FITFILE

Sharing United Health and Activity Data:

The Challenge of Pseudonymisation vs. Anonymisation

fitfile.com

The Challenge

The most important healthcare challenge of our times is to break down health and activity data silos, making united data easily and readily available to healthcare professionals, researchers, planners, policy makers, and industry.

The use of these united data was demonstrated during the COVID-19 pandemic, where united data has led to safe and effective vaccinations and therapeutics such as dexamethasone, allowing academic and industry research teams to collaborate and produce solutions in record time.

The legal and ethical frameworks make different provisions on how data is processed for a number of purposes, including:

- Direct care and administration
- Commissioning and planning
- Regulatory/ public health
- Research

For data to be used for the different purposes listed above, it may need to be consented, or pseudonymised, or processed using a variety of means and techniques. The only consistent means of delivering Real World Evidence[1] derived from united data for any purpose other than direct patient care is anonymisation.

Definitions

The **General Data Protection Regulation** (GDPR) is a legal framework on data protection and privacy, applicable in the EU and the UK, which defines individual data subjects' rights with regard to their data and governs how personal data of individual data subjects is processed.

United data are data concerning the same individual from multiple sources, united in a single record. Example components can be data from healthcare (e.g. hospital), social (e.g. local government), activity (e.g. heart rate monitor) and medical device (e.g. glucose meter).

Pseudonymisation is the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information such as a key that exists elsewhere[2].

Anonymisation is the processing of personal data in a manner such that the person-specific data do not relate to an identified or identifiable natural person. By anonymisation, personal data is rendered anonymous so that the individual data subject is not or no longer identifiable[3].

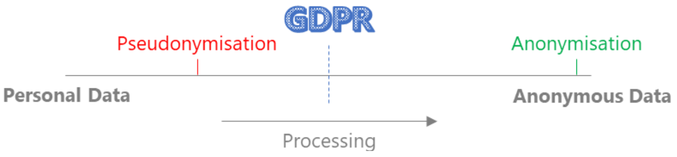
[1] https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence

[2] https://www.legislation.gov.uk/eur/2016/679/article/4

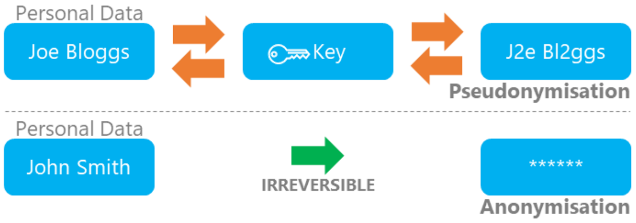
[3] https://eur-lex.europa.eu/eli/reg/2016/679/oj

Overview

Pseudonymisation and Anonymisation are ways in which information can be processed to increase the privacy of personal data.



Pseudonymised data have been processed to replace personal identifiers with alternatives (e.g. with a token that can be reversed with a key), while **anonymised data** have been processed to remove those personal identifiers altogether (irreversible de-identification).



The Facts

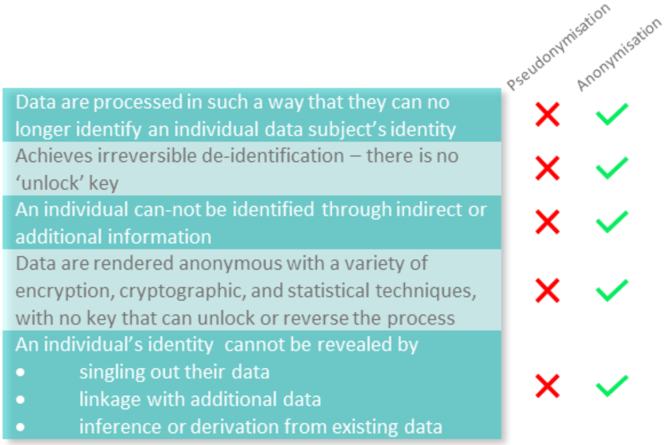
- Pseudonymisation is not an anonymisation technique. The main reasons are that unaltered data can be used to re-identify data subjects via matching techniques, and that reused pseudonyms permit the linking together of different records relating to the same individual[4]
- Pseudonymised data still fall within the auspices of the GDPR: “Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information, should be considered to be information on an identifiable natural person[5]
- Pseudonymisation is only a security measure[6]. It cannot independently prevent re-identification
- Because individuals can no longer be (re-) identified, anonymised data do not concern the GDPR for any purpose, including e.g. statistical and research[7]
- Current techniques to produce united data typically rely on using consent-based identifiable or pseudonymised data (in token- or key-based processes). The reason behind this is that until now, it has not been possible to unite data across sources without sharing actual or pseudonymised personal data. Doing so without consent (or a legitimate purpose such as direct patient care) would be against the law.

[4] https://www.dataprotection.ie/sites/default/files/uploads/2019 06/190614%20Anonymisation%20and%20Pseudonymisation.pdf

[5] https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/what-is-personal-data/what-is-personal-data/

[6] https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/what-is-personal-data/what-is-personal-data/#pd2

[7] https://www.legislation.gov.uk/eur/2016/679

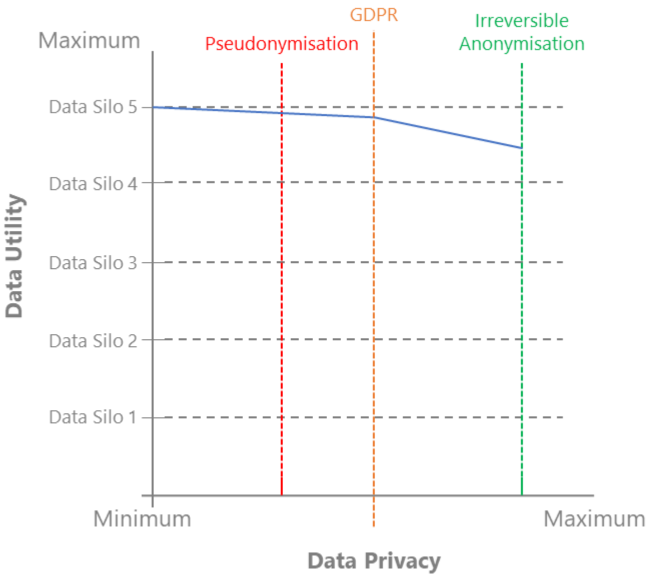


	Pseudonymisation	Anonymisation
Data are processed in such a way that they can no longer identify an individual data subject's identity	✗	✓
Achieves irreversible de-identification – there is no 'unlock' key	✗	✓
An individual can-not be identified through indirect or additional information	✗	✓
Data are rendered anonymous with a variety of encryption, cryptographic, and statistical techniques, with no key that can unlock or reverse the process	✗	✓
An individual's identity cannot be revealed by • singling out their data • linkage with additional data • inference or derivation from existing data	✗	✓

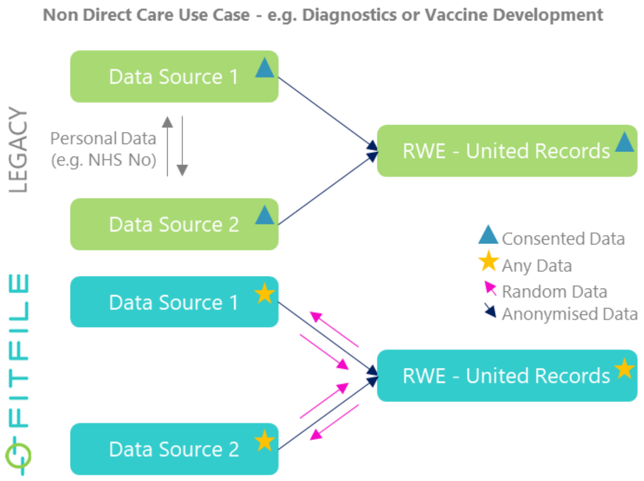
The Conundrum

The generation of patient-level evidence has, to-date, needed to accept a clear trade-off between best-possible privacy (irreversible anonymisation) and largest-possible utility (united data). Conversely, the utility of data may technically decrease a little more with anonymisation than pseudonymisation.

At **FITFILE,** we believe that individuals' rights and privacy are absolutely fundamental, not only to comply with the law, but also to satisfy our strict ethical standards. Hence, we believe that the current trade-off between best-possible privacy (irreversible anonymisation) and largest-possible utility (united data) should be resolved with a process that unites irreversibly anonymised data.



In order to be possible to unite, the data could only be transmitted across sources in identifiable or pseudonymised form, not in irreversibly anonymised form. The issue with this legacy approach could be summarised as the inability to unite non-consented data from individual data sources.



Anonymisation at FITFILE

As mentioned before, data are considered anonymised when data subjects are no longer identifiable. To be precise, recital 26 of the GDPR (UK) states that to determine whether or not the individual is identifiable, one should take into account '…all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly'[8].

At **FITFILE**, we move the privacy needle significantly, by removing all re-identifiable data whenever we irreversibly anonymise and unite data for use in health research and planning. To anonymise, we employ a host of randomisation and generalisation techniques, advanced mathematics, leading edge cryptography, and statistical disclosure controls. However, we do not stop there.

We recognise that:

a) Data processing techniques, advanced statistical analysis and machine learning are improving
b) Even anonymised data must have a defined lifetime
c) Released anonymised datasets create additional accountable stakeholders

To make sure we produce the safest anonymised data, we:

a) Track, assure the quality and monitor the lineage across the whole of the data lifecycle
b) Operate strict data retention rules, even for fully and irreversibly anonymised data
c) Never release datasets publicly, as well as enforce legally binding contractual commitments that prevent onward release and re-identification

Why FITFILE?

At **FITFILE**, we are **defenders of individuals' rights and privacy**. What we demand as individuals, as patients, as human beings, we deliver as professionals.

To safely enable united data to be used in improving outcomes, we have developed privacy-preserving, secure, patented technology that fully and irreversibly anonymises data, whilst deterministically uniting any individual's data across different sources. The result is an unparalleled, quality-assured and uniquely complete set of data signals.

Please email us at contact@fitfile.com for further information.

[8] https://eur-lex.europa.eu/eli/reg/2016/679/oj

FITFILE Group Limited
Bourne Business Park
5 Dashwood Lang Road
Weybridge KT13 0PT
United Kingdom

fitfile.com