UNIVERSITY OF OSLO

GDPR: Identifiability, pseudonmyisation and anonymisation

Peter Davis
Stipendiat, NRCCL, University of Oslo

p.a.e.davis@jus.uio.no

2 December 2021

Agenda

- Article 4(1) primer
- Threshold for identifiability
  - *Case example: smart billboards / facial detection*
- Anonymisation techniques
  - *Case example: Transport for London*
- Practical considerations for anonymisation

A29WP Building Blocks

Opinion 04/2007 on the concept of personal data

1. Any information
2. Relating to – content / purpose / result elements
   - Joined Cases C-141/12 and C-372/12: YS. and M. and S
     - Privacy Commissioner v Telstra Corporation Limited [2017] FCAFC 4
   - Cf. Case C-434/16: Nowak
3. **Identified or identifiable**
4. Natural person

3rd Building Block: Identified criterion

- 'Identified' – two purposes
  - One of two alternate criteria for satisfying 3rd building block
  - Conceptual touchstone for assessing *identifiability* criterion
- When is someone 'identified'?
  - A29WP 04/2007: 'within a group of persons, he or she is "distinguished" from all other members of the group'
  - European Agency for Fundamental Rights & CoE, Handbook on European DP Law 2008: 'identification… requires elements which describe a person in such a way that he or she is distinguishable from all other persons and recognisable as an individual'
  - *Unique* civil identity or something less?

3rd Building Block: Identifiable criterion

• **Article 4(1) GDPR**:

… an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person

• **Recital 26 GDPR**

To determine whether a natural person is identifiable, account should be taken of 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.

- Also see Recital 30, elaborating on types of identifiers

• ***Case C-582/14, Breyer*** at 45-46:

'a means reasonably likely to be used… would not be the case if the identification of the data subject was prohibited by law or practically impossible on account of the fact that it requires a disproportionate effort in terms of time, cost and man-power, so that the risk of identification appears in reality to be insignificant.

Significance of *Breyer*

- First CJEU embrace of ‘absolute’ (vs ‘relativist’) approach
  - Though see Mourbry & others 2018, *Are ‘pseudonymised’ data always personal data?*
  - ‘Protecting against each and all possible third parties at any time, who are not necessarily the intended recipients of the data, is problematic and unrealistic. Such a requirement eliminates the need for any risk management because it compels the data controller to always make the worst possible assumptions even if they are not relevant to the specific context.’
    - El Emam & Alvarez 2015, *A critical appraisal of the Article 29 Working Party Opinion 05/2014 on data anonymization techniques*
- Contributing to GDPR’s march towards a ‘law of everything’
  - Purtova 2018, *The Law of everything. Broad concept of personal data and future of EU data protection law*

Smart Billboards and Facial Detection

- Facial detection = technology that uses sensors (typically one or more cameras) …that facilitate the detection of faces, and perceived characteristics thereof, of people that come within its view.
- Technical process
    1. Modelling faces of passers-by
    2. Compression (to assist comparison)
    3. Comparison
    4. Deletion or anonymisation
- Purposes
    - Targeted advertising
    - Analytics (pricing and demographic targeting)
- Is PD being processed?

DPA Confusion on Smart Billboards

PD is processed

- **Norway**: Datatilsynet guidance, ‘Tracking in Public Spaces’
- **Italy**: Garante Decision no. 551, 11 December 2017
- **Netherlands**: DPA guidance 25 June 2018

PD is not processed

- **Norway**: Datatilsynet oral advice to ProntoTV, c.2017
- **Sweden***: Press release 27 June 2019
- **Ireland**: Press release 15 May 2017

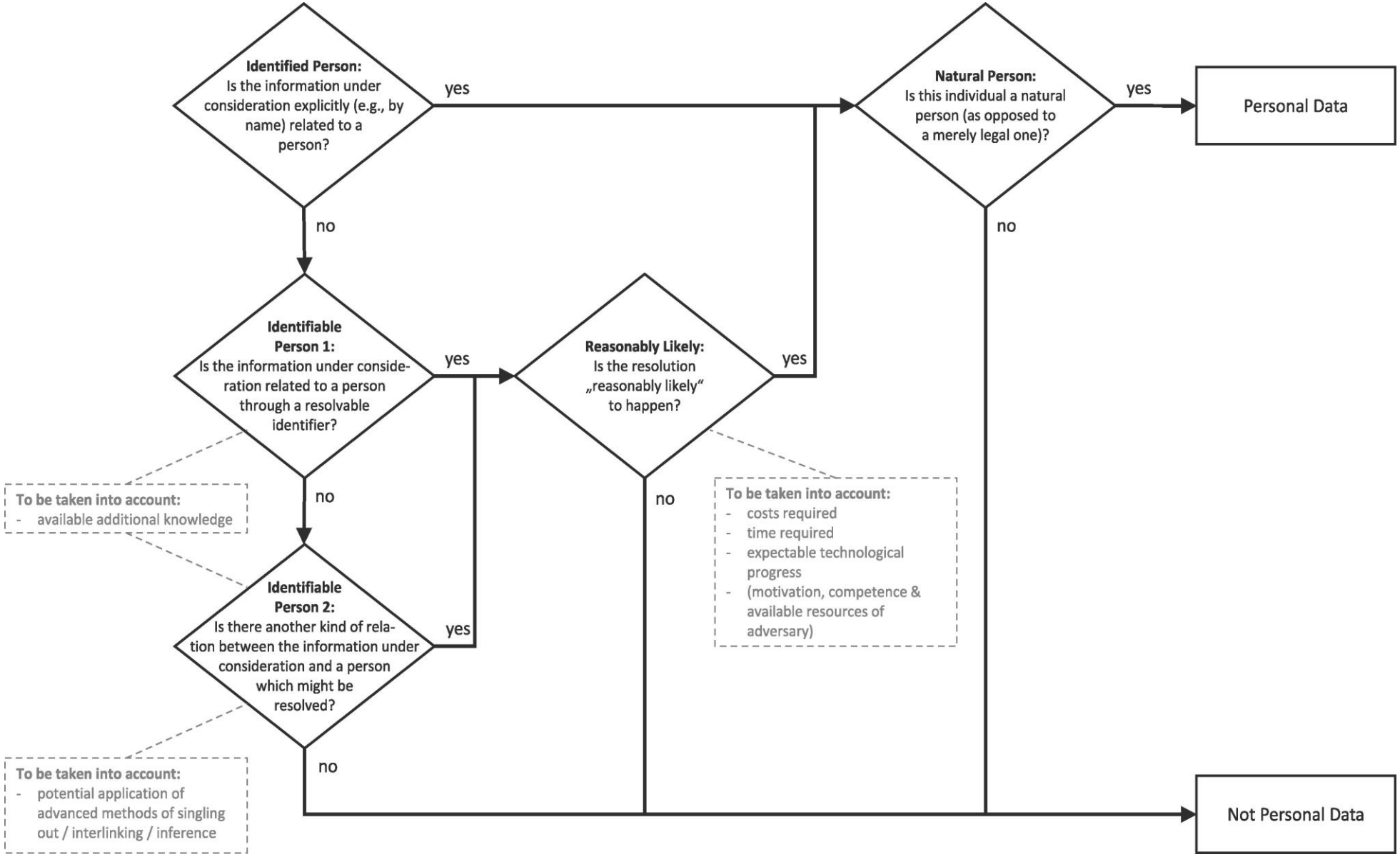
Anonymisation – why?

- Compliance with e.g. data minimisation / storage limitation principles
  - Anonymisation as alternative to erasure
- For analytics purposes
  - Where anonymised data is sufficient
  - Where difficult to procure/identify legal basis for processing PD
    - Though note potential for Art 6(4) (further processing) & 89 (research/statistical purposes)
  - Ethics (e.g. medical / research)

Anonymisation Techniques

A29WP Opinion 05/2014 on Anonymisation Techniques

- Two species of techniques
  - Randomisation
    - E.g. noise addition, differential privacy, homomorphic encryption*
  - Generalisation
    - Aggregation / K-Anonymity
- Note arms race vs ‘rapid progress of (re)identification technologies’ (Purtova, 2018)
  - (Increasing) difficulty of achieving anonymisation in practice
- EFTA case Joined Cases E-11/19 and E-12/19, Adpublisher AG v J & K (EFTA court)
  - Non-disclosure vs anonymisation
  - Referring court assumed ‘J & K’ was anonymisation


Identified Person:
Is the information under consideration explicitly (e.g., by name) related to a person?
yes
Natural Person:
Is this individual a natural person (as opposed to a merely legal one)?
yes
Personal Data
no
no
Identifiable Person 1:
Is the information under consideration related to a person through a resolvable identifier?
yes
Reasonably Likely:
Is the resolution „reasonably likely" to happen?
yes
no
To be taken into account:
- available additional knowledge
To be taken into account:
- costs required
- time required
- expectable technological progress
- (motivation, competence & available resources of adversary)
Identifiable Person 2:
Is there another kind of relation between the information under consideration and a person which might be resolved?
yes
no
To be taken into account:
- potential application of advanced methods of singling out / interlinking / inference
Not Personal Data

Transport for London – WiFi tracking

- Scope of WiFi pilot:
  - 1 month in November/December 2017
  - >509 million MAC addresses
  - 54 stations
  - Data included:
    - MAC addresses
    - Date and time
    - Device manufacturer
  - From the data, 'constructed 42 million journeys from five million devices during the pilot'

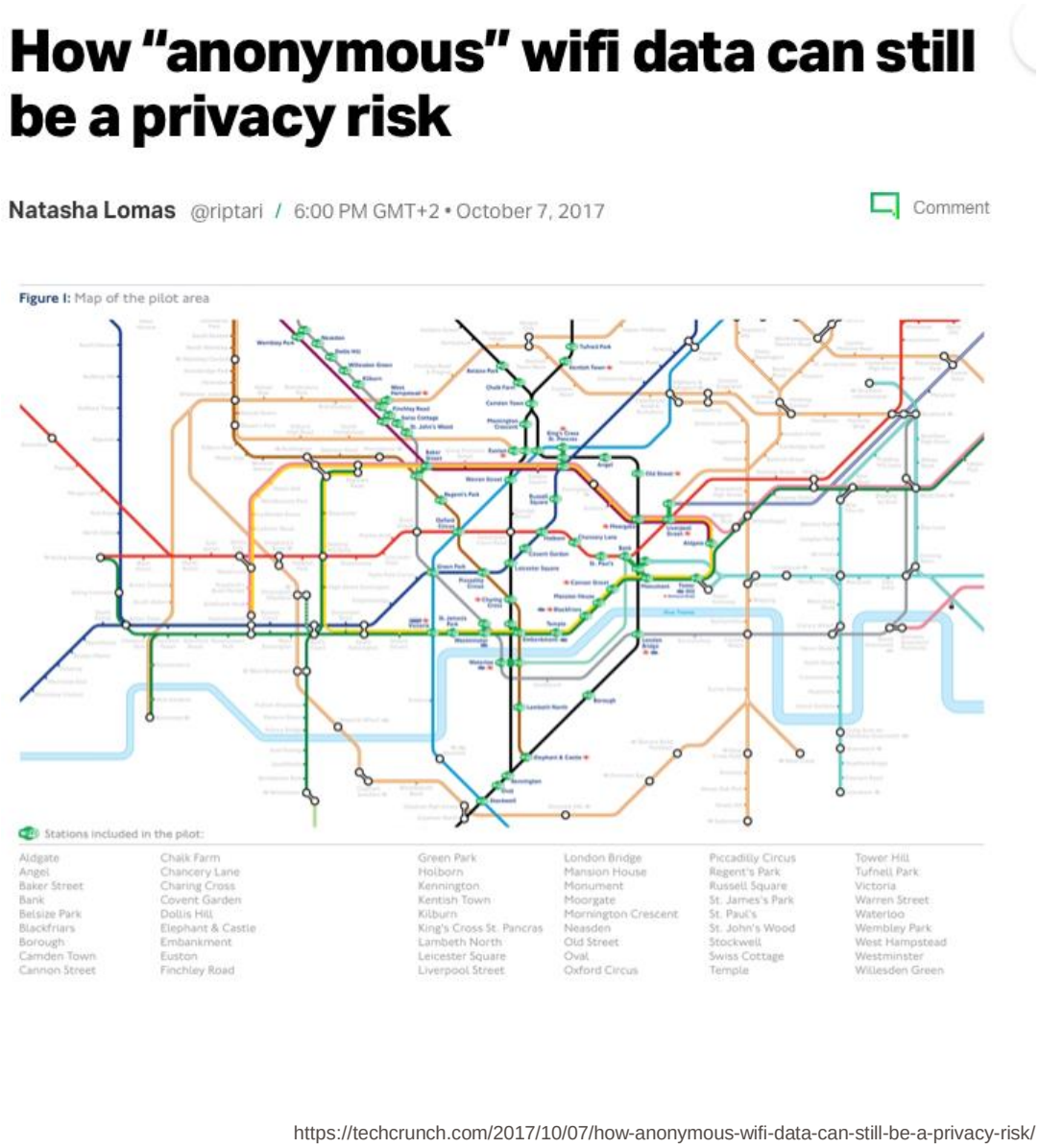
Transport for London – WiFi tracking (2)

- Steps taken by TfL
  - Notifying passengers in advance
    - Passengers could ‘opt out’ by turning off WiFi on devices
  - DPIA
  - Met with ICO
  - Applied ICO
  - ‘Anonymisation’ techniques applied
    - MAC addresses were salted and hashed
      - Salt apparently not changed throughout month
    - ‘we consider the data to be anonymous and are unable to identify any specific device.’

Ahead of and during the pilot, posters kept customers informed

Transport for London – WiFi tracking (3)

- FOI request made for ‘anonymous’ data
- Response:
  - ‘Although the MAC address data has been pseudonymised […] given the possibility that the pseudonymised data could, if it was matched against other data sets, in certain circumstances enable the identification of an individual, it is personal data. The likelihood of this identification of an individual occurring would be increased by a disclosure of the data into the public domain, which would increase the range of other data sets against which it could be matched.’
  - See further Veale, Binns & Ausloos 2015, *When data protection by design and data subject rights clash*



https://techcrunch.com/2017/10/07/how-anonymous-wifi-data-can-still-be-a-privacy-risk/

Considerations when adopting anonymisation techniques

- Note that anonymisation = processing operation
- Does ‘anonymisation’ decrease risk, or just create new risks?
- Increased value or sensitivity of data -> increased motivation for attacker -> more ‘reasonably likely’ that data subjects will be identified
- Risk-mitigating measures
  - DPIA?
  - Allow data subjects to opt-out *ex ante*?
    - Trūata/Mastercard example
  - Organisational and technical measures to protect ‘anonymised’ data
    - What if breach of ‘anonymised’ or pseudonymised data?
      - Notification obligations?

Conclusions

- Concept of personal data – overinclusive but underdetermined
  - GDPR as ‘law of everything’
  - But uncertainty in ‘edge’ cases – e.g. face detection
- Achieving (useful) anonymisation is difficult in practice
  - Broader conversation about anti-competitive effects of GDPR
- Do anonymisation - and other PbD techniques - actually help privacy?
  - Veale, Binns & Ausloos 2015: privacy-as-confidentiality vs privacy-as-control

Discussion points

- How best to respond to ‘edge’ cases (i.e. where application of GDPR uncertain)
- Is useful anonymous data achievable in your field?
- Have you considered 3rd party / off-the-shelf anonymisation solutions? What are the risks?
- Is the GDPR trending towards a ‘law of everything’?