UNECE Expert Meeting on Statistical Data Confidentiality

Poznan, 01-03 December 2021

Shedding light on the legal approach to aggregate data under the General Data Protection Regulation & the Free Flow Data Regulation

Emanuela Podda

MSCA PhD Candidate
Alma Mater Studiorum – Università di Bologna
University of Luxembourg
Università di Torino

European Legal Framework Data Flow

General Data Protection Regulation (EU) 2016/679 - (GDPR)

Free Flow Data Regulation (EU) 2018/1807 - (FFDR)

Taxonomic Analisys

Personal Data
Art. 4 General Data Protection Regulation

any information related to an identified or identifiable natural person

Non-Personal Data
Art. 3 Free Flow Data Regulation

data other than personal data as defined in the GDPR

two mutually exclusive definitions of data based on its nature

Any kind of different data represents a declination of these two categories

What about statistical data, namely when used by private entities?

Data Processing for Statistical Purposes (Art. 89 GDPR)

Processing for […] statistical purposes, shall be subject to appropriate safeguards, in accordance with this Regulation, for the rights and freedoms of the data subject.

Statistical Purpose (Recital 162 GDPR)

*any operation of collection and the processing of personal data necessary for statistical surveys or for the production of statistical results. […] the statistical purpose implies that the result of processing for statistical purposes is not personal data, but **aggregate data**, and this result or the personal data are not used in support of measures or decisions regarding any particular natural person.*

OUTPUT OF PROCESSING FOR STATISTICAL PURPOSES

AGGREGATE DATA

First Processing of PII & Creation of a Dataset Compliant with Consent and the Initial Purpose justifying the First Collection of PII

Further Processing...

Aggregation

Anonymization

Output of data processing for Statistical Purposes:
Aggregate Data
NON-PERSONAL DATA

Output of Anonymization Data Processing:
Anonymous Data
NON-PERSONAL DATA

Outside of the scope of application of the GDPR

It means that...

+

- aggregate data are equalized to anonymized data, and somehow associated,
- as such data is not used in support of measures or decisions regarding any particular natural person.

HOWEVER...
The state of the art confirms that, as for anonymization, aggregation is strictly dependand of the context of application

Private entities performing aggregation of personal data may not have in place the same ethical framework of public entities

+

It appears reasonable to question:

- how large the aggregate should be before the data ceases to be "personal"?

As a consequence,

- how to tackle the risks linked to re-uses, re-purpose and re-contextualization of aggregate data, especially when developing ML models and AI products?

Art. 89 GDPR

Derogations from the mandated disclosures (data subject rights)

⬇

Risk Containment for Data Aggregation in the Private Sector

+ **(in the light of the context of application, as the state of the art confirms that the GDPR has an oversimplified approach)**

Thank you