The law of everything?
Broad scope of personal data and future of EU data protection law

Nadya Purtova, TILT

PI.lab annual conference

7 December 2017

European Research Council

Established by the European Commission

Introducing INFO-LEG

"Understanding information for legal protection of people against information-induced harms"

Funded by the 2016 ERC Starting Grant

Started on 1 March 2017

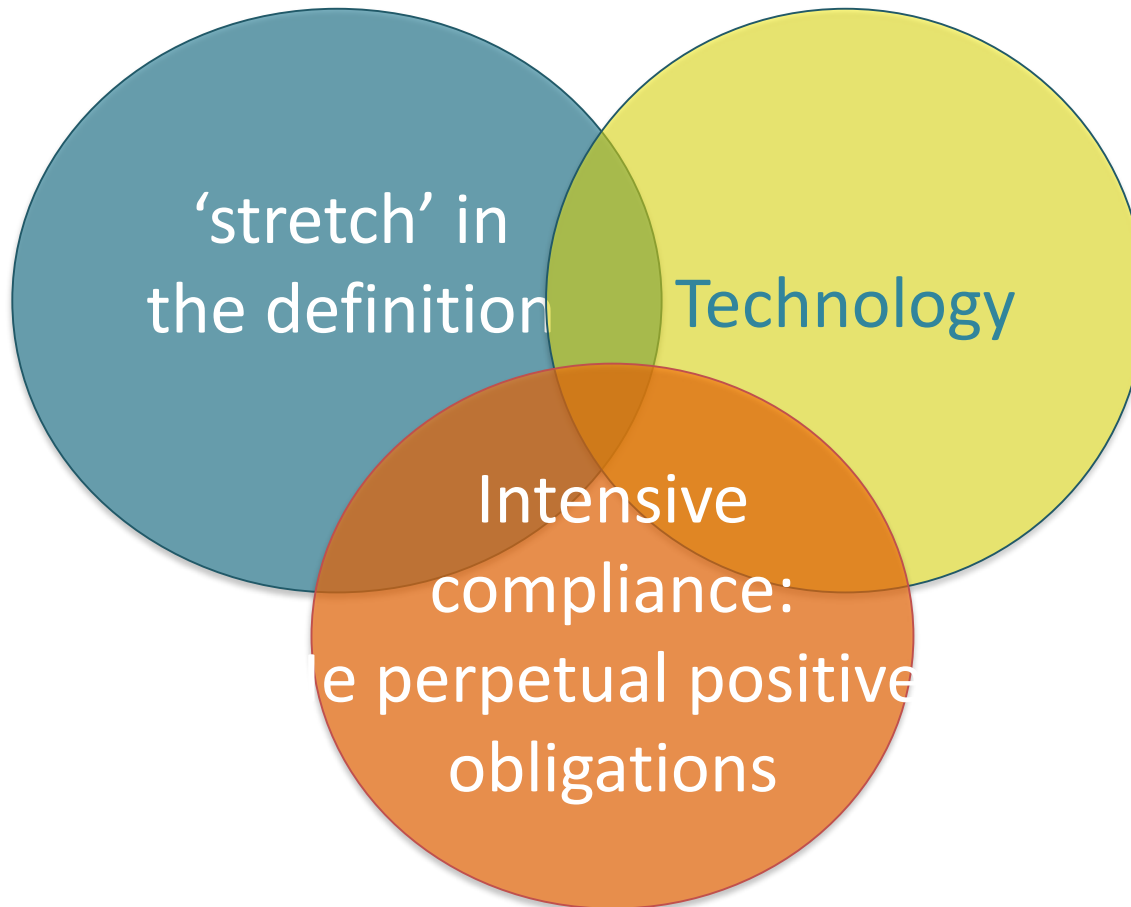# Legal protection organized around 'personal data'

LEGAL PROTECTION

Personal Data

Protection based on 'personal data' will overstretch and collapse

Time to look for alternatives

Interplay of factors

'Personal data' is meant to stretch far

- Scope of data protection law is meant to be broad (WP29, ECJ case-law eg *Lindqvist, Google Spain,*);

- 'Principle of complete protection' (e.g. *Google Spain*);

- Proportionality is a secondary consideration (after determining the scope)

Personal data

Art 4(1) GDPR (similar in Data Protection Directive)

any information relating to a natural person ('data subject') who is identified or identifiable;

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.

Broad approach to identifiability

Recital 26, WP136 + *Breyer*:

- Direct & Indirect (via overlapping non-unique factors)
- Possibility / "reasonable likelihood" of identification, taking into account
- state of art at time of processing
- all other objective factors
- Regardless whether or not you have the identification means at your disposal (the 2016 *Breyer* case)
- Mere purpose to identify → positive conclusion on identifiability

Technology moves towards perfect identifiability

- 2000:
  - ZIP code + date of birth + gender ➔ identification of 87% of the US population (Sweeny *"Simple demographics … "*).
- 2009:
  - stripped-down graphs of social relationships with non-anonymised graphs from other social networks ➔ re-identification of 'anonymous' users of Twitter (Narayanan et al).
- 2014:
  - location of credit card holders on 4 occasions + amounts spent + shop type +code representing each person ➔ re-identify 90% of 3 months of credit card transactions of 1.1 million people in 10,000 shops
  - + the amounts spent on *these 4 occasions* ➔re-identification of nearly all of 1.1 million card-holders. (Bohannon (2015) in *Science*)

Relating to [a natural person]

Broad understanding in WP136:

- *Content* (about DS)
- *Purpose* (processed with the purpose to evaluate of influence DS)
- *Result* (data processing likely to impact DS)

Cautious CJ / inconclusive:

- Case-law does not explicitly mention WP29;
- *YS and others*: PD is data *about* a person?
- AG Kokott in *Nowak:* relating in terms of a <u>purpose</u> to assess
- Judgment expected 20.12.2017 (hopefully a broad interpretation)

The more data, the more spurious correlations → 'remote relationships'

Datafication:

Everything is being converted to data

A provocation (?): weather in *Stratumseind 2.0*

Stratumseind 2.0: naar een veilige en leefbare uitgaansstraat

Tekst | Freke Sens
Foto's | Bart van Overbeeke

Stappen veiliger maken en de sfeer tijdens het uitgaan verbeteren. Niet door toegangspoortjes of een hek, zoals eerder dit jaar geopperd werd, maar door de inzet van licht of geur. Dat is wat er gebeurt op Stratumseind. De Eindhovense kroegenstraat is namelijk omgevormd tot een 'living lab'. Daarin wordt onderzoek gedaan naar het gedrag van bezoekers, maar ook geëxperimenteerd. Volgende maand vinden de eerste lichtexperimenten plaats. Ook de TU/e levert een bijdrage.

- 'Smart infrastructures';
- 'Datafication' of the going out area, incl.
  - Noise, mobile phone data, direction of wind, amount of rain per sq m
- Objective: to predict, prevent, and react to 'deviant behavior', incl. by changing street lighting / involving the police

Ongoing research of Maša Galic @TILT

How far / close to reality?

- Cavoukian and Castro re identifiability:
  - high identifiability only in "high-dimensional data" ('Big Data and Innovation, Setting the Record Straight: De- identification Does Work' (2014))

- Narayanan and Felten:
  - "high-dimensional data is now the norm, not the exception. … [T]hese days it is rare for useful, interesting datasets to be low-dimensional." ('No silver bullet: De-identification still doesn't work' *randomwalker.info*, published 9 July 2014)

Boundaries of PD will be tested

- Art 20 GDPR right to data portability:
  - What data shall be subject to portability requests & obligations?
- EU policy initiatives (2017 proposal for Regulation on free movement of non-personal data)
  - Is there non-personal data?

Problem exacerbated

Continuous compliance

- *ex ante,* i.e. before processing takes place (purpose statement, data protection by design; proportionality assessment, etc.)

- *ex post*, i.e. after processing (accountability, yet rooted in measures taken ex ante)

- *during* processing: continuous data protection impact assessment feedback loop

Lon Fuller, *The Morality of Law*

Good law is the law which is possible to obey

- *a crime "to cough, sneeze, hiccough, faint or fall down in the presence of the king ... [and] not to understand, believe in, and correctly profess the doctrine of evolutionary, democratic redemption."*

Omnibus & hence impossible to comply with regulation is bad law.

Options

1. Status quo
   - Risking appearance of compliance, not real protection
   - Standards, seals to exude the image of compliance / manage uncertainty

2. Limit the concept of personal data (e.g. to data *about* a person)
   - Limited protection;
   - Where to draw borders?

3. Look for another option:
   - Based on a better understanding of information, data & how it relates to people
   - Given that all data is personal, what should trigger protection of various degrees of intensity?
   - On top of PD, or instead.

INFO-LEG aims to reinvent protection
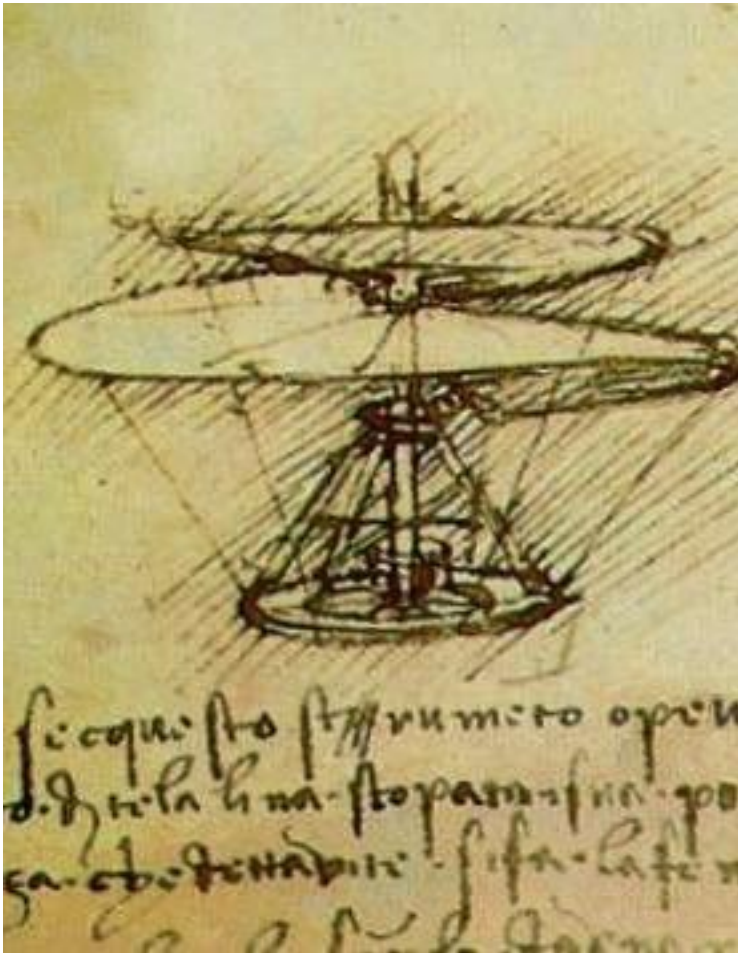
Find new notion instead of personal data to build legal protection around

- *'organizing notion(s)'.*

All alternatives so far are based on 'personal data'

Caveats

- I do not plea for a limited definition of personal data (eg akin to the US PII);

- We should stick to the broad definition until a better way is found;

- The fact that compliance is burdensome does not necessarily mean that the law is bad and needs to be reformed.

Integrated interdisciplinary understanding of information
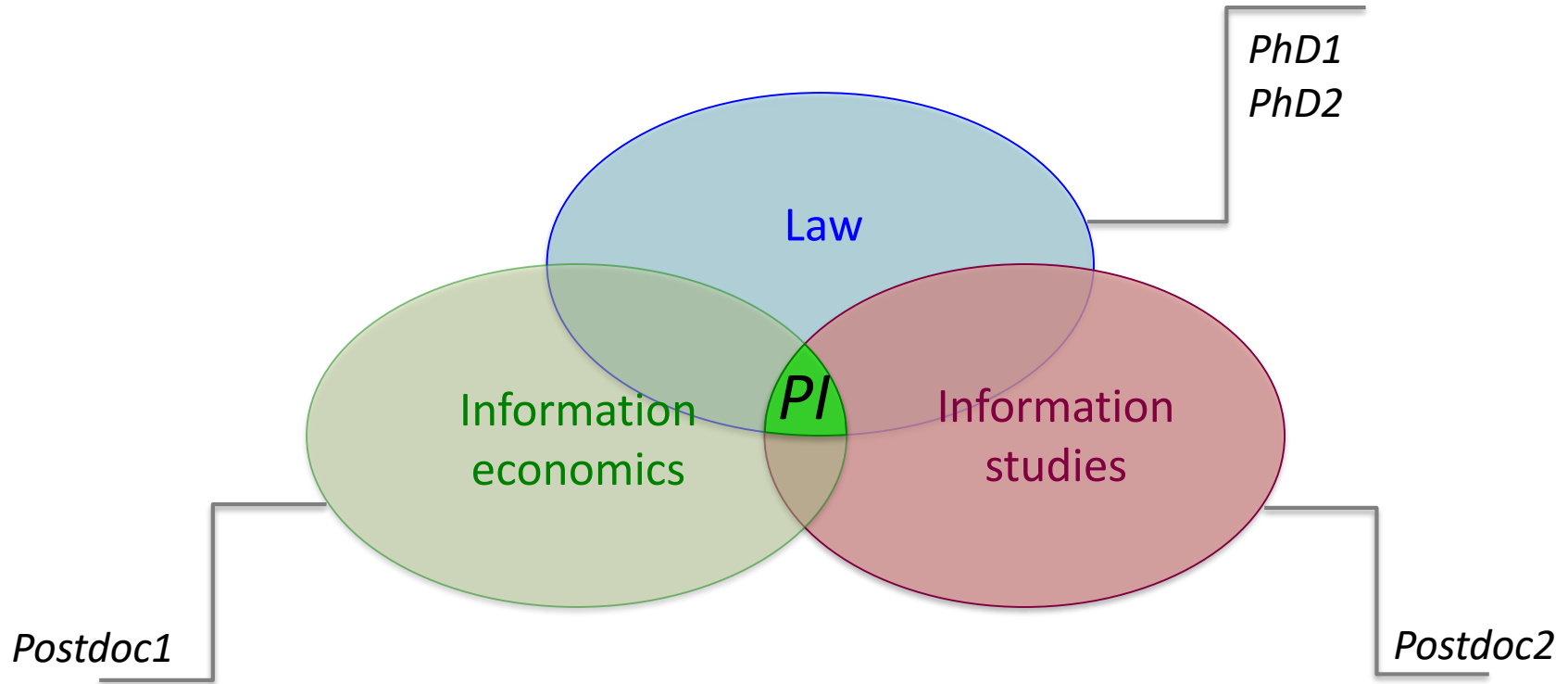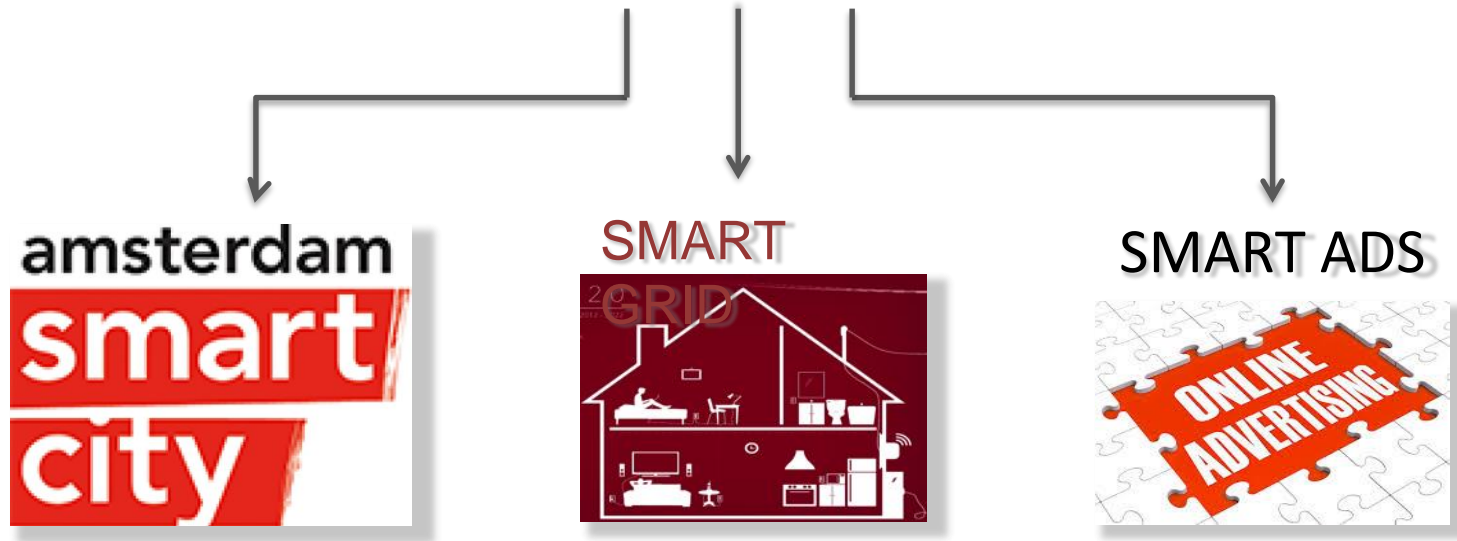
Empirical research

- to keep up with rapid technology developments;
- to capture *what is not in the state-of-art*, via:
  - document analysis
  - Interviews
  - Delphi surveys

3 case studies

amsterdam smart city

SMART GRID

SMART ADS

Why these cases?

- Social significance
- Rapid development → new notions
- Access to documentation & people

What will we gain?

- Imminent need for a new notion instead of 'personal data':
    - Personal data in everything - daily practice in 5-10 years;
    - Compliance and enforcement unworkable;
- Fundamental research now to address regulatory necessity later.

Acknowledgements

European Research Council

Established by the European Commission