

Building European data economy in the Digital Single Market : Data protection and the data economy: friends or foes?

Dr. Raphaël Gellert, Assistant Professor, Faculty of law, Radboud Universiteit

Towards a Holistic Regulatory Approach for the European Data Economy: Why the Illusive Notion of Non-Personal Data is Counterproductive to Data Innovation

by

Inge Graef, Raphaël Gellert and Martin Husovec

*Reprinted from European Law Review
Issue 5 2019*

INFO-LEG Project

Funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 716971).



European Research Council



TILT Data Portability Fellowship 2019

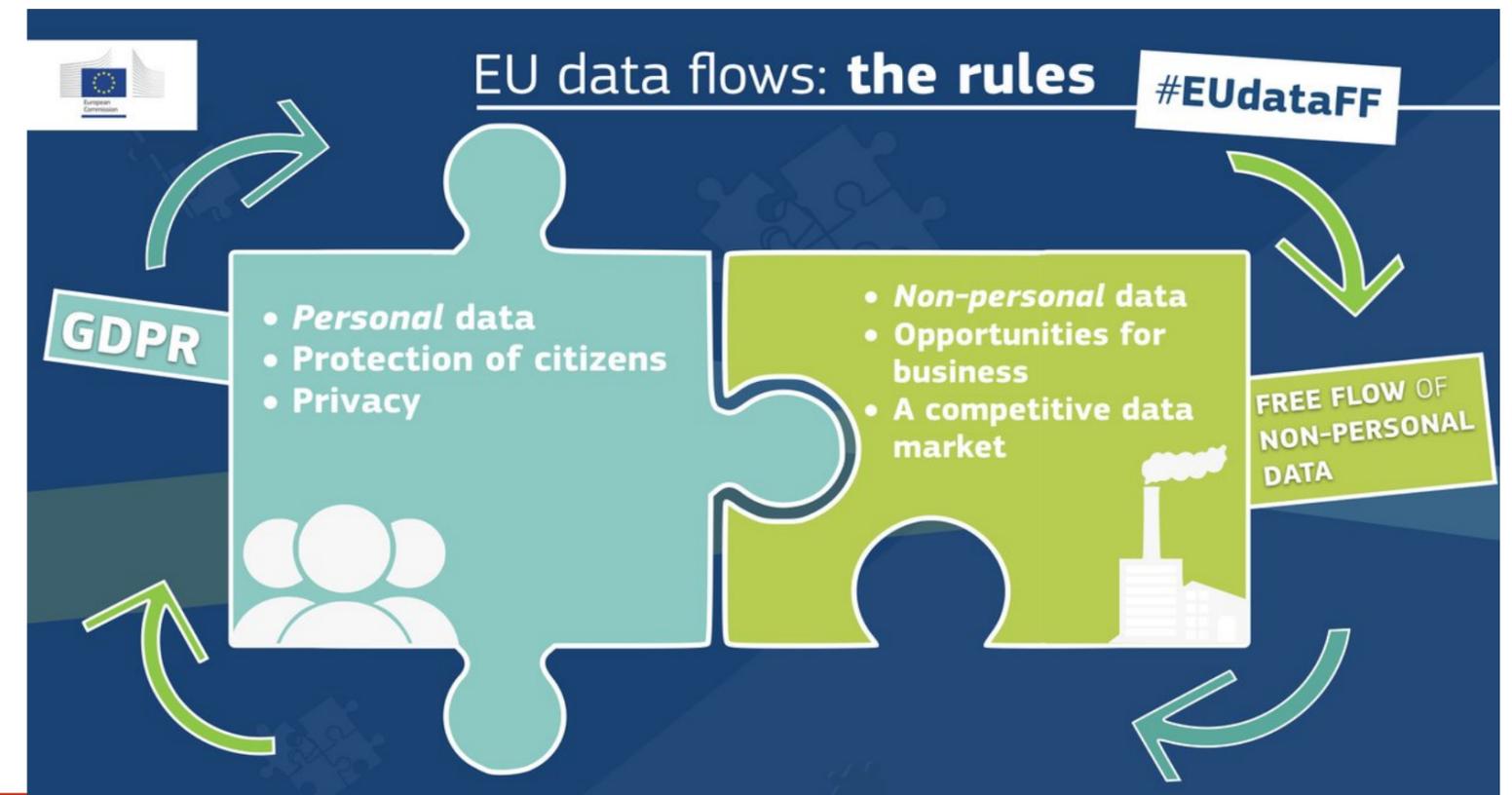
In 2019, TILT will welcome up to two research fellows working in the area of data ownership, re-use, access and portability (broadly, data-enabled innovation). We particularly encourage candidates working in the area of non-personal data and the European data economy to apply. The selected fellows will become a part of an ongoing project of the institute and will be able to join the upcoming TILTing Perspectives conference that takes place from 15 to 17 May 2019.

European Data Economy initiative

- January 2017 Commission Communication [Building a European Data Economy](#)
- September 2017 proposal for a Regulation on the [free flow of non-personal data](#)
- April 2018 Commission Communication [Towards a common European data space](#)
- April 2018 Commission Staff Working Document [Guidance on sharing private sector data in the European data economy](#)
- 19 February 2020: [A European strategy for data](#)

Notion of non-personal data: Regulation on FFNPD

- European Commission:
 - SMEs and start-ups experience obstacles in reusing data from third parties, in particular as regards **machine-generated, non-personal data**
 - Machine generated data = non-personal data
 - Regulation (EU) 2018/1807 on a framework for the free flow of non-personal data in the European Union
 - GDPR and ePrivacy Directive ‘fully regulate the processing of personal data’



Dynamic and moving notion of personal data

- Notion of personal data: (i) any information; (ii) relating to; (iii) an identified or identifiable; (iv) natural person
- Relating to: in content, purpose or impact
 - intended and unintended impact of data processing
 - increasingly difficult to distinguish between data that will and will not impact a natural person
- Identified or identifiable: possibility of identification (i.e., single out)
 - Threshold: all the means reasonably likely to be used
 - Technical state-of-the-art
 - means at the disposal of the data controller or 'another person'
 - Identification is socio-technical
 - Governance, and purpose of actors

Can data really be anonymous?

- 2000:
 - ZIP code + date of birth + gender → identification of 87% of the US population.
- 2008:
 - Netflix: Anonymised movie recommendations database re-identified by crossing it with open-access IMDB database → 68% re-identification success
- 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
- 2019 (Rocher et al.):
 - 99.98% of Americans would be correctly re-identified in any dataset using 15 demographic attributes
 - **“even heavily sampled anonymized datasets are unlikely to satisfy the modern standards for anonymization set forth by GDPR**
 - **and seriously challenge the technical and legal adequacy of the de-identification release-and- forget model.”**

Personal data = broad notion

- **Contextual**
- **Probabilistic**
- **Low threshold**

- 2018: Purtova: “everything is PD”
- 2019: Fink & Pallas:
 - Probabilistic notion
 - Too difficult to make definitive judgment call

- PD = **dynamic and context-dependent elements** which cannot be established for one data item in an absolute way

Points of interaction between the two regimes (1)

- **Parallel** application in cases of mixed datasets:
 - Mixed dataset: contains PD and Non-PD
 - how to disentangle several data points to be able to comply with different requirements of the two regimes?
- art.2(2) FFNPD: “Where personal and non-personal data in a data set are **inextricably linked**, this Regulation shall not prejudice the application of Regulation (EU) 2016/679”.
 - Recital 10 FFNPD: “this Regulation does not impose an obligation to store the different types of data separately”
 - Inextricably linked: is it even possible to distinguish in the 1st place?
- 2019: Guidance on FFNPD:
 - the GDPR applies fully to the whole mixed dataset when the non-personal and personal data parts are inextricably linked—even when personal data represent only a small part of the dataset.
 - → **What’s the point of FFNPD?**

Points of interaction between the two regimes (2)

- Subsequent application:
 - Exact point of switch between the regimes will be difficult to determine
- EDPS: 2017 Opinion on the draft EU Regulation on integrated farm statistics
- Creation of statistical data shows the moving nature of personal data
 - Original dataset: personal data
 - Enriched with additional data to make it meaningful
 - Transformation: output statistical data through pseudonymization
 - Resulting statistical data:
 - Can be non personal if:
 - Keys are destroyed
 - Additional anonymization measures

Points of interaction between the two regimes (3)

- **Undermining** each other's policy objectives

| GDPR | Data economy |
|--|---|
| <ul style="list-style-type: none">• Free movement of data• Fundamental right to data protection | <ul style="list-style-type: none">• Free movement of data• “Data innovation” |

Extension of free flow to non-personal data

- GDPR is less strict to data localisation restrictions than FFNPD
 - GDPR only prohibits restrictions of free movement of personal data for reasons relating to data protection (Art. 1(3) GDPR)
 - Free flow proposal contains general prohibition of restrictions of free movement of non-personal data, with the exception of justifications on grounds of public security (Art. 4(1) proposal)

Comparison of approaches

GDPR

Restrictions of free movement of personal data connected with data protection

No limits to such restrictions imposed for reasons other than data protection

FFNPD

Restrictions of free movement of non-personal data generally

Justification for such restrictions imposed for reasons of public

Why creating N-PD?

- Elephant in the room?
 - GDPR can create hurdles to seamless sharing of data
 - Legitimate basis
 - Purpose limitation

Looking at data sharing arrangements

- 2018 Guidance on data sharing:
 - Proportionality
 - Purpose limitation
 - “Do no harm”
 - → // GDPR requirements
- Even stricter:
- 2018 Guidance on data sharing:
 - Conditions for data re-use

- sharing can only take place concerning explicitly agreed upon purposes
- easy access, review, retrieval, extraction,
 - and correction of the data for the data originator
- Unconditionally opt out of the contract



- What if?
- Mistake:
 - Machine generated data = Non-Personal Data
- Instead:
 - End of a two-tiered regime
 - Build a single regime for ALL data
 - By default: GDPR compatible

Residual Q?

- Do we still want a special status for machine generated data?
 - Why?
 - Added value?
 - How?

Thank you!
R.Gellert@jur.ru.nl