
## Deliverable 4.9.2

# Survey of Relevant Legal Context for the Secondary Use of Health Data

Funded by
Bundesministerium
für Bildung
und Forschung

Funded by
the European Union
NextGenerationEU

**https://anomed.de**

| | |
|---|---|
| UAP | 4.9.2 |
| Date | June 2024 |
| Version | 1.0 |
| Status | Final |
| Distribution | PU |
| Responsible Author (© by affiliation) | Harald Zwingelberg (ULD) |
| Additional Contributors (© by affiliation) | Bud P. Bruegger (ULD) |

# Table of Contents

# 1 Purpose

Legal aspects are of high relevance for the AnoMed project for two reasons:

- Any in AnoMed developed technical solution must satisfy legal requirements. In particular, solutions for anonymization of personal data must conform to the legal notion of anonymous. Such legal requirements are stated in the law (in particular in the GDPR). How the legal notion of anonymous has to be interpreted is largely determined by authoritative interpretations, for example by the European Data Protection Board (EDPB) or in actual court decisions. Since interpretation by the EDPB and courts are currently being developed, monitoring the correct interpretation of the law is directly relevant to AnoMed.
- Legal aspects are also directly important in context of technology transfer (see also Deliverable D 4.9.4) that is necessary to bring research results from AnoMed into practice. Here, new technologies are typically introduce into our society by a political strategy followed by the issuance of legal acts. A prime example that is highly relevant to AnoMed is the *European Data Strategy*. It has first been conceived in the form of a policy document by the European Commission[1]. To implement this strategy, a number of legal acts have been and will be issued. Examples are the *Data Governance Act*[2] (issued regulation) and the *European Health Data Space*[3] (proposed regulation). These legal acts thus describe the technical artefact that need to be implemented by technical solutions. In this context, three technology transfer activities are relevant for AnoMed:

  a. Identifying AnoMed solutions that are able to implement technical artefacts described in legal acts (transfer initiating on the technical side);

  b. Create awareness in the technical community of the technical artefacts described in legal acts as a pre-requisite for the previous step (transfer initiating on the legal side); and

  c. In support of the drafting process of legal acts, to compare the legal and technical conceptualization of the problem space and on this basis take steps to assure that the legal descriptions (abstractions) of technical artefacts are realistic (i.e., indeed have technical solutions) and can accommodate state of the art technical solutions ("reality-check" feedback originating on the technical side). (Deliverable D 4.9.4 contributes to this activity).

[1] COM/2020/66 final, CELEX 52020DC0066, *COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS A European strategy for data*, 19/2/2020, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0066

[2] Regulation (EU) 2018/1724 (Data Governance Act), CELEX 32022R0868, Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32022R0868

[3] COM/2022/197 final, CELEX 52022PC0197, *Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the European Health Data Space*, 3/5/2022, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52022PC0197
A compromise version from the interinstitutional dialogue has been published on March 18 2024: https://www.consilium.europa.eu/media/70909/st07553-en24.pdf

The present deliverable is concerned with these legal aspects. The concrete contributions are listed in the *Outline* section.

# 2 Context and Outline

The interaction between the technical and legal worlds span several tasks (UAPs) in work package 4.9 "*Analysis of Data Protection Risks*". In particular, fostering understanding of the relevant legal notions (point 1 in *purposes* above) and aspects of comparing legal and technical conceptualizations (2.c. in *purposes* above) are supported by the workshops of UAP 4.9.1 and the terminology in UAP 4.9.4. The technology transfer activities (point 2 in *purposes* above) are supported by discussion between legally- and technically-oriented AnoMed partners for example in UAPs 4.9.3 and 4.9.5 and have directly influenced the terminologies (see Deliverables D 4.9.4 and D 4 9.5).

The role of the present Deliverable in this context is to identify the legal texts that are relevant for the purposes of AnoMed in Section 3, report on the monitoring the authoritative interpretation of relevant legal concepts in Section 4, and present important characteristics of the legal descriptions of technical artefacts to technical partners in Section 5. The deliverable is then concluded in Section 6.

# 3 Survey of Relevant Legal Texts

The AnoMed project addresses anonymisation and pseudonymization in the context of medical research. This section explores the applicable legal framework to consider for requirements, limitations and conditions for a better understanding of anonymisation and pseudonymization when dealing with medical data for research purposes.

Medical research with patient data evokes the classical conflict between fundamental rights of patients as data subjects on the one hand and researchers as controllers on the other. For patients as natural persons, Article 7 CFR establishes the right for respect for private and family life and Article 8 for the protection of personal data. Scientists act as controllers when processing personal data for research purposes. The academic freedom is also guaranteed as a fundamental right in Article 13 CFU.


*Figure 1 Balancing of Fundamental Rights*

The European and Member States' legislators have to cater for a balance between the rights and interests of all involved legal entities. Neither right may be negated in its essential core elements while exercising the right may not infer disproportionally in other persons fundamental rights. In the field of data protection, essential principles[4] have been developed to provide guidance on how to

[4] See Article 5 GDPR.

uphold data subject's rights and limit risks for their fundamental rights and freedoms. In proportion to the risks involved with the processing, technical and organisational measures need to be implemented to ensure appropriate implementation of data protection principles and levels of security.

The AnoMed project aims at improving the means anonymisation. The anonymisation of health data is often understood as a gold standard, allowing for both ideal protection of data subjects as well as allowing free exchange and access to data for research. The ongoing research within the work package aims at better understanding anonymity. Focal point for any analysis is the legal definition of "personal data" in Article 4(1) GDPR. The existing legal acts as well as legislative initiatives have been surveyed for relevance in relation to anonymity research in the AnoMed project in particular, for secondary use of health data and potential impact of the AnoMed results. An overview of legal acts and their relation to anonymity research is provided below for European and German legal acts followed by a table indicating the relevance for different actors.

**European Legal Texts**

**European Charter of Fundamental Rights (CFR):**
The Charter of Fundamental Rights establishes fundamental rights of natural persons. The right to data protection is codified in Art 7 and 8 CFR. Further, the right for researchers is guaranteed as freedom of sciences in Art 13 CFR. While the fundamental rights apply directly for both data subjects and researchers the CFR has not been identified as "applies directly" as in legal practice the first reference must be held to the most specific applicable legal acts. The fundamental rights and freedoms will become relevant for the application and interpretation of these specific laws.

**The General Data Protection Directive (GDPR):**
The GDPR is the central governing act for all processing of personal data by public and private entities in the member states. Subsequent acts had been issued "without prejudice" to the GDPR,[5] thus are not meant to restrict rights of data subjects granted in the GDPR or grant additional permissions to controllers, unless explicitly specified otherwise. In consequence, any requirement for processing of personal data and research based on such data must meet the requirements of both legal acts.
For research on anonymity, the GDPR contains the central definition of personal data. It is the starting point of any assessment on whether data does still relate to an identified or identifiable natural person.

**Regulation (EU) 2018/1725:**
This regulation is applicable to the processing of personal data by EU institutions, bodies and agencies. Definitions of personal data and other relevant wording are identical with the GDPR. The Regulation is applicable to research done by EU institutions. Oversight body for these institutions and in charge of the enforcement of this regulation is the European Data Protection Supervisor (EDPS).

**Data Governance Act (DGA):**
The DGA is a pillar of the EU's data strategy. It aims at enabling sharing and reuse of existing data: a) with mechanisms to facilitate the reuse of certain data held by public sector entities such as data that cannot be shared as open data, b) by establishing intermediaries to function as organizers also for the planned data spaces, c) measures to make sharing of data for the common good easier (data altruism) and d) further measures to enable sharing across sectors and

[5] See e.g. Article 1(3) DGA.

borders. The DGA does not intend to undermine protections provided by the GDPR. For the scope of this survey, the DGA holds a series of relevant definitions such as “secure processing environment”. Subsequent acts such as the EHDS reference these definitions.

**Data Act (DA):**
The DA complements the DGA. Connected products and related services (essentially what has been understood as internet of things, IoT) produce massive amounts of data. The DA sets a framework for sharing of such data. Users of connected devices have a right to access data stemming from their connected products from manufacturers and service providers. The DA introduces mandatory sharing of available data in business to business and business to government relations including those with research institutions.
The work done in the AnoMed project[6] may be relevant for the interpretation of the DA as the understanding anonymity, risks of re-identification and the resulting qualification as anonymous or personal data is more essential where sharing of data becomes mandatory. While right to data protection explicitly remains untouched by the DA and the GDPR determines legal basis and limitations, the DA does introduce ‘non-personal data’.[7]

**European Health Data Space (EDHS):**
The EDHS establishes a data space for health data in the European Union. It governs the primary use of health data including cross-border aspects as well as providing the foundation for a broader availability of data for secondary in research and other areas. For specifics of the EHDS, the role of access bodes, the procedure as to how data access requests are handled and data access is granted see sections 5.2 and 5.4 below.

**Artificial Intelligence Act (AI Act):**
The AI Act contains specific rules for AI systems that create a high risk to the health and safety or fundamental rights of natural persons.[8] These rules complement the requirements stipulated in the GDPR – thus both sets of requirements must be met. For medical and anonymity researchers it should be mentioned that the act classifies medical devices including medical apps as high-risk AI systems and stipulates specific requirements in Title III (Articles 6 et seq.). These include risk management, documentation, continuous conformity assessments, record-keeping, and transparency requirements. Manufacturers, representatives and importers are mandated with the compliance (Article 24 et seq.). The AI Act has been ratified in June 2024.

**German Legal Texts**

**Gesundheitsdatennutzungsgesetz (GDNG):**
The German “Gesundheitsdatennutzungsgesetz” (Engl. Health Data Usage Act) is a federal law. It aims at enabling and simplifying the access to health data for secondary purposes and proactively implements aspects of the EHDS in national law. A coordination office is established

[6] S ee ULD, „Identity Reduction - The Technical Perspective”, 2024, https://www.datenschutzzentrum.de/uploads/projekte/anomed/Identity-Reduction_v0_9_2.pdf, and Bruegger, "Towards a Better Understanding of Identification, Pseudonymization, and Anonymization", 2021, https://uld-sh.de/pseudoanon.
[7] Article 2(4) DA, Recital 7 DA.
[8] See section 5.2.3 in Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), https://eur-lex.europa.eu/search.html?scope=EURLEX&text=ai+act&lang=en&type=quick&qid=1719910265819.

as national access point. A permission and conditions for the linkage of existing databases with health data is provided.

The GDNG also stipulates a prohibition to use data provided under the act for (re-)establishing a link to a patient or involved health professional. Infringement of the prohibition is punished with a financial penalty or imprisonment up to one year. The prohibition may interfere with anonymity research. In particular testing re-identification attacks to verify the strength of anonymisation may well require careful scrutiny and should not be based on data obtained under the GDNG but rather on data for which the patients have provides an appropriate consent.

**Sozialgesetzbuch V, SGB V**

Volume V of the German Social Insurance Code (SGB V) regulates the statutory health insurance. The parliamentary act by which the GDNG has been passed contains amendments to SGB V for the access to, merging of and research with data in existing registers held by actors in the social security system. Notably it also clearly grants a permission to anonymize data in reaction to a debate, whether anonymisation for secondary purposes is a processing activity that requires a legal ground beyond that for the primary processing.

**Bundesdatenschutzgesetz (BDSG)**

The German Federal Data Protection Act applies to the processing of data by bodies of the Federation. § 27 BDSG implements the opening clause in Article 89 GDPR by permitting the processing of special categories of data for research purposes where the interests of the controller in processing substantially outweigh those of the data subject. Data must be anonymized as soon as the research purpose allows it.

**Legal Acts of the federal States**

State Data Protection Acts (Landesdatenschutzgesetz, LDSG), Higher Education Acts (Hochschulgesetz) and Hospital Acts (Klinikgesetz) of the German Federal States contain research clauses which often permit the secondary use of health data. These Acts apply to public authorities and bodies in the respective federal state and where more than one act may apply the more specific one prevails. As medical research primarily takes place in hospitals or universities the clauses[9] in Hospital and Higher Education Acts usually are applicable to many research projects. In absence of specific acts or where public actors carry out the research, the more generic Sate Data Protection Acts contain research clauses.[10]
The Acts vary and differ in scope, extent of the permission and other details. This obstructs the operation of research projects that operate across the borders of individual federal states. The German data protection authorities therefore encouraged a coherent and nationwide harmonized regulation.[11]

[9] For an overview of research clauses and the extent to which secondary use of health data is permitted see Dierks, Kircher, Husemann et al. „Lösungsvorschläge für ein neues Gesundheitsforschungs-datenschutzrecht in Bund und Ländern“, 2019, https://www.dierks.company/de/publikationen/losungsvorschlage-fur-ein-neues-gesundheitsforschungsdatenschutzrecht-in-bund-und-landern/.
[10] For a list of research clauses see Weichert, „Datenschutzrechtliche Rahmenbedingungen medizinischer Forschung“, 2022, https://www.mwv-open.de/site/books/m/10.32745/9783954667000/.
[11] Germany’s Conference of Independent Federal and Länder Data Protection Authorities, commonly referred to as the DSK or “Datenschutzkonferenz” (Data Protection Conference), “Petersberger Erklärung zur datenschutzkonformen Verarbeitung von Gesundheitsdaten in der wissenschaftlichen Forschung”, 2022, https://www.datenschutzkonferenz-online.de/media/en/20221124_en_06_Entschliessung_Petersberger_Erklaerung.pdf.

The following table provides links to the mentioned legal texts and an overview of their relevance for different actors. A legend of the symbols is provided below the table.

**Table 1: Laws applicable for research with medical data (primarily anonymity research but also other research)**

| Legal Acts \ Actors<br><br>(black = in force; gray = proposed) | Natural persons as data subjects | Anon. Research. | Doctors Primary users | Medical researchers secondary users | Access bodies | Private actors, companies | Public bodies, authorities | Manufacturers, importers |
|---|---|---|---|---|---|---|---|---|
| CFR[12] | ✓✓ | ✓ | ✓ | ✓ | ○ | ✓ | ○ | ○ |
| GDPR[13] | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | × |
| Regulation (EU) 2018/1725[14] | ✓✓ | ✓✓ | × | × | ✓✓ (EU) | × | ✓ | × |
| Data Governance Act (DGA)[15] | ✓✓ | ✓ | ○ | ○ | ✓✓ | ✓ | ✓ | × |
| Data Act (DA)[16] | ✓✓ | ✓ | ○ | ○ | ✓✓ | ✓ | ✓ | ✓✓ |
| EHDS[17] | ✓✓ | | | ✓✓ | ✓✓ | ✓ | ✓ | × |
| Directive (EU) 2011/24[18] | ✓✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | × |
| AI Act19 | ✓✓ | ✓✓ | ○ | ✓ | ✓ | ✓ | ✓ | ✓✓ |
| GDNG[20] | ✓✓ | ✓✓ | ✓ | ✓✓ | ✓✓ | ✓ | ✓ | × |
| SGB V | ✓✓ | ○ | ✓✓ | ✓ | ✓ | × | ✓ | × |
| German Federal Data Protection Act (BDSG) | ✓✓ | ○ | × | ✓ | × | × | ✓✓ | × |
| Legal Acts of the federal states | ✓✓ | ○ | × | ✓✓ | × | × | ✓✓ | × |

| | |
|---|---|
| ✓✓ | Applies directly to anonymisation research or medical research and the listed group of actors are direct addresses or beneficiaries. |
| ✓ | Applies in general for the group of actors with limited impact for anonymity and medical research |
| ○ | May apply for certain types of processing or for specific circumstances |
| × | Law does not typically address the group of actors in this role. However, the acts may be applicable nevertheless for other activities or types of processing. E.g. while manufacturers of data processing devices are commonly not addressees of the GDPR in this role, they are for processing activities such as services provided to their customers. |

[12] https://eur-lex.europa.eu/eli/treaty/char_2012/oj
[13] https://eur-lex.europa.eu/eli/reg/2016/679/2016-05-04
[14] https://eur-lex.europa.eu/eli/reg/2018/1725/oj
[15] https://eur-lex.europa.eu/eli/reg/2022/868/oj
[16] https://eur-lex.europa.eu/eli/reg/2023/2854/oj
[17] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52022PC0197
[18] https://eur-lex.europa.eu/eli/dir/2011/24/2014-01-01
[19] Commission Proposal. The Act has been ratified and publication in the official journal is pending at the time of writing: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206.
[20] https://www.recht.bund.de/bgbl/1/2024/102/VO.html.

# 4 Collaboration with Relevant Working Groups

Many technical requirements in AnoMed must be derived from legal concepts. This is for example necessary for "*anonymous*" and "*anonymization*" which are both legal notions in need of a technical interpretation and/or implementation. The technical fulfillment of legal requirements is only possible if the legal concept is interpreted correctly.

The task of interpreting the legal concept is by no means is trivial or easy. Even in the legal interpretation of a legal concept, i.e., before the technical interpretation even starts, there is ample diversity. Evidence for this can for example in the following quotation from a law professor: "A favourite aphorism, which has settled into a folk form of truth says, where there are two lawyers, there are at least three legal opinions."[21]

Which of the several legal interpretations is the correct one and thus suitable to be used as the input for the technical interpretation? In the legal world, not all interpretations are equal. In particular, in certain areas, specific bodies ("Gremien") have been instituted by law with the mandate to harmonize legal interpretation in their area of competence and thus harmonize the application of law. In addition, in all areas, court decisions interpret the law for individual cases with appellate and supreme courts providing more broadly transferable interpretations. For the interpretation of legal acts of the EU, the European Court of Justice issues final decisions. Court decisions are typically informed by, but supersede decisions by harmonizing bodies.

In the area of data protection, the relevant harmonizing bodies are the *European Data Protection Board* (EDPB), that harmonizes legal interpretation of the GDPR in Europe,[22] and the German "*Data Protection Conference*" ("Datenschutzkonferenz", DSK), that harmonizes the interpretation across the German federal states (Länder) and the Federation (Bund); The final interpretation of the GDPR as EU regulation resides with the European Court of Justice (ECJ).

Figure 2 visualizes the situation. The EDPB issues the initial interpretation of legal concepts, typically in the form of guidelines. Optionally, and in the context of a concrete case, courts can create interpretations which supersede the EDPB's opinion. The highest authority in regard of data protection in Europe is the European Court of Justice.

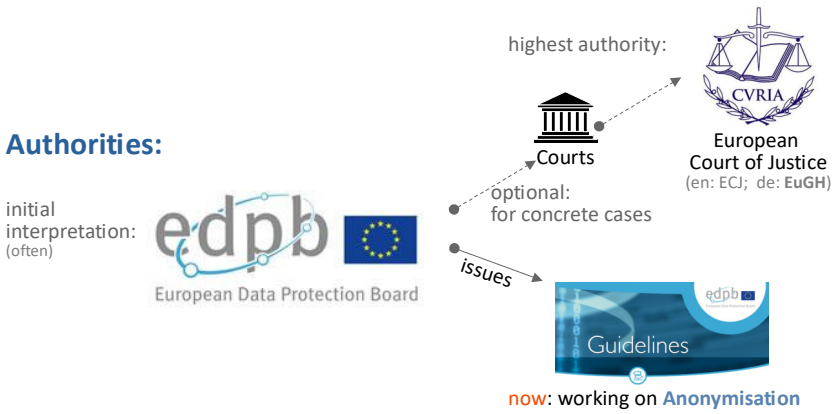


*Figure 2: Development of legal interpretations in data protection.*

To obtain an up-to-date and correct legal interpretation of relevant legal concepts, it is thus necessary to monitor work done by the relevant harmonization bodies and courts. It is not possible

[21] Lucia Žitňanská, transcript of 3rd lecture, http://www.upms.sk/en/prednasky/prepis/3.-prednaska-dvaja-pravnici-tri-pravne-nazory-alebo-ako-sa-tvori-pravo/

[22] See Article 70 GDPR.

at the time of proposal writing to foresee which of the relevant topics will be worked on by which body or court. A first step of the described work is therefore to identify when relevant legal interpretations are being created. In the reporting period, two cases of very high relevance have been identified:

i. The *EDPB* who is currently writing guidelines on anonymization; and

ii. the *European Court of Justice* (ECJ) that will be deciding in a concrete situation whether certain data can be considered to be anonymous.

(i) In more detail, the EDPB's *technology expert subgroup* has formed a drafting team to write and eventually publish *Guidelines on Anonymization* in its Guideline Series[23] on important legal concepts. It replaces and evolves the interpretation ("opinion") previously created by the Article 29 Data Protection Working Party, the predecessor of the EDPB, on "anonymization techniques"[24].

(ii) At the time of writing, the European Court of Justice is working on an appeal by the European Data Protection Supervisor in Case C-413/23 P[25]. In particular, the EDPS appeals the decision of a lower court that the concrete data in the case was to be considered *anonymous*. The EDPS reasons that the data is *pseudonymous* and therefore *personal*, and thus cannot be considered *anonymous*. The outcome of this process will have a significant impact on the correct legal interpretation of the concept of anonymity; it can either confirm the interpretation by the EDPB (of which the EDPS is a member) or can significantly change that interpretation.

At the time of writing, the EDPB is preparing a Statement in Intervention in support of the EDPS for this case. Its main content is the legal interpretation of *anonymous* and *pseudonymous*. Again, the statement of intervention is confidential until finally approved and submitted to the court.

# 5 Legal Context of Anonymization and Pseudonymization

While the previous section focused on core concepts related to *anonymization* that are treated in the GDPR, the present section focuses on the more advanced concepts of technical artefacts that are described in other legal acts. This section reports on relevant technical artefacts and how they are described in the legal text.

This section thus aims at providing a "legal transfer" to technical project partners. In particular, it should help technical partners to identify technical solutions that are suited for the implementation of legally described artefacts.

[23] https://edpb.europa.eu/our-work-tools/our-documents/publication-type/guidelines_en.

[24] ARTICLE 29 DATA PROTECTION WORKING PARTY, WP216, *Opinion 05/2014 on Anonymisation Techniques*, Adopted on 10 April 2014, https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf.

[25] Case C-413/23 P, CELEX 62023CN0413, *Appeal brought on 5 July 2023 by the European Data Protection Supervisor against the judgment of the General Court (Eighth Chamber, Extended Composition) delivered on 26 April 2023 in Case T-557/20, Single Resolution Board v European Data Protection Supervisor*, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:62023CN0413

## 5.1 Relevant Legal Acts

The following briefly describes the legal acts that were identified to be most relevant for the intended "legal transfer".

The most important political initiative that requires pseudonymization and anonymization at large scale is the *European Data Strategy*[26]. Figure 3 shows the legal acts that are relevant for implementing the data strategy.

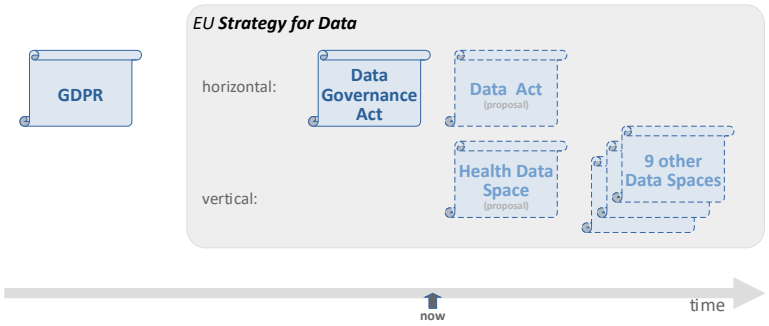


*Figure 3: Legal Acts of the European Data Strategy.*

A multitude of legal acts will implement the European Data Strategy. Considered the "horizontal" (i.e., cross-sectorial) acts are the *Data Governance Act*[27] and the *Data Act*[28]; the "vertical" acts are many data spaces[29] that concentrate of sector-specific types of data. All relevant acts are regulations and thus directly applicable (i.e., without the need for implementation in national law) in all Member States of the EU.

At the time of writing, the *Data Governance Act* (*DGA*) is already in force (since 24/9/2023); the *Data Act* (*DA*) currently exists as a proposal by the EC; the *European Health Data Space* (*EHDS*)[30] is the first of the planned ten data spaces that currently exists as a proposal by the EC.

Since personal data are an integral part of the strategy, the GDPR lays out the basic rules of how to process personal data including medical data. Every later act thus refers to the GDPR. This is most obvious in Art. 1(3) DGA that states: "In the event of a conflict between this Regulation and Union law on the protection of personal data [i.e. the GDPR] […], the [GDPR] […] shall prevail." (Editing in [] added for clarity).

An analysis of these acts showed that the DA has little relevance for the intended "legal transfer" and that the DGA is the most relevant of the horizontal acts of the data strategy. The EHDS, being the first of the data spaces and being concerned with health data, was found to be the most relevant vertical legal act for AnoMed.

[26] See footnote 1 above.
[27] See footnote 2 above.
[28] COM/2022/68 final, *Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on harmonised rules on fair access to and use of data (Data Act)*, CELEX 52022PC0068, 23/2/2022, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2022:68:FIN.
[29] https://digital-strategy.ec.europa.eu/en/policies/data-spaces
[30] See footnote 3 above.

The EHDS requires implementation by Member State's legislators and government. In Germany the federal Legislator passed the Gesundheitsdatennutzungsgesetz (GDNG)[31] in March 2024 which anticipates parts of the implementation. At the time of writing, the EHDS has not been passed and some changes to the German law can be expected after its enactment.

The legal acts analyzed in more detail in this section are therefore the DGA and the EHDS.

## 5.2 The major Actors in the EHDS

The major actors in the EHDS are visualized in Figure 4. In particular, *data holders* such as hospitals and other health care providers collect the health data for the primary use of health care. A subset of this data is then made available through *data access bodies* to *data users*. Data users can pursue a number of purposes that are listed in Article 34 EHDS "Purposes for which electronic health data can be processed for secondary use". Scientific research is probably the purpose most relevant to AnoMed.

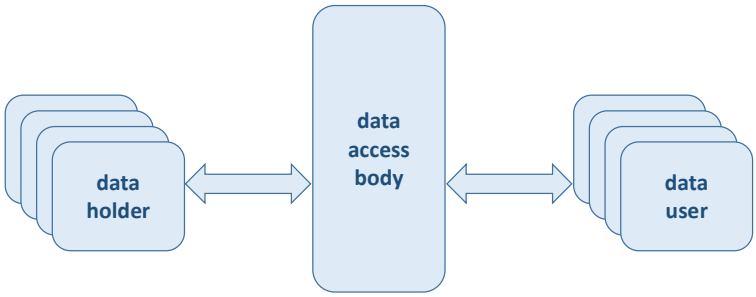


*Figure 4: The main actors in the EHDS.*

## 5.3 Pseudonymization and Anonymization in the EHDS

For the secondary use of data in the EHDS, pseudonymization and anonymization of data are most relevant. This is most evident in its Article 44 titled "*Data minimisation and purpose limitation*". In particular, where possible, anonymized data shall be used. This is stated in Article 44(2) that reads "The health data access bodies shall provide the electronic health data in an **anonymised** format, where the purpose of processing by the data user can be achieved with such data, […]."

Only where this is not possible, access to pseudonymized data shall be granted. This is states in Article 44(3) that reads "Where the purpose of the data user's processing cannot be achieved with anonymised data, […], the health data access bodies shall provide access to electronic health data in **pseudonymised** format." (Emphasis added by author).

The use of **anonymized** data is further regulated in Article 47 EHDS on "***Data request***". Here, according to Article 47(1), "A health data access body shall only provide an answer to a data request in an anonymised statistical format […]". Data requests will be discussed further below.

The use of **pseudonymized** data is further regulated in Article 45 EHDS "**Data access applications**" that will be discussed further below. *Data access applications*, if granted by the *data access body*, yield (remote or on-premise) access to pseudonymized data within a *secure processing environment*.

[31] Gesundheitsdatennutzungsgesetz , Bundesgesetzblatt 2024 Tiel I, Nr. 102, online: https://www.recht.bund.de/bgbl/1/2024/102/regelungstext.pdf?__blob=publicationFile&v=2.

The results of such processing shall then be made public (see Article 46(11) EHDS). These results "shall only contain anonymised data". Evidently, this requires **anonymization**.

## 5.4 Data Access Requests in the EHDS

Figure 5 illustrates *data access requests* according to Article 45 EHDS.

Here, *health data access bodies* enable *data users* to process pseudonymous data originating from *data holders*. For this purpose, data access bodies operate a *secure processing environment*. As a prerequisite for *access*, data users need to submit a *data access request* to the competent data access body. This is denied or granted in the form of a *data permit*. The data permit then enables *access* to the data in the secure processing environment. According to Article 2(13) DGA, "'access' means data use, […], without necessarily implying the transmission or downloading of data". Article 5(3)(b) and (c) DGA further clarify that access can happen "remotely within a secure processing environment" or "within the physical premises in which the secure processing environment is located".

In more detail, following the numbering in the figure, the EHDS describes the following interactions:

1. Arbitrary *data users* can issue *data access applications* to a *health data access body* as long as one of the purposes foreseen in Article 34 EHDS are pursued (see Article 45(1) EHDS). The content of a data access application is prescribed in Article 45(2) and (4) and is the basis on which data access bodies make their access decision.

2. "A health data access body shall issue or refuse a data permit [usually] within 2 months of receiving the data access application." (Article 46(2) EHDS). (Comment added by author).

   "Following the issuance of the data permit, the health data access body shall immediately request the electronic health data from the data holder." (Article 46(4) EHDS).

3. "The data holder shall put the electronic health data at the disposal of the health data access body [usually] within 2 months from receiving the request from the health data access body." (Article 41(4) EHDS). (Comment added by author).
   "The health data access bodies shall ensure that electronic health data can be uploaded by data holders and can be accessed by the data user in a secure processing environment." (Article 50(2) EHDS).

4. According to Article 44(3) EHDS, the uploaded data must be rendered pseudonymous before being accessed by data users. This paragraph further states that "The information necessary to reverse the pseudonymisation shall be available only to the health data access body." This seems to imply that the pseudonymization of the data must actually be performed by the health data access body. According to Article 44(3) EHDS, "Data users shall not re-identify the electronic health data provided to them in pseudonymised format."

5. According to Article 50(1)(a) EHDS, health data access bodies "restrict access to the secure processing environment to authorised persons listed in the respective data permit".

6. According to Article 38(3) EHDS, data users can inform data access bodies of "a finding that may impact on the health of a natural person".

7. "[T]he health data access body may [then] inform the natural person and his or her treating health professional about that finding." Since health data access bodies are in possession of

reversing the pseudonymization (see point 4 above), they are the only party that can actually inform the concerned natural person and treating health professionals.

8. According to Article 46(11) EHDS, "Data users shall make public the results or output of the secondary use of electronic health data, […], no later than 18 months after the completion of the electronic health data processing." "Those results or output shall only contain anonymised data." Results are then made "public on health data access bodies' websites". Evidently, this implies the need for effective anonymization of processing results.

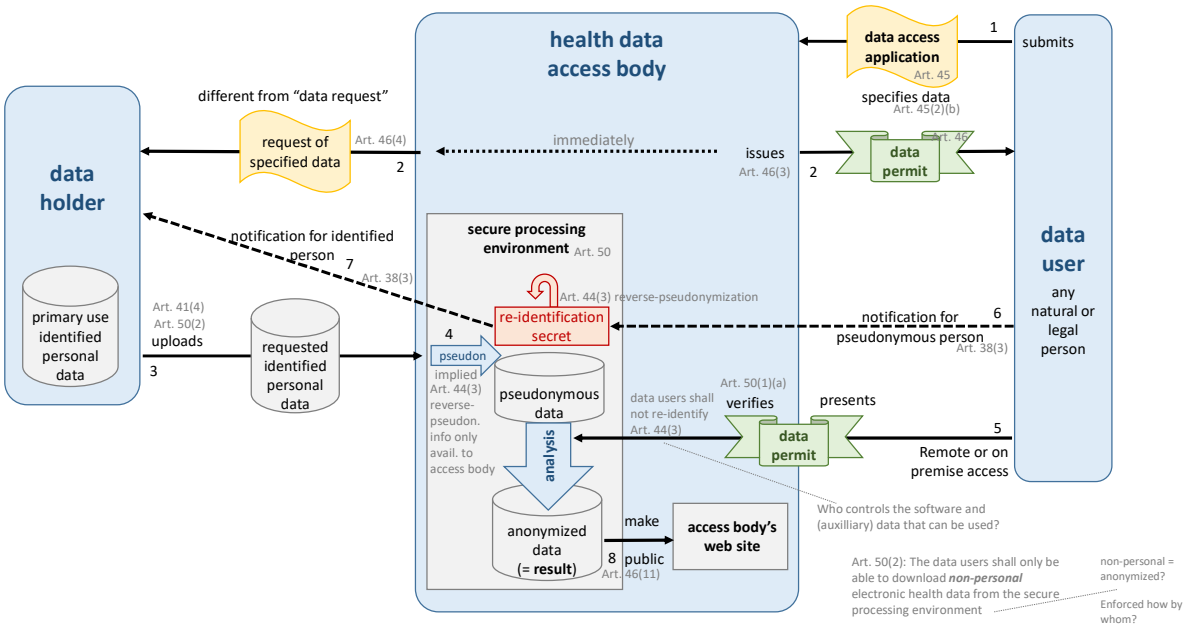


*Figure 5: Data Access Requests in the EHDS.*

## 5.5 Secure Processing Environments

The notion of *secure processing environment* is central to understanding data access requests. It is defined in Article 2(20) DGA as follows: "'secure processing environment' means the physical or virtual environment and organisational means to ensure compliance with Union law, such as [the GDPR], […]," This holds "in particular with regard to data subjects' rights, […] and statistical confidentiality, integrity and accessibility, […]."

A secure processing environment must "allow the entity providing the secure processing environment to **determine and supervise all data processing actions**, including:

- the display,
- storage,
- download and
- export of data and
- the calculation of derivative data through computational algorithms"

(Bullets added by author).

The prototype for secure processing environments seems to be access to statistical microdata. This is apparent in Recital 7 DGA that states: "There is experience at Union level with such secure processing

environments that are used for research on statistical microdata on the basis of Commission Regulation (EU) No 557/2013." Commission Regulation (EU) No 557/2013[32] is an implementing Regulation on European Statistics as regards access to confidential data for scientific purposes.

From a technical point of view, it seems evident that the personal data that is controlled inside a secure processing environment must not be allowed to exit. Considering that the results of processing must be published in an anonymized form (see above), it therefore seems to imply that the entity providing the secure processing environment is responsible to assure that results which leave the environment are indeed anonymous.

Article 7(1) of Commission Regulation (EU) No 557/2013 seems to make this explicit by stating that "Access to secure-use files may be granted provided that **the results of the research are not released without prior checking** to ensure that they do not reveal confidential data." (Emphasis added by author).

A possible technical interpretation of a secure processing environment is shown in Figure 6. Secure processing environment is abbreviated by SPE. It consists of several layers:

- The **base infrastructure** consisting of the physical premises, networking, etc.
- The actual **computing hardware** (such as servers) avails itself of the services of the base infrastructure. It is typically provided by the SPE, but when specialized computing (such as machine learning) is required, it is thinkable that SPE providers offer housing of specialized hardware (such as GPUs) to data users.
- The hardware is then used by a **base software environment**. It consists of the operating system, possibly also data base management systems and libraries.
- To do the actual analysis that pursues purposes listed in Article 34 EHDS is executed by **application software**. SPE providers can offer (and pre-install) general purpose application software such as statistics or data analysis packages. This can then be accessed by data users by scripts or simply invoking commands from an interactive shell. Since within the foreseen purposes, data users should have a free choice of the kind of data analysis, the provided application software may not suffice. In this case, data users may provide the necessary application software. This is unavoidable if the software is a custom development at the forefront of the scientific state of the art. Such user-provided software may in turn require a certain software environment (such as libraries) or even specialized hardware (such as GPUs for CUDA-based machine learning).
- It is common for data analysis to also require **auxiliary data** beyond that provided in the SPE. It may therefore be necessary for data users to upload data into the SPE.

One way how SPE providers can fulfill their obligation to "determine and supervise all data processing actions, including the display, storage, download and export of data and the calculation of derivative data through computational algorithms" is through **entry** and **exit gateways** of the SPE.

[32] COMMISSION REGULATION (EU) No 557/2013, CELEX 32013R0557, *implementing Regulation (EC) No 223/2009 of the European Parliament and of the Council on European Statistics as regards access to confidential data for scientific purposes and repealing Commission Regulation (EC) No 831/2002, 17 June 2013, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32013R0557.*

In the different entry gateways, anything that enters the SPE must be approved; in the exit gateway, everything that leaves the SPE must also be approved. Examples for the approval in entry are application software that determines the analysis that is conducted, or the approval of a user-provided software environment that must satisfy certain security requirements. Examples for the approval in exit are anonymized results and storage devices that have to be wiped of any sensitive data.

The gateway approach fits well with a batch-oriented processing by data users. It is less applicable to (possibly remote) interactive use. This would for example be the case when data users interactively invoke commands of some provide statistics package or perform system maintenance[33] for the computing environment they provided.

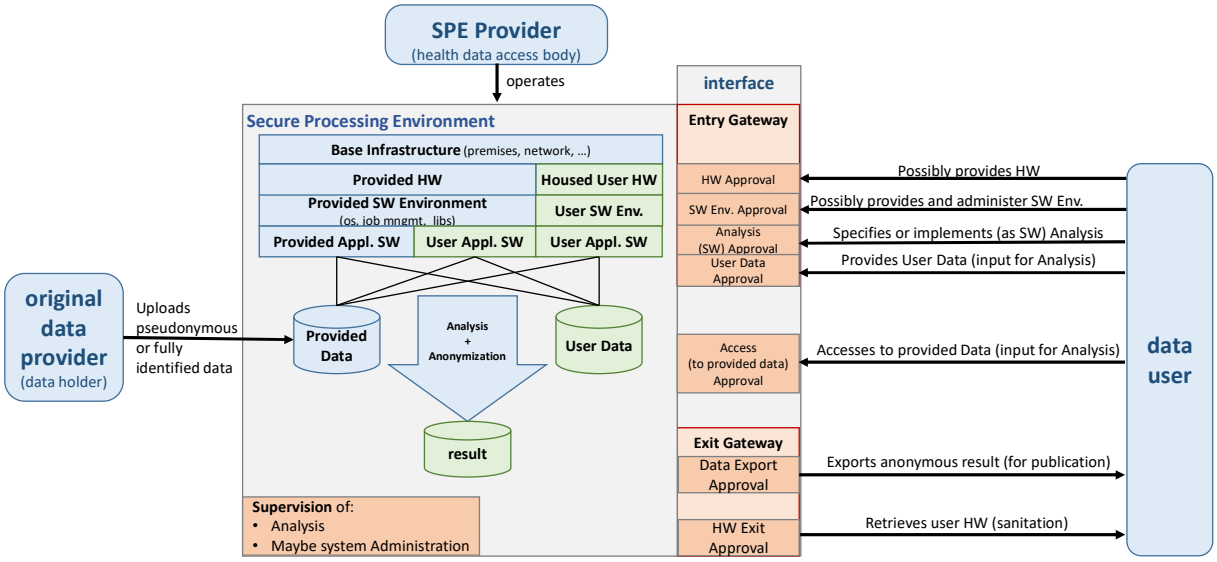


*Figure 6: Possible technical interpretation of a secure processing environment.*

According to Article 51 EHDS, the SPE provider, i.e., the health data access body, and the data user are deemed joint controllers. They are thus both responsible for the processing.

## 5.6 Data Requests in the EHDS

Were it is not possible in regard to the purposes of processing to use anonymized data, *data access requests* for access to SPEs are necessary (see discussion above). In contrast, where (in accordance with Article 44(2) EHDS) the purposes can be achieved with anonymized data only, data users can directly be provided anonymized data based on a *data request* (see Article 47 EHDS).

According to Article 47(1) EHDS, arbitrary data users can submit *data requests* to data access bodies.

According to Article 47(2) EHDS, the request must specify:

- a "description of the requested electronic health data, their format and data sources" (Article 45(2)(b) EHDS),
- "a description of the result expected" (Article 47(2)(a) EHDS) and
- "a description of the statistic's content" (Article 47(2)(b) EHDS).

[33] Note that system maintenance if instrumental in keeping software secure.

According to Article 47(3) EHDS, "[…] the health data access body shall […], where possible, provide the result to the data user within 2 months." It means that usually, data users directly receive anonymized data. Again, there is a need for anonymization this time by the data access body.

For *data access requests*, the proposed legal text specifies that after its approval, the data access body requests the data from data holders. It is not clear from the proposed legal text when the data necessary for *data requests* is requested or obtained from data holders. It is therefore unclear whether a preventive data retention by data access bodies is necessary to handle *data requests*.

Another open question is how far data access bodies can automate the handling of data requests. Considering that anyone can submit data requests (see Article 47(1) EHDS) and that data access bodies are required to respond within a certain time period (see Article 47(3) EHDS), the volume of requests may be high and challenge the resources of data access bodies. Far-reaching automation seems possible since the necessary assessment seems far less complex than that of *data access requests*. In particular, the only elements that can be assessed is the requested anonymized data and the intended use including purposes (according to Article 45(2)(a) EHDS).

Considering the potentially high number of requests and the obligation to satisfy acceptable requests, the question of whether a privacy budget can be managed must be posed. This is even more relevant considering that data access bodies cannot restrict data disclosure to a fixed set of statistics, but instead, every data user can specify the requested anonymized data.

## 5.7 Some Technical Questions relative to the EHDS

The following will briefly list technical questions which deserve further technical analysis and discussion.

- When and how data flows from data holders to data access bodies is not all the way clear.
    - For data access requests an upload is requested at the first time it is needed (based on a data access request). Whether the data are then stored or deleted as soon as the data access request is completed does not follow from the proposed legal text.
    - In contrast to the case of data access requests, the legal text doesn't seem to address the case of data requests (see also above).
- Can "distributed computing" replace a physical transfer of data from data holders to data access bodies? In particular, for analysis tasks such as statistics or machine learning, it would be technically possible that data holders transfer a partial computing result much rather than the original data. Such an approach may bear benefits in terms of data minimization as well as liability of data access bodies.
- The legal text states that data access bodies need to possess the information necessary to reverse pseudonymization. An alternative approach would be that the data are already pseudonymized by data holders who then also refrain from transferring the data necessary to reverse the pseudonymization to data access bodies. The re-identification mentioned in Article 44(3) EHDS would then have to be performed by data holders instead. They would also have to contact the affected persons. In this alternative approach, data access bodies would then communicate only the pseudonym of affected persons to data access bodies. Again, such an approach may bear benefits in terms of data minimization as well as liability of data access bodies.

- The legal text does not specify which actors should use the same pseudonyms for data subjects. It could be interpreted in a way where the data access body and all data users could share the same pseudonyms, rendering the linking possible. In support of purpose limitation it may be beneficial to use different pseudonyms for all actors, unless explicitly required by the purposes. This could be achieved with 2nd-level pseudonymization[34] executed by the data access body for every data user or even request.

# 6 Conclusions

This deliverable has provided a survey of the legal landscape and its relevance for the project. In addition, it has reported about the collaboration with relevant working groups that was conducted as part of Task 4.9.2. A more detailed analysis of the legal texts that are most relevant for the work conducted in Work Package 4.9 is provided in section 5. It is used, for example, in the collaboration with technical partners to describe the problems in search of good technical solutions, and as a basis for the work in Tasks 4.9.7 and 4.9.8.

[34] See the Pseudonymization Terminology in Deliverable D4.9.4 for more information of this concept. The concept was introduced in the Terminology to facilitate discussion of this use case.