COUNCIL OF EUROPE

CONSEIL DE L'EUROPE

Strasbourg, 7 November 2016 T-PD-BUR(2015)12Rev4

BUREAU OF THE CONSULTATIVE COMMITTEE OF THE CONVENTION FOR THE PROTECTION OF INDIVIDUALS WITH REGARD TO AUTOMATIC PROCESSING OF PERSONAL DATA

(T-PD-BUR)

DRAFT GUIDELINES ON THE PROTECTION OF INDIVIDUALS WITH REGARD TO THE PROCESSING OF PERSONAL DATA IN A WORLD OF BIG DATA[1]

Directorate General of Human Rights and Rule of Law

[1] The draft guidelines were prepared by the expert Alessandro Mantelero, Tenured Aggregate Professor at Politecnico di Torino (Italy).

I. Introduction

Big Data represent a new paradigm in the way in which information is collected and analysed. Big Data - which benefit from the interplay with other technologies such as Internet of Things and cloud computing - are a source of significant value and innovation for society, enhancing productivity, public sector performance, and social participation.

The valuable insights provided by Big Data change the manner in which society can be understood and organised, with a direct impact on individuals and their rights with regard to the automatic processing of personal data. This led the Council of Europe to draft these Guidelines, which provide a general framework for the Parties to devise appropriate policies and measures to make effective the principles and provisions of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (CETS 108) (hereafter Convention 108) in the context of Big Data.

These guidelines have been drafted on the basis of the principles of Convention 108, in the light of its on-going process of modernisation, and are primarily addressed to rule-makers, data controllers and data processors, as defined in section III.

Considering that it is necessary to secure the protection of personal autonomy based on a person's right to control his or her personal data and the processing of such data, the nature of this right to control should be carefully addressed in the Big Data context.

Control requires awareness of the use of personal data and real freedom of choice. These conditions, which are essential to the protection of fundamental rights, and in particular the fundamental right to the protection of personal data, can be met through different legal solutions. These solutions should be tailored according to the given social and technological context, taking into account the lack of knowledge on the part of individuals.

The complexity and obscurity of Big Data applications should therefore prompt rule-makers to consider the notion of control as not circumscribed to mere individual control. They should adopt a broader idea of control over the use of data, according to which individual control evolves in a more complex process of multiple-impact assessment of the risks related to the use of data.

II. Scope

The present Guidelines recommend measures which Parties, data controllers and data processors should take to prevent the potential negative impact of the use of Big Data on human dignity, human rights, and fundamental individual and collective freedoms, mainly with regard to personal data protection.

Given the nature of Big Data and its uses, the application of some of the traditional principles of data processing (e.g. the principle of data minimisation, purpose limitation, fairness and transparency, and free, specific and informed consent) may be challenging in this technological scenario. These guidelines therefore suggest a specific application of the principles of Convention 108, to make them more effective in practice in the Big Data

context.

The purpose of these guidelines is to contribute to the protection of data subjects regarding the processing of personal data in the Big Data context by spelling out the applicable data protection principles and corresponding practices, with a view to limiting the risks for data subjects' rights. These risks mainly concern the potential bias of data analysis, the underestimation of the legal, social and ethical implications of the use of Big Data for decision-making processes, and the marginalisation of an effective and conscious involvement by individuals in these processes.

Given the expanding breadth of Big Data in various sector-specific applications, the present Guidelines provide a general guidance, which may be complemented by further guidance and tailored best practices on the protection of individuals within specific fields of application of Big Data (e.g. healthcare, financial sector).

Nothing in the present Guidelines shall be interpreted as precluding or limiting the provisions of Convention 108 and the safeguards for the data subject recognised by the Convention and by the European Convention on Human Rights.

III. Terminology used for the purpose of these guidelines

a) Big Data: there are many definitions of Big Data, which differ depending on the specific discipline. Most of them focus on the growing technological ability to collect, process and extract new and predictive knowledge from great volume, velocity, and variety of data.[2] In terms of data protection, the main issues do not only concern the volume, velocity, and variety of processed data, but also the analysis of the data using software to extract new and predictive knowledge for decision-making purposes. For the purposes of these guidelines, therefore, the definition of Big Data encompasses both Big Data and Big Data analytics.[3]

b) Data Controller: the natural or legal person, public authority, service, agency or any other body which, alone or jointly with others, has the decision-making power with respect to data processing.

c) Data Processor: a natural or legal person, public authority, service, agency or any other body which processes personal data on behalf of the Data Controller.

d) Parties: the parties that are legally bound by the Convention for the Protection of

[2] The term "Big Data" usually identifies extremely large data sets that may be analysed computationally to extract inferences about data patterns, trends, and correlations. According to the International Telecommunication Union, Big Data are "a paradigm for enabling the collection, storage, management, analysis and visualization, potentially under real-time constraints, of extensive datasets with heterogeneous characteristics" (ITU. 2015. Recommendation Y.3600. Big data – Cloud computing based requirements and capabilities).

[3] This term is used to identify computational technologies that analyse large amounts of data to uncover hidden patterns, trends and correlations. According to the European Union Agency for Network and Information Security, the term Big Data analytics "refers to the whole data management lifecycle of collecting, organizing and analysing data to discover patterns, to infer situations or states, to predict and to understand behaviours" (ENISA. 2015. Privacy by design in big data. An overview of privacy enhancing technologies in the era of big data analytics).

Individuals with regard to Automatic Processing of Personal Data (CETS 108).

e) Personal Data: any information relating to an identified or identifiable individual (data subject).[4]

f) Sensitive Data: special categories of data covered by Article 6 of Convention 108, which require complementary appropriate safeguards when they are processed.[5]

g) Supervisory Authority: the authority established by a Party and responsible for ensuring compliance with the provisions of Convention 108.

IV. Principles and guidelines

1. Ethical and socially aware use of data

1.1 According to the need to balance all interests concerned in the processing of Personal Data, and in particular where information is used for predictive purposes in decision-making processes, Data Controllers and Data Processors should adequately take into account the likely impact of the intended Big Data processing and its broader ethical and social implications to safeguard human right and fundamental freedoms, and ensure the respect for compliance with data protection obligations as set forth by Convention 108.

1.2 Personal Data processing should not be in conflict with the ethical values commonly accepted in the relevant community or communities and cannot prejudice societal interests, values and norms, including the protection of human rights. While defining prescriptive ethical guidance may be problematic, due to the influence of contextual factors, the common guiding ethical values can be found in international charters of human rights and fundamental freedoms, such as the European Convention for the Protection of Human Rights.

1.3 If the assessment of the likely impact of an intended data processing described in section IV.2 highlights a high impact of the use of Big Data on ethical values, data controllers may establish an ad hoc ethical committee, or rely on existing ones, to identify the specific ethical values that shall be safeguarded in the use of data. The ethical committee should be an independent body composed by members selected for their competence, experience and professional qualities. The committee and its members should carry out their duties impartially and objectively.

2. Preventive policies and risk-assessment

2.1 Given the increasing complexity of data processing and the transformative use of Big Data, the Parties should adopt a precautionary approach in regulating data protection in this field.

2.2 Data Controllers should adopt preventive policies concerning the risks of the use of Big

[4] According to this definition, Personal Data are also any information used to single out from data sets people identified to take decisions affecting them on the basis of group profiling information.

[5] According to this definition, Personal Data that do not directly reveal sensitive information, but may provide such information when further processed or combined with other data, are Sensitive Data.

Data and its impact on individuals and society, to ensure the protection of Personal Data and taking into account the rights and freedoms of the data subjects.

2.3 Since the use of Big Data may affect not only individual privacy and data protection, but also the collective dimension of these rights, preventive policies and risk-assessment shall consider the legal, social and ethical impact of the use of Big Data, including with regard to the right to equal treatment and to not be discriminated.

2.4 According to the principles of legitimacy of data processing and quality of data of Convention 108, and in accordance with the obligation to prevent or minimise the impact of data processing on rights and fundamental freedoms of data subjects, a risk-assessment of the potential impact of data processing on fundamental rights and freedoms is necessary to balance the protection of those rights and freedoms against the different interests affected by the use of Big Data.

2.5 Data Controllers should examine the likely impact of the intended data processing on the rights and fundamental freedoms of data subjects in order to:

1) Identify and evaluate the risks of each processing activities involving Big Data and its potential negative outcome on individuals' rights and fundamental freedoms, in particular the right to the protection of Personal Data and the right to non-discrimination, taking into account the social and ethical impacts
2) Develop and provide adequate solutions, such as by-design and by-default solutions,[6] to mitigate these risks
3) Monitor the adoption and the effectiveness of the solutions provided.

2.6 This assessment process shall be carried out by persons with adequate professional qualifications and knowledge to evaluate the different impacts, including the legal, social, ethical and technical dimensions.

2.7 With regard to the use of Big Data which may affect fundamental rights, the Parties should encourage the involvement of the different stakeholders (e.g. individuals or groups potentially affected by the use of Big Data) in this assessment process and in the design of data processing.

2.8 When the use of Big Data may impact on the rights and fundamental freedoms of data subjects, data controllers can consult Supervisory Authorities to receive advice to mitigate the risks referred to in paragraph 2.5.

2.9 Data Controllers shall regularly review the results of the assessment process.

2.10 Data Controllers shall document the assessment and the solutions referred to in paragraph 2.5.

[6] In the context of data protection, the terms "by design" and "by default" refer to appropriate technical and organizational measures taken into account throughout the entire process of data management, from the earliest design stages, to implement legal principles in an effective manner and build data protection safeguards into products and services. According to the "by default" approach to data protection, the measures that safeguard the rights to data protection are the default setting, and they notably ensure that only personal information necessary for a given processing is processed.

2.11 Supervisory Authorities should provide guidance to Data Controllers on risk management in data processing and on the assessment process.

2.12 It is recommended that the measures adopted by data controllers to mitigate the risks referred to in paragraph 2.5 are taken in to account by Supervisory Authorities in the evaluation of any possible sanction, when such a power is foreseen by law.

3. Purpose specification and transparency

3.1 Given the transformative nature of the use of Big Data and in order to comply with the requirement of free, specific, informed and unambiguous consent and the principles of purpose limitation, fairness and transparency, Data Controllers should also identify the potential impact on individuals of the different uses of data and inform data subjects about this impact.

3.2 According to the principle of transparency of data processing, the results of the assessment process described in section IV.2 shall be made publicly available, without prejudice to secrecy safeguarded by law. In the presence of such secrecy, Data Controllers provide any confidential information in a separate annex to the assessment report. This annex shall not be public, but may be accessed by Supervisory Authorities.

3.3 Where the data gathered are further processed for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, they shall be stored in a form that permits identification of the data subjects for no longer than is necessary. In some of these cases, appropriate safeguards may include restriction to access and/or public availability of data where, according to the law, there is no public or individual legitimate interest to access such information.

4. By-design approach

4.1 On the basis of the assessment process described in section IV.2, Data Controllers and, where applicable, Data Processors shall adopt adequate by-design solutions at the different stages of the processing of Big Data.

4.2 Data Controllers and, where applicable, Data Processors carefully consider the design of their data analysis, in order to minimise the presence of redundant or marginal data, avoid potential hidden data biases and mitigate any risk of discrimination or negative impact on the rights and fundamental freedoms of data subjects, in both the collection and analysis stages.

4.3 When it is technically feasible, Data Controllers and, where applicable, Data Processors test the adequacy of the by-design solutions adopted on a limited amount of data by means of simulations, before their use on a larger scale. This would make it possible to assess the potential bias of the use of different parameters in analysing data and provide evidence to minimise the use of information and mitigate the potential negative outcomes identified in the risk-assessment process described in section IV.2.

4.4 Regarding the use of Sensitive Data, by-design solutions shall be adopted to avoid as

much as possible non-sensitive data being used to infer sensitive information and, if so used, to extend the same safeguards to these data as adopted for Sensitive Data.

5. Consent

5.1 The free, specific, informed and unambiguous consent shall be based on the information provided to the data subject according to the principle of transparency of data processing. Given the complexity of the use of Big Data, this information shall be comprehensive of the outcome of the assessment process described in section IV.2 and might also be provided by means of an interface which simulates the effects of the use of data and its potential impact on the data subject, in a learn-from-experience approach.

5.2 When data have been collected on the basis of the data subject's consent, they cannot be processed in a manner incompatible with the initial purposes. Data Controllers and, where applicable, Data Processors shall provide easy and user-friendly technical ways for data subjects to withdraw their consent and to react to data processing incompatible with the initial purposes.

5.3 Personal Data shall be stored for specified and legitimate purposes and not used in a way incompatible with those purposes. Personal data should not be further processed in a way that the data subject might consider unexpected, inappropriate or otherwise objectionable. Data are further processed in an unexpected manner when the use of data exposes data subjects to risks greater, or other than, those contemplated by the initial purposes.

5.4 Consent is not freely given if there is a clear imbalance of power between the data subject and the Data Controllers or Data Processors, which affects the data subject's decisions with regard to the processing. The Data Controller shall bear the burden of proof that this imbalance does not exist or does not affect the consent given by the data subject.

6. Anonymisation

6.1 In the Big Data context, the fact that efforts have been made to anonymise the data should not always exclude the application of the principles of data protection, due to the possible risk of re-identification.

6.2 Anonymisation as a technical measure may be combined with legal or contractual obligations to prevent possible re-identification of the persons concerned.

6.3 On the basis of the risk of re-identification, the Data Controller should demonstrate and document the adequacy of the measures adopted to anonymise data and to ensure the effectiveness of the de-identification. This assessment of the risk of re-identification takes into account both the nature of the data and the costs of implementation of the available anonymising technologies.

6.4 Data controllers shall regularly review the assessment of the risk of re-identification, in the light of the technological development with regard to anonymisation techniques.

7. Role of the human intervention in Big Data-supported decisions

7.1 The use of Big Data should preserve the autonomy of human intervention in the decision-making process.

7.2 Decisions based on the results provided by Big Data analytics should take into account all the circumstances concerning the data and not be based on merely de-contextualised information or data processing results.

7.3 Where decisions based on Big Data might affect individual rights significantly or produce legal effects, a human decision-maker should, upon request of the data subject, provide her or him with detailed motivation.

7.4 On the basis of reasonable arguments, the human decision-maker should be allowed the freedom not to rely on the result of the recommendations provided using Big Data.

7.5 Where there are elements from which it may be presumed that there has been direct or indirect discrimination based on Big Data recommendations, Data Controllers and Data Processors should demonstrate the absence of discrimination.

7.6 The subjects that are affected by a decision based on Big Data have the right to challenge this decision before a competent authority.

8. Open data

8.1 Given the availability of Big Data analytics, public and private entities should carefully consider their open data policies concerning Personal Data.

8.2 When Data Controllers adopt open data policies, the assessment process described in section IV.2 should take into account the effects of merging and mining different data belonging to different open data sets, also in light of the provisions referred to in paragraph 6.

9. Data processing for archiving, research or statistical purposes

9.1 Where the Parties provide specific derogations to the provisions concerning the transparency of data processing and the rights of the data subject with respect to data processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, such derogations should be accompanied by appropriate safeguards.

9.2 Derogations shall be limited to the extent strictly necessary and not be applied unless expressly provided for by the law.

9.3 Derogations cannot prejudice fundamental rights, the principle of non-discrimination, and the right of data subjects to challenge before a competent authority decisions taken on the basis of automated data processing.

10. Education

To help individuals understand the implications of the use of information and Personal Data in the Big Data context, the Parties consider digital literacy as an essential educational skill.