UNIVERSITY OF
Southampton
School of Electronics
and Computer Science

The Anonymisation Decision-Making Framework (ADF).

Kieron O’Hara
18 January 2017

UKAN

UKAN
UK ANONYMISATION NETWORK

ico.
Information Commissioner's Office

With thanks to the ICO for 2 years' seed funding

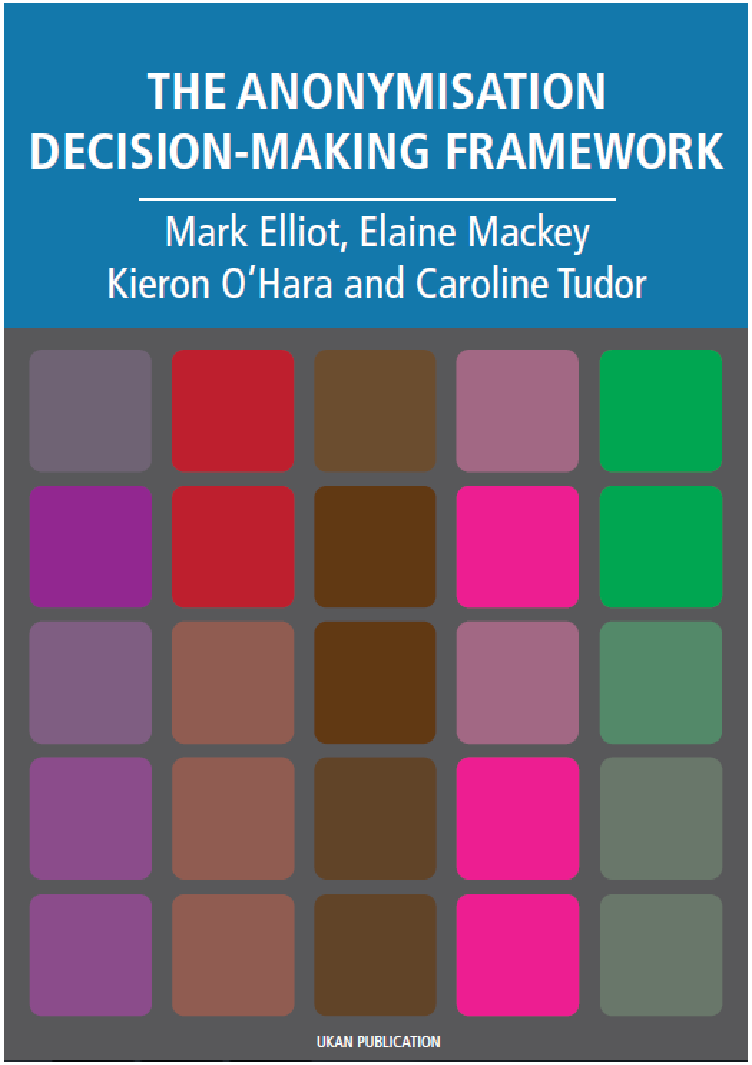
Office for National Statistics

UNIVERSITY OF Southampton

ODI open data institute

MANCHESTER 1824
The University of Manchester

UKAN Anonymisation Book



THE ANONYMISATION DECISION-MAKING FRAMEWORK

Mark Elliot, Elaine Mackey
Kieron O'Hara and Caroline Tudor

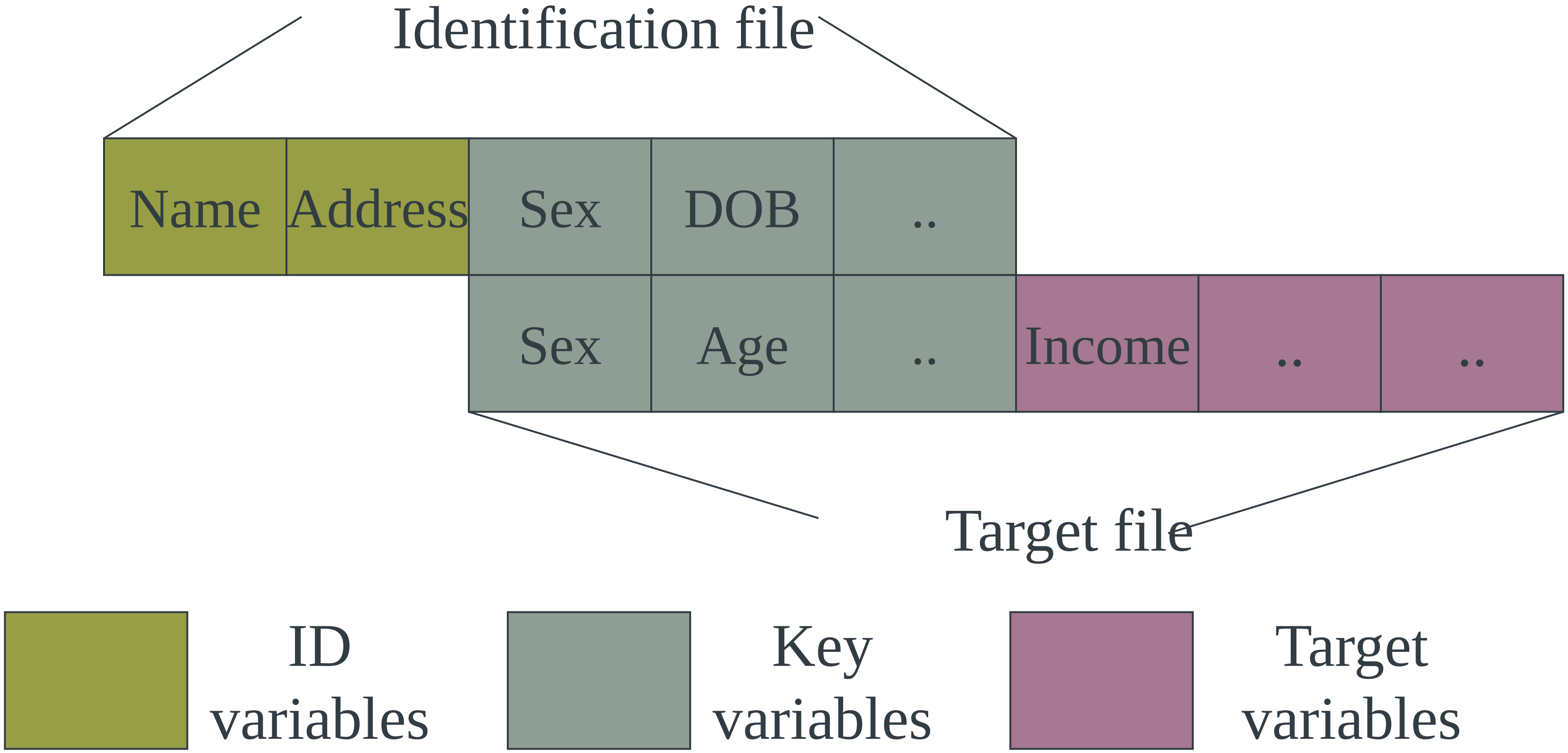
UKAN PUBLICATION

- A practical guide
  - Complementing ICO CoP
- http://ukanon.net/ukan-resources/ukan-decision-making-framework/

Types of Anonymisation (I)

- Deidentification – to prevent identification directly from the data
  - Remove direct identifiers
- Pseudonymisation – to allow limited reidentification of deidentified individuals
  - Replace identifiers
  - I don't know who this is, but I know she is the same as her
- Statistical disclosure control
  - Manipulate the data to quantify risk

Deanonymisation



Identification file

Name	Address	Sex	DOB	..			
		Sex	Age	..	Income	..	..

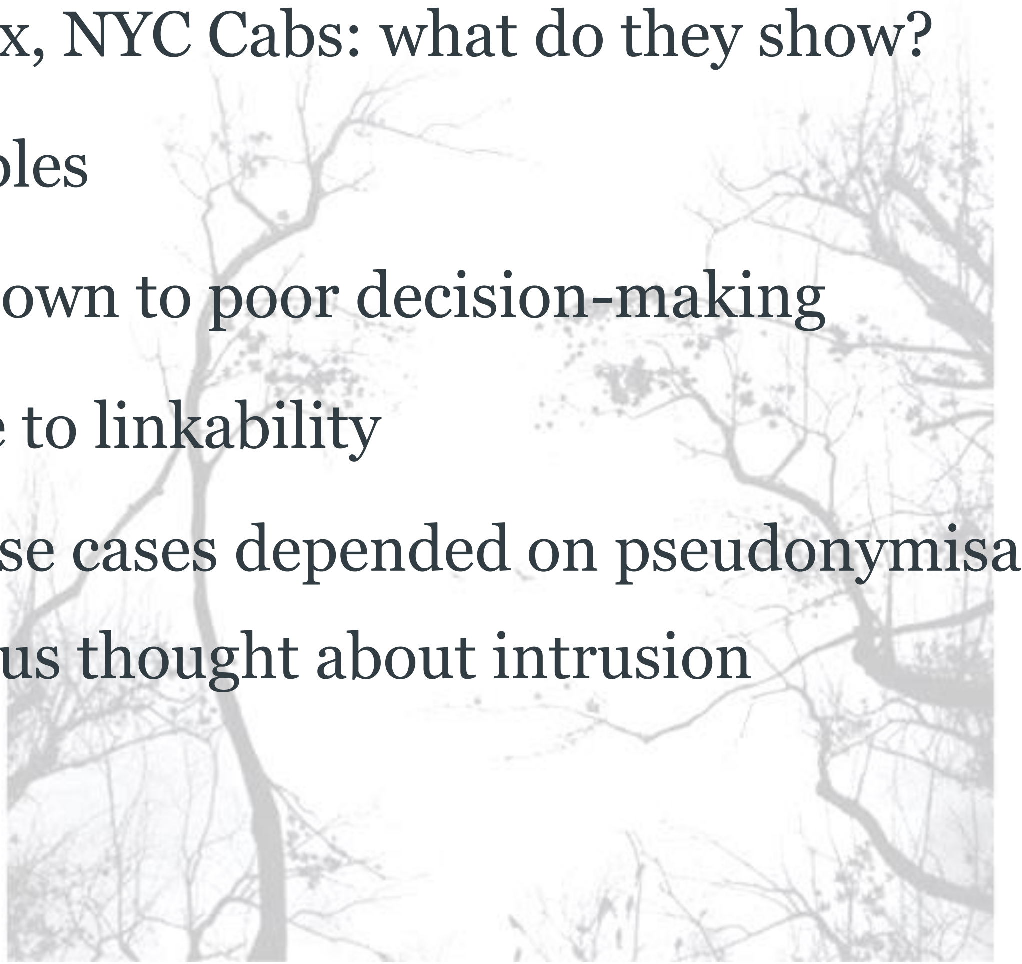
Target file

ID variables

Key variables

Target variables

Examples



- AOL, Netflix, NYC Cabs: what do they show?
- BAD examples
- Generally down to poor decision-making
- Largely due to linkability
  - But all use cases depended on pseudonymisation
  - No serious thought about intrusion

Risk Within the Data

- Remove variables
- Remove records
- Aggregation
- Suppressing unique values
- Sampling
- Barnardisation

- Data swapping
- Adding noise
- Microaggregation on k-partition
- Detecting verbal tics
- Identifying and pixellating faces

Only Goes So Far

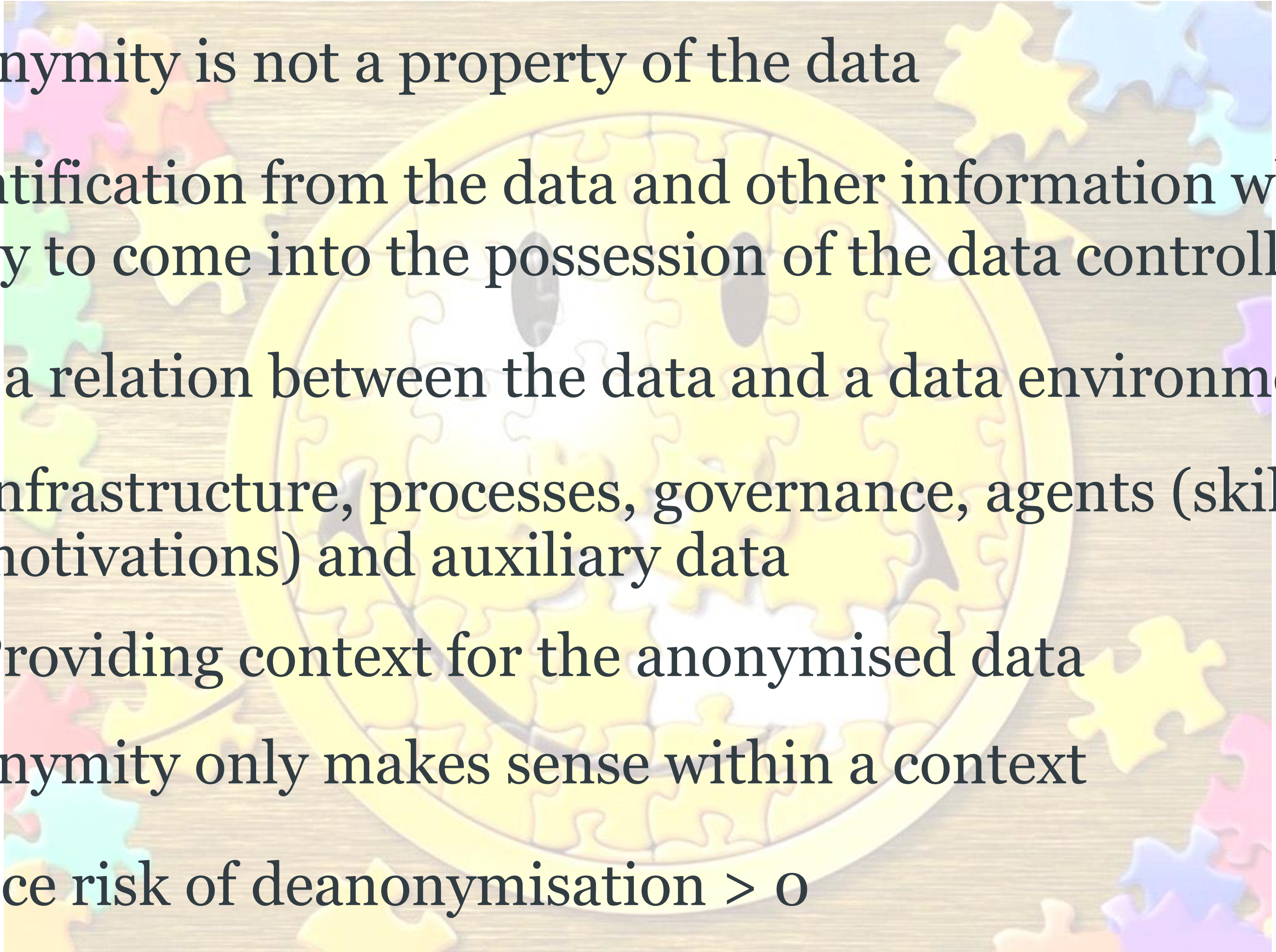
- E.g. k-anonymity
  - "every combination of quasi-identifier values occurring in the dataset must occur at least k times."
  - Hence the set of quasi-identifiers has to be defined in advance
- The success criteria of anonymisation are pre-defined
- But the risk of breach depends on intruder's information
  - Cannot be known in advance

Sex	Address	Age	Nat
F	SO1 3BB	23	UK
F	SO1 5MD	23	UK
F	SO1 9QQ	21	Fr
F	SO1 2DH	27	UK
F	SO1 2DH	27	UK

Sex	Address	Age	Nat
F	SO1	21-30	EU
F	SO1	21-30	EU
F	SO1	21-30	EU
F	SO1	21-30	EU
F	SO1	21-30	EU

Sex	Address	Age	Nat	Pay
F	SO1	21-30	EU	£25k
F	SO1	21-30	EU	£17k
F	SO1	21-30	EU	£21k
F	SO1	21-30	EU	£32k
F	SO1	21-30	EU	£750k

Axioms of the ADF



- Anonymity is not a property of the data
- Identification from the data and other information which is likely to come into the possession of the data controller
- It is a relation between the data and a data environment
  - Infrastructure, processes, governance, agents (skills, motivations) and auxiliary data
  - Providing context for the anonymised data
- Anonymity only makes sense within a context
- Hence risk of deanonymisation > 0

Two Views

- Irreversibility
  - Reidentification must be impossible
  - BUT we know is it always possible
  - Data is anonymous or useful, but not both
- Risk management
  - Raise costs of reidentification above benefits
- BOTH views present in GDPR
- ADF provides a methodology for the second view

Misunderstanding

- ‘Anonymisation’ appears as a success word
  - Cf. ‘murder’, ‘scoring a goal’
- But anonymisation is a context-relative *process*
  - Means likely to be used by intruder will change over time
  - Anonymous now ≠ anonymous tomorrow
  - Anonymous here ≠ anonymous there
- Can’t just anonymise, release and forget about it

Types of Anonymisation (II)

- Functional anonymisation
  - Prevent identification indirectly from the data and other information
  - Trace and remove identifying information
  - Technical/legal/managerial means
  - Risk management
- Example of Privacy by Design
  - See Cavoukian principles
  - Though GDPR problematic!!

Risk Outside the Data

- Motivation
- Consequences (is this goal achievable with other means?)
- Governance (who gets to see the data, under what conditions?)
- Provenance
- Other available data (time series, open data, commercial data, data in the same domain)
- Data quality

Who’s the Attacker?

- Spontaneous recognition
  - Researcher recognises someone in the data
- General attack
  - Reidentify as many as possible in the data
- Fishing attack
  - Looking for a specific person in the data
- Fishing attack with response knowledge
  - Looking for a specific person *known* to be in the data

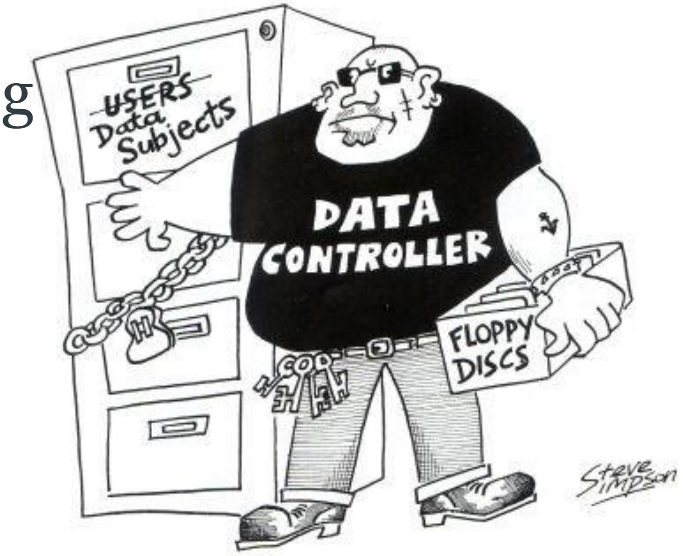
Alter the Context

- Access control
  - – Who is trusted to have access?
  - – What constraints do we have?
- Query control
  - – Differential privacy
- Secure environments
- Restricting the analysis
  - – Project approval
  - – Publishing agreements

Responsibilities of Data Controllers

UNIVERSITY OF Southampton
School of Electronics and Computer Science

- Understand how a privacy breach may occur
- Understand the possible consequences
- Address the risk of a breach occurring
  - What do you do when it does?
- Understand the environment
- Never release-and-forget
  - Anonymising is an ongoing commitment



http://www.stevesimpson-art.co.uk/illustration/136972_data-controller.html

ADF

1. Describe your intended data situation
2. Understand your legal responsibilities
3. Know your data
4. Understand the use case
5. Meet your ethical obligations
6. Identify the processes you will need to go through to assess disclosure risk
7. Identify the disclosure processes that are relevant to your data situation
8. Identify your stakeholders and plan how you will communicate with them
9. Plan what happens next once you have shared or released the data
10. Plan what you will do if things go wrong

UKAN
UK ANONYMISATION
NETWORK

UKAN Services

- Website ukanon.net
- Clinics
- Consultancy
- Engagement
- Dissemination of best practice via case studies
- ADF guidance
- admin@ukanon.net

UKAN
UK ANONYMISATION
NETWORK

Conclusions

- You cannot decide whether data is anonymous only by looking at the data
- Anonymisation aims at producing data that is useful (as well as safe)
- Zero risk is not an option
- Anonymisation methods should be proportional to the risk
- The ADF will take the data controller on a journey through these issues

Disclaimer

- Texts, marks, logos, names, graphics, images, photographs, illustrations, artwork, audio clips, video clips, and software copyrighted by their respective owners are used on these slides for non-commercial, educational and personal purposes only. Use of any copyrighted material is not authorized without the written consent of the copyright holder. Every effort has been made to respect the copyrights of other parties. If you believe that your copyright has been misused, please direct your correspondence to: kmo@ecs.soton.ac.uk stating your position and I shall endeavour to correct any misuse as early as possible.