVO BDSG Art 5 DSGVO 2.

data processing: “accountability”.[111] Additionally, the burden of proof is on the controller as the central actor of the GDPR, who must be able to demonstrate the compliance with data protection law.[112] However, due to the quite broad scope of accountability, this imposition of duty should not be interpreted too extensively and must be judged according to the principles of proportionality in regard of the nature and scope of the data processing and also the risks, which are likely to occur (“risk-based-approach of the GDPR).[113] Also, due to the undetermined formulation, the wide scope of the principles and the intertwined character of the principles it is hardly possible to process in full accordance with Art 5, which of course leads to application uncertainties by the controller and fosters legal uncertainty for the data subjects.

Since this chapter shall point out what lawfulness of data processing means, it must be evaluated, what lawfulness within the GDPR regime in concrete terms concludes. The principle of lawfulness and the general concept of processing only on the base of a legal ground can be seen as indications for the notion of lawfulness. The notion “lawfulness” is not further defined in the GDPR.

Art 5 (1) lit a GDPR lays down the principle of ”lawfulness, fairness and transparency”.

“*Personal data shall be processed lawfully, fairly and in a transparent manner in relation to the data subject (‘lawfulness, fairness and transparency’);”*

Also, the principle of lawfulness can either be understood exceedingly broad due to its generic formulation “*personal data should be processed lawfully (…)*”[114] or in a narrower sense. The narrow understanding of lawfulness contains the lawfulness of the data processing itself in the sense of a legal base as a requirement for the processing activity as it is also referred to in the recital 40:

“*In order for processing to be lawful, personal data should be processed on the basis of the consent of the data subject concerned or some other legitimate basis, laid down by law, either in this Regulation or in other Union Member State law as referred to in this Regulation, including the necessity for compliance with the legal obligation to which the controller is subject or the necessity for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract.”*

Otherwise, it could be conceivable, that the principle of lawfulness can be interpreted in a more generic understanding. The obligation of lawfulness also includes acting in accordance with the legal order and therefore to comply with all obligations and requirements laid down within the GDPR, which however shouldn´t be explicitly addressed in a legal act in a separate provision, which is why this broad interpretation must rather be rejected.[115] The immanence of any legal act is achieving compliance with the rules it is based on.

---

[111] *Hötzendorfer/Tschohl/Kastelitz* in Knyrim, DatKomm Art 5 DSGVO 57 (7.5.2020, rdb.at).
[112] *Hötzendorfer/Tschohl/Kastelitz* in Knyrim, DatKomm Art 5 DSGVO 58 (7.5.2020, rdb.at).
[113] *Hötzendorfer/Tschohl/Kastelitz* in Knyrim, DatKomm Art 5 DSGVO 59 (7.5.2020, rdb.at).
[114] Art 5 (1) lit a GDPR.
[115] *Hötzendorfer/Tschohl/Kastelitz* in Knyrim, DatKomm Art 5 DSGVO (7.5.2020, rdb.at) 12.

So, lawfulness within Art 5 GDPR requires presence of a legal base in order to process data lawfully. The paradigm of the GDPR assumes that any nature of data processing is generally prohibited, except the controller can rely on a legal ground according to Chapter II GDPR. An additional legal base is required for processing special categories of personal data.

Art 6 (1) GDPR is the general provision regarding the lawfulness of the processing, it constitutes:

"*1. Processing shall be lawful only if and to the extent that at least one of the following applies:*

*(a) the data subject has given **consent** to the processing of his or her personal data for one or more specific purposes;*

*(b) processing is necessary **for the performance of a contract** to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract;*

*(c) processing is **necessary for compliance with a legal obligation** to which the controller is subject;*

*(d) processing is necessary in order **to protect the vital interests** of the data subject or of another natural person;*

*(e) processing is necessary **for the performance of a task carried out in the public interest or in the exercise of official authority** vested in the controller;*

(f) *processing is **necessary for the purposes of the legitimate interests** pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child*"

Furthermore, Article 7 GDPR lays down certain conditions for consent, e.g. the possibility of withdrawal the consent. Art 8 GDPR statutes further conditions for a child´s consent, since for the protection of children stricter requirements should apply.

Art 9 GDPR assumes that there is information, which needs due to its immanence a higher level of protection, the so called "special category of personal data" e.g. "*data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation*". The processing of sensitive data requires according to the common opinion an additional legal base from Art 9 (2) GDPR pursuant the data protection framework.

Art 10 and 11 GDPR refers to data processing relating to criminal convictions and offences, and which do not require the identification of the data subject.

3.2. Specificities in the context of scientific research

European Law in general is known for its drive for harmonisation and the aim to reduce disparities between the different legal systems of the member states, so the internal market can be strengthened while removing or at least reducing barriers or in the case of data protection obstacles to flow of personal data within the European Union.[116] This harmonisation of the different legal systems should ensure a consistent and high protection for personal data and therefore the data protection provisions should be equivalent in all member states.

However, due to the principle of conferral and subsidiarity, European law cannot act in a fully harmonising manner, which is why a middle way is sought in many regulatory areas.[117] European law opts also for a middle course in data protection law, which is indeed regulated jointly by the European Union and the member states; it follows, that both national and supranational provisions apply in these areas, which of course can lead to divergences in application instead of preventing these and also immanently reducing the uniform level of protection in many cases.

Data protection law is thus characterised through its co-regulation between the Member States and the European Union. Therefore, in the *verba legalia* of the GDPR a multiplicity of opening clauses was enacted, whereby the member states have a regulatory participation competence, partly also an obligation to regulation. The range of regulatory participation competence of the member states is largely indeterminate but usually the national legislation should specify or concretise European law and not replace it, but the range of this specification and concretisation remains largely unanswered. The nature of these opening clauses can be divided in obligatory and facultative, which means some of them must be implemented and for some of them it is up to the national legislator if additional provisions besides the GDPR will be enacted or not.

Particularly when it comes to data processing in research, the fundamental right to data protection of the individual and the fundamental right to freedom of research are contrary to each other, which must be reconciled by regulatory mechanisms. Especially in the area of research national provisions are pertinent. This chapter should provide a brief overview of the relevant opening clauses in the context with scientific research data processing.

The national implementation – if existent –must be assessed separately by the project partners and will not be content of this deliverable.

[116] *Hoffmann/Miscenic*, The Role of Opening Clauses in Harmonization of EU Law: Example of the EU's General Data Protection Regulation (GDPR), EU and comparative law issues and challenges series (ECLIC), 2020, 44-61.
[117] Art 5 TEU.

Regarding XAIface the following opening clauses could be of interest:[118]

Provision	Content	Wording
Art 6 (4) GDPR	Processing for a purpose other than that for which the personal data have been collected	"Where the processing for a purpose other than that for which the personal data have been collected is not based on the data subject's consent or on a Union or Member State law (…) ."
Article 9 (2) (j) GDPR	Processing for the purposes of archiving, scientific or historical research or statistics	"Paragraph 1 shall not apply if one of the following applies: (j) processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law (…) ;"
Article 9 (4) GDPR	Conditions and restrictions for processing of genetic, biometric and health data	"Member States may maintain or introduce further conditions, including limitations, with regard to the processing of genetic data, biometric data or data concerning health."
Article 49 (5) GDPR	Legal limitations to transferring specific categories	"In the absence of an adequacy decision, Union or Member State law may, for important reasons of public interest, expressly set

[118] For a comprehensive listing, see: *Chakarova* Kristina, European Union Law Working Papers – No. 41 General Data Protection Regulation: Challenges Posed by the Opening Clauses and Conflict of Law Issues: https://law.stanford.edu/publications/no-41-general-data-protection-regulation-challenges-posed-by-the-opening-clauses-and-conflict-of-laws-issues/ (10.06.2022); and also *Feiler* Lukas, Öffnungsklauseln in der Datenschutz-Grundverordnung - Regelungsspielraum des österreichischen Gesetzgebers, jusIT 2016/5/93: https://lesen.lexisnexis.at/_/oeffnungsklauseln-in-der-datenschutz-grundverordnungregelungssp/artikel/jusit/2016/5/jusIT_2016_05_093.html (10.06.2022).

		limits to the transfer of specific categories of personal data to a third country or an international organisation."
Article 85 (1) GDPR	Reconciling the right to personal data protection with the right to freedom of expression and information	"Member States shall by law reconcile the right to the protection of personal data pursuant to this Regulation with the right to freedom of expression and information, including processing for journalistic purposes and the purposes of academic, artistic or literary expression"
Article 85 (2) GDPR	Processing for journalistic, academic, artistic or literary purposes	"For processing carried out for journalistic purposes or the purpose of academic artistic or literary expression, Member States shall provide for exemptions or derogations from (…) if they are necessary to reconcile the right to the protection of personal data with the freedom of expression and information."
Article 89 (2) GDPR	Derogations when processing for scientific or historical research purposes or statistical purposes	'Where personal data are processed for scientific or historical research purposes or statistical purposes, Union or Member State law may provide for derogations from the rights referred to in Articles 15, 16, 18 and 21 (…)"

3.3. Processing of Data(sets) which have been obtained from a Third Party

After having established the bases of lawful processing in the context of scientific research, we have to address the specific issue of processing personal data that have already been obtained, e.g. a face image data set that has been created by a third party that could be used for the training of an AI.

In the current collaborative digital business environment, it appears strange, that this legal issue has not already been addressed by various authors, given that many development activities depend on the processing of data by other parties. The existing literature on processing especially in the context of processing for scientific purposes focuses primarily on the processing of personal data that have been obtained by the primary controller himself, but not on ("existing") data from third parties that are now used for another purpose by another controller. Therefore, we have to analyse if and how the way data that should be used in a scientific project have been originally obtained (i.e. in a lawful manner or not) and in how far the controller of said project would have to take measures to ensure that these data have been obtained lawfully.

In order to address this topic with an open and scientific approach, we will consider different scenarios, starting with the most "extreme" case in which data have been obviously obtained unlawfully and working towards the most common (at least in a world-wide scientific research community) scenario wherein Controller B is not completely sure whether or not Controller A has obtained the data lawfully or not.

3.4. Processing Data that have been obtained unlawfully

As a starting point, the most obvious question would be: could personal data that have originally been obtained unlawfully (e.g., in a manner that is non-compliant with the GDPR) be processed lawfully under the GDPR?

While one would initially be inclined to answer that this could not be possible, there are some aspects that should be considered, that indicate that the processing of even unlawfully obtained data could, in some cases be considered lawful.

Rather recently, the Belgian Data Protection Authority (DPA) has ruled against the use of personal data as evidence in civil law proceedings, if these data have been obtained without

a legal basis under Art 6 GDPR.[119] In this case further processing of personal data, even in the context of legal proceedings, was deemed unlawful and was therefore prohibited.

While some practices of obtaining data used as evidence are dubious, in this case it appears to have been a simple mistake. Apparently the first defendant sent an e-mail to a third person as a result from his habit of sending e-mails to both the plaintiff and the second defendant, whereas he could have sent an e-mail to both the plaintiff and the second defendant concerning their notary practice and a separate e-mail to the plaintiff only concerning her personal company.[120] The DPA argued that even if unintentional, the transfer of personal data constituted processing of personal data and since the first defendant could have sent a separate e-mail, not including the second defendant, he did not adhere to the data minimisation principle and therefore the processing activity was deemed unlawful.[121]

The second defendant, who was the recipient of these e-mails, then went on to send these to his legal counsel. The DPA neither accepted the argument that the communication was privileged as attorney-client correspondence, nor did it deem the further processing of the data in pending or future legal proceedings possible without violation of the principles of lawfulness, loyalty and transparency.

While this decision of a national DPA cannot represent the EU as a whole, the opinion of a national DPA can still have an EU-wide influence if adopted by the European Data Protection Board. Nevertheless, at the current state it is still a singular decision that is contrasted to different national case law of the member states.

Another example would be the decision of the Maltese court that, in a similar case, stated that the “fruit of the poisonous tree” doctrine adopted by the US jurisprudence according to which unlawfully obtained evidence was rendered inadmissible in court proceedings, was alien to Maltese law.[122] More importantly the Maltese Court stated, that data protection law, even under the GDPR, does not create or provide for such an “exclusionary rule” in case if illegally obtained evidence.[123] Also, the Maltese Court referenced the European Court of Human Rights’ case law according to which the use of illegally obtained evidence is not in breach of the right to a fair trial (Art 6 ECHR).

The GDPR does in fact, not explicitly state, that in order to process data lawfully, the controller must ensure that the processed data are to be obtained lawfully (e.g., in a manner

---

[119] DPA (BE) Décision quant au fond n° 07/2021 du 29 janvier 2021, available under https://www.autoriteprotectiondonnees.be/publications/decision-quant-au-fond-n-07-2021.pdf [25.03.2022]; see also Van Beal & Bellis, Data Protection Newsflash, 25th March 2021, available under: https://www.vbb.com/media/Insights_Articles/Data_Protection_Newsflash_-_DPA_prohibits_use_of_personal_data_by_legal_counsel.pdf [25.03.2022].

[120] Ibid.

[121] Ibid.

[122] *Bugeja*, Can evidence obtained in breach of GDPR be lawfully used as evidence?, Times of Malta 30th June 2009, available under: https://timesofmalta.com/articles/view/can-evidence-obtained-in-breach-of-gdpr-be-lawfully-used-as-evidence.718126 [25.03.2022].

[123] Ibid.

compliant with the GDPR). This could indicate that this is not a requirement for lawful processing of personal data, i.e. that it could also be lawful, to (further) process data that has initially been obtained unlawfully. This would also fit in with the fact that the legal grounds for processing data in Art 6 (1) GDPR refer to the current processing activity (and not the previous processing activities) as well as the fact that the existence of Art 14 GDPR indicates, that personal data must not always be obtained by the data subject itself.[124]

It could, however, also be implicitly taken for granted that the initial collection of personal data is required to be lawful under the GDPR and that therefore would not have been worth mentioning explicitly within the GDPR. The GDPR does require the personal data to be "*processed lawfully, fairly and in a transparent manner in relation to the data subject ('lawfulness, fairness and transparency')*".[125] This seems indicate a general requirement of lawful processing of personal data throughout their "life-cycle" (i.e. from obtaining the data through every processing activity until its deletion).

Also, the fact that personal data have been unlawfully obtained would have to be considered within the balancing of interests under Art 6 (1) (f) GDPR. In that regard, the fact that personal data have been illegitimately obtained would have to be taken into account and would most likely shift the balance considerably against the further processing of such data by another controller.

So as an interim conclusion it appears that the further processing of personal data that has been initially obtained in a manner that can be considered unlawful under the GDPR (i.e. that cannot be based on one of the legal bases provided in Art 6 (1) GDPR) and is in many cases itself unlawful. In some cases, however, this may not be the case.

If we consider the legal bases under Art 6 (1) GDPR, the further processing of personal data (that initially have been unlawfully obtained) could not be based on consent given by the data subject[126] or the performance of a contract with the data subject[127], seeing as this concerns the initial collection of the data and not the further processing. It could nevertheless be possible, that the national law contains a legal obligation[128] to process personal data even if they have been initially obtained unlawfully.[129] It could be argued that such a provision (or its interpretation in that manner) would not fulfil the requirement of Art 6 (3) GDPR, according to which legal obligations in the sense of Art 6 (1) (c) and (e) GDPR shall meet an objective of public interest and be proportionate to the legitimate aim pursued, but such a

[124] Art 12-15 GDPR contain specific provisions on the fulfilment of the transparency principle in Art 5(1)(a) GDPR; there are different provisions regarding information that is to be provided to the data subject by the controller, depending on whether the data have been collected by the data subject or obtained from sources other than the data subject (e.g. third parties, "the internet", etc.).
[125] Art 5(1)(a) GDPR.
[126] Art 6(1)(a) GDPR.
[127] Art 6(1)(b) GDPR.
[128] Art 6(1)(c) GDPR.
[129] Consider, for example, national criminal law and the obligation to initiate criminal proceedings if evidence suggests that there have been a crime; the Code of Criminal Procedure in Austria contains such an obligation (§ 2(1) Strafprozeßordnung 1975 (StPO; *Austrian Code of Criminal Procedure*); „Amtswegigkeit" [*prosecution ex officio*]).

strict interpretation would impose a serious restriction for national legislators and would be in stark contrast to Art 23 GDPR.

3.4.1. Consequences of processing unlawfully obtained Data

The central provision, which specifically deals with unlawfully processed data, is Article 17 GDPR. The provision is also known as the „right to be forgotten“. According to the article, a data subject has the right to obtain erasure of their personal data, if certain conditions are met. One of the legal grounds, on which the right to erasure may be based is Article 17 par 1 lit d, which concerns data, that have been unlawfully processed.

In the context of scientific research, such a request could be detrimental to the results of the research project. The deletion of certain data could significantly alter the results in unintended ways. Therefore, it is in the interest of every researcher to ensure, that data has been lawfully processed, so that no such request will be lodged. However, especially in case of the use of public databases, it may be unfeasible or impossible to ensure, that all data was obtained lawfully by the third party. In many cases, no indication of unlawful processing will be available, even if parts of the data were not obtained lawfully.

Hence, the legislator implemented certain exceptions to the general right. According to Art 17 par 3 lit d), the right does not apply to the extent the processing is necessary “for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) in so far as the right referred to in paragraph 1 is likely to render impossible or seriously impair the achievement of the objectives of that processing”.

Insofar as the data subject can benefit from the appropriate safeguards according to Art 89 GDPR, which protect the rights and freedoms of said data subject, the right does not apply, if the objectives, which must be scientific research or a similarly protected public interest, cannot be obtained or the endeavor is seriously impaired.[130] Importantly, such safeguards can only be provided for by national law and must therefore be evaluated by each research group individually.

4. Database Protection & Licenses

While the main section of the analysis focuses on the protection of the data within a database, it should also be noted, that databases themselves can be protected under copyright law and a specific *sui generis* system. The framework of database protection stems from international treaties, union law as well as national laws.

[130] For details see *Herbst* in *Kühling/Buchner*, DSGVO/BDSG (2018)[2] Art 17 cif 82.

4.1.1. International Law

Under the rules of international law, protection of databases is primarily established by the Berne Convention.[131] Since 181 states are contracting parties, which includes Austria, France, Switzerland and Portugal[132], the treaty can serve as a foundation for the analysis. The Berne Convention aims to provide a minimum standard for protection of intellectual property. At its core are three basic principles. Firstly, works that originate in contracting state A must benefit from the same protection in contracting state B as contracting state B provides to the works of its nationals.[133] Secondly, works are automatically protected without any further formal procedure. Thirdly, the protection is independent of the level of protection in contracting state A. The minimum standard of rights includes the right to make adaptations, the right to communicate such works to the public, the right to make reproductions and the right to use the work as a basis for an audio-visual work. Exemptions may, however, be provided in special cases under national law.[134] With regard to the topic of the analysis, however, the scope of the Berne Convention is limited to "creative" databases, meaning that the arrangement and selection of the content in itself has to constitute an intellectual creation[135] in order for protection to be awarded to such a work. This is the case even if the creation of the database required a significant investment of resources.

Further treaties were signed by several states. The two central documents are the WTO/TRIPS Agreement[136] and the WIPO Copyright Treaty (WCT)[137]. Within the TRIPS Agreement, protection for databases is established within Art 10 par 2: "*Compilations of data or other material, whether in machine readable or other form, which by reason of the selection or arrangement of their contents constitute intellectual creations shall be protected as such. Such protection, which shall not extend to the data or material itself, shall be without prejudice to any copyright subsisting in the data or material itself*." The article again only refers to "intellectual creations". However, it extends the level of protection by making use of the terms "data" and "material", which is broader than the terms used in the Berne Convention. While this detail may seem insignificant at first glance, it results in the protection of databases consisting of "simple" data and material, rather than works, which in themselves are copyrightable. Furthermore, the article highlights two central principles of many database protection frameworks. The data within a database is generally not protected by the framework, but rather by other instruments. Additionally, the protection of the data or material within a compilation is independent of the rights granted through the treaties.

---

[131] Berne Convention for the Protection of Literary and Artistic Works of September 9, 1886, completed at Paris on May 4, 1896, revised at Berlin on November 13, 1908, completed at Berne on March 20, 1914, revised at Rome on June 2, 1928, revised at Brussels on June 26, 1948, and revised at Stockholm on July 14, 1967 (with Protocol regarding developing countries). ["Berne Convention"].
[132] See: https://wipolex.wipo.int/en/treaties/ShowResults?search_what=C&treaty_id=15 [23.06.2022].
[133] See Art 5 Berne Convention.
[134] Art 9 par 2 Berne Convention.
[135] Comp Art 2 par 5 Berne Convention.
[136] Agreement on Trade-Related Aspects of Intellectual Property Rights, Apr. 15, 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1C, 1869 U.N.T.S. 299, 33 I.L.M. 1197 (1994) ["TRIPS Agreement"].
[137] WIPO Copyright Treaty, Dec. 20, 1996 S. Treaty Doc. No. 105-17 (1997); 2186 U.N.T.S. 121; 36 I.L.M. 65 (1997).

The WIPO Copyright Treaty, a special agreement within the meaning of Article 20 of the Berne Convention[138], again deals with the protection of works, but focuses on the "new" digital environment. According to Art 5 of the treaty, databases are explicitly mentioned. However, similar notions on the limitation of the scope as in the TRIPS Agreement apply. From the perspective of the European legislator, this outcome seemed insufficient.

4.1.2. EU Law – Directive 96/9/EC

Therefore, in order to secure the functioning of the internal market, the EU implemented a directive on the legal protection of databases.[139] The aim was to harmonise the different degrees of legal protection of databases under national law, especially with regard to the provision of online databases.[140] Even though some databases were already protected by copyright law, the scope and modalities varied greatly.[141] The directive did not abolish existing copyright protection, but rather extended the rights of creators of databases through the instrument of an additional *sui generis* right.[142] This extension specifically took into consideration the protection of databases, which cannot be considered intellectual creations.

4.1.2.1. Material Scope

Thusly, Article 1 of the directive states, that the directive concerns the protection of "databases in any form". A database within the sense of the directive is a collection of independent works, data or other materials.[143] Within the framework of XAIface, potential databases consist of images of faces, which can mostly be subsumed under the term "data". However, the application of the term "independent work" cannot be ruled out. Even though substantial efforts may have been made by authors in obtaining the data, the databases usually do not constitute an intellectual creation. Hence, the focus should be the sui generis right according to the directive, since the right applies irrespective of such criteria.[144]

Article 3 states, that the protection of the database itself does not extend to the contents and that the protection of the content must be viewed independently of the directive.[145] A database will benefit from protection if the make of a database can demonstrate, that a significant investment has been made in the creation of the database. The investment may be qualitative or quantitative and relate to obtaining, verifying or presenting the contents in question.[146]

---

[138] See Art 1 WCT.
[139] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, OJ L 77, 27.3.1996, p. 20–28 ["Directive 96/9/EC"].
[140] Rec. 2 Directive 96/9/EC.
[141] Rec. 4 Directive 96/9/EC.
[142] Rec. 27 Directive 96/9/EC.
[143] Art 1 par 2 Directive 96/9/EC.
[144] See Art 7 par 3 Directive 96/9/EC.
[145] Art 3 par 2 Directive 96/9/EC.
[146] Art 7 par 1 Directive 96/9/EC.

4.1.2.2. Territorial Scope

The directive protects the rights of the "authors" of a database. An author in the sense of the directive is a "*natural person or group of natural persons who created the base or, where the legislation of the Member States so permits, the legal person designated as the rightholder by that legislation.*"[147] However, creators of the databases in question are not necessarily nationals of Member States, but frequently of third countries. Nevertheless, third country nationals may benefit from the protection of the directive, insofar as those third countries offer a comparable level of protection of databases to nationals of member states.[148]

4.1.2.3. Rights

If the database is protected under the directive, the maker of the database must have the right under national law to prevent extraction and/or re-utilisation of the whole or of a substantial part of the database.[149] A part may be considered to be substantial even if only a small part of the database is used, if this part constitutes a qualitatively makes up a significant part of the database. According to Art 7 par 2 "extraction" means "*the permanent or temporary transfer of all or a substantial part of the contents of a database to another medium by any means or in any form;*", whereas "re-utilisation" refers to "*any form of making available to the public all or a substantial part of the contents of a database by the distribution of copies, by renting, by on-line or other forms of transmission.*" For the project XAIface, the acts in question will mostly constitute an extraction rather than a re-utilisation.

4.1.2.4. Licensing

Even though the sole right to extract from or reutilise a database lies with the author, they may transfer, assign or grant these rights under a contractual licence.[150] It is common practice to provide online-databases under the framework of Creative Common Licences. The CC framework is generally a suitable licensing system, since it includes copyright as well as sui generis database rights.[151] Broader licences, such as CC0 Public Domain Dedication, may be advantageous in relation to the author or licensor of the database. However, the scope of the licences under the CC framework is limited, not unlike the database directive itself. The licenses only apply to the database structure and contents, insofar as those are copyrightable and to the sui generis database rights. Additionally, restrictions to the licences are permitted. Even though, according to the pre-licensing Guidelines of the CC framework[152] an author should obtain the necessary rights for the

[147] Art 4 par 1 Directive 96/9/EC.
[148] See Rec. 56; Art 11 Directive 96/9/EC
[149] Art 7 par 1 Directive 96/9/EC.
[150] See Art 7 par 3 Directive 96/9/EC.
[151] See: https://wiki.creativecommons.org/wiki/data [22.06.2022].
[152] See: https://wiki.creativecommons.org/wiki/Considerations_for_licensors_and_licensees#Considerations_for_licensors [22.06.2022].

provision of the data or clearly indicate, that such rights were not obtained, there is naturally no guarantee. The licenses rather usually contain waivers in the notice section. According to those waivers, no warranties are given. Contents may still be subject to national and international privacy, publicity or moral rights[153], which is a downside of using freely available data. Such waivers are, however, usually permissible, since the content is provided without compensation. The database FairFace for example is provided under the CC BY 4.0 licence, which contains a disclaimer and waiver of liability in section 5 and explicitly mentions that privacy, publicity and personality rights are not within the scope of the licence in section 2a.[154] In turn, the license grants a worldwide, royalty-free right to reproduce and share the material.

5. Results

When determining which face image data sets should be used (e.g. for training and evaluation purposes) a collection that is as compliant as possible with the GDPR.
To help determine the compliance of the dataset with the GDPR, a checklist is provided below.

In the context of scientific research (i.e., the purpose is exclusively research), it is arguable to use data sets where it is not clear whether they were collected in a manner, that is compliant with the GDPR or not (e.g., with the consent of the data subject), as long as the data protection principles are still upheld.

In order to be able to conclusively assess the permissibility of processing for research purposes, the implementation of the opening clauses with regard to **scientific research** (e.g. under Art 89 GDPR) **in respect to national data protection law** must be considered separately.

In Austria for example, there is a very broadly formulated exception for the use of personal data for scientific research purposes, the so-called "*research privilege*", i.e. a far-reaching general legal authorisation to process personal data for the purpose of research under the Federal Data Protection Act ("Österreichisches Datenschutzgesetz" – DSG) and the Federal Research Organisation Act ("Österreichisches Forschungsorganisationsgesetz" – FOG).

Since each Member State has implemented **their own specific "research privilege**" on the basis of Art 89 GDPR, the concrete assessment also raises the question of **which national law applies**.
This is not explicitly regulated in the GDPR and is currently disputed in the literature.

It should also be highlighted, that there has to be a **distinction between** the **collection and the use of the data** set on the one hand and the **subsequent dissemination** of the results on the other. Even if the use of the dataset for research purposes is permissible, data usage for publication must be assessed separately. This is especially true for datasets that were

[153] Comp.: § 78 Bundesgesetz über das Urheberrecht an Werken der Literatur und der Kunst und über verwandte Schutzrechte (Urheberrechtsgesetz) BGBl. I 2021/244.
[154] See: https://creativecommons.org/licenses/by/4.0/legalcode [22.06.2022].

collected in a potentially unlawful manner. In this case, an alternative data set should be selected.

Checklist – Choosing from existing face image data sets

Name of the Data Set: _____

Name Project Partner:
_____

Date of Assessment: _____

Interim Result: ☐ will be used ☐ will not be used

1.) General Considerations: it must be considered that the image data sets within XAIface:

- ☐ are processed in a secure manner,
- ☐ organisational measures will be or are implemented,
- ☐ do not concern children,
- ☐ do not concern special categories of personal data,
- ☐ will not be republished,
- ☐ are processed solely for scientific research purposes of the project,
- ☐ can be processed on a valid legal ground according to Art 6 and following,
- ☐ data subject rights will be considered, if there is no exemption applicable;

2.) The **initial collection** of personal data can be deemed lawful, if there is a

- ☐ detailed description of
  - initial controller,
  - process of collection,
  - purpose of the collection of personal data,
  - valid legal basis of the processing (especially in the sense of Art 6 GDPR)
- ☐ there are no indications giving rise to doubts that these descriptions are not genuine.

3.) If the collection has been conducted outside of the EU jurisdiction, it can be deemed lawful, if there is

☐ an adequacy decision by the Commission available (e.g. Switzerland, Japan,..)[155]
☐ other national data protection law applicable and the collection is in accordance with it (e.g. Chinese PIPL)

If no valid information on point 2.) and 3.) can be determined, the processing of these image data sets may still be processed lawfully under Art 6 (1) (f) GDPR if the following considerations have been taken into account to the best of the partners knowledge and beliefs:

☐ there is no available image data set that fulfils the above-mentioned requirements and has a detailed description,

☐ the image data set – within those data sets that have the necessary attributes to achieve the scientific research purpose – at least meets the requirements as best as possible

☐ the effort to generate an new data set by the own consortium is disproportionate

☐ appropriate safeguards are provided for by national law

In these cases, the following must be reconsidered by the partners:

Access to the databases?
- ☐ Is the access to the database restricted?
  - ☐ Access only for scientific research purposes
  - ☐ Access only for scientific research institutions

Can the initial controller(s) be determined?
- ☐ is this controller a well-known and generally trustworthy institution?
  - if the collection has been conducted by third parties (processors),
  - do codes of conduct apply?

Can the process of collection be determined?
- ☐ data have been collected with the (implied) consent of the data subject
- ☐ data have been collected with the probable knowledge of the data subject
- ☐ data subject can reasonably expect the processing?

Can the purpose of the collection be determined?
- ☐ data set has been collected for further processing for research purposes

Can the intensity of intervention be reduced to a minimum?

It should be highlighted that this evaluation does not allow to determine the lawfulness of the collection. It is also, in the views of the authors, **only applicable to the specific scientific research purpose** and under the specific circumstances as described above.

155 https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en (11.03.2022).

This checklist simply provides an indication of the best alternative, if there are only data sets available that do not fulfil the above requirements for the initial collection. It should not be concluded that, if there would only be data sets available with a score of zero points from questionable origin, that all of these data sets could – in lack of a better alternative – be processed under Art 6 (1) (f) GDPR. On the other hand, even if a lot of the checkboxes are ticked, the processing of these data sets could be considered unlawful under different circumstances (e.g. if the future processing purposes are privacy intrusive).

It should also be considered that processing of personal data still requires a legal ground for the processing of personal data that is contained in the data sets, within the GDPR regime. The processing of the partners itself must always be in accordance with the GDPR, even if opening clauses might be applicable. National law and GDPR must be read simultaneously. The national provisions do not supersede European law. Privileges for scientific purposes can be laid down in the national legislation, but the opening clauses should solely complement and specify rather than replace the provisions in the GDPR. The assessment of the lawfulness of the data processing and the final decision, which data set will be used, should be conducted by each controller (including their respective DPO who can provide guidance on the specific national provisions).