PANELFIT

PARTICIPATORY APPROACHES TO A NEW ETHICAL AND LEGAL FRAMEWORK FOR ICT

Guidelines on Data Protection Ethical and Legal Issues in ICT Research and Innovation.

THE GDPR – MAIN PRINCIPLES

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 788039. This document reflects only the author's view and the Agency is not responsible for any use that may be made of the information it contains.

1 Principles

Bud P. Bruegger (ULD)

Acknowledgements: The author thankfully acknowledges the review and suggestions by Giuseppe D'Acquisto, Senior Technology Advisor, Italian Data Protection Authority (Garante per la Protezione dei Dati Personali).

This section of the Guidelines has been validated by José Luis Piñar, former president of the Spanish Data Protection Agency and currently Cátedra Google on Privacy, Society and Innovation Universidad CEU-San Pablo, Madrid

The Section "Understanding data protection: the EU regulation in a nutshell" above has given an overview of the GDPR. It has thus also introduced the *principles* of data protection, as contained in Chapter 2 "Principles" of the GDPR and there in particular in Art. 5 "Principles relating to processing of personal data". While *Understanding data protection: the EU regulation in a nutshell* has chosen a structure that motivates the content of the GDPR in terms of power, the present section follows the structure of Art. 5 GDPR. It discusses each principle in further detail.

The principles express the following structure:

- **Conditions on the purposes** of processing: What kind of *purposes* pursued by the processing of personal data are allowed is described in **Art. 5(1)(a)** and **5(1)(b)** GDPR. Processing of personal data for purposes that fail to satisfy these conditions is not allowed. The conditions are:
    - ***Lawfulness*** (Art. 5(1)(a) GDPR);
    - ***Legitimacy*** (Art. 5(1)b) GDPR).
- **Conditions on the implementation** of processing: Where the purpose meets the above criteria, to be permitted, the implementation of the processing must in addition meet certain conditions. These are described in Art. 5(1)(a) though 5(1)(f); namely the implementation:
    - must be ***fair*** (Art. 5(1)(a) GDPR);
    - must be ***transparent*** (Art. 5(1)(a) GDPR);
    - must be ***limited to the stated purposes*** (Art. 5(1)(b) GDPR);
    - must use the **minimum of data** that is necessary for the purposes (Art. 5(1)(c) GDPR);
    - must use **only accurate data** (Art. 5(1)(d) GDPR);
    - must use the **minimum degree of identification** of data subjects that is necessary for the purposes (Art. 5(1)(e) GDPR);
    - must be **secure** (Art. 5(1)(f) GDPR).

In addition, according to Art. 5(2) GDPR, for controllers to **comply** with the GDPR means that their **processing**:

- **satisfies all the above conditions** and
- the controllers are able to **demonstrate it**.

To aid readers to understand the GDPR, the detailed discussion of the above principles uses the structure provided by the law. This means, that one point of the GDPR is discussed at a time. **Each point** of Art. 5(1) and Art. 5(2) are then called a ***principle***. The name of the principle that is provided by the GDPR corresponds to the titles use for the following sections. In some cases, several of the above stated conditions fit into a single principle.

There are two exceptions to structuring the following discussion by paragraph of Art. 5 GDPR. They are motivated by increased clarity and discuss statements provided in one paragraph of the GDPR under the principle (i.e., main meaning) provided in another paragraph. Namely, the exceptions are that:

- the requirement that purposes must be *specified*, *explicit* and *legitimate* (provided in Ar. 5(1)(b) GDPR) is discussed together with *lawfulness, fairness, and transparency* (of Art. 5(1)(a) GDPR), and
- the statement about the storage period pertaining to certain kinds of processing (provided in Art. 5(1)(e) GDPR) is discussed together with data minimization (of Art. 5(1)(c) GDPR) since arguably, the storage period is pertinent to the data being (temporarily) "*limited to what is necessary in relation to the purposes*".

The following table gives an overview of how principles relate to letters of Article 5 GDPR.

	Art. 5(1)(a)	**Art. 5(1)(b)**	**Art. 5(1)(c)**	**Art. 5(1)(d)**	**Art. 5(1)(e)**	**Art. 5(1)(f)**	**Art. 5(2)**
Legitimacy and **Lawfulness**							
Fairness							
Transparency							
Purpose limitation							
Data minimization							
Accuracy							
Storage limitation (minimization of identification potential)							
Integrity and **Confidentiality**							
Accountability							

The discussion of each principle is structured as follows:

- An abstract **description** of the principle,

- a brief discussion of **related articles and recitals of the GDPR** suited to provide a deeper understanding of the principle, and
- examples of concrete **technical or organizational measures** that can be used to implement the principle.

The description attempts to capture the essence of the principle. The section on related articles and recitals points to places in the GDPR that describe in more detail how the principle needs to be concretely applied. This section may be appreciated in a first reading and consulted when a deeper understanding is desired. The section on measures provides a non-exhaustive list of examples of how each principle can be implemented in practice.

The remainder of this chapter describes the principles listed in Art. 5 GDPR using the described structure.

1.1 Lawfulness, fairness and transparency

Bud P. Bruegger (ULD)

Acknowledgements: The author thankfully acknowledges the contribution by Iñigo de Miguel Beriain (UPV/EHU) who wrote an analysis of this principle as input to the here presented description.

The following discusses the principle of *lawfulness, fairness and transparency* that is defined in Art. 5(1)(a) GDPR.

***Lawfulness, fairness and transparency* at a glance:**

According to the GDPR, processing must be *lawful* and in pursuit of *legitimate purposes*. It further has to be *fair* and *transparent*.

Lawfulness is defined very precisely in the GDPR and is achieved if the purpose of processing falls into one of the six categories (aka. *legal bases*) listed in Art. 6(1) GDPR.

Legitimate is a much wider concept, meaning compliance with the letter of the law the spirit of the law, the values of society (in particular, the *European Charter of Fundamental Rights*), and the principles of *ethics*.

Fairness is used in its common understanding. It prohibits for example manipulative practices on part of the controller, such as nudging. Arguably, most articles of the GDPR are about fairness. To name the principle explicitly may be a fallback for the case where a consequence of fairness may not be spelled out explicitly in the GDPR. This prevents any loop holes.

Transparency of processing is a main strategy to balance power between controller and data subject. It works by pulling everything into the light and thus open it up to scrutiny.

It is spelled out in the GDPR as detailed requirements of information that has to be provided by the controller to both, data subjects and supervisory authorities.

1.1.1 Description

In "Understanding data protection: the EU regulation in a nutshell" above, most of the properties required in this principle were discussed in terms of balancing the power between controller and data subjects. This is summarized in the following: Both, *lawfulness* and *legitimacy* of the purposes is presented as a pre-requisite for the processing to be permissible. See "For which purposes is processing allowed for detail". *Fairness* was not discussed in the introduction. Arguably, by balancing the power between controller and data subjects, the whole GDPR is about fairness. *Transparency* was presented as a pre-requisite for accountability. See "Controllers are fully accountable" for detail.

The GDPR defines the principle as follows:

Definition in Art. 5(1)(a) GDPR:

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

Lawfulness, fairness, and transparency are discussed in more detail in the following.

1.1.1.1 Prerequisite to lawfulness: specified, explicit purposes

Lawfulness is a requirement for the purposes of processing[1]. It is therefore impossible to reason about it without first knowing the precise purposes that are pursued by the processing. For this reason, the requirement from **Art. 5(1)(b)** that purposes must be specified and explicit is discussed here as a prerequisite:

Personal data shall be collected for **specified, explicit** and legitimate **purposes**

Specified purposes:

The *Article 29 Data Protection Working Party* writes[2]:

> "**Purpose specification lies at the core of the legal framework** established **for the protection of personal data**. In order to **determine whether data processing complies with the law**, and to establish what data protection

[1] It is outside the scope of this document to provide a thorough legal analysis of the concept of purpose beyond its meaning in common language. It shall solely be pointed out that purposes of processing usually are related to an objective that the controller pursues. Such objectives should be concrete (rather than theoretical) and it is often possible to determine whether the objective has been reached or measure to which degree it has been reached.

[2] Highlighting added by the author, for the citation, see page 15 of: Article 29 Data Protection Working Party, 00569/13/EN, WP203, Opinion 03/2013 on purpose limitation, Adopted on 2 April2013, https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf (last visited 27/05/2020).

safeguards should be applied, it is a **necessary precondition to identify the specific purpose(s)** for which the collection of personal data is required."

The specification can be seen as the first task of the conceptualization of a processing activity that guides all subsequent decisions including:

- whether the **processing is permissible**, i.e., lawful and legitimate,
- **what the implementation** of the processing that needs to achieve the purposes **entails**, and
- what data protection **safeguards** should be applied.

The *Working Party* further states[3]:

> "The **purpose** of the collection must be **clearly** and **specifically** identified: it must be **detailed** enough to determine what kind of processing is and is not included within the specified purpose, and to allow that compliance with the law can be assessed and data protection safeguards applied."

and

> "For these reasons, **a purpose that is vague or general**, such as for instance 'improving users' experience', 'marketing purposes', 'IT-security purposes' or 'future research' will – without more detail - usually **not meet the criteria of being *specific.*"**

Explicit purposes:

The Working Party further states[4]:

> "Personal data must be collected for **explicit purposes**. The purposes of collection must **not only** be specified **in the minds** of the persons responsible for data collection. They must also be made explicit. In other words, they must be **clearly revealed, explained or expressed in some intelligible form**."

Note that the requirement to make the purposes explicit is closely related to informing data subjects about the purposes of processing (see Art. 13(1)(c) and 14(1)(c) GDPR).

Based on the pre-requisite of specified explicit purposes, legitimacy and lawfulness can be discussed.

1.1.1.2 Legitimacy and lawfulness

While Art. 5(1)(a) GDPR speaks only of *lawfulness*, the closely related requirement of *legitimacy* is stated in Art. 5(1)(b) GDPR. Since both express requirements regarding the purposes of processing, they are discussed here together.

Art. 5(1)(b) GDPR states:

Personal data shall be collected for *specified, explicit* and **legitimate purposes** and […]

[3] WP203, page 15, highlighting added by the author.
[4] WP203, page 17, highlighting added by the author.

The GDPR fails to provide a definition for *legitimacy*; but the *Article 29 Data Protection Working Party* provides the following[5]:

> The requirement of ***legitimacy*** means that the purposes must be **'in accordance with the law' in the broadest sense**. This includes **all forms of written and common law**, **primary and secondary legislation**, **municipal decrees**, judicial **precedents**, **constitutional principles**, **fundamental rights**, **other legal principles**, as well as **jurisprudence**, as such 'law' would be interpreted and taken into account by competent courts.

Legitimacy is thus a very **broad requirement**. This becomes even more significant when considering that certain legislation, such as the *Clinical Trial Regulation*[6], include also ***ethical requirements***. But even where ethics is not prescribed by the law, there is a danger that purposes that are clearly unethical may be considered to also be illegitimate. For example, this may be the case where processing takes place in disregard of a disapproval by a research ethics committee.

In contrast to *legitimacy*, ***lawfulness*** is indeed defined in the GDPR. Namely, **Art. 6(1)** GDPR reads:

> Processing shall be **lawful** only if and to the extent that at least one of the following applies: […]

In the omission represented by […], six possible so called *legal bases* are listed. They can be seen of categories of purposes. These are described in more detail in the section "Related articles and recitals" below.

1.1.1.3 Fairness

Arguably, all of data protection and thus the GDPR is about fairness towards data subjects. The GDPR can be seen in spelling out what *fair* actually and concretely means.

So its explicit mention as a principle may be considered to be a "fall-back clause" for the case where a concrete requirement of fairness has not been explicitly stated in the GDPR. Even if this case, the *fairness* principle would prevent any "loophole" in the GDPR.

While the whole GDPR can be considered to be about fairness, the section "Related articles and recitals" below gives some examples where fairness is particularly evident.

[5] WP203, page 20, , highlighting added by the author.

[6] REGULATION (EU) No 536/2014 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 16 April 2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC, https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-1/reg_2014_536/reg_2014_536_en.pdf (last visited 27/05/2020).

1.1.1.4 Transparency

Transparency is a well-understood concept and is a key pre-requisite for accountability in the GDPR. The main focus of transparency is to inform **data subjects** up-front[7] of the existence of the processing and its main characteristics. Other information (such as the data about the data subject) is available on request. Data subjects also have to be informed of certain events, most notably data breaches (in the case where the data subject is exposed to high risk). Transparency is also supported by controllers designating a Data Protection Officer who acts as single point of contact for concerns by data subjects. In the GDPR, data subjects are empowered to be the main guardians of their own rights and freedoms. Evidently, transparency is a pre-requisite for detecting and intervening in case of non-compliance.

Supervisory authorities, as obvious from their name, are also guardians of the compliance with the GDPR, even if their involvement is often triggered by complaints lodged by data subjects[8]. There are transparency requirements for controllers that are specifically targeted at supervisory controllers, including the records of processing (see "Documentation of Processing" in the section "Main Tools and Actions" of Part II) and Data Protection Impact Assessments (see the section with the same name in "Main Tool and Actions", Part II of these Guidelines). Controllers being answerable[9] to supervisory authorities and having to permit on-premise[10] investigations and audits[11] further implement transparency.

1.1.2 Related articles and recitals

1.1.2.1 Lawfulness

The definition of lawfulness is given in Art. 6(1) GDPR. It reads as follows:

> Processing shall be ***lawful*** **only if and to the extent that** at least one of the following applies:
>
> (a) the data subjects have given **consent** to the processing of their 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**

[7] Up-front here means that data subjects should be aware of the processing before it takes place. It does not imply a certain method of providing information or exclude dynamic ways of providing the necessary information.
[8] See Art. 57(1)(f) GDPR.
[9] See Art. 58(1)(a) GDPR.
[10] See Art. 58(1)(f) GDPR.
[11] See Art. 58(1)(b) GDPR.

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.

Point (f) of the first subparagraph shall not apply to processing carried out by public authorities in the performance of their tasks.

Whereas purposes of processing must be specified and explicit (see Art. 5(1)(b), and therefore also sufficiently narrow and specific, the above are clearly **categories of purposes**. (Where the word purpose was used explicitly, it is therefore written in italics). They are commonly called ***legal bases***[12] and are references by their position in Article 6; for example, *consent* would then be the *legal basis* of Art. 6(1)(a).

The GDPR provides two Articles that state **further requirements for lawfulness** for two different cases: **sensitive data** and data concerning **criminal convictions**. In particular these are the following:

Art. 9 GDPR states that the processing of particularly sensitive data is in principle prohibited and lists 10 exceptions to that rule. The exceptions are comparable in structure to the legal bases of Art. 6. The Article specifies that data are particularly sensitive, if they reveal:

- racial or ethnic origin,
- political opinions,
- religious or philosophical beliefs,
- trade union membership,

or are:

- 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.

For these data, more stringent requirements apply in order for their processing to be considered lawful. For example, instead of just consent of Art. 6(1)(a), the processing of such sensitive data requires a more demanding level of consent called ***explicit consent*** (see Art. 9(2)(a) GDPR).

Like Art. 9 does for particularly sensitive data, **Art. 10 GDPR** further restricts the processing of "data relating to **criminal convictions and offences** or related security

[12] The term *legal basis* is used extensively in the GDPR and is recommended here as preferential term. Alternatively, the GDPR also contains the term *legal ground*. In the literature, the term *lawful basis* is also used.

measures". In particular, to be lawful, the processing must be either "carried out only under the **control of official authority** or when [it] is authorized by Union or Member State law provid[e] for appropriate safeguards for the rights and freedoms of data subjects".

There are several Articles and Recitals in the GDPR that specify the **concept of consent** (of Art. 6(1)(a) GDPR) in further detail. The most important are the following:

- **Art. 4(11)** which **defines consent**;
- **Art. 7** which lists **conditions for consent**; and
- **Art. 8** which regulates **conditions applicable to child's consent in relation to information society services.**

Considering that consent is a complex concept, the **European Data Protection Board** has issued authoritative ***Guidelines** 05/2020 **on consent** under Regulation 2016/679*[13].

Besides *consent*, also the concept of ***legitimate interest pursued by the controller*** (of Art. 6(1)(f) GDPR) is difficult to fully understand. What is crucial here is the restriction of "**except** where such interests are overridden by the interests or fundamental rights and freedoms of the data subject". This means, that the legitimate interest of the controller must be balanced with the interests of data subjects. To determine, whether this is the case, the controller has to conduct a so-called ***balancing test***. How to do this is described in "Main Tools and Actions" within Part II of these Guidelines. It is predominantly based on the *Article 29 Working Party*'s authoritative *Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC*[14]. While this opinion is based on the Data Protection Directive that pre-dated the GDPR, it is in general applicable to the interpretation of Art. 6(1)(f) GDPR. It is recommended for **further reading** on the subject.

1.1.2.2 Fairness

Arguably, the whole GDPR is about fairness. The following points out some articles of the GDPR that illustrate this particularly well.

One area where fairness is evident regards the requirements of transparency. Here, **Art. 12(1)** states that controllers shall provide information "to the data subject in a **concise**, transparent, **intelligible** and **easily accessible** form, using **clear and plain language**, in particular for any information addressed specifically to a child." Evidently, this prohibits the unfair practice to provide the required information in a form that is inaccessible to data subjects.

Similarly, **consent** cannot be implicit, but rather requires an explicit "statement or by a **clear affirmative action**" (see **Art. 4(11)** GDPR). The same article further states that

[13] EDPB, Guidelines 05/2020 on consent under Regulation 2016/679, Version 1.0, Adopted on4 May2020, https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-052020-consent-under-regulation-2016679_en (last visited 22/05/2020).

[14] Article 29 Data Protection Working Party, 844/14/EN, WP217, Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC, Adopted on 9 April 2014, https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf (last visited 22/05/2020).

consent must be **freely given**, **specific**, **informed** and **unambiguous**". Further, **at any time**, without need for justification, a data subject must be able to **withdraw** consent **as easily as it was given**. These stringent requirements for consent directly prohibit many manipulative practices, including the "nudging"[15] of data subjects.

Several **data subject rights** can directly be associated with fairness. These include:

- The **right to rectification** (Art. 16 GDPR) to prevent data subjects to suffer negative consequences due to inaccurate data;
- The **right to restriction of processing** (Art. 18 GDPR) that prevents controllers from further using data that have been reported to be inaccurate or pertain to processing the data subject has objected to;
- The **right to data portability** (Art. 20 GDPR) that prevents lock-in situations and a possible loss (e.g. of investment[16]) when users change their relationship with the controller;
- The **right to object** (Art. 21 GDPR) where in the case of a legal basis of Art. 6(1)(f) GDPR, data subjects can present **their specific situations** under which their interest prevail over the legitimate interests of the controller;
- The **right not to be subject to a decision based solely on automated processing** (Art. 22 GDPR), that also provides the **right to obtain human intervention** on the part of the controller (see paragraph 3).

Another indication of fairness is where the controller must take the data subjects' point of view into consideration. This is for example evident in Recital 50 GDPR that requires to consider the reasonable expectations by data subjects when determining whether a purpose is compatible according to Art. 6(4). It also appears in Data Protection Impact Assessments (Art 35 GDPR), where controllers, where appropriate, shall seek the views of data subjects or their representatives (Art. 35(9) GDPR).

1.1.2.3 Transparency

Several articles in the GDPR provide further detail on the principle of *transparency*. They include the following:

- **Articles 12 through 14** describe in detail the **information** that controllers must provide **up-front** to data subjects.
- **Art. 15** describes the information that needs to be provided on request by data subjects, including full access to their data.
- **Art. 34** describes how data subjects need to be informed of data breaches, where it is likely to result in a high risk.

[15] See for example, Weinmann, M., Schneider, C. & Brocke, J.v. Digital Nudging. Bus Inf Syst Eng 58, 433–436 (2016). https://doi.org/10.1007/s12599-016-0453-1 (last visited 22/05/2020).
[16] A prime example for a possible loss of investment is the collection of personal photos.

- **Art. 38(4)** designates the *Data Protection Officer* at the controller as access point for data subjects.
- **Art. 12** and **19** describe the information that controllers must provide to data subjects who exercise one of their rights.
- **Art. 30** ***records of processing*** and **35** ***Data Protection Impact Assessment*** describe the information that needs to be provided to supervisory authorities. (The latter only if the processing likely results in a high risk).
- **Art. 58(1)** specifies how controllers must be transparent towards supervisory authorities by being answerable (point a), allow inspections and audits (point b), and grant access to their premises (point f).
- **Art. 33** describes breach notifications towards supervisory authorities.

Considering the importance of transparency in the GDPR, the **European Data Protection Board** has provided an authoritative interpretation of related obligations in their **Guidelines on Transparency** under Regulation 2016/679 (wp260rev.01)[17]. This is recommended for further reading.

1.1.3 Related technical and organizational measures

Examples of measures to implement different aspects of the principle are provided in the following.

1.1.3.1 Legitimacy and lawfulness

- At least where the verification and demonstration of **legitimacy** requires **formal steps**, these can be considered organizational measures in support of legitimacy. A prime example are the **request and approval** of certain medical research through the competent **research ethics committee**.
- A pre-requisite for evaluating both, legitimacy and lawfulness is the **specification of explicit purposes**. This in itself can be considered a measure, in particular when it goes hand in hand with **considerations** about how to make the specification **as specific** and **narrow as possible**. In this case, also such analysis can be considered part of this measure.
- The main measure in support of lawfulness is to identify one or several **legal bases** of **Art. 6(1)** GDPR. In many cases, a processing activity uses multiple legal bases. A use case[18] published by the *Data Privacy Vocabulary Community Group* of the W3C provides an easily accessible example.
- Where Art. **6(1)(a)** GDPR, i.e., *consent,* was chosen as a legal basis, an **analysis** that justifies that the stringent **requirements** of the GDPR **for (freely given,**

[17] EDPB, Guidelines on Transparency under Regulation 2016/679 (wp260rev.01), https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=622227 (last visited 22/05/2020).
[18] Bruegger, Schlehahn & Zwingelberg, Data Privacy Vocabulary Community Group, Data Protection Aspects of Online Shopping – A Use Case, https://www.w3.org/community/dpvcg/2019/12/12/data-protection-aspects-of-online-shopping-a-use-case/ (last visited 25/05/2020).

informed) consent have been met is an important measure. This can for example include tests to see whether the information provided as basis for consent are indeed understandable to data subjects and whether the withdrawal of consent is indeed as easy as giving it.

    - In addition, where **children** or other **vulnerable data subjects** are affected, this analysis should put special focus on safeguards relative to Art. 7 GDPR.

- Where Art. **6(1)(f)** GDPR, i.e., *legitimate consent by the controller,* was chosen as a legal basis, measures include a precise specification of the legitimate interests, as well as a ***balancing test*** (see section of the same name in "Main Tools and Actions" within Part II of these Guidelines) to ascertain that these indeed prevail over the interests, rights, and freedoms of data subjects.
- With any legal basis, where controllers intend to **process** certain data **further,** beyond the initial purposes, for **compatible purposes** (see Art. 5(1)(b) GDPR), the analysis based on the criteria of Art. 6(4) for demonstrating that these additional purposes are indeed compatible, is a measure that demonstrates the lawfulness of such processing.
- If special categories of data (i.e., sensitive data) or data relating to criminal convictions are processed, further measures must be taken in addition to those relating to Art. 6(1) GDPR. In particular, in the former case, the condition of Art. 9(2) GDPR, why an exception to the prohibition of processing sensitive data applies, must be found and documented. In the latter case, the conditions that make the processing permissible according to Art. 10 GDPR shall be implemented and documented.

1.1.3.2 Fairness

- As has been reasoned above, all requirements of the GDPR can be considered a matter of fairness; several data subject rights were presented as particularly relevant, however. Prime measures in support of fairness are thus an adequate **implementation of data subject rights**.

1.1.3.3 Transparency

- Implementation of the requirements of Art. 12 through 14 GDPR to provide adequate and easy understandable **information to data subjects** is a prime measure to support transparency.
- The same goes for documents prepared to inform supervisory authorities, in particular the ***records of processing*** (according to Art. 30 GDPR) and a ***data protection impact assessment*** (according to Art. 35 GDPR). A further measure is the partial publication of this impact assessment.
- Any analysis that evaluates the effectiveness and accessibility of the provided information—possibly in regard with special categories of data subjects such as children—can be considered a measure in itself.

- The appointment of a Data Protection Officer can in part be seen as a measure to increase transparency both towards data subjects and the supervisory authority.

1.2 Purpose limitation

Bud P. Bruegger (ULD)

*Acknowledgements: The authors thankfully acknowledges the contribution by Iñigo de Miguel Beriain (UPV/*EHU) who wrote an analysis of this principle as input to the here presented description.

The following discusses the principle of *purpose limitation* that is defined in Art. 5(1)(b) GDPR.

***Purpose limitation* at a glance:**

Data that was **collected for specified "initial" purposes** shall **only be further processed**:

- for these **initial purposes**, or for
- **compatible purposes**.

For the general case, the GDPR gives **criteria** for how to **determine the compatibility** of purposes (see Art. 6(4)). In addition, some purposes are **preapproved as compatible** by the GDPR (see Art. 5(1)(b)) as long as appropriate safeguards are implemented (see Art. 89). Namely, these are:

- **archiving** in the **public interest**,
- **scientific or historical research**, and
- **statistics**.

1.2.1 Description

In "Understanding data protection: the EU regulation in a nutshell" above, *purpose limitation* was motivated by limiting the use of the gained power exclusively to reaching the declared and legitimate purposes. (See section "Restricting the controllers to use the power solely for reaching the declared legitimate purposes" for detail).

The GDPR defines the principle as follows:

Definition in Art. 5(1)(b) GDPR:

Personal data shall be collected for specified, explicit and legitimate purposes and **not further processed in a manner** that is **incompatible with those purposes**; […] (‘*purpose limitation*’);

Note that the first half of this sentence has already been discussed under the previous principle. In particular, the requirement that purposes must be ***specified*** **and** ***explicit*** was a **prerequisite for** being able to speak of ***lawfulness***; the requirement of legitimacy regards purposes and was therefore discussed together with *lawfulness*.

What is discussed here in more detail is the essence of this principle, namely the **limitation to processing compatibly with the purposes**. This is a requirement regarding the implementation of the processing activity, not the purposes.

1.2.1.1 Not processed in a manner that is incompatible with those purposes

The essential part of this principle is thus contained in the half-sentence “not further processed in a manner that is incompatible with those purposes”. The following analysis this sentence in more detail.

The sentence speaks about compatibility with **purposes**. It is clear from the first half of the sentence that these are the purposes **that have been explicitly specified**[19] (see section above on “Prerequisite to lawfulness: specified, explicit purposes”). The part of Art. 5(1)(b) that will be discussed below also uses the concept of “compatibility with ***initial purposes”***. The *initial purposes* therefore seem to be the same as those specified (during the conception of the processing activity).

Art. 5(1)(b) thus expresses, that processing shall be compatible with:

- the **initial purposes themselves**, or
- **other purposes** that are **compatible** with these initial purposes.

The former follows from the reasoning that purposes are always compatible with themselves.

The wording of Art. 5(1)(b) speaks of “**further** processed”. While this could be understood temporarily, i.e., in a sense of “after the initial purposes have been achieved”, the temporal aspect seems to be irrelevant for this principle. Instead, “further” has the meaning of “beyond” without temporal significance and refers purely to the purposes.

The situation is visualized in Figure:

[19] These are also the purposes that are communicated to data subjects as required by Art. 13 and 14 GDPR).

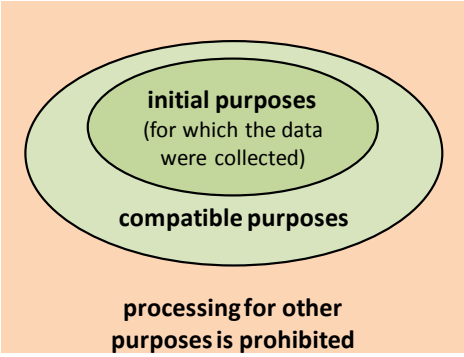

Figure 2: Processing is allowed for the initial and compatible purposes.

It is important to know that no additional legal basis is necessary to further process for compatible purposes. This is stated explicitly in Recital 50 GDPR (2nd sentence). Referring to further processing for compatible purposes, it states:

> In such a case, no legal basis separate from that which allowed the collection of the personal data is required.

1.2.1.2 Use for incompatible purposes

This raises the question how it can happen to process personal data for incompatible purposes and what its consequences are.

Understanding how processing can happen is important to be able to avoid it. The following three examples illustrate the issue without claim to comprehensiveness:

- **Function creep**: It is common for processing activities to evolve over time. It is also common that they then acquire new functionality or "features" that correspond to additional or modified processing. In cases where the controller fails to exercise sufficient control over such evolution, the processing can move unnoticed beyond the initial or compatible purposes.
- **Lack of separation**: Assume that a controller operates multiple independent processing activities that pursue distinct purposes. If the controller fails to implement adequate measures to separate the different processing activities, it is easy that data collected for one set of purposes is used for other purposes. This is illustrated in Figure 3:

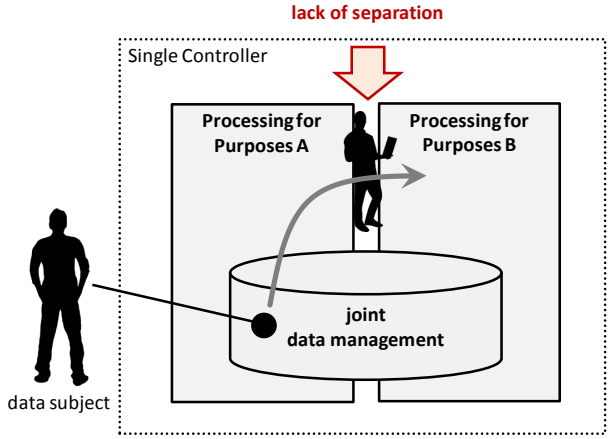

Figure 3: A lack of separation leads to the use of data for incompatible purposes.

- **Recipients who pursue their own purposes**: Recipients are persons or organizations to whom personal data is disclosed (see definition in Art. 4(9) GDPR). Recipients can for example be:
    - **employees** who access data *legitimately* on instruction by the controller to fulfill compatible purposes of the processing, or
    - **external attackers** who illegitimately accesses the data through a breach[20].

  In the latter case, it is obvious that the recipient uses the personal data for other purposes. It is these very purposes that likely motivated the attack in the first place. But even employees can have other interests in the data than pursuing the stated purposes of their employer. A prime example for that is where the employee already knows the data subject and learns information that would not otherwise be accessible.

With the understanding gained from these examples that illustrate how data can be used for other purposes, the question of the possible consequences must be asked.

In all cases, the **basic principle**s of *lawfulness* and *legitimacy* are likely **violated**. According to these principles, processing is prohibited unless it is justified by a demonstrated lawfulness and legitimacy of the purposes. This is obviously not the case when processing happens for incompatible, and thus unjustified purposes.

The use of data outside and beyond the justified purposes also **permits** rogue **controllers** to **accumulate power**. This can happen for example when controllers combine the data sets of persons across distinct processing activities, keep and accumulate data when they are no longer necessary for the purposes, and possibly even acquire data from other sources in order to gain more power over their data subjects. Such accumulated power evidently exceeds the power gain that was justified by a demonstrated lawfulness and legitimacy of the initial purposes.

It is evident that beyond the sole violation of data protection principles, depending on the purposes for which the data is (ab)used, **data subjects** can also experience **material**

[20] Controllers are not responsible for the actions of attackers but only to prevent attacks through adequate security measures.

or immaterial damage. For example, knowledge of certain health data may significantly affect relationships when accessible to acquaintances or prevent employment opportunities when accessible to potential employers. When used for criminal purposes, some kinds of data may be the basis for blackmail.

1.2.1.3 When are purposes compatible?

The following discusses how to determine whether potential additional purposes are considered compatible. It is predominantly based on Art. 6(4) GDPR.

In the case where a **legal basis** of ***consent*** (see Art. 6(1)(a) GDPR) was chosen for the processing, further processing for **additional purposes** other than the preapproved compatible ones (see below) are **deemed incompatible**[21]. This is because consent is always specific to specified[22] purposes. To "widen" the purposes of processing beyond the specified ones purposes that a data subject has consented to, would be clearly unfair and nontransparent.

Art. 6(4) then provides the following **criteria** to be used by controllers for determining whether an additional purpose is compatible (reworded slightly compared to the GDPR):

> (a) Any **link between** the **initial purposes** and the **additional purposes** under consideration;
>
> (b) the **context in which** the personal **data have been collected**, in particular regarding the **relationship** between **data subjects** and the **controller**;
>
> (c) the **nature of the personal data**, in particular whether they include **special categories of** (i.e., sensitive) personal **data** or personal data related to **criminal convictions** and **offences** are processed;
>
> (d) the **possible consequences** of the intended further processing **for data subjects**;
>
> (e) the **existence of appropriate safeguards**, which may include **pseudonymization**.

Further guidance including examples of applying these criteria is available from the *Article 29 Data Protection Working Party*[23]. While this opinion refers to the *Data Protection Directive* (i.e., the predecessor or the GDPR), many aspects are still equally applicable today.

To simplify the determination whether additional purposes are compatible, the **GDPR preapproves some** of the most common additional **purposes** pursued in further processing. Namely, Art. 5(1)(b) includes the following:

> [F]urther processing for **archiving** purposes **in the public interest**, **scientific or**

[21] Note that Art. 6(4) GDPR about compatible purposes explicitly excludes that it is applicable when the legal basis is consent.

[22] In particular, these purposes are specified in the dialog that asks for consent and the specification is an important aspect of the informedness of consent.

[23] Article 29 Data Protection Working Party, 00569/13/EN, WP203, Opinion 03/2013 on purpose limitation, Adopted on 2 April 2013, https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf (last visited 28/05/2020).

historical research purposes or **statistical** purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes.

The mentioned Art. 89(1) requires the presence of additional safeguards.

Here, the mentioned Art. 89 GDPR mandates that further processing for these preapproved purposes is only admissible if adequate safeguards are in place.

1.2.2 Related articles and recitals

The **essence** of the principle of purpose limitation is described in **Art. 5(1)(b)** GDPR and also contains the listing of **preapproved compatible purposes**.

Art. 5(1)(e) GDPR provides further details on the possible **storage period** of data relating to further processing **for the preapproved compatible purposes**.

Recital 50 GDPR provides guidance for the interpretation of further processing for compatible purposes. Of particular interest is the second sentence that states that **no additional legal basis** separate from that which allowed the collection of the personal data **is required**.

Art. 89 GDPR mandates that when processing further for **preapproved compatible purposes**, controllers must implement **adequate safeguards**. It also opens to the possibility that in this context, Union or Member State law may provide for **derogations from certain data subject rights**.

1.2.3 Related technical and organizational measures

The following provides examples for technical and organizational measures in support of *purpose limitation*:

- A precise clear specification of the initial and potentially compatible purposes is a prerequisite for any reasoning about purpose separation.
- Understanding data protection as a process that includes **regular reviews** during the whole life cycle of the processing activity is important to avoid processing data for incompatible purposes, e.g., due to **function creep**. Note that regular review is mandated in the context of *data protection by design* (Art. 25(1) GDPR), *data protection impact assessments* (Art. 35(11) GDPR) and *security* (Art. 32(1)(d) GDPR).
- The **verification of the compatibility of purposes** according to Art. 6(4) can be considered an organizational measure in support of purpose limitation.
- **Analysis** of how **authorized personnel** may use personal data **for other purposes** is another organizational measure. Such analysis aims at identifying possible **motivations**, **conflicts of interest** (such as personnel processing data of relatives and acquaintances), and measures to **prevent**[24] **or mitigate** such situations (e.g., the possibility that an employee can signal a conflict of interest for an assigned case and pass it to another employee without conflict of interest).

[24] Another example to prevent conflicts of interest is when a large company processes in offices far from the affected data subjects in order to reduce the probability that employees process data of acquaintances.

- Another measure is an analysis of the **motivations** that **external attackers** may have to obtain the data for other purposes. This is an important part of risk assessment and a prerequisite for implementing adequate safeguards in support of purpose limitation.
- Any organizational or technical measure to implement **separation between distinct processing activities** pursued by the same controller are in direct support of purpose limitation.
- Any measure (such as encryption) in support of **confidentiality** prevents that unauthorized parties use data for illegitimate purposes.
- Any measure to ensure that **authorized personnel** acts **only on instruction and as instructed** by the controller (see Art. 29 and 32(4) GDPR) ensures that the processing does not go beyond that necessary to achieve the specified purposes.
- A secondary measure that mitigates the damage after a breach is **pseudonymization**. The drastically reduced possibility of identifying data subjects and linking to other data sets may in many cases effectively prevent the use of the leaked data for other purposes.

1.3 Data minimization

Bud P. Bruegger (ULD)

Acknowledgements: The author thankfully acknowledges the contribution by Andrès Chomczyk Penedo (VUB) who wrote an analysis of this principle as input to the here presented description.

The following discusses the principle of *data minimization* that is defined in Art. 5(1)(c) GDPR.

***Data minimization* at a glance:**

Data minimization restricts the data that is collected and used to those **adequate**, **relevant** and **limited** to what is **necessary in relation to the purposes**. The limitation to the necessary has two aspects:

- data volume (or more precisely, information content) and
- duration of storage.

Consequently, as little data as necessary shall be processed (and stored) for as short a time period as possible while still achieving the stated purposes.

1.3.1 Description

In "Understanding data protection: the EU regulation in a nutshell" above, *data minimization* was motivated by minimizing the power gain of the controller to that what is minimally necessary to fulfill the declared, legitimate purposes. In particular, it addressed the minimization of information content present in the processed personal data. This complements the minimization of the degree of association that the data have with the data subject, and the limitation of access to power. See "Minimization of power to what is necessary to fulfill the declared purposes" for detail.

The GDPR defines the principle as follows:

Definition in Art. 5(1)(c) GDPR:

Personal data shall be **adequate**, **relevant and limited to what is necessary in relation to the purposes** for which they are processed ('*data minimization*');

Evidently, this is only possible if these purposes are specified and explicit (as required in Art. 5(1)(b) GDPR).

1.3.1.1 Adequate, relevant and limited

Adequate and ***relevant*** are easy to understand: Data that is inadequate, i.e., unfit for the purposes, cannot be collected or processed; the data must also be relevant, i.e., it must serve the purposes.

To understand the ***limitation*** aspect, a more precise look at what *data* actually means is necessary. In particular, It is intuitive clear that not just the number of data elements is concerned here, but the actual **information content** of the data. The following shall illustrate this in relation to the purposes:

- **Selection**: Where a set of possible data elements is under consideration, **select** those that are necessary for the purposes. Note that if data is already stored, selection can also be understood as **deletion** of unnecessary data elements. Otherwise it is concerned with data that is actually collected.
- **Resolution**: Where data is available at multiple possible resolutions, **limit the resolution** to what is minimally necessary for the purposed. For example:
    - **Values**: express **values** at the **coarsest scale** that still supports the purposes,
        - for example, use an **age category** (40-59 years old, 20 year resolution) **instead of** a **date of birth** (one day resolution),
    - **Locations**: express **locations** in terms of the coarsest geographic subdivision possible,
        - for example, use **administrative units** such as postal code zones or provinces or **grid cells** instead of precise coordinates (of meters in resolution),
    - **Time Series**: express **time series** of data at the coarsest sampling rate that still supports the purposes,

        - this may require a resampling of the data obtained from some sensor,
    - **Fingerprints**: If you need to **only** compare data sets for **equality**, consider just processing some "**fingerprint**" of the data.
        - For example, a "cryptographic hash value" (aka. "digest") of the data may be sufficient to detect change[25].

- **Level of Aggregation**: Where possible, chose an adequate **level of aggregation**. Most of the data values we deal with are a form of aggregation, even if this may not be evident since it may be done "invisibly" by some sensor or data collection method. Aggregation is a way of **substituting several data elements by a single one**. Prime examples come from statistics and include the average, median, minimum, and maximum. In the context of data protection, two kinds of aggregation have to be distinguished:
    - **Single Person**: Aggregation of data elements pertaining to a **single person**:
        - Taking for example a person's average income over a year reduces the information content pertaining to that person.
    - **Multiple Persons**: Aggregation of data elements pertaining to a **multitude of persons**:
        - Taking for example the average yearly income over group of persons also reduces the overall information content (data minimization). In addition, it also weakens the degree of association between a data element and a given person. This kind of aggregation is therefore also pertinent to storage limitation (see section 1.5)

1.3.1.2 Temporal aspect

Data minimization clearly also has a **temporal aspect**. Most importantly, "limited to what is necessary in relation to the purposes" also means that it is no longer justified to store data when the purposes have already been fulfilled. Data therefore has to be **deleted as soon as it is no longer necessary**.

In practice, this may be even **more diversified**: Of the purposes (plural), some may be fulfilled earlier than others. Also, after the "main processing"[26], "further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes"[27] may take place. To model this we distinguish several **phases of processing**. The following figure attempts to visualize this situation.

[25] For further information about cryptographic digests, see for example, https://en.wikipedia.org/wiki/Cryptographic_hash_function (last visited 15/5/2020).
[26] The term "main processing" is used here to distinguish from "further processing".
[27] The wording was directly copied from Art. 5(1)(b) GDPR.

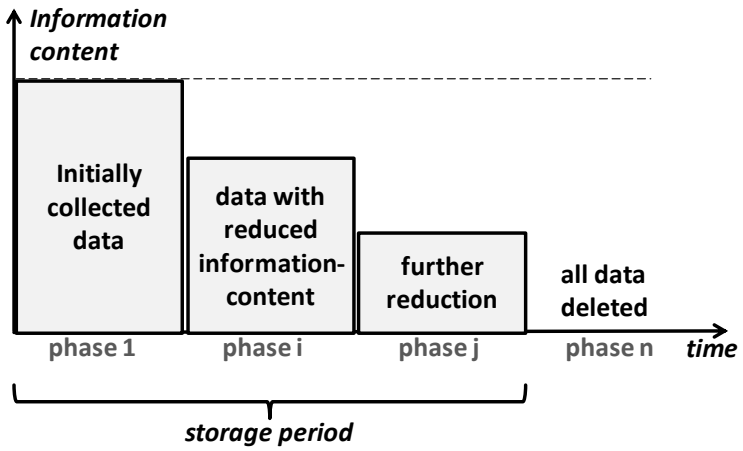

Figure 1: Reduction of information content in multiple steps.

In particular, the figure shows an example with four phases. Any number of phases is possible. Since every phase is associated with a subset of purposes, at the end of each phase, when the respective purposes have been fulfilled, certain data is no longer necessary. Consequently, at the **end of each phase**, certain data can be **either deleted** (selection), or its **information content can be reduced** (reduction of resolution or increase of level of aggregation). It is evident that such a diversified approach minimizes data further than a single-phase approach that keeps the full information content until all purposes have been fulfilled.

1.3.2 Related articles and recitals

Beyond the definition of *data minimization* given in Art. 5(1)(c), the second part of Art. 5(1)(e) "storage limitation" GDPR states explicitly states that:

> [P]ersonal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organizational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject;

This refers to the further processing for compatible purposes after fulfilling the initial purposes described in Art. 5(1)(b) GDPR[28].

Since it is concerned with storage of personal data, it is considered here to be pertinent to data minimization since the statement "limited to what is necessary in relation to the purposes" is not restricted to only data volume but clearly must also be understood to

[28] Namely, Art. 5(1)(b) contains the following statement: "further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes".

address the temporal aspect of data. Also, data minimization concerns all aspects of processing (such as *collection* and *disclosure*) and therefore also addresses *storage*.

For these reasons, the second part of Art. 5(1)(e) GDPR is considered here to provide guidance on how to interpret the principle of data minimization in the context of further processing for compatible purposes after fulfilling the initial purposes.

Beyond this, the GDPR emphasizes the importance of the principle in various contexts:

In Art. 25(1) GDPR on ***Data Protection by Design***, It emphasizes how ***data minimization*** shall be **considered in every phase of the life cycle** of a processing activity. This includes for example the analysis and conception phase of a processing activity where the purposes of processing are determined: Evidently, the more precise and narrow the purposes are specified, the clearer it becomes which data are actually necessary and the more data can be recognized as unnecessary. Similarly in a later life cycle phase, measures can be taken to implement effective deletion or reduction of information content.

Art. 89(1) and Recital 156 GDPR emphasize the **importance of *data minimization*** for the case where after fulfilling the initial purposes, data is processed further for "compatible purposes"[29]. In particular, "**archiving** purposes in the public interest, **scientific** or **historical research** purposes or **statistical purposes** shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes"[30]. Art. 89(1) GDPR (2nd sentence) explicitly mandates that for this further processing, "technical and organizational measures are in place in particular in order to ensure respect for the principle of data minimization".

1.3.3 Related technical and organizational measures

The following provides examples of technical or organizational measures in support of data minimization. It is not intended to be complete but rather to render the principle more concrete:

- **Know what data are necessary for the purposes**: Knowing which data is actually necessary is only possible with a precise and narrow definition of the purposes. To work out what is really needed is a measure in support of data minimization that is typically implemented during the conception or design phase of a processing activity.
- **Collect only necessary data**: During the design phase and the selection, implementation, and/or configuration of software, data acquisition, for example through input forms or dialogs, shall be designed such as to collect only the necessary data at the necessary level of detail.

[29] See Art. 5(1)(b) GDPR.
[30] Wording taken from Art. 5(1)(b) GDPR.

- **Delete data and reduce information content between phases of processing**[31]: Plan and implement the functionality to delete unnecessary data at the end of processing phases or otherwise reduce their information content.
- **Protect against exceeding the maximal storage period**: As a second line of defense, define a *maximal storage period*[32] and implement a procedure that alerts you about the presence of data that has exceeded this period. This measure protects against failures of deletion, for example those caused by a software bug that manifests in certain cases, a system crash during the deletion operation, or the restauration of data from a backup after a system crash although the data was previously already deleted.

1.4 Accuracy

Bud P. Bruegger (ULD)

Acknowledgements: The author thankfully acknowledges the contribution by Frédéric Tronnier (GUF) who wrote an analysis of this principle as input to the here presented description.

The following discusses the principle of *accuracy* that is defined in Art. 5(1)(d) GDPR.

***Accuracy* at a glance:**

Accuracy of data addresses both, **factual correctness** and being **up to date**. Is is prohibited to use inaccurate data that are unfit for purpose or that have negative consequences for data subjects. The main measure to implement this principle is to adequately support the data subjects' **right to rectification**.

1.4.1 Description

In "Understanding data protection: the EU regulation in a nutshell" above, *accuracy* (along with integrity) was motivated by the fact that accuracy of data is necessary in order to be fit for the declared purposes. Any processing that fails to be fit for purpose cannot justify a gain of power over a data subject. See "Prohibition of processing that fails to be fit for purpose" for detail.

[31] Note that this statement is relative to the overall data held by the controller. It is also assumed here that data are collected only once from/about data subjects and that no later data collection (e.g., as the need arises) takes place. The statement does not exclude that different phases or processing steps use only a subset of the overall data.

[32] Note this could be directly "the period for which the personal data will be stored" according to Art. 13(2)(a) or if the storage period depends on criteria, the maximal time when it can be expected that these conditions must have been met.

In addition to fitness for purpose, the processing of inaccurate data may have negative consequences for data subjects. These may range from an increased effort that is necessary to exercise one's rights, over the negation of rights and opportunities, up to negative financial or legal consequences. While processing that is affected by such flaws is arguably not fit for purpose, in addition it would violate the principle of *fairness* (see 1.1.1.3 above).

The GDPR defines the principle as follows:

Definition in Art. 5(1)(d) GDPR:

Personal data shall be **accurate** and, where necessary, **kept up to date**; **every reasonable step** must be taken to ensure **that personal data that are inaccurate**, having regard to the purposes for which they are processed, are **erased or rectified** without delay ('*accuracy*');

The following discusses various aspects of *accuracy* in further detail:

1.4.1.1 How can accuracy be assessed?

The concept of accuracy must be objective. It must be possible to verify whether data is accurate or not without doubt, and different verifiers must arrive at the same assessment. This is only possible when the data represents **verifiable facts**. This is for example not the case for data that represents an expression or a person's opinion.

The verification of the accuracy of data therefore typically involves the verification of facts that underlie the data. For example, to verify that a mobile phone number actually belongs to a person, a test message with a random code could be sent and received back over another channel.

In some situations, it may be the data subject who provides the controller with the necessary documentation of facts that permit a verification. For example, a data subject may supply a certificate of residency issued by a trusted authority in order to support the verification of an address of residence.

1.4.1.2 What does "up to date" mean?

When assessing whether data is up to date, the purposes of processing have to be taken into account. For example, a vendor may store the delivery address of a data subject whereas the data subject has since moved to a new residence. If the purpose of processing is to actually deliver goods to the data subject, the address is evidently out of date and the data is unfit for purpose. If the purpose of processing is billing for already delivered goods, however, the old address must be considered to be up to date.

1.4.1.3 How is inaccuracy of data discovered?

Inaccurate (including out-of-date) data must be rectified or deleted by the controller without delay. But how is inaccuracy in the data actually discovered and what responsibilities do controllers?

The probably most important mechanism for controllers to detect inaccuracy in their data is by being **notified by the concerned data subject**[33]. In particular, data subject must be aware of the processing (see Art. 13 and 14 GDPR) and can access the data used by the controller (see Art. 15 GDPR). On this basis, they can verify the accuracy of their data and, if necessary, invoke their **right to** request **rectification** of their data (see Art. 16 GDPR). In this case, a controller fulfills the obligation to ascertain accuracy by adequately supporting the right to rectification in their processing.

When data is collected directly from the data subjects, it is most reasonable for a controller to assume that the obtained data are accurate (at least at the time of collection). The situation may be different when the data is collected from another source. In this case, it is the controller's obligation to verify the accuracy of the obtained data, at least in respect of fitness for the declared purposes of processing and to any negative consequences that inaccuracies may have for data subjects.

For some data elements, the fact that they were directly collected from the data subjects may not be sufficient for a controller to assume accuracy. This is the case when a potentially inaccurate claim leads to benefits for the data subject. In these cases, the controller may need to conduct a verification of the data up front as an integral part of data collection. This is possible for example by requesting data subjects to provide certification by a trusted authority of the claimed facts.

1.4.2 Related articles and recitals

The GDPR article most closely related to the principle of *accuracy* is **16 *right to rectification***. Its relevance has already been discussed in section 1.4.1.3 above on "How is inaccuracy of data discovered?". Adequate **information** that creates awareness of the processing among data subjects (**Art. 13 and 14 GDPR**) and the **right to access** the data in possession of the controller (**Art. 15**) can be seen to be necessary to enable the right to rectification.

When a controller cannot instantly act on a request for rectification (according to Art. 16 GDPR), but requires adequate time to verify the accuracy of the data in discussion, it may be necessary to **restrict the processing** of the data (see **Art. 18(1)(a)** GDPR). After the verification of the accuracy and effectuated rectification, the controller must **inform the data subject** according to **Art. 12(3)** GDPR. Should the controller find that the data is indeed accurate and does not need rectification, the **data subject** must be **informed** according to **Art. 12(4))** GDPR. If the processing was restricted, the data subject can then **consent to lifting the restriction** even without rectification (see **Art. 18(2)** GDPR). In absence of such consent, the controller can either **delete the data** (see **Art. 5(1)(d)** GDPR) or have its Data Protection Officer **consult the Supervisory Authority** on the issue (see **Art. 39(1)(e)** GDPR).

In the case that the controller disclosed the data to **recipients**, these **must also be made aware** of the inaccuracy (according to **Art. 19** GDPR). In particular, controllers are obliged to notify recipients of the rectifications that were carried out. Considering that

[33] Other mechanisms include for example consistency checks, excessive variance, or a lack of expected correlation.

the verification of accuracy can depend on the purposes of processing (see 1.4.1.2 above), it may be useful and more timely to voluntarily notify recipients already of the request of rectification. Such an extended approach then also covers the case where the data is accurate for the controller, but requires rectification at one of the recipients.

Data subjects also have the **right to request to be informed about such notifications** (see 2nd sentence of **Art. 19** GDPR). This information includes the naming of individual recipients[34].

1.4.3 Related technical and organizational measures

Any organization or technical measure to support the detection of inaccuracies or timely rectification (or deletion) of data supports the principle of *accuracy*. To understand when accuracy is particularly important and stronger measures are required, an analysis is necessary, on how inaccuracies relate to the fitness for purpose and how they can adversely affect data subjects.

Examples of possible measures in support of accuracy include:

- an organizational measure at design time is the analysis of the minimal level of accuracy required to be fit for purpose;
- an organizational measure at design time is the analysis of the possible adverse impacts that inaccurate data can have on data subjects;
- a design-time measure is the analysis of the accuracy of data obtained from sources other than the data subjects themselves;
- another one is the analysis of whether certain data elements require up-front verification (see 1.4.1.3 above);
- another design-time measure is to formulate requirements for the support of the rights to information (Art. 13 or 14 GDPR), the right to access (Art. 15 GDPR), and most importantly, the right to rectification (Art. 16 GDPR;
- the same goes for the implementation of notifications of recipients (Art. 19 GDPR) about inaccuracy and rectification;
- at the time of operating the processing activity, the designation of staff to possible manual intervention necessary for verifying accuracy or effectuating rectification is a possible organizational measure;
- the same goes for preparing the Data Protection Officer to effectively deal with rectification requests.

1.5 Storage limitation

Bud P. Bruegger (ULD)

[34] This is interesting, since in Art. 13(1)(e) and 14(1)(e), it is sufficient to inform about categories of recipients.

The following discusses the principle of *storage limitation* that is defined in Art. 5(1)(e) GDPR.

***Storage limitation* at a glance:**

Storage limitation (even if not implied by its name) considers the degree to which data subjects are **identified** by the data, i.e., how easy data subject can be associated with the data. The degrees of identification foreseen in the GDPR are ***directly identifying data*** that contain *identifiers*, ***pseudonymous data***, and ***anonymous data***. Data shall be collected with the lowest degree of identification possible and pseudonymization and anonymization shall be used to further reduce the identification as soon as possible over time.

1.5.1 Description

In "Understanding data protection: the EU regulation in a nutshell" above, *storage limitation* was motivated by minimizing the power gain of the controller to what is minimally necessary to fulfill the declared, legitimate purposes. In particular, it addressed the minimization of the degree to which the personal data is associated with the data subject. This complements the minimization of the information content and the limitation of access to power. See "Minimization of power to what is necessary to fulfill the declared purposes" for detail.

The GDPR defines the principle as follows:

Definition in Art. 5(1)(e) GDPR:

Personal data shall be **kept in a form which permits identification of data subjects for no longer than is necessary for the purposes** for which the personal data are processed; […]
('*storage limitation*');

Clearly, the main concept of this principle is concerned with the ***identification***, i.e., the association of the personal data with its data subject. The remainder of this section therefore mostly analyzes what identification actually means.

Note that in the definition box above, the omitted part that is represented by [...] has been discussed under the principle of *data minimization* (see section 1.3.2 "Related articles and recitals" in "Data "). It is concerned with the **temporal limitation of the storage** which is arguably one aspect of the general **concept of limitation** expressed for data in the principle of *data minimization*.

From this point of view, the name *storage limitation* is misleading since it implies solely the temporal aspect of data minimization but fails to refer to identification altogether. Calling it *minimization of identification potential* may be clearer.

1.5.1.1 Identification of data subjects

To better understand what is meant by identification, we refer to Art. 4(1) GDPR. The second half-sentence[35] reads as follows:

> [A]n 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;

For better understanding, this sentence is split into the following two parts:

Direct identification by reference to an identifier:

> [A]n 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~~;

Indirect identification by reference to one or more factors specific to the identity of a natural person:

> [A]n 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**;

The **examples for identifiers** are[36]:

- A name,
- an identification number,
- location data,
- an online identifier.

Note particularly *location data* that may not commonly be thought of as an identifier that supports direct identification, even if its highly identifying character is indeed intuitive.

The examples for **factors specific to the identity of a natural person** concern the following aspects:

- Physical,
- physiological,
- genetic,

[35] A part of a sentence that is separated from the rest with semicolons is here referred to as "half-sentence".

[36] Note that Recital 30 GDPR provides in addition examples for "online identifiers": internet protocol addresses, cookie identifiers or other identifiers such as radio frequency identification tags.

- mental,
- economic,
- cultural,
- social.

This distinction of direct and indirect identification now allows diversifying the concept of *a form which permits identification of data subjects*.

1.5.1.2 Types of data distinguished in the GDPR

The GDPR distinguishes three kinds of data with different degrees of association with data subjects:

(i) **directly identifying personal data**[37],

(ii) **pseudonymous personal data**, and

(iii) **anonymous data**.

(i) Directly Identifying Personal Data: The first evidently must contain **identifiers**, since it permits direct identification of data subjects. Most personal data sets contain not only identifiers, though. The other data must then be considered to be all **factors specific to the identity of a natural person** since they all describe different aspect that are linked to the identity of the data subject.

(ii) Pseudonymous Personal Data: Art. 4(5) GDPR defines the related concept of "pseudonymization". Its wording can be adapted as follows to define pseudonymous personal data:

Pseudonymous personal data is personal data that **can no longer be attributed to a specific data subject without the use of additional information**.

This must be interpreted in the following manner:

- Pseudonymous personal data **cannot support direct identification**.
- It therefore **must not contain identifiers**.
- **Additional data**, in this context, is data that permits to **associate factors specific to the identity of a natural person with identifiers.**

(iii) Anonymous Data: Anonymous information are defined in Recital 26 GDPR (fifth sentence). Using *information* and *data* synonymously, its wording can be adapted as follows:

Anonymous data is either • data which does not relate to an identified or identifiable natural person or • personal data rendered anonymous in such a manner that the data subject is not

[37] The term "*directly identifying personal data*" is not used in the GDPR but cloned by the author.

or no longer identifiable.

Note that identifiable here comprises both, direct and indirect identification. Even with additional information, it is not possible to attribute anonymous data to a specific data subject.

Note that according to the Recital 26 (sentence 6), the GDPR does not apply to anonymous data. This is also clear since it does not match the definition of personal data (see Art. 4(1) and Recital 26 GDPR).

Having distinguished these types of data, "kept in a form which permits identification of data subjects for no longer than is necessary for the purposes" can now be understood more precisely, considering also the temporal aspect of the principle.

1.5.1.3 Temporal aspect

Art. 5(1)(e) clearly addresses the temporal aspect by mandating, that a form which permits identification shall be kept **for no longer** than is necessary for the purposes. This temporal aspect is discussed here in a diversified manner. The following two criteria define this diversification:

- **Identification** can be either **direct** or **indirect**.
- **Identification** can be **accessible to everyone** or **to a restricted group of people.**

Based on these distinctions, it is possible to distinguish four different cases. These are shown in Figure 2 represented as "phases". It is possible to transition from one phase to any later phase. This can be done either sequentially, or by omitting intermediate phases. In every phase, the degree of identification of the data with the data subject is reduced. The principle of *storage limitation* states that at any moment, only **the minimal degree of identification that is necessary to fulfill the purposes must be used**.

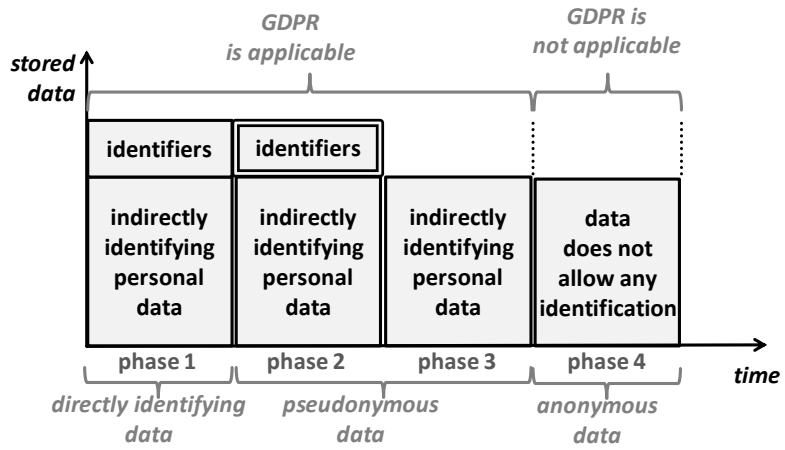

Figure 2: Data with different degrees of association with data subject.

Note that the principle of *storage limitation* is shown in its pure form: Purely the degree of association with the data subject is reduced between consecutive phases. In practice, storage limitation is typically combined with *data minimization*. In a combined scenario, the height of the boxes shown in the figure would also be reduced.

The phases of the figure are described in more detail in the following:

Phase 1 shows the data that contain both, **identifiers** and **factors specific to the identity of a natural person**. For brevity, the latter are called *indirectly identifying personal data*. The identifiers support direct identification. It is **accessible** to **anybody to whom the data is disclosed**.

Phase 2 shows a manner of processing called "**pseudonymization**"[38]. Here, the **identifiers** are still stored, but **kept separately** and **protected** in a manner that enables **access only** under **well-specified conditions,** using **pre-defined procedures**, for achieving **precisely defined purposes**, with access restricted to a **predefined set of authorized persons**[39]. These restrictions are depicted by a double border around the identifiers. The **access to direct identification** is thus **closely controlled** and available only to few designated persons.

Indirect identification that uses additional information is still possible based on the indirectly identifying personal data. It requires additional information, however. The controller implements measures to prevent the availability of such additional information to the persons who access this data during the processing activity. This means that **for the large part of processing** (and an important subset of purposes), and the majority of employees, **identification is no longer possible**.

Phase 3 shows the situation where the **purposes no longer require the possibility of direct identification** of data subjects, not even in exceptional cases. In this case, the *identifiers* that allow direct identification can be deleted altogether. Consequently, with adequate protection measures in place, **the controller itself** (including all staff) **is no longer able to identify the data subjects**. This evidently reduces the degree of identification further compared to phase 2.

Phase 4 shows that only **anonymous data** are used. The figure implies that these are the result of an anonymization of the data of phase 3 (or earlier phases). By definition[40], anonymous data cannot be attributed to a data subject, not even with the use of additional information. This data is therefore no longer personal data and thus not subject to the GDPR (and successful anonymization therefore has the same effect as deletion). **Anonymous data therefore completely eliminates the possibility of identification**.

Some readers may know the concept of "***unlinkability***"[41] that is closely related to that of storage limitation. This becomes clear when considering that direct identification can be seen as an identifier establishing a link to the data subject; and that the use of

[38] See Article 4(5) GDPR.
[39] See Recital 29 GDPR, 2nd sentence.
[40] See Recital 26 GDPR.
[41] German Conference of the Independent Data Protection Authorities of the Federation and the Länder, 17. April 2020, The Standard Data Protection Model, https://www.datenschutz-mv.de/static/DS/Dateien/Datenschutzmodell/SDM-Methode_V20b_EN.pdf (last visited 28/05/2020).

additional information for indirect identification requires to link data records that belong the same person in the two data sets.

1.5.2 Related articles and recitals

As has been shown, several concepts that are defined outside of Art 5 GDPR are relevant for the understanding of the principle of storage limitation. In particular, these are:

- *Direct* and *indirect identification* defined in Art. 4(1) GDPR,
- *pseudonymization* that is defined in Art. 4(5) GDPR, and
- *anonymous data* that is defined in Recital 26 GDPR.

In Art. 11(1), the GDPR states:

> If the purposes for which a controller processes personal data do not or do no longer require the identification of a data subject by the controller, the controller shall not be obliged to maintain, acquire or process additional information in order to identify the data subject for the sole purpose of complying with this Regulation.

This provides guidance about the importance the principle of storage limitation has in comparison to other concepts in the GDPR: Storage limitation has a clear precedence over other obligations of the GDPR in the sense that a controller shall not collect or store identifiers for the sole purpose to comply with these obligations.

In Art 11(2) GDPR[42], this is then stated explicitly for the obligations of the data subject rights of Articles 15 to 20:

> Where, in cases referred to in paragraph 1 of this Article, the controllers are able to demonstrate that they are not in a position to identify the data subject, the controllers shall inform the data subject accordingly, if possible. In such cases, Articles 15 to 20 shall not apply except where the data subjects, for the purpose of exercising their rights under those articles, provide additional information enabling their identification.

Beyond this, the GDPR emphasizes the importance of pseudonymization in in various contexts:

Art. 89(1) emphasizes the **importance of *pseudonymization*** for the case where after fulfilling the initial purposes, data is processed further for "compatible purposes"[43]. In particular, "**archiving** purposes in the public interest, **scientific** or **historical research** purposes or **statistical purposes** shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes"[44]. Art. 89(1) GDPR (2nd sentence) explicitly mandates that for this further processing, "technical and organizational measures need to be in place and lists pseudonymization as sole example for such measures (3rd sentence). It further states (4th sentence): "Where those purposes can be fulfilled by further processing which does not permit or no longer permits the

[42] See also Art. 12(2) GDPR that further discusses this case.
[43] See Art. 5(1)(b) GDPR.
[44] Wording taken from Art. 5(1)(b) GDPR.

identification of data subjects, those purposes shall be fulfilled in that manner." This seems to be a direct application of the principle of storage limitation.

Art. 6(4)(e) further underlines the role of pseudonymization when a controller determines, whether an additional purpose is compatible with the purposes for which the data was collected.

Art. 25(1) lists pseudonymization as sole example for a measure that can be implemented during data protection by design.

Also Art. 32(1)(a) lists pseudonymization together with encryption as a measure in support of security. While this further underlines the importance of pseudonymization and thus storage limitation, it may be questioned however, whether pseudonymization does indeed support one of the common protection goals of IT security, namely *confidentiality*, *integrity*, and *availability*.

1.5.3 Related technical and organizational measures

The following provides some examples of concrete measures that support the principle of storage limitation:

- At the time of designing a given processing activity, an organizational measure is to **verify whether** directly **identifying data needs to be collected at all** to fulfill the stated purposes.
- **Pseudonymization** and **anonymization** of data between processing steps are prime technical measures. They require the verification whether the remaining purposes after the completion of the processing step still require the same degree of identification of data subjects.
- When planning to issue authentication credentials to data subjects, an organizational measure is to verify whether it is sufficient to **issue pseudonymous credentials**. For example, issuing a random one-time-password during data collection may be sufficient to later support the right to withdraw consent.
- Designing a web site such that it **refrains from setting cookies** outside of the areas that require authentication **avoids one way of identifying data subjects** across sessions and can be considered a measure in support of storage limitation (see "Setting Cookies and Writing a Cookie Policy"). Concretely, this may be done via an appropriate configuration of the web application (such as a content management system and its plugins) or web server.
- Operating an Internet-based service in a manner that **permits** users to connect via an **anonymizing overlay network such as TOR**[45] avoids identifying data subjects via their (network) IP address and thus is a measure in support of storage limitation.

[45] See for example, https://en.wikipedia.org/wiki/Tor_(anonymity_network) (last accessed 18/5/2020).

- Equipping a **WiFi-enabled user device** with **MAC address randomization**[46] such as to prevent data subject from broadcasting unique identifiers.

1.6 Integrity and confidentiality

Bud P. Bruegger (ULD)

Acknowledgements: The author thankfully acknowledges the contribution by Frédéric Tronnier (GUF) who wrote an analysis of this principle as input to the here presented description.

The following discusses the principle of *integrity and confidentiality* that is defined in Art. 5(1)(f) GDPR.

***Integrity and confidentiality* at a glance:**

The principle refers to the classical protection goals of IT **security**, namely ***confidentiality***, ***integrity*** and ***availability*** (CIA). ***Resilience*** can be considered an aspect of availability. The main focus is to protect *assets* against *risks* caused by *undesirable events*. In stark contrast to IT security, these **assets and risks** are not those of the controller (an organization), but those **of the data subjects**. From this point of view, it is also clear why *data portability* fits to *availability* within this principle: It protects data subjects from losing an asset (represented by the data) when changing controller (mostly provider).

1.6.1 Description

In "Understanding data protection: the EU regulation in a nutshell" above, *integrity* (along with accuracy) was motivated by the fact that accuracy of data is necessary in order to be fit for the declared purposes. Any processing that fails to be fit for purpose cannot justify a gain of power over a data subject. See "Prohibition of processing that fails to be fit for purpose" for detail. *Confidentiality* on the other hand was motivated by the limitation of access to power. See section 1.6.5.3 "Limitation of the access to power" for detail. *Availability* was motivated by protecting the data subject's assets. See section "1.6.6 Protection of the data subject's assets" for detail.

The GDPR defines the principle as follows:

[46] See for example, https://en.wikipedia.org/wiki/MAC_spoofing#MAC_Address_Randomization_in_WiFi (last accessed 18/5/2020).

Definition in Art. 5(1)(f) GDPR:

Personal data shall be processed in a manner that **ensures appropriate security** of the personal data, including protection against **unauthorized** or **unlawful processing** and against **accidental loss, destruction** or **damage**, using appropriate technical or organizational measures ('*integrity and confidentiality*').

1.6.1.1 The structure of Art. 5(1)(f) and security risks

What is evident from the wording of Art. 5(1)(f) is that the GDPR speaks of **undesired events**, namely:

- unauthorized or unlawful processing, and
- accidental loss, destruction or damage.

Clearly, these events are not part of the processing as planned; ideally, they should be prevented altogether. Since in security, this is never possible with 100% certainty, there is a residual **likelihood that such events do occur**.

It is also evident, that occurrences of such events have **undesirable consequences**.

Readers familiar with IT security will have recognized that this discussion has introduced the elements used in the definition of *risk*. This is made explicit in the following:

Security risk = likelihood of undesirable event * severity of undesirable consequences

This is an "individual" risk and the total risk is then a sum over all applicable individual risks.

Careful readers may have noted that the terminology used here somewhat differs from that common in IT security[47]. In particular, the term "security risk" was used, rather than just "risk" and similarly, "severity of undesirable consequences" was used instead of "damage". The motivation for this choice of terms is explained in the following:

1.6.1.2 Main difference from other risks in the GDPR and from risks in IT security

The GDPR refers to at least two fundamentally different kinds of risk (but without making this distinction explicit). The following therefore introduces two different terms to make this distinction explicit. Namely, they are *security risk* and *data protection risk*.

In the GDPR, ***security risk*** is implicit in both, Articles 5(1)(f) and 32. As apparent from the previous subsection, its definition derives from the existence of **undesirable events** that are **not part of the planned processing operations**.

In contrast to this, the GDPR clearly also considers risks arising from the data processing itself--in absence of any undesirable events--i.e., during undisturbed processing as planned. We call this kind of risk *data protection risk.* It is present, even

[47] See for example https://en.wikipedia.org/wiki/IT_risk#Measuring_IT_risk (last visited 19/05/2020).

if security was perfect and all possible undesirable events could be prevented with 100% certainty.

Therefore, it is important to understand that *security risks* are only a subset of the risks that controllers are obliged to mitigate through the implementation of appropriate technical and organizational measures.

After distinguishing security risk form data protection risks, let us compare the GDPR's security risks with those of IT security. Since its definition provided in the box in the previous subsection has the same structure, can it be concluded that *security risks* in the GDPR are the same as risk in IT security?

This points to the choice of the second term, namely *severity of undesirable consequences* instead of *damage*.

In **IT security**, ***damage*** is a quantification of the undesirable consequences as compared to the **mission and values of the organization** who operates the processing activity. It is often quantified in terms of a **monetary value**, **consistent with** an organization whose mission is to produce **profit**.

In stark contrast to this stand the **severity of undesirable consequences** inherent in the principle of integrity and confidentiality **in the GDPR.** This measure **refers to** the **rights and freedoms of natural persons** as they are laid out in the European Charter of Fundamental Rights. The undesirable effect may thus consist in impeding or negating the free exercise of one's rights and freedoms[48]. Such effects can typically not be measured in terms of monetary values. It is also typically impossible to quantify them, and they can be only expressed on an ordinal scale of measurement (for example that consisting of *low*, *medium* and *high*).

So, the **difference** between **IT security** and **security according to Art. 5(1)(f)** GDPR is the **assessment of the undesirable consequences**, even if the undesirable events may be the same. In many cases, an event that inflicts only minor consequences for the mission of the controller's organization, may inflict severe interference in the rights and freedoms of an affected individual (and vice versa).

1.6.1.3 Protection goals inherent in Art. 5(1)(f)

The GDPR names this principle defined in Art. 5(1)(f) solely ***integrity*** *and* ***confidentiality***. These are two of the three well-known protection goals of IT security. The third is ***availability***. This trinity of protection goals is often referred to simply by the acronym *CIA*.

While the name of the principle given in the GDPR seems to suggest that availability is excluded, both the exact wording of Art. 5(1)(f) and Art. 32 "Security of processing" suggest otherwise. In particular:

- the wording "protection against accidental loss" can clearly be associated with *availability*, and

[48] Felix Bieker, Benjamin Bremert, Identifizierung von Risiken für die Grundrechte von Individuen, in: ZD, 2020, p. 7 et seq. (in German, abstract in English).

- Art. 32(1)(b) mandates controllers to "ensure the ongoing *confidentiality*, *integrity*, *availability* and ***resilience*** of processing systems and services".

Resilience is named here as the fourth protection goal. It is also clearly accepted as an objective of IT security, often treated as an aspect of *availability*.

In conclusion, Art. 5(1)(f) GDPR makes reference to the full spectrum of protection goals known from IT security. They will all be discussed here without restricting the discussion to only the two that are part of the principle's name.

For an in-depth discussion, see ENISA's publications on the topic[49],[50]. The following will only give a brief description of each protection goal.

1.6.1.4 Integrity

Integrity refers to the aspect of Art. 5(1)(f) that requires protection of personal data "against accidental damage", for example due to a transmission error. It thus aims at preventing any kind of event that could "corrupt" the data in any way that renders them unfit for the purposes of processing.

1.6.1.5 Confidentiality

Confidentiality refers to the aspect of Art. 5(1)(f) that requires protection of personal data "against unauthorized or unlawful processing". It is important to note that in the GDPR, *processing* also encompasses *disclosure* of data (see Art 4(2) GDPR). So confidentiality requires to protect personal data from undesired disclosure while at rest, in transit and in use[51]. In addition, it requires that no unauthorized person can interact with the processing operation, for example by inputting decisions that concern a person, by modifying or deleting personal data, or triggering any other operation that is reserved for authorized personnel that work according to precise instructions from the controller.

1.6.1.6 Availability, resilience and portability

Availability refers to the aspect of Art. 5(1)(f) that requires protection of personal data "against accidental loss or destruction", for example due to the failure of a storage component.

Resilience seems to be defined in Art. 32(1)(c) as "the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident". It is thus clearly an aspect of availability and is related to the well-known measure of disaster recovery.

Arguably, another aspect of *availability* is the *portability* of data as it is defined in Art. 20 GDPR. While availability is usually understood at protecting data subjects from losing their data while they are processed by a given controller, data *portability* protects

[49] ENISA, Guidelines for SMEs on the security of personal data processing, January 27, 2017, https://www.enisa.europa.eu/publications/guidelines-for-smes-on-the-security-of-personal-data-processing (last visited 19/05/2020).
[50] ENISA, Handbook on Security of Personal Data Processing, January 29, 2018, https://www.enisa.europa.eu/publications/handbook-on-security-of-personal-data-processing (last visited 19/05/2020).
[51] Art. 32(2) GDPR uses the expression "transmitted, stored or otherwise processed".

data subjects against loss when moving from one controller (e.g., in the role of service provider) to another. Portability entails that data subjects can obtain their data in a machine readable format (see Art. 20(1) GDPR) and, if feasible, to have them transmitted directly from one controller to another (see Art. 20(2) GDPR).

1.6.2 Related articles and recitals

While Art. 5(1)(f) GDPR states abstractly, that "appropriate technical or organizational measures" shall be used to implement the above mentioned security protection goals, **Art. 32 GDPR provides further detail.**

Art 32(1), states that when deciding on appropriate measures, controllers shall take into account "the **state of the art** and the **costs of implementation**", as well as "the nature, scope, **context** and purposes **of processing**". In particular, the context of processing is of relevance here, since it can be argued that the current **threat landscape** is an aspect thereof. As expected, the controller shall also take into account "**the risks for the** rights and freedoms of **natural persons**".

So the required level of protection clearly depends on the severity of the possible undesirable consequences that data subjects are exposed and a threat model that estimates the likelihood of undesirable events. Security is thus only a means, not an objective in itself. The level of security is sufficient, when the risks for data subjects are mitigated down to an acceptable level. The selection of measures depends both on what the market has to offer and how cost-effective these measures are.

Art. 32(1)(d) GDPR states the well-accepted concept that **security is a process**, not an objective that is reached once. In particular, the GDPR requires "a process for regularly testing, assessing and evaluating the effectiveness of technical and organizational measures for ensuring the security of the processing".

Art. 32(2) GDPR provides marginal **additional detail about what the protection goals entail**, enumerating "accidental or unlawful destruction, loss, alteration, unauthorized disclosure of, or access to personal data transmitted, stored or otherwise processed".

Art 32(3) GDPR suggests that "[a]dherence to an **approved code of conduct** or an **approved certification mechanism** may be used as an element by which **to demonstrate compliance**" with the principle of *integrity and confidentiality*.

Art. 32(4) GDPR clarifies that an important element of security is to **ensure that employees act only on instruction and as instructed by the controller**. This is necessary to establish clear responsibility and accountability. It is also necessary to ensure the requirement of Art. 5(1)(f) to "protection against unauthorized or unlawful processing".

From **Art. 25** GDPR, it follows that all requirements posed by the GDPR, including security, have to be considered **throughout the life cycle** on the processing activity. The GDPR thus also requires **security by design and default**. Security thus has to be considered also at the start of the life cycle, for example through according requirements used for a tender; and at the end of the life-cycle, for example when migrating operations to a new processing system and dismantling the old one.

Art. 30(1)(g) GDPR requires to specifically list the technical and organizational ***security*** **measures** in the ***records of processing*** that are targeted at supervisory authorities.

1.6.3 Related technical and organizational measures

The following examples of technical and organizational measures shall further concretize the concept of security in the GDPR.

1.6.3.1 Measures in support of integrity

- One of the classical technical measures to support integrity is **transactional processing**. It is best known from data base management systems, but is also possible in other settings[52]. Transactions are important when an operation that takes the system from one consistent state to another is composed of multiple processing steps (i.e., it is not "atomic"). A transaction then makes sure that either all these steps or none are applied, even if the system should crash in the middle. It thus guarantees that the system always remains in a consistent state.
- Inconsistencies can arise due to transmission errors in noisy communication lines. The technical measure of **forward error correction**[53] that is built into modern communication protocols thus supports the integrity of data during transfer.
- A common technical measure to detect undesirable changes in data sets uses **checksums** (aka. hash or digest). In particular, a checksum of a set of data is computed when it is known to be in a consistent state. At later points in time, the checksum of the data set can be newly computed and compared to the initial one in order to detect changes and corruption.
- Integrity is an important issue in the distribution of software—in particular if software is downloaded automatically over a network. Automatic updates of operating systems are a prime example. To support integrity of the software, technical measures such as **authentication of sources** on the network and **digital signature of software** are often used. Digital signature is often also used for data files.

1.6.3.2 Measures in support of confidentiality

- A design-time organizational measure in support of confidentiality is an **analysis** of the **consequences** undesired disclosures to various parties can have **for data subjects**. This is comparable to IT security where the critical assets of the organization that need particular protection are identified.

[52] For examples of transactional processing outside DBMS, see for example https://en.wikipedia.org/wiki/Tuxedo_(software) and https://docs.oracle.com/cd/E13222_01/wls/docs81/jta/trxejb.html (both last visited 20/05/2020).
[53] See for example, https://en.wikipedia.org/wiki/Forward_error_correction (last visited 20/05/2020).

- Confidentiality mandates that the controller implements measures to protect against unauthorized processing (see Art. 5(1)(f) GDPR). As emphasized in Art. 29 and 32(4) GDPR, this includes that employees only process personal data on instruction and as instructed by the controller. There are a multitude of organizational measures that support this requirement, including the following:
    - **Vetting** of new employees to ensure the necessary skills to execute the controllers instructions;
    - Legal means that "ensures that **persons authorized** to process the personal data have **committed themselves to confidentiality** or are under an appropriate statutory obligation of confidentiality". (The wording is taken from Art. 28(3)(b) that refers to persons working for processors, but is equally applicable to persons working for the controller).
    - In this sense, also the **contracts with possible processors** (see Art. 28(3) GDPR) that pass on confidentiality requirements must be considered as measures.
    - **Training** of employees on how to execute instructions;
    - **Internal contact points** for employees who want to clarify how to execute instructions;
    - Manuals that describe the instructions (**process manuals**);
    - **Supervision** and **quality control**.
- What holds for instructions to human resources also holds for **instructions for technical resources**, i.e., software. Implementing measures to protect against unauthorized processing means that controllers have to ascertain that the software actually corresponds to their instructions. There are several measures for this purpose, including the following:
    - **Specification of precise requirements** as input for tenders or for custom development of software;
    - Formal **acceptance testing** by the controller;
    - **Analysis of new versions** of software to ascertain that changed functionality still corresponds to the controller's instructions and that no additional functionality has creeped (**function creep**) that corresponds to processing that has not been authorized by the controller.
- An important technical measure is **access control** that enforces that only authorized personnel can access the systems and data for authorized purposes. Access control can entail a multitude of measures, including the following:
    - Issuance of **authentication credentials**.
    - Configuration of **access rights** and conditions.

- Management of the **life cycle of credentials** and **access rights**, including expiry and renewal, revocation (e.g., when employees leave), granting and revoking temporary access rights (e.g., when employees are sick).
- Regular **audits** of the overall effectiveness of the access control system.

- There is a wealth of technical measures aimed at preventing unauthorized (internal or external) persons to access data. Usually, they are referred to as **protection of data at rest, in transit, and in use.** The former two aspects typically require **encryption**.
- There is a wealth of measures to prevent unauthorized persons to gain access to systems and networks. Examples include the following:
  - **Hardening** of operating systems;
  - Timely application of **security-critical patches and updates**;
  - **Firewalls**;
  - Installation of **anti-malware** software;
  - Operation of **intrusion detection systems**;
- When **developing software**, many measures are available to prevent unauthorized access to software and systems, including input sanitation, prevention measures for known kinds of attacks such as cross site scripting, methods that prevent buffer overflows, memory randomization, etc.
- Some measures are unable to directly prevent unauthorized processing, but acts as a **deterrents** by helping to **detect** such action, clearly **determine responsibility**, and enable to **hold persons** who acted without authorization **accountable**. Such measures typically involve **logging** or the creation of **audit trails**.
- An important measure associated with the **end of life** of storage components include the complete and **secure destruction** of all data before **disposal**.

1.6.3.3 Measures in support of availability and resilience

- A design-time organizational measure is the analysis of the impact of accidental loss on data subjects. This aims at identifying the assets that have to be protected by availability measures.
- Another design-time measure pertains data portability and investigates the availability of suitable standardized machine-readable formats that are available and possibilities to automatically transfer the data to another controller (see Art. 20(2) GDPR).
- A very common kind of measure in support of availability is the **redundancy of storage**. Well-known examples include the following:
  - RAID storage;
  - Backups;

- Remote storage in support of disaster recovery.

- Beyond data storage, **redundancy** may also be important **in processing systems**. According measures include the following:
  - Master/Slave configurations with fail-over;
  - Server farms and cloud configurations;
  - Virtualization-based process migration strategies.

1.7 Accountability

Bud P. Bruegger (ULD)

Acknowledgements: The author thankfully acknowledges the contribution by Johann Čas and Walter Peissl (both OEAW) who wrote an analysis of this principle as input to the here presented description.

The following discusses the principle of *accountability* that is defined in Art. 5(2) GDPR.

***Accountability* at a glance:**

Accountability consists of two requirements for controllers:

- **Compliance** with the principles of the GDPR;
- **Demonstration of compliance.**

Compliance is achieved by implementing ***technical and organizational measures*** that are adequate compared to the risks to the rights and freedoms of data subjects, correspond to the state of the art of technology, and are cost-effective. Every description of the principles has provided examples of such technical and organizational measures. For a systematic application of these measures, controllers can create ***data protection policies***. ***Approved codes of conduct***, where available, are similar but are pre-approved and usually address an entire sector. Compliance is not a state that is reached once, but **a continuous process** that spans the whole life cycle of a processing activity.

Demonstration of compliance is predominantly achieved by **documentation** (see the section "Documentation of Processing" in "Main Tools and Actions"). Documentation should be continuous like the process of compliance. Every implemented measure, including data-protection-relevant considerations and decisions, should be documented. The GDPR requires two formal documents as part of demonstrating compliance towards *supervisory authorities*: the ***register of processing*** (see "Documentation of Processing" for detail) and, where the risks are likely to be high, a ***data protection impact assessment*** (see the section with the same name in "Main Tools and Actions" in Part II for detail). *Certification* can support the demonstration of compliance.

1.7.1 Description

In "Understanding data protection: the EU regulation in a nutshell" above, full *accountability* of controllers was stated as the first of several measures taken by the GDPR to limit the power gained by the controller through processing and balance it with the power of data subjects. See section 1.6.1 "Controllers are fully accountable" for detail.

The GDPR defines the principle as follows:

Definition in Art. 5(2) GDPR:

The **controller** shall be **responsible for**, and be able to **demonstrate compliance** with, paragraph 1 ('*accountability'*).

Paragraph 1 here refers to the principles that were discussed in the six previous sections, namely

- Lawfulness, fairness and transparency;
- Purpose limitation;
- Data minimization;
- Accuracy;
- Storage limitation; and
- Integrity and confidentiality.

To rephrase Art. 5(2), a **controller** is fully **responsible for** two things:

- **Compliance** with these six principles,
- **Demonstrating compliance**.

Accountability is thus not a new principle that controllers need to comply with, but it instructs controllers **how the six principles must be applied**.

Note that having to be able to demonstrate compliance is a big step beyond just having to comply. In particular, it puts the "burden of proof" on the controller; a controller who is unable or unwilling to demonstrate compliance, is in violation of the GDPR.

1.7.1.1 What does it mean to comply?

While Art. 5(2) only speaks of compliance with the six principles, in fact it must be extended to the **whole GDPR**. This is motivated by the fact that all the other articles are intended to provide detail to the principles or describe in more detail how they have to be implemented.

There is one way stated all over the GDPR about how compliance has to be achieved; namely, through the implementation of **technical or organizational measures**. In Art.

24 which describes the obligations of a controller, the first paragraph explicitly states that this is how controllers comply (and demonstrate compliance) with the GDPR; Art. 25(1) states that data protection by design boils down to implementing such measures throughout the life cycle of the processing activity; Art 25(2) similarly emphasizes the use of such measures for data protection by default; Art. 28(1) states that also processors much implement such measures; Art. 32 states that also compliance with security requirements is achieved through the implementation of such measures; and Art. 89(1) states that the safeguards necessary for the "processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes" ensure that such measures are in place.

Since technical and organizational measures are so central to achieve compliance, the discussion of every of the six principles above ended with examples of such measures.

Compliance with data protection requirements can be seen as a process. Following the concept of *data protection by design* (see Art. 25(1) GDPR), in every life-cycle phase of the processing activity, the risks to the rights and freedoms of natural persons are assessed and appropriate mitigation measures implemented. The GDPR uses a very broad definition of the term *technical and organizational measures*. It basically includes everything a controller does to comply with the GDPR. Therefore, even the above mentioned assessment step can be considered to be a measure in itself.

1.7.1.2 What does it mean to demonstrate compliance?

Considering that compliance is achieved through the implementation of appropriate measures, it is not surprising that the **demonstration of compliance documents such measures**.

This is evident for example from Art. 30(1)(g) that mandates to list the measures pertinent to security in the ***records of processing***. It is also central in Art. 35 on the **Data Protection Impact Assessment** which is arguably the main tool foreseen by the GDPR for demonstrating compliance. In particular, Art. 35(7)(d) asks controllers to declare the measures they implemented to ensure the protection of personal data and to demonstrate compliance with the GDPR.

A **more detailed discussion** of **Documentation of Processing** in general, and Data Protection Impact Assessments in particular, can be found in "Main Tools and Actions" below. Both these sections further emphasize the importance of technical and organizational measures.

1.7.1.3 Economy of scale for compliance and its demonstration

As argued above, compliance is achieved by implementing technical and organizational measures. It is evident from the discussion above that compliance may require a significant number of such measures. This can render it more difficult to assess the actual protection offered by these measures and whether this protection is applied uniformly and consistently.

To mitigate this difficulty, the GDPR offers some kinds of "abstraction mechanisms" that permit to consider a set of related measures as a single unit. In particular, the GDPR

foresees two such mechanisms in its Art. 24 that describes the "Responsibility of the controller":

- ***Data protection policies*** (see Art. 24(2) GDPR), and
- ***approved codes of conduct*** (see Art. 24(3) and 40).

A ***data protection policy*** is a mechanism to render the application of measures systematic. This guarantees a uniform and consistent set of measures in similar situations. For example, instead of having to assess which security measures are appropriate for each of many highly similar servers, a single policy can be written once and applied to all servers. Evidently, particularly in complex and extensive processing operations, this brings a potentially very significant economy of scale which can even span multiple independent processing activities of the same controller.

The mechanism of ***approved codes of conduct*** extends this economy of scale beyond a single controller to an entire processing sector. These codes of conduct are prepared by **associations** and other bodies **representing categories of controllers or processors** (see Art. 40(2) GDPR). Where a code of conduct does not relate to processing activities in several Member States, the competent *supervisory authority* can **approve** it (see Art. 40(5) GDPR) and subsequently register and publish it (see Art. 40(6) GDPR). Where a draft code of conduct relates to processing activities in several Member States, a similar process is used that involves the European Data Protection Board (see Art. 40(7) GDPR). Codes of conduct evidently provide also an economy of scale to supervisory authorities who have to monitor compliance with the GDPR.

Both, *approved codes of conduct* and *certification* (according to Art. 42 GDPR) can help controllers in the demonstration of compliance (see Art. 24(3) GDPR).

1.7.2 Related articles and recitals

Accountability is about compliance and demonstration of compliance. It directly references the six principles of data protection defined in Art. 5(1) but indirectly extends to the entire GDPR.

Art. 24 GDPR provides details on how a controller has to achieve compliance and demonstrate it. Art. 25(1) on data protection by design illustrates how compliance (and consequently also its demonstration) must be considered to be a continuous process that spans all life cycles of a processing activity. The *codes of conduct* and the *certification* that can help with compliance and its certification are described in Art. 40 and 42 GDPR, respectively.

Articles particularly pertinent to the demonstration of compliance are 30 *records of processing* and 35 *data protection impact assessment*.

1.7.3 Related technical and organizational measures

Measures pertinent to *accountability* address how to go about compliance and its demonstration, rather than what needs to be done to comply.

The following "meta" measures address ways of **achieving compliance**:

- ***Data protection by design and default*** (see Art. 25 GDPR),
- The ***Data protection impact assessment*** (see Art. 35 GDPR) in its function as a continuous process that guides the controller in assessing the risks and to identify appropriate technical and organizational measures for their mitigation.
- The creation and application of ***data protection policies*** (see Art. 24(2) GDPR).
- The adherence to ***approved codes of conduct*** (see Art. 24(3) GDPR).
- The adherence to ***approved certification mechanisms*** (see Art. 24(3) GDPR).

The following "meta" measures address ways of **documenting compliance**:

- The ***data protection impact assessment*** (see Art. 35 GDPR) in its function as a report. Where the risk is not likely to be high and such an impact assessment is therefore not required, the documentation of how this risk estimate was established should be documented (see section on "Data Protection Impact Assessment*"* in "Main Tools and Actions" in Part II of these Guidelines for detail).
- The ***records of processing*** (see Art. 30 GDPR).