Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data

www.coe.int/data-protection

COUNCIL OF EUROPE
CONSEIL DE L'EUROPE

COUNCIL OF EUROPE

CONSEIL DE L'EUROPE

Strasbourg, 23 January 2017 T-PD(2017)01

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

(T-PD)

GUIDELINES[1] ON THE PROTECTION OF INDIVIDUALS WITH REGARD TO THE PROCESSING OF PERSONAL DATA IN A WORLD OF BIG DATA[2]

Directorate General of Human Rights and Rule of Law

[1] Out of the 50 voting members consulted by written procedure: Denmark, Liechtenstein and Luxembourg abstained, Germany and Ireland objected.

[2] The initial draft and subsequent versions were prepared by 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, combined and analysed. Big Data - which benefit from the interplay with other technological environment such as internet of things and cloud computing - can be 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. Not all data processed in a big data context concern personal data and human interaction but a large spectrum of it does, with a direct impact on individuals and their rights with regard to the processing of personal data.

Furthermore, since Big Data makes it possible to collect and analyse large amounts of data to identify attitude patterns and predict behaviours of groups and communities, the collective dimension of the risks related to the use of data is also to be considered.

This led the Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (CETS 108, hereafter "Convention 108") to draft these Guidelines, which provide a general framework for the Parties to apply appropriate policies and measures to make effective the principles and provisions of 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, controllers and 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 that Parties, controllers and 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, in particular 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 informed 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. health sector, financial sector, public sector such as law enforcement).

Furthermore, in light of the evolution of technologies and their use, the current text of the Guidelines may be revised in the future as deemed necessary by the Committee of Convention 108.

Nothing in the present Guidelines shall be interpreted as precluding or limiting the provisions of Convention 108 and of the European Convention on Human Rights.

III. Terminology used for the purpose of the 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.[3] 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 regarding individuals and groups. For the purposes of these Guidelines, the definition of Big Data therefore encompasses both Big Data and Big Data analytics.[4]

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

d) **Processing**: any operation or set of operations which is performed on personal data, such as the collection, storage, preservation, alteration, retrieval, disclosure, making available, erasure, or destruction of, or the carrying out of logical and/or arithmetical operations on such data.

e) **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.

f) **Open data**: any publicly available information that can be freely used, modified, shared, and reused by anyone for any purpose, according to the conditions of open licenses.

g) **Parties**: the parties that are legally bound by Convention 108.

h) **Personal data**: any information relating to an identified or identifiable individual (data subject).[5]

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

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

[3] 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).

[4] 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).

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

[6] In a big data context, this is particularly relevant for information relating to racial or ethnic origin, political opinions, trade-union membership, religious or other beliefs, health or sexual life revealed by personal data further processed, or combined with other data.

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, controllers and 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 should not 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 on 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, controllers could establish an ad hoc ethics committee, or rely on existing ones, to identify the specific ethical values to be safeguarded in the use of data. The ethics committee should be an independent body composed by members selected for their competence, experience and professional qualities and performing 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 Controllers should adopt preventive policies concerning the risks of the use of Big Data and its impact on individuals and society, to ensure the protection of persons with regard to the processing of personal data.

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 non-discrimination.

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 the 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 with the different interests affected by the use of Big Data.

2.5 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 appropriate measures, such as "by-design" and "by-default" solutions,[7] to mitigate these risks.
3) Monitor the adoption and the effectiveness of the solutions provided.

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

[7] In the context of data protection, the terms "by design" and "by default" refer to appropriate technical and organisational 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.

2.7 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 significantly impact on the rights and fundamental freedoms of data subjects, controllers should consult the supervisory authorities to seek advice to mitigate the risks referred to in paragraph 2.5 and take advantage of available guidance provided by these authorities.

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

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

2.11 The measures adopted by controllers to mitigate the risks referred to in paragraph 2.5 should be taken into account in the evaluation of possible administrative sanctions.

3. Purpose limitation and transparency

3.1 Personal data shall be processed 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. Exposing data subjects to different risks or greater risks than those contemplated by the initial purposes could be considered as a case of further processing of data in an unexpected manner.

3.2 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, controllers should also identify the potential impact on individuals of the different uses of data and inform data subjects about this impact.

3.3 According to the principle of transparency of data processing, the results of the assessment process described in Section IV.2 should be made publicly available, without prejudice to secrecy safeguarded by law. In the presence of such secrecy, controllers provide any confidential information in a separate annex to the assessment report. This annex shall not be public, but may be accessed by the supervisory authorities.

4. By-design approach

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

4.2 Controllers and, where applicable, processors should carefully consider the design of their data processing, in order to minimise the presence of redundant or marginal data, avoid potential hidden data biases and the 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, controllers and, where applicable, processors should 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.

4.5 Pseudonymisation measures, which do not exempt from the application of relevant data protection principles, can reduce the risks to data subjects.

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, controllers and, where applicable, processors shall provide easy and user-friendly technical ways for data subjects to react to data processing incompatible with the initial purposes and withdraw their consent.

5.3 Consent is not freely given if there is a clear imbalance of power between the data subject and the controller, which affects the data subject's decisions with regard to the processing. The controller should demonstrate that this imbalance does not exist or does not affect the consent given by the data subject.

6. Anonymisation

6.1 As long as data enables the identification or re-identification of individuals, the principles of data protection are to be applied.

6.2 The controller should assess the risk of re-identification taking into account the time, effort or resources needed in light of the nature of the data, the context of their use, the available re-identification technologies and related costs. Controllers should demonstrate the adequacy of the measures adopted to anonymise data and to ensure the effectiveness of the de-identification.

6.3 Technical measures may be combined with legal or contractual obligations to prevent possible re-identification of the persons concerned.

6.4 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 the reasoning underlying the processing, including the consequences for the data subject of this reasoning.

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 indications from which it may be presumed that there has been direct or indirect discrimination based on Big Data analysis, controllers and 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 since open data data might be used to extract inferences about individuals and groups.

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. Education

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

Big Data is changing the manner in which the society can be understood. It provides valuable insights and offers opportunities for innovation, enhancing productivity and social participation.

The Guidelines are interested in Big Data involving the processing of personal data and are meant to assist policy makers and organisations processing such data, to ensure that persons are placed at the centre of our digital economies and that their rights and fundamental freedoms are upheld.

The nature of Big Data may impact the application of traditional principles of personal data protection, such as purpose limitation or data minimisation and the Guidelines aim at providing safeguards for the persons concerned. It is for instance crucial to ensure that personal autonomy and the right to control personal data are guaranteed and can be exercised by individuals in a Big Data context.

PREMS 013617

ENG

www.coe.int

The Council of Europe is the continent's leading human rights organisation. It comprises 47 member states, 28 of which are members of the European Union. All Council of Europe member states have signed up to the European Convention on Human Rights, a treaty designed to protect human rights, democracy and the rule of law. The European Court of Human Rights oversees the implementation of the Convention in the member states.

COUNCIL OF EUROPE

CONSEIL DE L'EUROPE