# TU Dublin Guidelines for Data Anonymisation / Pseudonymisation

## Purpose and Scope

The purpose is to provide guidance on the anonymisation and/or pseudonymisation of (strictly) confidential and/or personal information contained, processed or transmitted within hardware and software resources that are owned and/or operated by the University. It applies to all University students, staff, and any external parties (e.g. contractors, etc.) with access to information hardware and software resources. It covers all (strictly) confidential/personal data at rest (data stored), and in transit (transmission of information).

## Anonymisation and Pseudonymisation

Anonymisation and Pseudonymisation are two methods of processing (strictly) confidential data, in such a manner that the confidential data in question cannot be traced back to the individual to whom it originally pertained. The key difference between these methods as defined under GDPR, is whether the original data subject can be re-identified.

**Anonymisation** renders the data subject unidentifiable, even to the party that carries out the anonymisation of data. If the data is truly anonymised and identifying the subject is impossible, then the data falls outside the remit of GDPR.

**Pseudonymisation** renders the data subject unidentifiable without the use of additional information. Once the "additional information" and the pseudonymised data are held separately, the data processor/controller can use the data more freely, as the rights of the data subject under GDPR remain intact.

## Anonymisation

When anonymising data, the University must be certain that all information is assessed, and the risk of re-identification is evaluated. This entails examining whether other information is available that, in combination, is likely to facilitate re-identification of the anonymised information. Re-identification is most likely to occur where circumstances described by the combined information are unusual or the population sizes in question are very small.

A "motivated intruder" test should be carried out as a method to check whether the information has been anonymised effectively. This test checks whether a competent individual with the aim of de-anonymising the data could do so successfully.

This test involves discovering whether easily available online/physical information exists that can be used in combination "a jigsaw attack" to re-identify the data subjects to whom the anonymised data pertains. Such resources could include social media, library archives, press archives, electoral register etc.

Re-identification of a data subject would lead to the unauthorised disclosure of (strictly) confidential information and thus constitute a data breach. Any such event should be reported as soon as possible to the DPO.

Members of staff, students and external affiliates of the University should only have access to the level of identifiable information that is necessary for them to complete their assigned activity. However, through effective anonymisation, these activity owners are able to make use of anonymised data for a range of secondary purposes.

Effective anonymisation is achievable via a range of techniques, depending on the nature of the dataset in question and how suitable the chosen technique is. Samples of techniques include:

- Removing personal identifier(s).
- Using identifier ranges (e.g. age range instead of age, partial postcode, age at time of activity event instead of date of birth, output area instead of full address etc.)
- Aggregation – information is only viewed as totals rather than individual data values.
- Randomisation – informational "noise" injected into the dataset to prevent data identification (e.g. the age/height of the individual being increased or decreased by a small amount to avoid identification).

De-identified or anonymised information that goes down to the level of the individual should still be stored and used within a secure environment that has restricted access privileges.

### Pseudonymisation

Pseudonymisation is the process of distinguishing identities. The aim of such a process (vs anonymisation) is to be able to collect additional data relating to the same individual without having to know the identity.

When pseudonymisation techniques are consistently applied, the same pseudonym is provided for individuals across different datasets and over time. This allows datasets and other information to be linked in ways that would not be possible if person identifiable information was removed completely.

This will often be achieved by attaching codes or other unique references to information so that the data will only be identifiable to those who have access to the key or index.

To effectively pseudonymise information, the following actions must be taken:

- Each field of person identifiable information must have a unique pseudonym;
- Pseudonyms to be used in place of identifiable information (e.g. date of birth etc.) and similar fields must be of the same length and formatted on output to ensure readability. For example, in order to replace date of birth in existing record formats, the output pseudonym should generally be of the same field length, but not of the same characters, to avoid confusion with real person identifiable information.
- Generalisation – Diluting information so that identification of individuals is impossible (e.g. instead of date of birth, use year of birth etc.).

- Consideration needs to be given to the impact on existing systems, both in terms of the maintenance of internal values and the formatting of reports;

- Where used, pseudonyms for external use must give different pseudonym values in order that internal pseudonyms are not compromised;

- The secondary use output must, where pseudonyms are used, only display the pseudonymised data items that are required;

- Pseudonymised information should have the same security as person identifiable information.