



How to define anonymization and pseudonymisation of Health Data

Rocco Panetta

Secretary General of Italian Compliance Forum

Barcelona, April 6, 2017

Anonymous and pseudonymous data: scope

Under EU data protection law, there are three broad categories of data:

Personal data

Personal data is defined as any information relating to an identified or identifiable natural person. An identifiable person is one who can be identified both directly and indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.

Anonymous/anonymized data

Anonymous data is any information from which the person to whom the data relates cannot be identified, whether by the company processing the data or by any other person/legal entity. The threshold for anonymization under EU data protection law is very high and absolute, meaning that the company's intent is not relevant. Data can only be considered anonymous if re-identification is impossible for any party and by all means likely reasonably to be used for this purpose.

Anonymized data is no longer considered personal data and is thus outside the scope of EU data protection law.

Pseudonymous data

Pseudonymisation is a form of de-identification, in which information remains personal data. The EU General Data Protection Regulation ("GDPR") defines pseudonymisation as *"the processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information."* The key legal distinction between anonymised and pseudonymised data is its categorization as personal data.

Pseudonymous data still allows some form of re-identification (even indirect and remote) and so **falls within the scope of application of EU data protection law.**

Anonymous data in WP29 interpretation

An opinion issued by the Article 29 Working Party, in 2014, in relation to anonymization exemplifies some relevant techniques, such as:

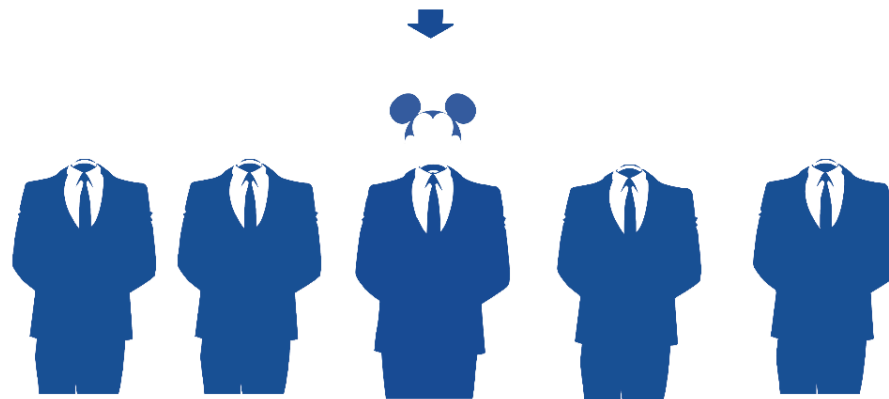
- ✓ **Noise addition.** This means that an imprecision is added to the original data. For example, a doctor may measure an individual weight correctly, but after noise addition it shows a weight bandwidth of +/- 10lb.
- ✓ **Substitution.** Information values of the original data are replaced with other parameters. For example, instead of indicating a patient's height comprised between 180 - 185 cm, this value is substituted by the word "blue." If a patient's height ranges between 160 - 165 cm, it is registered as "yellow." Substitution is often combined with noise addition.
- ✓ **Aggregation/Generalisation.** In order not to be singled out, an individual is grouped with several other individuals that share some or all personal data, like referring to the inhabitants of Northern California instead of San Francisco, sharing the same characteristics. The process impedes re-identification by removing some information but letting the data be intact for future use. *K-anonymity* is a form of aggregation. One method of k-anonymity is data suppression. You can suppress data by replacing a value with a place holder. For example, instead of "age 29," the value is "X." Another method is by generalizing the data. Instead of "age 29," the input is "between 25 and 35".
- ✓ **Differential privacy.** This comes into play when a company gives a third party access to an anonymized data set: a copy of the original data remains with the company, and the third-party recipient only receives an anonymous data set. Additional techniques such as noise addition are applied prior to the data set transfer. Differential privacy is applied when an authorized third party is requesting data.

Pseudonymous data

Pseudonymisation is the separation of data from direct identifiers so that linkage to an identity is not possible without additional information that is held separately. Pseudonymisation, therefore, may significantly reduce the risks associated with data processing, while also maintaining the data's utility. For this reason, the GDPR creates incentives for controllers to pseudonymise the data that they collect.

In sum, it is a privacy-enhancing technique where directly identifying data is held separately and securely from processed data to ensure non-attribution.

Although Recital 28 recognizes that pseudonymisation *"can reduce risks to the data subjects,"* it is not alone a sufficient technique to exempt data from the scope of the Regulation.



*I can't identify you... But I can single you out!
Does the difference really matter?*

Where pseudonymisation is applied

The WP29 provides the following selected examples of pseudonymisation techniques:

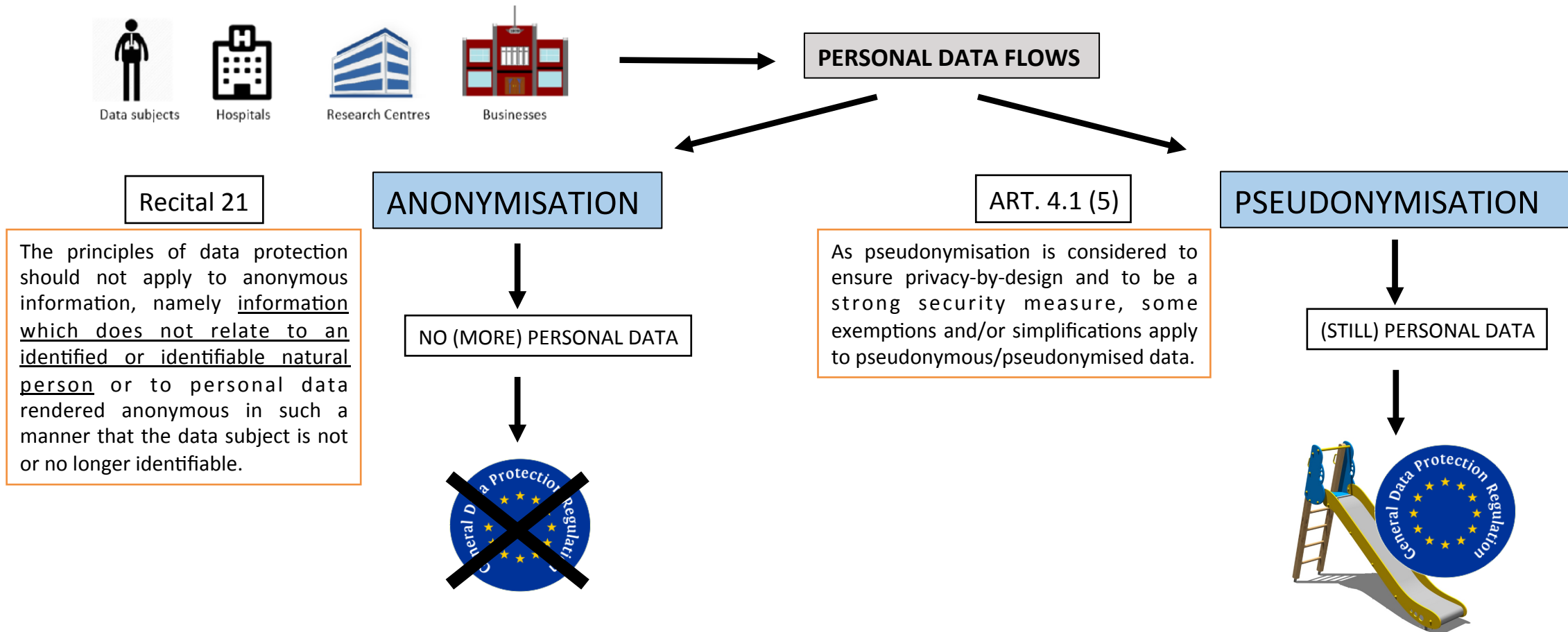
- *hash functions*: hashes are a popular tool because they can be computed quickly. They are used to map data of any size to codes of a fixed size. For example, the names Rocco Panetta and Edwin Morley-Fletcher can be hashed to “01” and “02”. However long the name, the hash value will always be two digits;
- *tokenization*: which is a process by which certain data components are substituted with a non-sensitive equivalent. That equivalent is called the token. The token has no exploitable value, but it serves as an identifier. It is a reference that traces back to the original data.

When the data are pseudonymised, the GDPR grants a wider room for manoeuvre for relevant data processing:

- ✓ **Privacy-by-design**: pseudonymisation shall be considered as an appropriate technical and organisational measures to ensure compliance with data-protection principles;
- ✓ **Accountability**: pseudonymisation strongly helps data controllers to demonstrate that adequate safeguards to protect the data have been put in place;
- ✓ **Scientific research**: processing for scientific research purposes shall be subject to appropriate safeguards aimed at ensuring that technical and organisational measures are in place, such as pseudonymisation, in light of the principle of data minimisation.



Anonymous vs. pseudonymous data: different legal regimes

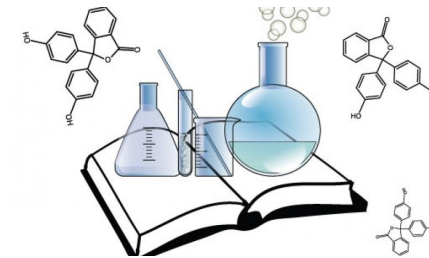


The purpose of health and medical research (i)

Principles

When the aim pursued by data controllers relates to scientific research, some further exemptions may apply, such as:

Scope of scientific research: It is often not possible to fully identify the scope of data processing for scientific research purposes at the time of data collection. Therefore, data subjects should be allowed to give their consent to certain areas of scientific research when in keeping with applicable recognised ethical standards. Data subjects should have the opportunity to give their consent only to certain areas of research or parts of research projects to the extent allowed by the intended purpose.



Broad interpretation: the processing of personal data for scientific research purposes should be interpreted in a broad manner including, for example, technological development and demonstration, fundamental research, applied research and privately funded research, also in the light of the EU objective – under Article 179(1) TFEU – of achieving a European Research Area. Scientific research purposes should also include studies conducted in the public interest in the area of public health.

The purpose of scientific research (ii) Relevant exemptions



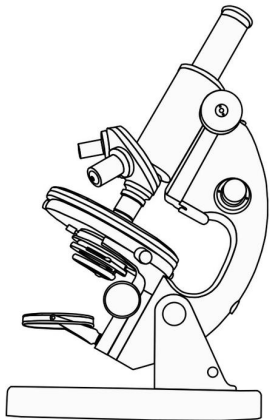
Information obligation: when personal data have not been obtained from the data subject, no information notice must be given to the data subjects when fulfilment of this requirement proves impossible or would involve a disproportionate effort, as in case of scientific research, so long as appropriate security measures have been implemented.

Sensitive data: prohibition of processing genetic and health data, or data concerning sex life, shall not apply when such processing is necessary for scientific purpose and appropriate security measures are in place.



The purpose of scientific research (iii) – Additional provisions

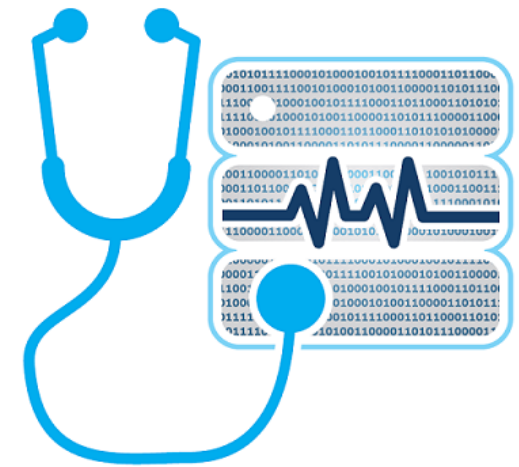
Statistic results: the results obtained through any operation of collection and the processing of personal data for statistical purpose may further be used for scientific research objectives.



Compatible purpose: personal data must be collected for specified, explicit and legitimate purposes and not be further processed in a manner that is incompatible with those purposes. Nonetheless, further processing for scientific research must not be considered to be incompatible with the initial purposes, where adequate security measures are in place for data subjects' rights.

The purpose of scientific research (iv) – Additional provisions

Data retention: personal data shall be kept in a form which permits identification of data subjects for no longer than is necessary to achieve those purposes for which they have been collected and processed. However, personal data may be stored for longer periods as long as they will be processed for scientific research only.



Right to erasure: “right to be forgotten” shall not apply to the extent that the relevant processing is necessary for scientific research, in so far as said right is likely to render impossible or seriously impair the achievement of such purpose.

DIRECTIVE 95/46 VS GDPR: SANCTIONS

DIRECTIVE CURRENT REGIME:

SANCTIONS VARY ON A NATIONAL MEMBER STATE LEGISLATION BASIS.

THE EU HIGHEST SANCTION EVER ISSUED BY A DATA PROTECTION AUTHORITY (“DPA”) FOR A PRIVACY RULES VIOLATION UP TO DATE IS EQUAL TO EUROS **5MILLIONS**.

THE ITALIAN DPA, ALSO KNOWN AS THE GARANTE, HAS ISSUED SUCH A SANCTION IN MARCH 2017

GDPR NEXT REGIME:

FROM MAY 2018 GDPR WILL ENTER INTO FORCE AND SANCTIONS MAY BE UP TO 4% OF THE GROUP ANNUAL TURNOVER OF THE DATA CONTROLLER MEASURED ON A GLOBAL WORLDWIDE BASIS.

IF THE DATA CONTROLLER IS A PUBLIC BODY WITH NO TURNOVER, THE MAXIMUM PENALTY IS FIXED AT **20 MILLIONS EUROS**



Avv. Rocco Panetta
Secretary General of ICF
+39 340 2220023
r.panetta@panetta.net

**MANY
THANKS**