Comments on the document
Guidelines 4/2019 on Article 25 Data Protection by Design and by Default

Source: DP Security Consulting SAS. France

Date: December 9, 2019

"Privacy by design and by default" is a well known terminology. Article 25 is using the wording "Data protection by design and by default". This is not the same: protecting (personal) data is certainly necessary but is insufficient to guarantee the privacy of an individual. Unfortunately, it can be observed that the GDPR is using nowhere the word "privacy".

1. Privacy principles

The current writing of Article 25 is the following:

> (...) the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures (...) which are designed to implement data-protection principles,

The current interpretation of the text is too close to the text of Article 25, since only two examples ('pseudonymisation' and 'data minimization') have been expanded.

The current writing of Article 25 should be understood as :

> (...) the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures (...) which are designed to implement **privacy principles**,

ISO/IEC 29100 Privacy Framework sets forth the following eleven privacy principles which apply to four types of actors who can be involved in the processing of PII: PII principals, PII controllers, PII processors **and third parties**:

1. consent and choice,
2. purpose legitimacy and specification,
3. collection limitation,
4. data minimization,
5. use,
6. retention and disclosure limitation,
7. accuracy and quality,
8. openness,
9. transparency and notice,
10. individual participation and access, accountability,
11. information security and privacy compliance.

Adhering to the "consent and choice principle" means in particular:

- informing individuals, before obtaining consent, about their rights under the individual participation and access principle;
- presenting to the individual the choice whether or not to allow the processing of their personal data;
- obtaining the opt-in consent of the individual for collecting personal data.

Adhering to the "collection limitation" principle means limiting the collection of personal data to that which is within the bounds of applicable law and strictly necessary for the specified purpose(s).

Whereas "collection limitation" refers to limited data being collected in relation to the specified purpose, "data minimization" strictly minimizes the processing of personal data.

Adhering to the use, retention and disclosure limitation principle means:

- limiting the use, retention and disclosure of personal data to that which is necessary in order to fulfil specific, explicit and legitimate purposes;
- limiting the use of personal data to the purposes specified by the data controller prior to collection;
- retaining personal data only as long as necessary to fulfil the stated purposes.

Adhering to the openness, transparency and notice principle means:

- providing individuals with clear and easily accessible information about the data controller's policies, procedures and practices with respect to the processing of personal data;
- including in notices the fact that personal data is being processed, the purpose for which this is done, the types of privacy stakeholders to whom the personal data might be disclosed, and the identity of the data controller including information on how to contact the data controller.

Adhering to the information security principle means in particular:

- protecting personal data with appropriate controls at the operational, functional and strategic level to ensure the integrity, confidentiality and availability of the personal data, and protect it against risks such as unauthorized access, destruction, use, modification, disclosure or loss throughout the whole of its life cycle.

Adhering to the privacy compliance principle means:

- verifying and demonstrating that the processing meets data protection and privacy safeguarding requirements by periodically conducting audits using internal auditors or trusted third-party auditors;
- having appropriate internal controls and independent supervision mechanisms in place that assure compliance with relevant privacy law and with their security, data protection and privacy policies and procedures; and

- developing and maintaining privacy risk assessments in order to evaluate whether program and service delivery initiatives involving personal data processing comply with data protection and privacy requirements.

As explained in this document:

> "In accordance with the principle of data minimisation, by default, only the amount of personal data that is necessary for the processing shall be processed".

The data minimization principle shall be used only once the collection limitation principle has been applied, otherwise a data controller would be allowed to collect all the personal data of an individual and would claim that it only uses a part of it (without saying which part of it).

The collection limitation principle shall be used only once the consent and choice principle has been applied.

The consent and choice principle shall be applied once the openness, transparency and notice principle has been applied.

Referencing only the data minimization principle is insufficient to protect the rights of the individuals. In addition to the data minimization principle, the following privacy principles need also to be taken into consideration: openness, transparency and notice; consent and choice; and collection limitation.

2. The major difference between "shall not" and shall not be able"

Basically, Article 25 states that a data controller **shall not** process more personal data than the minimum necessary.

However, individuals would rather prefer that a data controller **shall not be able** to process more personal data than the minimum necessary. In order to reach such a goal, the cooperation of the individuals is needed.

The GDPR defines 'personal data' as:

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

When using the Internet, many or most accesses are performed using a web browser. This has two consequences:

1) In order to send back a response, the IP address of the workstation supporting web browser must be known. Using specific tools like a VPN (Virtual Private Network) or Tor (The Onion Router), it is possible to hide this IP address towards a data controller but the vast majority of users is not aware of these tools.

Tor is a free web browser for enabling online anonymity that *hides the true IP address*. Tor directs Internet traffic through a free, worldwide, volunteer network consisting of more than four thousand relays to **conceal a user's location** or usage from anyone conducting network surveillance or traffic analysis.

2) In order to adapt the information to be displayed to the size of the display, tools are used in the background. Whereas it would be sufficient to indicate the size of the display, much more information is made available to the hardware/software supporting the data controller using the *User-Agent request-header field* which is defined in RFC 7231.

The User-Agent request-header field contains a characteristic string that allows the network protocol peers to identify the application type, operating system, software vendor or software version of the requesting software user agent. RFC 7231 states: a user agent SHOULD send a User-Agent field in each request unless specifically configured not to do so.

In order to know the information released by that field, the following URL may be used: https://www.whoishostingthis.com/tools/user-agent/

As an example, the information released when using Chrome on my PC is:

Mozilla/5.0 (Windows NT 6.1; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/78.0.3904.108 Safari/537.36

This information, used in combination with the IP address, can allow data controllers to establish links between their accounts.

When using a Tor web browser, the information released is:

Mozilla/5.0 (Windows NT 10.0; rv:68.0) Gecko/20100101 Firefox/68.0

When using Mozilla, the browser information is:

JavaScript Enabled:	Yes
Cookies Enabled:	Yes
Device Pixel Ratio:	1
Screen Resolution:	1600 px x 900 px
Browser Window Size:	1584 px x 756 px
Local Time:	3:40 pm
Time Zone:	-1 hours

while, when using Tor, the browser information is the same except:

Local Time:	2:42 pm
Time Zone:	0 hours

The time zone is set to UTC, so the location that cannot be inferred from the time zone.

When Tor is being used, data controllers get less personal data, but still much more than what would be absolutely necessary. The vast majority of users is not aware of the dangers of using conventional web browsers, including the non-deletion of *cookies* when closing these web browsers.

Individuals should be educated to use tools that do not present to a data controller more personal data than what is necessary and in this way **data controllers will not be able to use more personal data than necessary**.

3. Linkeage of individual accounts between several data controllers

Article 25 is using the wording "the controller" instead of "the controllers" (plural).

However, the processing of personal data usually involves several different controllers in the real life. The relationships between these data controllers need to be investigated. One data controller does not necessary have access to the same personal data about one individual and thus the sharing (or the non sharing) of that personal data between these data controllers is of primary importance.

The current text does not consider several data controllers, so it is missing to identify the threats implied by one data controller cooperating with another data controller or by the cooperation between several data controllers. Aggregating personal data from several data controllers is a major concern which also allows to link accounts from the same individual managed by these different data controllers.

As previously indicated, the use of the real IP address combined with the use of *User-Agent request-header field* may allow data controllers to link the accounts of individuals placed on different data controllers.

The current text does not take this aspect into consideration.

Two specific additional privacy principles should be supported:

- **user-unlinkeability** between different Service Providers, where a Service Provider in collusion with another Service Provider using the data exchanged in the protocol is unable to know whether accesses are performed by the same user or not.
- **session-unlinkeability** towards a single Service Provider, where the Service Provider, once a session has been closed, is unable to know whether another session is opened for the same user or not.

4. Big Brother is watching you

For practical reasons, an intermediary may provide to an individual a "token" which contains some attributes from the individual and which restricts the use of these attributes to one or more services. Hence comes another concern: if the intermediary can identify the service(s) to which the token is restricted, then it is able to **act as Big Brother**: the intermediary will be able to know **where** the token may/will be used and then will be able to trace all the accesses of the individuals that have requested an access token.

In some cases, a statement indicates that intermediaries ***will not log*** that information. Can an individual trust such a statement ? Obviously not. The design of the system should be such that intermediaries ***shall not be able to log*** that information.

These intermediaries are usually identified as "**Attribute Providers**".

When attribute Providers are used, three additional specific privacy principles should be supported:

- **target-untraceability**, when it is infeasible for an Attribute Provider to trace the location(s) where the attributes that it issues may be successfully used.
- **action-untraceability**, when it is infeasible for an Attribute Provider to trace which actions will be performed by a user using the attributes that it issues.
- **attribute-auditability**: when it is possible for an individual to verify which attributes have effectively been issued by an Attribute Provider before an access is being performed to a Service Provider.

5. Data minimization

Data minimization is addressed on line 69 on page 19: "Controllers must first of all determine whether they even need to process personal data for their relevant purposes".

Such a sentence may easily be misinterpreted.

The following sentence on line 75 page 32 highlights even more the confusion: "It is vital that the controller knows exactly what personal data the company processes and why".

The personal data that a company processes will be dependent upon the kind of action that an individual is willing to perform. It can only be determined at the time of the action, i.e. once the action that the individual is willing to perform is being known.

Most service providers are evaluating the personal data that will be necessary to perform *any kind of action* supported by the service instead of evaluating which "minimum personal data" is needed to perform a *given action*.

In other words, the necessary attributes are requested at the time of the log-in to the Service Provider instead of once the requested action has been identified.

Guidance should be given to prevent such a misinterpretation.

6. Static, dynamic and computed attributes

As described, readers may think that data minimization principle can simply be met by processing an appropriate subset of *static* personal data when a user is performing an access. However, this would be a limited point view.

When an access is being performed to a service, the key question is what kind of data, including personal data, is being processed.

Personal data may include *dynamic* personal data, like where from the individual is performing an access or/and the (local) time at which the individual is performing an access.

This personal data may include *computed* attributes instead of *static* personal information. As an example, it may be sufficient to demonstrate to a service that an individual is over 18. In such a case, an intermediary can derive that computed attribute from the date of birth of the individual. This highlights the fact that, in this case, the attribute that is being used is not a subset of the personal data of the individual.

The current text does not take these aspects into consideration.

7. Anonymization

Anonymization is addressed on line 70 on page 19: "If the purpose of the processing does not require the final set of data to refer to an identified or identifiable individual (such as in statistics), but the initial processing does (e.g. before data aggregation), then the controller **shall** anonymize personal data as soon as identification is no longer needed".

Such a recommendation could be understood by an individual as the following: the controller **will** anonymize personal data as soon as identification is no longer needed. Can an individual trust such a statement ? Obviously not. The design of the system should be such that data controllers ***shall not need to*** anonymize personal data. The initial processing should instead require the use of pseudonyms so that individuals cannot be identified at all.

8. About certification

Certifying a company should not be confused with the certification a product. Should a company be "Article 25 certified" this would not mean that any of its products or services meets the requirements implied by Article 25.

Line 81 illustrates the confusion: "where a controller has been awarded a certification, ..."

What needs to be demonstrated is not a good understanding of Article 25 by a *single* data controller but the fact that *a whole system involving several data controllers* has been designed taking into consideration a **methodology** where a trade off between privacy requirements and other constraints like the legislation, security measures, costs or ease of use has been done. Such a methodology should be based on a Deming wheel approach using several Plan-Do-Check-Act cycles.

The current text is missing to indicate that **a methodology shall be followed**.

The second bullet of line 86 states: "A processing operation may be certified for DPbDD."
This would need to be interpreted in the following way:

> A processing operation may be certified for DPbDD, *if the details of the methodology that has been used to build the overall system are disclosed to the auditors so that they can be aware of the trade off that has been performed between privacy requirements and other constraints (including security).*