Anonymisation: theory and practice

Mark Elliot
Manchester University
UK Anonymisation Network
www.ukanon.net
Outline

• What is anonymisation?
• The anonymisation decision making framework
What is Anonymisation?

Anonymisation is a **process** by which personal data are rendered non-personal.
DPA definition of personal data

- Data which relate to a *living individual* who can be identified:
  - From those data, or
  - From those data and other information which is in the possession of, or is likely to come into the possession of, the data controller

- Other legislation and jurisdictions do not concern themselves with whether the individuals are living.
Anonymisation and de-identification

• Deal with different parts of the DPAs definition of personal data

• Deidentification tackles:
  – “Directly from those data”

• Anonymisation tackles:
  – “Indirectly from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller…”
Anonymisation types

• Absolute Anonymisation
  – Zero possibility of re-identification under any circumstances

• Formal Anonymisation
  – De-identification (including pseudonymisation)

• Statistical Anonymisation
  – Statistical Disclosure Control

• Functional Anonymisation
Some principles

• Anonymisation is not about the data.
• Anonymisation is about data situations.
• Data situations arise from data interacting with data environments.
Data environment definition

The set of formal and informal structures, processes, mechanisms and agents that either:

i. act on data;

ii. provide interpretable context for those data or

iii. define, control and/or interact with those data.

Elliot and Mackey (2014)
Data environments in practice

• Consist of
  – Agents (people)
  – Infrastructure (particularly security)
  – Governance processes
  – Other data

• Layered

• Partitioned
Some principles

• Anonymisation is not about the data.
• Anonymisation is about data situations.
• Data situations arise from data interacting with data environments.
• You cannot decide whether data are safe to share /release or not by looking at the data alone
Some principles

• Anonymisation is a process to produce safe data but it only makes sense if what you are producing is safe useful data.
Some principles

• Anonymisation is a process to produce safe data but it only makes sense if what you are producing is safe *useful* data.

• *Zero risk is not a realistic possibility if you are to produce useful data.*
Some principles

• Anonymisation is a process to produce safe data but it only makes sense if what you are producing is safe *useful* data.

• Zero risk is not a realistic possibility if you are to produce useful data.

• *The measures you put in place to manage risk should be proportional to that risk and its likely impact.*