

24.02.2015

Statement by the German Data Forum

on select aspects of the current draft

of the **General Data Protection Regulation (GDPR)**

- *This statement pertains to the European Parliament legislative resolution of 12 March 2014 on the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) – COM(2012)0011 – C7 - 0025/2012 – 2012/0011(COD) – First Reading*
- *This statement was drafted having regard to, but without taking into account the amendments made during the Italian Presidency to the Council document 15395/14 from December, 19 2014 – Interinstitutional File 2012/0011(COD) Limite*

Background

The „Directive on the protection of individuals with regard to the processing of personal data and on the free movement of such data“ proposed by the European Commission in January 2012, better known as the EU General Data Protection Regulation, will provide the European Union with a unitary data protection policy that is legally binding for all member states. It will replace the data protection directive 95/46/EG adopted in 1995 which was thus updated and amended to meet the demands of technological development.

The relevance of a harmonized data protection regulation is underscored by the over 3,000 suggested amendments to the Commission’s regulation draft that were submitted to the Committee on Civil Liberties, Justice and Home Affairs of the European Parliament.

Presently, the trialogue between the European Council, the European Commission, and the European Parliament on the new EU regulation has not yet begun. A Council working group will continue negotiations after a revised document has been submitted under the Italian EU Council presidency. We expect a consolidated version of the Council document soon, and the trialogue to begin before the summer break in 2015. Against this backdrop, the German Data Forum wants to amplify the perspective of the social and economic sciences on selected aspects of the draft in question.

The German Data Forum (*Rat für Sozial- und Wirtschaftsdaten*, RatSWD) is an independent expert committee. Since 2004, it has been advising the German Federal Government on the advancement of the research data infrastructure. The German Data Forum

has significantly contributed to the improvement research access to official statistical data and research data while safeguarding data protection.

Data Protection in the Social and Economic Sciences

Data protection, meaning the protection of people's personal data from disclosure and abuse, is essential to empirical research. The willingness of people to participate in scientific surveys and to share information on their personal situation is closely linked to the guarantee and realization of anonymity; this includes anonymous data analysis.

This basic premise of data protection also applies to secondary analysis of administrative data (e.g., tax data) or registers (e.g., population register). Violating data protection regulation contradicts the logic of action and the interests of scientific research because such conduct would not only be unethical, but would alienate the public thereby eroding the empirical foundation of research.

Against this backdrop, the German Data Forum welcomes the EU's plans to introduce the General Data Protection Regulation (GDPR, in the following) that harmonises generally binding protection of personal data EU-wide. The German Data Forum also welcomes that the GDPR is shaped by research-friendly aims and is in many ways similar to German data protection regulation. However, the GDPR also contains provisions which could hinder future empirical research in judicial practice. This is especially true for medical research and the use of health data. Important German academic organisations have already voiced their concerns regarding this matter. Since health data are increasingly collected and used in social¹ and economic research, the German Data Forum² supports these concerns.

In view of the analysis of social and economic data, the following remarks pertain to **problematic details** in the GDPR draft which have been overlooked in the discussion so far. Since these details could prove a barrier to empirical social and economic research in the future, we want to point out these problems and offer possible solutions.

Problems

The most important problem in both the Commission's and the Parliament's draft is the merely implicit, but potent definition of anonymisation.

The principle of anonymisation is of outstanding importance to empirical social and economic research because it makes it possible to provide data to researchers for secondary analysis³. So-called scientific-use-files (SUFs) contain less variables and are less

¹ A broad definition of the term 'social research' is applied in the context of this statement.

² Cf. Statement of the German Data Forum on the Current Draft of the EU Data Protection Regulation, 27th August 2014 [*German only*](www.ratswd.de/dl/RatSWD_Stellungnahme_EU_DATA.pdf).

³ Instead of collecting data as part of a research project, secondary analysis draws on existing data. This is not only a way of using (often publicly-financed) research resources economically, but is also crucial for academic discourse. Secondary analysis, or the re-use of research data, is therefore an important aim of science policy and is strongly supported by the German Research Foundation (DFG) and the Federal Ministry of Education and Research (BMBF).

specific than the original files in order to make re-identification highly unlikely. Moreover, SUFs are protected by legal and technical measures i.e. by stipulated guidelines for data processing. Examples of social studies that make scientific-use-files available to the scientific community include ALLBUS, NEPS, PIAAC, and SOEP.⁴ An increasingly important way of accessing sensitive microdata is also based on factual data anonymisation: research data centers provide visiting scholars with on-site access to particularly sensitive data. In this case, the data are primarily protected by organisational and technical measures (alongside legal measures).

In accordance with the German Federal Data Protection Act (in the following: BDSG), the current code of practice for handling empirical social and economic data includes

- (a) Deleting direct identifiers (name, address, etc.) after processing data for non-repeated surveys. If a survey is to be repeated (panel study, and longitudinal approach), the data should be **pseudonymised** i.e. the direct identifiers and the data are stored separately and used only when a repeated survey is carried out.
- (b) **Anonymizing** data for secondary analysis and re-use in research in accordance with BDSG §3 (6) i.e. in a way that re-identification would require **disproportionate efforts** (principle of **factual anonymity** based on the principle of **relative anonymity**).

In view of this long-standing practice, the GDPR is problematic:

- (1) In contrast to the BDSG, which contains a definition of factual anonymisation (§3 (6)) in its terminology section, the GDPR lacks such a definition.
- (2) Instead, GDPR Recital 23 adopts an indirect definition based on the concept of the determinability of a person.

*“(23) To determine whether a person is identifiable, account should be taken of all the means reasonably likely to be used either by the controller or by any other person to identify or single out the individual directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the individual, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, **taking into consideration both available technology at the time of the processing and technological development** (...).” (GDPR, European Parliament draft)*

(3) While the BDSG is based on the principle of *relative* anonymity (“principle of disproportionate means”), the GDPR’s definition of anonymity is not specific enough. Indeed, by mentioning costs and time involved with re-identification, the parliament’s draft applies a relative concept. Its lack of clearness, however, will result in considerable legal uncertainty. It is conceivable that the concept of absolute anonymity will prevail in judicial practice as the formulation in Recital 23 suggests:

⁴ See www.ratswd.de/forschungsdaten/fdz

- Account should be taken of all the means for **„reasonably likely to be used either by the controller or by any other person“**: A narrow interpretation of this formulation has to take into account not only the knowledge of a person working with the data, but the hypothetical knowledge of any other persons.
- **„reasonably likely to be used“**: It is unclear why the adjective ‚reasonable‘, used in this context in the EU data protection regulation as well as the GDPR, is changed from ‚reasonably‘ (Directive) to the significantly more specific phrase **‘reasonably likely to be used‘** (GDPR).
- **„both available technology at the time of the processing and technological development.“**: The requirement that anonymisation should take into account future technologies cannot be met in reality.

Absolutely anonymised data are not suitable for scientific analysis because their information content is insufficient for most research questions (coarsening or removal of key variables). For this reason, the gradual replacement of the notion of absolute anonymity with the notion of relative anonymity in the *BStatG* (Federal Statistics Act), *BDSG* (Federal Data Protection Act) and the *Sozialgesetzbuch* (Social Security Code) led to considerable progress in empirical research in Germany without affecting the data protection needs of the public.

Against this backdrop, the lack of clearness of the definition of anonymity is a problem. Legal uncertainty will severely limit secondary data analysis in Germany.

Secondary analysis is not only an economic way of using (often publicly-financed) research resources, it is also crucial for academic discourse. Especially if results from empirical research form the basis for political decisions, it is important to use different theoretical and methodological approaches when analysing data. Therefore secondary analysis and re-use of research data are essential for science as such. Against the backdrop of the increasing importance of secondary analysis, a narrow interpretation of Recital 23 [absolute anonymity] could render important data sources unusable for researchers and hinder the advancement of knowledge and science. This is all the more serious as access to research data in accordance with data protection has finally been facilitated and established as a cornerstone of scientific research and scientific policy advising in Germany – a process that was initiated and fostered by the German Data Forum.

Finally, we want to point out a further problem: Both the *BDSG* and the *GDPR* contain an explicit clause on research. The respective clause in the *BDSG* refers to data that are explicitly collected for research purposes and stipulates that these data may only be used for research purposes (equivalent to the principle of statistical secrecy which stipulates that statistical data may not be used for other purposes). Such a research clause protects research data containing personal information that deals with, for example, delinquent behavior, from state interference. The *GDPR*’s research clause (Article 83) does not contain such a provision on limitation of purpose.

The public's willingness to take part in scientific surveys and its consent for the use of secondary information is based on mutual trust between researchers and their subjects. The lack of commitment to the principle of purpose limitation of research data and the lack of a clear definition of domains could have a negative effect on research.

Recommendations of the German Data Forum on the EU-GDPR

- (1) The German Data Forum deems it necessary that the GDPR clearly recognizes the notion of relative anonymity.
- (2) The German Data Forum recommends including a clear definition of "anonymity" in the draft's terminology section. The definition included in the Regulation on European Statistics has proven of value. (REGULATION (EC) Nr. 223/2009).

In view of Recital 23, a slightly modified version of the Commission's draft (see below) is preferable to the European Parliament's draft:

To determine whether a person is identifiable, account should be taken of all the means reasonably to be used either by the controller or by any other person to identify or single out the individual directly or indirectly.

The great advantage of this phrasing is that this definition of anonymity has already been tried and tested. Moreover, it is largely similar to the Regulation on European Statistics (Regulation (EC) No. 223/2009)⁵ making it unnecessary to apply different anonymisation regulations to different types of data.

- (3) The German Data Forum recommends giving a clear definition of the concepts of personal, pseudonymous, and anonymous data. GDPR, Article 4 (2a) Parliament draft defines 'pseudonymous data' as follows:

(2a) 'pseudonymous data' means personal data that cannot be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution;

- Since pseudonymised data here are defined as personal data, it is the view of the German Data Forum that the guidelines on processing personal data laid out in Article 5 GDPR also apply to pseudonymised data.

A very narrow interpretation of the GDPR would imply that third-party researchers who use or edit pseudonymised data are obliged to inform the person concerned or obtain their consent if need be.

- Furthermore, this definition of anonymous data is equivocal since anonymous data „cannot be attributed to a specific data subject without the use of additional information“.

⁵ „To determine whether a statistical unit is identifiable, account shall be taken of all relevant means that might reasonably be used by a third party to identify the statistical unit“ REGULATION (EC) Nr. 223/2009; Article 3 (7)

(4) The German Data recommends amending the research clause (Article 83) with a clear limitation of purpose. The sole purpose of research data should be research.