How personal information can be made safe for reuse

This diagram shows the range of processes that can be used to make personal information safe to reuse in different ways.

Te Tari Taiwhenua
Internal Affairs

Least likelihood

Confidentialisation

Ages: 20-40
Hair: /
Height: 179cm

Ages: 40-60
Hair: blond
Height: 155cm

Confidentialisation
The statistical methods used to protect against confidential information being disclosed to people who are not authorised to have access to it, in a way that could identify an individual, household or organisation. The statistical methods used provide a level of protection against identification that cannot be obtained from de-identification.

Aggregation

Ages: 21-35
Hair: blond + black + brown
Height: 170-185cm

Ages: 41-60
Hair: blond + red
Height: 146-155cm

Aggregation
Data combined from several measurements, but without the additional use of statistical methods to protect against re-identification.

De-identification

Age: 32
Hair: blond
Height: 170cm

Age: 21
Hair: black
Height: 185cm

Age: 35
Hair: brown
Height: 179cm

Age: 60
Hair: blond
Height: 146cm

Age: 41
Hair: red
Height: 179cm

Age: 53
Hair: blond
Height: 173cm

De-identification
The process of removing information from microdata to reduce risk of a recognition of an individual. It typically includes removing names, exact dates of birth or death, and exact addresses.

Pseudonymisation

Name: ABC
DoB: 1-6-84
Hair: blond

Name: DEF
DoB: 23-2-96
Hair: black

Name: XYZ
DoB: 14-9-81
Hair: brown

Name: Oliver
DoB: 8-11-56
Hair: blond

Name: Aroha
DoB: 12-8-75
Hair: red

Name: Ceri
DoB: 20-4-63
Hair: blond

Pseudonymisation
The process of replacing direct identifiers with different ones in microdata to reduce the risk of a recognition of an individual.

Greatest likelihood

Individual client level data

Name: Taika
DoB: 1-6-84
Hair: blond

Name: Leo
DoB: 23-2-96
Hair: black

Name: Mei
DoB: 14-9-81
Hair: brown

Name: Fetu
DoB: 8-11-56
Hair: blond

Name: Asad
DoB: 12-8-75
Hair: red

Name: Sarah
DoB: 20-4-63
Hair: blond

Individual client level data
Information or data on individuals collected/used as part of providing public services. (Other terms used: raw data, microdata, transactional data etc.)

Likelihood of individuals being identifiable