

# Data analytics and the GDPR: more constraints or new opportunities?

Assoc. Prof. Sophie Stalla-Bourdillon, Director of Ilaws

Twitter: @SophieStallaB

Blog: <https://peepbeep.wordpress.com/>

[s.stalla-bourdillon@soton.ac.uk](mailto:s.stalla-bourdillon@soton.ac.uk)

What is the GDPR?



Not (exactly) the BBW!

The EU General Data  
Protection Regulation  
(25 May 2018)

OK but what about  
data analytics?

Some of the distinctive aspects of big data analytics are:

- the use of algorithms
- the opacity of the processing
- the tendency to collect 'all the data'
- the repurposing of data, and
- the use of new types of data.

ICO 2017

True but let's start  
with a preliminary  
remark...

‘many instances of big data analytics do not involve personal data at all’

ICO 2017



'Examples of non-personal big data include world climate and weather data; using geospatial data from GPS-equipped buses to predict arrival times; astronomical data from radio telescopes in the Square Kilometre Array; and data from sensors on containers carried on ships.'

ICO 2017

‘Also, big data analytics may not involve personal data **for other reasons**; in particular it may be possible to successfully **anonymise** what was originally personal data, so that no individuals can be identified from it.’

ICO 2017

So what does the  
GDPR say about  
data analytics?



**“Sorry about this, but I’m only following orders.”**

Some interesting  
stuff indeed!

Not to be  
misleading



## 2 data protection principles in particular:

- **Data minimisation** ('personal data shall be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed')
- **Purpose specification** ('personal data shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes')





The GDPR requires

love and patience!

Personal data  
processing as

a set of dynamic  
processes starting at  
 $t=n$  and ending at  
 $t=n+2, 3, 4\dots$

$t=n$

Data collection

t=n+1

‘General analysis’.

‘In order to create incentives to apply **pseudonymisation** when processing personal data, measures of **pseudonymisation** should, whilst **allowing general analysis...**’

Recital 26 GDPR

With more words



‘Where the processing for a purpose other than that for which the personal data have been collected is not based on the data subject's consent or on a Union or Member State law which constitutes a necessary and proportionate measure in a democratic society to safeguard the objectives referred to in Article 23(1), the controller shall, **in order to ascertain whether processing for another purpose is compatible with the purpose for which the personal data are initially collected**, take into account, inter alia:

- (a) any link between the purposes for which the personal data have been collected and the purposes of the intended further processing;
- (b) the context in which the personal data have been collected, in particular regarding the relationship between data subjects and the controller;
- (c) the nature of the personal data, in particular whether special categories of personal data are processed, pursuant to Article 9, or whether personal data related to criminal convictions and offences are processed, pursuant to Article 10;
- (d) the possible consequences of the intended further processing for data subjects;
- (e) the existence of appropriate safeguards, which may include encryption or **pseudonymisation**.’

There is this “implicit” idea that if you  
pseudonymise the dataset  
you should be able to undertake  
a general analysis of the dataset

What does  
pseudonymisation mean?

“pseudonymisation’ means the processing of personal data in such a manner that **the personal data can no longer be attributed to a specific data subject without the use of additional information,** provided that such **additional information is kept separately and is subject to technical and organisational measures** to ensure that the personal data are not attributed to an identified or identifiable natural person’

Art. 4(5) GDPR

Legal pseudonymisation  
means more than masking  
direct identifiers!

By the way

What is legal  
anonymisation?

‘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. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments.’

Recital 26 GDPR



What else?

For 'processing which does  
not require identification'

There are exemptions!

'where, ..., the controller is able to demonstrate that it is not in a position to identify the data subject, the controller shall inform the data subject accordingly, if possible. In such cases, Articles 15 to 20 shall not apply except where the data subject, for the purpose of exercising his or her rights under those articles, provides additional information enabling his or her identification.'

Art. 11(2) GDPR

Question time

So if a controller has  
pseudonymised data does  
it mean that it is not in a  
position to identify data  
subjects?



What else?



Under the GDPR

profiling requires special  
care!

“profiling’ means any form of automated processing of personal data consisting of the use of personal data to **evaluate certain personal aspects relating to a natural person**, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements.’

Art. 4(4) GDPR

1. **Right to object** in particular when profiling for direct marketing purposes (Art. 21 GDPR)
2. **Right not to be subject to a decision based solely on automated processing**, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her (Art. 22 GDPR) unless...
3. **Mandatory impact assessment** when systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person (Art. 35 GDPR)

So if the general analysis is in fact a profiling exercise (at  $t=n+1$ ), which is intended to be followed by individual decisions affecting profiled subjects (at  $t=n+2$ )

a data protection impact assessment should be conducted at  $t=n+1$  and the repurposing of the data should not necessarily be seen as compatible with the initial processing.

Conclusion

# Is the GDPR sound\*?

\*Hu, Runshan and Stalla-Bourdillon, Sophie and Yang, Mu and Schiavo, Valeria and Sassone, Vladimiro, Bridging Policy, Regulation, and Practice? A Techno-Legal Analysis of Three Types of Data in the GDPR (September 1, 2017). Available at SSRN: <https://ssrn.com/abstract=3034261>