Deliverable 4.9.1

Workshop on Legal Aspects of Anonymity and Pseudonymity

Funded by Bundesministerium für Bildung und Forschung

Funded by the European Union NextGenerationEU

https://anomed.de

UAP	4.9.1
Date	May 2024
Version	1.0
Status	final
Distribution	PU
Lead Contributors (© by affiliation)	Bud P. Bruegger (ULD)
Additional Contributors (© by affiliation)	Harald Zwingelberg (ULD)
Reviewers	Harald Zwingelberg (ULD)
License	CC-BY 4.0

Table of Contents

1 Objectives of Sub Work Package

AnoMed is an interdisciplinary research project in which ULD represents the legal discipline, in particular for data protection. A major success factor for interdisciplinary endeavors is a minimal mutual understanding. This workshop and its material aim at fostering the understanding of legal aspect of data protection.

AnoMed and its scientists aims at advancing the technology and methodology of anonymization. Clear awareness of the requirements helps to guide advances in a direction that is suited for practical use. The majority of requirements in the area are legal and originate in the General Data Protection Regulation. A major objective of the described work is therefore to create a general understanding of the **obligations and principles of the GDPR**.

To maximize impact, research of anonymization techniques should take the evolving European Data Strategy[1] and other relevant context into account. In particular, understanding the potential needs and roles of anonymization techniques in this context can help the flow of the state of the art into practical use. The data strategy is implemented in a first step in the form of legal acts. In particular, the Data Governance Act is already in force; the Commission has proposed the European Health Data Space (regulation); and the Data Act, plus 9 additional data spaces, are to follow. The workshop therefore addresses significant **current developments** in the legal space which shape the context of research.

As in all interdisciplinary projects, discipline-specific languages impede common understanding. A significant aspect of this is that different disciplines use different conceptualizations and thus terminology to speak about the same. The obligations of the GDPR as well as the wider context of the data strategy are expressed in a legal terminology. While a suitable **terminology** that bridges between the legal and technical/scientific languages is being developed in Sub Work Package 4.9.4, the objective of the present work is to make it accessible to project partners.

While the material developed within the present Sub Work Package is primarily targeted at project participants but is also **useful beyond the project**. The approach taken to enable wider use is to keep the slides detailed and self-explanatory. The approach taken for the terminology will be described in Deliverable 4.9.4.

2 Overview of Major Results

The task UAP 4.9.1 has produced the following results:

- For fostering a **basic understanding of the GDPR and its principles**:
    - **Workshop Module 1 "*Overview of the GDPR*"** (49 slides)
      The workshop was held on January 18, 2024. It was integrated in the AnoMed seminary series to maximize the attendance of project partners. Bud P. Bruegger and Harald Zwingelberg presented at the workshop. The **slides** are provided are available under a Creative Commons license for further use beyond the project at https://uldsh.de/anomed-gdpr-overview.
    - **Workshop Module 2 "Principles of the GDPR"** (45 slides)
      The workshop was held on February 1, 2024. It was integrated in the AnoMed

[1] https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en, last visited 24/01/2024.

seminary series to maximize the attendance of project partners. Bud P. Bruegger and Harald Zwingelberg presented at the workshop. The **slides** are provided are available under a Creative Commons license for further use beyond the project at https://uldsh.de/anomed-gdpr-principles.

- **Workshop Module3 "Pseudo/Anon Terminology"** (45 slides)
  The workshop was held on February 15, 2024 introduces the pseudo/anon terminology and the rational behind it. It was integrated in the AnoMed seminary series to maximize the attendance of project partners. Bud P. Bruegger and Harald Zwingelberg presented at the workshop. The **slides** are provided are available under a Creative Commons license for further use beyond the project at https://uldsh.de/anomed-termionology.

- In addition to the planned work, the **concepts** presented in the first two modules were **further evolved and validated** internally at ULD by reaching out beyond the research department of the data protection supervisory authority. The evolution is shown in the **appendix** and will be subject of a **future publication** and wider use beyond the project.

- One of the objectives of UAP4.9.1 is to inform project partners about significant **current developments**. Based on internal discussions, we decided that a one-time workshop was ill-suited for this purpose. We therefor decided that whenever there are significant legal developments that shape the context of AnoMed's research, a **newsbrief** is sent out by e-mail. The briefs are intentionally kept short in order to cater to busy researchers. When the brief wakes an interest for more, it contains follow-up links and ULD is also available for further information and discussion.

3 Rational behind Contents

While the main result of UAP4.9.1 are the workshops and their slides, the following explains the rational on why and how its content was newly developed in an attempt to cater to the needs of the project. This is discussed separately for (i) the obligations and principles of the GDPR and (ii) the pseudo/anon terminology.

3.1 On Understanding the Obligations and Principles of the GDPR

In many cases, **legal laypersons** such as technical professionals are responsible for making **decisions which affect data protection**. This is for example the case where persons from other disciplines than the legal profession make technical and organizational decisions about how to implement a processing activity. Such decisions obviously have to comply with the requirements laid down by the GDPR.

The **GDPR,** with its 99 Articles and 176 recitals, is **relatively large and complex**. Beyond the text of the GDPR itself, understanding it requires additional knowledge such as that about common interpretations (as for example, documented in legal commentaries) and court decisions. Together with its legal language, the size and complexity makes the GDPR **very difficult for a technical audience to read** and interpret correctly.

For this reason, there is a **demand for explaining the GDPR** and its requirements to legal laypersons. Such an **"explanation"** is typically a **simplified abstraction of the GDPR** that is easier to digest than the GDPR itself.

We call such simplifying abstractions "**models"**. They represent **only the aspect** of the GDPR that is **relevant for the audience**. For example, for a technical audience, the **obligations for controllers and processors** are relevant, while the tasks of supervisory authorities or the harmonization among supervisory authorities in Europe are not.

A **model** is described in terms of **concepts** and **relationships between them**. In the language describing the model, the concepts typically correspond to the key **terms** of the description. These terms are then often defined in a glossary. The **proximity to terminology** work is evident.

The necessary **reduction of complexity** of a model as compared to the GDPR itself is achieved by **reducing the number of concepts** and the number and complexity of their relationships. In particular, the **number of concepts must be small** compared to the 99 Articles and 176 Recitals with their complex interactions. Note that a reduction of concepts also drastically reduces the possible relationships (i.e., interactions) between concepts[2].

To **communicate** successfully, the used model needs to match the basic **way of thinking** and **reasoning** of the intended audience. In particular, what is "logical"[3] to one discipline may not be so for another; what is evident for legal professionals may not be so for techies. A discipline's professional way of thinking is often a**cquired in** a possibly extensive **professional education**. Sayings like "think like a mathematician" or "like a computer scientist" provide evidence for this.

This way of thinking can be interpreted as a certain way in which a discipline conceptualizes the world and thus how it builds its models. A discipline may use a vast amount of models but they all share some basic characteristics. Therefore, in order to explain the GDPR successfully to a technical audience, the **model of the GDPR** used should possess the **same model characteristics that the audience is used to and expects**.

As guiding principles, the **model characteristics expected by a technical audience** were therefore identified as follows:

1. Concepts should be either **mutually dis**joint or possess **well-defined relationships** with other concepts.
2. Concepts should convey only a **single aspect**; where multiple aspects are addressed by a given concept, this should be split into several distinct concepts.
3. **Relationships** between concepts often form a **hierarchical structure**. Typical relationships in the technical world that possess such properties include "part of" and "is a".
4. As consequence of a hierarchical relationship structure, many concepts are unrelated. This permits a "**divide and conquer**" approach to reasoning, that considers solely a sub-branch of the hierarchy without having to consider the rest of the tree. This underlies the technical meta-concepts of "**modularization**" and "**separation of concerns**".
5. The set of concepts of a model should be **complete** in the sense that the model is sufficient to answer all questions of a targeted domain.

[2] In particular, n elements have n(n-1) / 2 possible connections (binary relations). 99 Articles thus have 4851 possible binary-relations, while 7 protection goals used in the SDM (see footnote 4) have only 21.

[3] "Logical" is quoted here to express that it is relative to a certain discipline. "Legal logic" is thus different from "mathematical logic".

6. Ideally, how each concept of the model relates to the whole must be clear. Concretely, it should be clear how every concept of a data protection model actually is necessary and supports data protection.

ULD's research department has significant experience of introducing technical audience (typically project partners) to the GDPR. In past projects, two models were used for this purpose:

- The **Standard Data Protection Model**[4] (SDM) that has its origin at ULD and is the official model used in Germany[5] to do perform "technical data protection"; and
- the **principles of the GDPR**[6] as stated in its Article 5.

These models cater to multiple disciplines (legal, technical, managerial, financial, etc.) in the case of the SDM and legal professionals in the case of the principles. Their design objectives are therefore different from the model that is sought here that is optimized to optimize communications to a solely technical audience.

Not surprisingly, the previously used models do not possess the model characteristics that are typical for technical models. The following short examples shall illustrate this:

- The scope of the SDM does not cover all requirements of the GDPR. For example, there is no protection goal (which are the main concepts of the SDM) that requires controllers to find a valid legal basis (according the GDPR principle of *lawfulness*). Also, protection goals leave room for interpretation, how much is actually required by the GDPR. For example, *transparency* is a general concept that is used to demand extensive logging while the GDPR's notion of transparency may be limited to adequately informing data subject[7]. Further, the GDPR principles of accountability and transparency both map to the same SDM protection goal of transparency.
- The principle of *data minimization* is typically understood to also have a temporal component. This is for example evident in the EDPS glossary entry[8] that states that "[controllers] should also retain the data only for as long as is necessary to fulfil that purpose". This evidently overlaps with the principle of *storage limitation*. The relationship of these two concepts is more complex than *storage limitation* being simply the temporal aspect of *data minimization*. This is evident in the fact that storage limitation speaks of a "form which permits identification of data subjects", i.e. also addresses identifiability that is for example relevant for pseudonymization[9].

Based on this assessment, it was decided to develop a new model for AnoMed that attempts to improve the introduction of a technical audience to the GDPR as compared to similar activities in previous projects. This goes beyond a purely didactical objective since a better understanding of data protection requirement on the part of technical professionals seems to have the potential of significantly raising the status quo of data protection compliance overall.

---

[4] Version 3 of the SDM in German is available at https://www.datenschutzzentrum.de/uploads/sdm/SDM-Methode_V3.pdfand Version 2.0.b in English is available at https://www.datenschutzzentrum.de/uploads/sdm/SDM-Methodology_V2.0b.pdf.

[5] Different versions of the SDM were officially adopted by the German Conference of the Independent Data Protection Supervisory Authorities of the Federation and the Länder.

[6] https://gdpr-info.eu/art-5-gdpr/

[7] See Articles 12 through 14 GDPR.

[8] https://www.edps.europa.eu/data-protection/data-protection/glossary/d_en#data_minimization

[9] At least one of the possible interpretation of the legal text is that a controller shall pseudonymized data to significantly lower the risks for data subject as soon as the purposes of processing permit this.

3.2 Rational behind Pseudo/Anon Terminologies

The rationale behind the third workshop module on the Pseudo/Anon Terminology is described in detail in D4.9.4. The workshop focused on the rational much rather than the terminologies themselves. The latter are easier to consume in their highly graphical handout format than in a presentation.

4 Conclusions

This deliverable reported about the workshops that were held as part of UAP4.9.1 and the rationale behind the development of the material. All slides are available also outside the project under a Creative Commons license at the stated URLs.

5 Appendix

5.1 Evolution of the Concepts relative to GDPR Overview and Principles

In addition to the planned work, the **concepts** presented in the first two modules were further evolved and validated internally at ULD by reaching out beyond the research department of the data protection supervisory authority. The evolution is shown here and will be subject of a future publication and wider use beyond the project.

The status of the concepts at the time of the workshop presentation are shown in Figures 1 and 2.

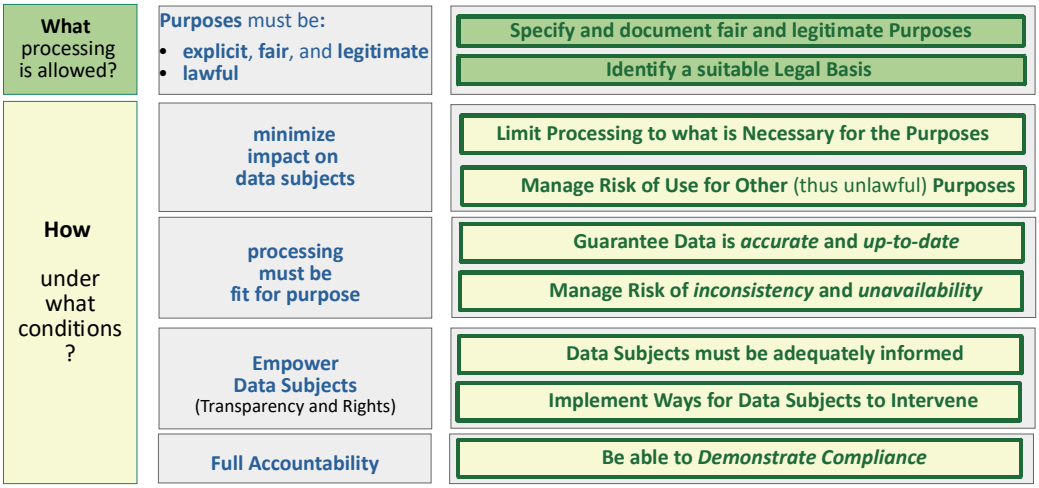

Figure 1: Conceptualization of the GDPR structure at the time of the workshop presentation.

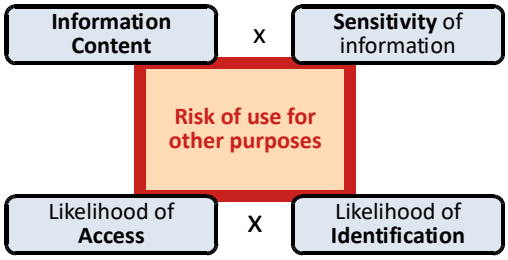

Figure 2: Conceptualization of risk factors at the time of the workshop presentation.

The status of the evolved and improved version is shown in Figures 3 and 4.

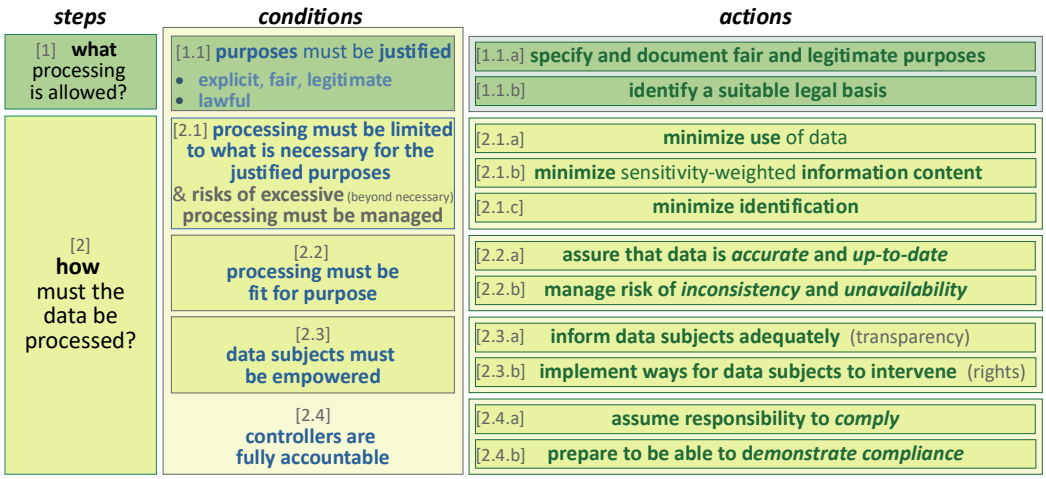

Figure 3: Evolved conceptualization of the GDPR structure.

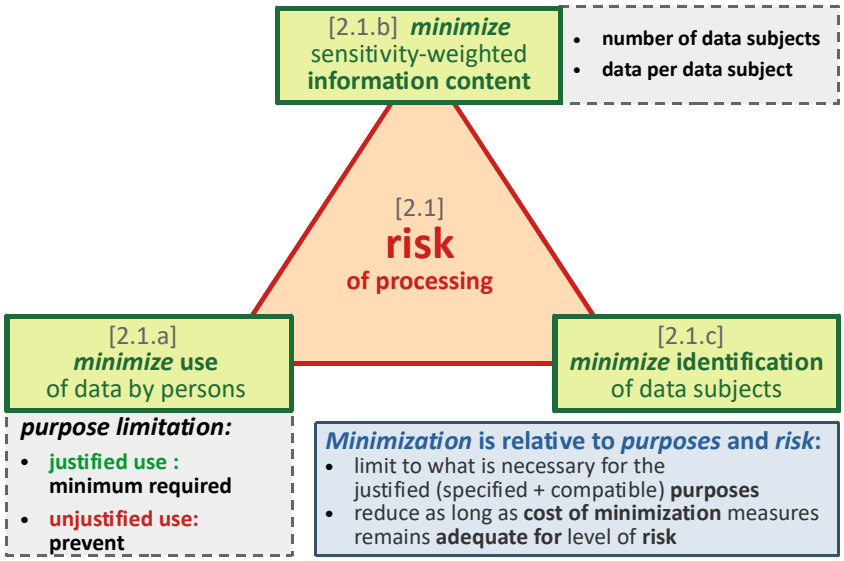

Figure 4: Evolved conceptualization of risk factors.