

Legal aspects of text mining publicly available data

Benjamin Bremert

IFIP Summer School 2017

„Privacy and Identity Management – the Smart World Revolution“

Ispra, 7 September 2017



Unabhängiges Landeszentrum für
Datenschutz Schleswig-Holstein



Agenda

- Introduction
- Data protection / GDPR
 - Personal data / Principles of data processing
 - Consent
 - Contract
 - Legitimate interest
 - Information obligations
- Copyright



Text and data mining

Extraction of patterns and knowledge from large (unstructured) amounts of data by applying methods of machine learning and statistics.



Article 2 (1) GDPR:

„This Regulation applies to the **processing of personal data wholly or partly by automated means** and to the processing other than by automated means of personal data which form part of a filing system or are intended to form part of a filing system.“



Personal data

Article 4 GDPR:

„... means any **information relating to an identified or identifiable natural person** ('data subject'); 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“



Principles of data processing

Article 5 (1) (a) GDPR:

„Personal data shall be ... processed **lawfully, fairly** and in a **transparent manner** in relation to the data subject (‘lawfulness, fairness and transparency’);“



Lawfulness of processing

Article 6 (1) GDPR:

„Processing shall be lawful only if and to the extent that at least one of the following applies:“

- Consent, Article 6 (1) (a) GDPR
- Necessity for performance of contract or prior to entering into a contract, Article 6 (1) (b) GDPR
- Necessity for purposes of legitimate interests, Article 6 (1) (f) GDPR



Article 6 (1) (a) GDPR

Article 4 (11) GDPR: „... any **freely given, specific, informed** and **unambiguous indication** of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her"



Article 6 (1) (a) GDPR

Article 6 (4) GDPR: „When assessing whether consent is freely given, utmost account shall be taken of whether, inter alia, the performance of a contract, including the provision of a service, is conditional on consent to the processing of personal data that is not necessary for the performance of that contract.“



Article 6 (1) (a) GDPR

Recital 42: „ For consent to be informed, the data subject should be aware at least of the identity of the controller and the purposes of the processing for which the personal data are intended. Consent should not be regarded as freely given if the data subject has no genuine or free choice or is unable to refuse or withdraw consent without detriment.“



Article 6 (1) (a) GDPR

Article 7 (2) GDPR: „ If the data subject's consent is given in the context of a written declaration which also concerns other matters, the request for consent shall be presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language. Any part of such a declaration which constitutes an infringement of this Regulation shall not be binding.“



Article 6 (1) (b) GDPR

“... processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract”



Article 6 (1) (f) GDPR

„1. Processing shall be lawful only if and to the extent that at least one of the following applies: (...)

(f) processing is **necessary** for the purposes of the **legitimate interests** pursued by the **controller** or by a **third party**, *except* where such interests are overridden by the **interests** or **fundamental rights** and **freedoms of the data subject** which require protection of personal data, in particular where the data subject is a child. (...)



Article 6 (1) (f) GDPR

Legitimate interests of the data controller or a third party

- vague term, not legally defined
- limitation only done by the adjective „legitimate“
- Should we take common interests into consideration?



Article 6 (1) (f) GDPR

Necessity of processing for pursuing legitimate interests

- Implementation of the principle of data minimisation, Article 5 (1) (c) GDPR
- A specific data processing is necessary, if the legitimate interests cannot be realised by a data processing with a lesser gravity of interference and simultaneously maintaining the level of realisation.



Article 6 (1) (f) GDPR

Opposing interests of data subject

- Opposing interests, fundamental rights and freedoms of the data subject are to be determined.
- What does it mean, that the data subject is not limited to „legitimate“ interests?
 - According to recital 47 the determination of interests has to take into consideration the „reasonable expectations of data subjects based on their relationship with the controller“.



Article 6 (1) (f) GDPR

Weighing of interests

- The standard for the weighing of interests is, that the more one interest is affected, the higher the importance of the opposite interest has to be.
- Standard is modified by Article 6 (1) (f) GDPR, which as a rule allows the data processing if the interests of the data subject do not prevail the legitimate interests of the data controller.
- Does it have to be taken into consideration, that the data controller is compliant with his obligations?



Article 12 GDPR

„The controller shall take appropriate measures to provide any information referred to in Articles 13 and 14 ... relating to processing to the data subject in a concise, transparent, intelligible and easily accessible form, using clear and plain language, in particular for any information addressed specifically to a child.“



Article 14 GDPR

- Identity of the data controller, Article (1) (a) GDPR
- Contact details of the data protection officer, Article 14 (1) (b) GDPR
- Purposes of the processing and the legal basis for processing, Article 14 (1) (c) GDPR
- Categories of personal data concerned (Article 14 (1) (d) GDPR
- ...



Article 14 GDPR

- Problem 1: Processing of additional and otherwise discarded information, just for the sole purpose of compliance (Article 11 (1) GDPR).
- Problem 2: Actual implementation of the notification of the data subject about data being processed.



Exception of Article 14 GDPR

Article 14 (5) (b) GDPR: "... the provision of such information proves impossible or would involve a disproportionate effort, in particular for processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes ... In such cases the controller shall take appropriate measures to protect the data subject's rights and freedoms and legitimate interests, including making the information publicly available"



Article 14 GDPR

The crawler could „leave“ the URL containing the information while crawling. For example in the useragent:

TestCrawler/1.0 – <https://gdpr.testcrawler.example>



Reuse Policy Framework





Reuse Policy Framework

reuse.txt

```
{public_posts}
allowed          all
policy          none
expire         7d
wait           5s
limit          1

{private_posts}
allowed          none

{science_posts}
allowed          all
policy          sci
wait           1s
limit          5

[/*]
ruleset         private_posts

[/blog/*.html]
(crawler: bad_crawler)
allowed         none

ruleset         public_posts
```

REST-API

```
"reuse" :{
  "allowed":"all",
  "policy":"none",
  "expire":"7d",
  ...
}
```




Copyright

Article 5 Directive 2001/29/EC: „**Temporary acts of reproduction** referred to in Article 2, which are **transient or incidental** [and] an **integral and essential part of a technological process** and whose **sole purpose** is to enable:

(a) a transmission in a network between third parties by an intermediary,

or (b) a **lawful use**

of a work or other subject-matter to be made, and which have **no independent economic significance**, shall be exempted from the reproduction right provided for in Article 2.“



Thanks for your attention! 😊



Supported by:



Federal Ministry
for Economic Affairs
and Energy

on the basis of a decision
by the German Bundestag