Regulating big data. The guidelines of the Council of Europe in the context of the European data protection framework

Alessandro Mantelero
In January 2017 the Consultative Committee of Convention 108 adopted its Guidelines on the Protection of Individuals with Regard to the Processing of Personal Data in a World of Big Data. These are the first guidelines on data protection provided by an international body which specifically address the issues surrounding big data applications.

This article examines the main provisions of these Guidelines and highlights the approach adopted by the Consultative Committee, which contextualises the traditional principles of data protection in the big data scenario and also takes into account the challenges of the big data paradigm. The discussion of the different provisions adopted focuses primarily on the core of the Guidelines namely the risk assessment procedure. Moreover, the article discusses the novel solutions provided by the Guidelines with regard to the data subject’s informed consent, the by-design approach, anonymization, and the role of the human factor in big data-supported decisions.

This critical analysis of the Guidelines introduces a broader reflection on the divergent approaches of the Council of Europe and the European Union to regulating data processing. Where the principle-based model of the Council of Europe differs from the approach adopted by the EU legislator in the detailed Regulation (EU) 2016/679. In the light of this, the provisions of the Guidelines and their attempt to address the major challenges of the new big data paradigm set the stage for concluding remarks about the most suitable regulatory model to deal with the different issues posed by the development of technology.

* Corresponding address: Department of Management and Production Engineering, Politecnico di Torino, C.so Duca degli Abruzzi, 24, 10120 Torino, Italy. E-mail address: alessandro.mantelero@polito.it.
1. Introduction

In February 2016 the Council of Europe began the process for drafting specific guidelines on the protection of personal information in the big data context, on the basis of Convention 108. After a year-long discussion of the draft proposal by the Parties, the final text of the “Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data” (hereinafter Guidelines) was adopted in January 2017.\(^1\)

These Guidelines concern big data in a broad perspective\(^2\) and their focus is not on the traditional 3Vs paradigm (volume, velocity, and variety), which is used to describe the growing technological ability to collect, process and extract new and predictive knowledge from large amounts of data. As described in the Guidelines, “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”.\(^3\) The Guidelines therefore concern both big data and big data analytics.\(^4\)

---

\(^1\) The author had the privilege to be appointed as consultant expert in drafting the text of the guidelines and to follow the discussion of the proposal by the representatives of the Parties to Convention 108 in the Bureau of the Consultative Committee of Convention 108 and the Plenary Meeting. The final version of the Guidelines, which is the result of this discussion and benefits from the various contributions of all the Parties, was adopted on January 23. The Guidelines were approved by the 50 voting members of the Council, with the abstention of Denmark, Liechtenstein and Luxembourg, while Germany and Ireland objected. The Guidelines are available at <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806ebe7a> accessed 4 February 2017. The drafting of the Guidelines followed a previous theoretical study conducted by Antoinette Rouvroy, see Antoinette Rouvroy, “Of Data and Men”: Fundamental Rights and Liberties in a World of Big Data’ (2016) <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806a6020> accessed 25 October 2016.

\(^2\) The Guidelines do not provide an authoritative definition of big data, since there are many definitions of big data, which differ depending on the specific field of application. 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, ‘Recommendation Y.3600. Big data – Cloud computing based requirements and capabilities’ (2015) <https://www.itu.int/rec/T-REC-Y.3600-201511-I/en > accessed 25 October 2016).


\(^4\) 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 (ENISA), 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”
Against this background, this article does not focus on the procedure that led to the adoption of the Guidelines or on the position of each of the Parties to the Convention during the discussion of the draft proposal, but sets out to highlight the novel approach adopted by the Council of Europe in its document. In this vein, this article critically analyses the main areas and provisions of the Guidelines.

Since the Guidelines are part of the broader European regulatory framework on data protection, the second section takes into account the approach adopted by the EU legislator in the new regulation on data protection and indicates the limits of Regulation (EU) 2016/679 in defining an adequate legal response to big data issues. The main reason for these limits lies in the decision to maintain the traditional paradigm in data protection regulation, which is based on the purpose limitation principle, the notice and consent model, and the individual dimension of data protection.

On the other hand, analysis of the provisions of the Guidelines adopted by the Council of Europe – in section three – points out how a tailored application of this traditional paradigm may produce a different regulatory outcome. In the light of this, the Guidelines take into account not only the individual dimension of data protection, but also its collective dimension, mainly with regard to the ethical and social impact of data uses. Moreover, the provisions of the Guidelines reveal an awareness of the limits affecting the data subject’s consent and encourage a participatory model in assessing the risks related to the use of personal data.

This risk assessment plays a central role in the architecture of the Guidelines and represents the main instrument used by the Consultative Committee of Convention 108 to go beyond the individual dimension of data protection. Potential risks are not restricted to well-known privacy-related prejudice, but also include other kinds of prejudice concerning the conflict between big data applications and ethical and social values.

Finally, big data regulation represents an opportunity to discuss the two different regulatory approaches existing in Europe: the EU detailed regulation on data protection (Regulation (EU) 2016/679) and the principle-based model adopted by the Council of Europe. In this vein, the provisions of the Guidelines and their attempt to address the major challenges of the

---


new big data paradigm provoke some concluding remarks on the most suitable regulatory model to deal with the different issues posed by the development of technology.

2. The data protection framework of the European Union

At global level, the EU model is probably the most successful regulation in the field of data protection, in terms of its influence on third countries’ regulations. Nevertheless, different regulatory approaches are supported by other countries and at international level (APEC, Council of Europe, OECD). Hence, the EU data protection reform could be placed in the broader context of a global “competition” over the international regulation of data protection, which involves different economic areas and organizations. From this perspective, the new Regulation (EU) 2016/679 can be seen as an attempt to reinforce the EU leadership in setting a future global regulatory standard in data protection.

The European Union’s intention to retain its prominent role in the international scenario, achieved by means of Directive 95/46/EC, may be one of the reasons that induced the EU legislator to remain on the path defined in 1995, without reconsidering the main pillars of its model. Another reason can be found in the legislative process used to build a uniform EU data protection framework: Directive 95/46/EC created a convergence that should be reinforced by Regulation (EU) 2016/679, in a manner which is necessarily consistent with the founding principles defined in the Directive.

In light of the above, as it was reasonable to expect, the new Regulation did not change the main elements of the previous regulatory model. Personal data are still primarily protected by individual rights. Moreover, the “notice and consent” model remains an important legal ground for data processing and the principles of purpose limitation and data minimization are reaffirmed.

This mainly conservative attitude of the EU legislator fails to keep pace with developments in technology, which represent one of the key forces that

---


traditionally concur in shaping society and, as a consequence, in influencing law. In this sense, the continuity between the Directive and the Regulation in terms of the regulatory model can be justified from a legislative perspective, but the development of technology often acts independently from legislative strategy and poses new challenges.

More specifically, the technology scenario of the ‘90s is far removed from our daily experience. At that time, data were still collected and processed for a limited range of purposes (e.g. loyalty programs) and the level of complexity of data analysis was still understandable by users. Today, powerful analytics make it possible to extract unpredicted inferences from large amounts of data, which are extensively collected for generic purposes and processed in a manner that is often obscure and usually unintelligible. This necessarily raises the question of whether the data protection paradigm as defined in Directive 95/46/EC – and confirmed in Regulation (EU) 2016/679 – is still adequate to address the challenges of this new scenario.

2.1 The limits of the EU approach in regulating big data: the purpose specification principle

Over the last decade, the concurring effects of the development of different technologies (cloud computing, big data analytics) and changes in the digital environment (sensors, IoT) have created a completely new context. Data are abundant, in some cases excessive and redundant, information is constantly reused, analytics are able to extract predictive information from data sets and algorithms are increasingly becoming part of decision making processes.

All of these changes undermine the foundational idea that data subjects’ self-determination is expressed through their meaningful consent to data processing, because data subjects are not able to cope with the complexity,

---


11 On the relationships between cloud computing and big data, see also ITU (n 2), section 6.3.

12 See Bollier (n 3).

and often obscurity, of today’s information uses.\textsuperscript{14} The big data paradigm also undermines the very notion of “specified purpose”, in terms of the scope of data processing which should be known and defined at the moment of data collection.

The descriptions of the purposes of data collection have become vague or extremely broad. Data are used on the basis of the inferences that arise in an unpredictable manner when analytics mine databases; as a result, the specific uses of data cannot always be known or expected at the moment of data collection.\textsuperscript{15} For this reason, the data subject’s consent, when required, becomes an easy way to legitimise any sort of data processing and further re-use of information.

The manner in which the purpose specification principle is affirmed and its link with the notice and consent model therefore represent the first main limitation in regulating big data on the basis of Regulation (EU) 2016/679. In this sense, the EU legislator seems not to ignore the extreme tension that characterises the application of this principle in the big data context and the weaknesses that affect the data subject’s self-determination.

Legal scholars have tried to suggest alternatives, which, in some cases, have led to a radical revision of the existing framework. For instance, one proposal was to replace the purpose specification principle with a broad notion of legitimate interest,\textsuperscript{16} which, ultimately places the power to assess and balance the different interests in the hands of corporations.

A different perspective expressly accepts the idea that data are collected for multiple purposes, which are only generically defined at the beginning of data processing. This model focuses on the different specific uses of information and the prior assessment of the potential risks of each use.\textsuperscript{17} Consequently, the risk assessment of data uses becomes a way to improve the definition of the purposes of these uses and mitigate their potential negative impact both ethically and socially. This should lead rule-makers to reconsider and reduce the role of the user’s self-determination in data

\textsuperscript{14} See Alessandro Mantelero, ‘The future of consumer data protection in the E.U. Rethinking the “notice and consent” paradigm in the new era of predictive analytics’ in this Review (2014), vol 30, issue 6, 643–660.


\textsuperscript{16} See Lokke Moerel and Corien Prins, ‘Privacy for the homo digitalis. Proposal for a new regulatory framework for data protection in the light of Big Data and the Internet of Things’ (2016) 2, 43-61, 75-77, 86 <https://ssrn.com/abstract=2784123> accessed 20 January 2017 (“the principle of purpose limitation will have to be abandoned as a separate criterion. The time has come to recognize the legitimate interest principle as the principal (and only) test for all the various phases of the life cycle of personal data” and “this approach would result in a regulatory framework within which data collection and processing is permissible, unless it is deemed to be ‘unfair’ or ‘not legitimate’ and therefore unlawful”); Lokke Moerel, ‘Big Data Protection: How to Make the Draft EU Regulation on Data Protection Future Proof’ (Tilburg: Tilburg University, 2014) <http://www.debrauw.com/wp-content/uploads/NEWS%20-%20PUBLICATIONS/Moerel_oratie.pdf> accessed 15 October 2016.

\textsuperscript{17} See Mantelero (n 14) 654-659.
processing, and to reinforce the role of independent authorities (i.e. data protection authorities) in the prior assessment of the risk of data uses.\textsuperscript{18}

Other authors have adopted a similar use-based approach to data protection regulation, but without attributing a prominent role to independent authorities. They suggest a shift away from “rituals” of consent “to deliberate assessment procedures ex ante – not just of the benefits but also the potential risks and harms for individuals associated with a particular data use – and the necessity to devise and implement concrete mitigation strategies”. In this sense Regulation (EU) 2016/679 has been seen as “a stepping stone, pointing towards the need to evolve data protection beyond the old paradigm, yet not fully committed to doing so”.\textsuperscript{19}

These theoretical approaches show the existing tension in the academic arena on how to go beyond the limits of Regulation (EU) 2016/679. It is highly improbable that these suggestions will be taken into account by European legislation in the near future, due to the huge effort invested in reaching a compromise between the different stakeholders on the final text of the new Regulation. Nevertheless, it is important that awareness of these limits becomes evident and that the debate among privacy scholars and experts continues.

### 2.2 The limits of the EU approach in regulating big data: the scope of risk assessment

Despite the decision of the EU legislator to maintain the traditional paradigm as framed in 1995, Regulation (EU) 2016/679 introduces specific provisions which, although not directly focused on big data, have an impact on this context, reinforcing data subjects’ safeguards.

In this sense, the progressive shift of the regulatory focus from individual self-determination to forms of accountability based on risk-assessment\textsuperscript{20}

\textsuperscript{18} ibid 655-659. This would lead to a safer digital environment in which data subjects’ interests are protected by default and, for this reason, their active self-determination is mainly expressed by an opt-out option concerning specific data uses. See also Maria Eduarda Gonçalves, ‘The EU data protection reform and the challenges of big data: remaining uncertainties and ways forward’ (2017) Inf. & Comm. Tech. Law, 18, Published online: 28 Feb 2017, where the author suggests an interesting comparison with the Environmental Impact Assessment (EIA).

\textsuperscript{19} See Viktor Mayer-Schönberger and Yann Padova, ‘Regime Change? Enabling Big Data through Europe’s Data Protection Regulation’ (2016) XVII Colum. Sci. & Tech. L. Rev. 315, 326, 332. On the other hand, on the basis of the legal framework defined by Regulation (EU) 2016/679, the authors suggest a strategy “focused on the exemption of data processing for statistical purposes”, limiting big data applications to statistical purposes. Nevertheless, in this case, using analytics for decision making directly affecting a particular individual “would be outside the meaning of ‘statistical purposes’ “. See also Fred H. Cate and Viktor Mayer-Schönberger, ‘Notice and consent in a world of Big Data’ (2013) 3 (2) International Data Privacy Law 67, 69.

(e.g. data protection impact assessment, prior consultation of data protection authorities, standardization)\(^{21}\) partially reduces the difficulties in regulating big data on the basis of the traditional approach to data protection.\(^{22}\) In this light, the provisions of Regulation (EU) 2016/679 about risk management in data processing represent an important evolution towards a risk-based approach and offer a partial remedy to the potential negative outcomes of the use of big data analytics.

Nevertheless, the main limit of these provisions is the existing relationship between risk assessment and purposes of data processing.\(^{23}\) The criticisms that characterise the application of the purpose limitation principle in the big data context therefore necessarily affect risk assessment.

Indeed, any assessment is related to the use of data for a specific purpose and, according to Regulation (EU) 2016/679, data processing purposes should be "specific, explicit and legitimate", and defined at the moment of

---

\(^{21}\) More specifically, accountability is based on the data protection impact assessment, the role played by data protection officers and, when required by law, the prior assessment process conducted by data protection authorities. In this sense, compared with the previous data protection Directive, Regulation (EU) 2016/679 undoubtedly moves towards a risk-based approach.

\(^{22}\) Nevertheless, this ongoing transition is still incomplete and, with regard to big data analytics, it seems difficult to properly address the social and technological challenges they pose without a complete redraft of the architecture defined in the '90s. See also Alessandro Mantelero, 'Personal data for decisional purposes in the age of analytics: from an individual to a collective dimension of data protection' in this Review (2016) vol 32, issue 2, 238-255.

\(^{23}\) See Article 35 (1) of Regulation (EU) 2016/679 ("Where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons") and 35(7)(b) ("[The assessment shall contain at least] an assessment of the necessity and proportionality of the processing operations in relation to the purposes").
data collection. Nevertheless, this is not consistent with the transformative use of data by private and public bodies through big data analytics.

A further limit concerns the nature of the risk-assessment required by the Regulation. In this regard, the notion of risk adopted in the new Regulation focuses on “material or non-material damages” that prejudice the “rights and freedoms of natural persons”. This is in line with a rights-based approach in risk management, which focuses on rights protection and not on a general trade-off between risks and benefits.

According to this approach, when a risk of prejudice exists and cannot be mitigated or excluded, data processing becomes unlawful, despite the presence of any legitimate grounds, such as the data subject’s consent. In light of this, Recital no. 75 of the Regulation provides a long list of cases where data processing is considered unlawful.

It should be pointed out that this recital does not limit these hypotheses to the security of data processing, but also takes into account the risk of discrimination and “any other significant economic or social disadvantage”. This notion of risk impact, which is echoed in Article 35 of the Regulation, represents an important step in the direction of an impact assessment of data processing that is no longer primarily focused on data security but evolves into a more robust and broader assessment of the different implications of data use.

Attention to the economic and social implications of data use assumes significant relevance in the big data context, where analytics become part of decision-making processes and may have negative impacts on individuals,

24 See Article 5(1)(b) of Regulation (EU) 2016/679.
25 See Recital no. 75 of Regulation (EU) 2016/679.
26 According to the risk/benefit approach, the assessment should be based on the comparison between the importance of benefits and the sum of all risks, without any distinction regarding the nature of risks and benefits. For instance, economic benefits may prevail over individual rights. On the other hand, the risk mitigation approach assumes that some interests (e.g. fundamental rights) are prevailing and cannot be compared with other interests that have a lower relevance. As a consequence, the risk mitigation approach focuses on the potential prejudice to fundamental rights and suggests adequate measures to reduce this risk or, where feasible, to exclude it. On the different classifications of risks related to privacy and data protection, see also David Wright and Charles Raab, ‘Privacy principles, risks and harms’ (2014) 28(3) International Review of Law, Computers & Technology 277-298.
28 See also Article 29 Data protection Working Party (n 20), 15 (“the reference to “the rights and freedoms” of the data subjects primarily concerns the right to privacy but may also involve other fundamental rights such as freedom of speech, freedom of thought, freedom of movement, prohibition of discrimination, right to liberty, conscience and religion”); Article 29 Data protection Working Party, ‘Statement on the role of a risk-based approach in data protection legal frameworks’ (2014) <http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp218_en.pdf>. 
in terms of discrimination\textsuperscript{29} rather than in terms of data security.\textsuperscript{30} However, the provisions of the Regulation do not offer an adequate framework for the assessment of this kind of negative outcome.

The risk-mitigation approach adopted by the Regulation still seems far from the idea of a Privacy, Ethical and Social Impact Assessment (PESIA),\textsuperscript{31} which is a multiple and participative risk-assessment process where the potential negative outcomes of data processing are not only measured in terms of information protection, but also encompass the societal consequences of data uses and their impact on the application of ethical values\textsuperscript{32}.

The lack of this broader perspective represents a limit, since the use of big data analytics in decision-making processes raises important questions regarding the values that should drive the future algorithmic society. Moreover, focusing on the collective dimension, rule-makers should also reflect on the role that the different social stakeholders can play in assessing the societal impacts of data use.\textsuperscript{33}

3. The data protection framework of the Council of Europe

Compared with the European Union, the Council of Europe differs in terms of compositions and nature.\textsuperscript{34} This has direct consequences on the

\footnotesize


\textsuperscript{31} See below section 3.3. Regarding the PESIA model, see also the H2020 project “VIRT-EU: Values and ethics in Innovation for Responsible Technology in Europe” <http://www.virteuproject.eu/> (accessed February 21, 2016). With regard to the ethical assessment in research and innovation, see Clare Shelley-Egan et al. (n 6) 42-44 (“Ethical impact assessment of research and innovation typically considers potential societal harms, risks and implications for fundamental rights, justice, well-being of citizens and the common good. Such assessments may require a consideration of potential impacts on health, the environment, work, leisure, social relations, politics, values, and so on. To achieve this, ethical impact assessment often combines ethical analysis with social impact analysis, futures studies, scenario analysis, and technology assessment. Engagement with stakeholders and public dialogue are other actions within ethical impact assessment, as stakeholders can help to anticipate utilizations and impacts, and can voice their concerns and interests as part of the process of ethics assessment”).

\textsuperscript{32} See below section 3.3.

\textsuperscript{33} See below section 3.4. See also Mantelero (n 21) 249-251.

\textsuperscript{34} In terms of global “competition” regarding data protection standards, Convention 108 represents the only “realistic alternative” to the EU model. See Greenleaf (n 7) 91-92 and
regulatory approach. On the one hand, the European Union is the result of a process of progressive cohesion among various European countries and has an extensive legislation, which has binding nature for member states. On the other hand, the Council of Europe does not pursue the same purpose in terms of economic and political union. The internal cohesion among the members of the Council of Europe is therefore lower than among the EU members.

In regulatory terms this means that the regulations adopted by the Council of Europe are addressed to the states, in line with to the standards of international conventions. This implies a different and weaker binding nature, compared to EU directives and regulations. Moreover, with regard to the existing data protection regulations, there are more differences among the members of the Council of Europe than among EU members.

Finally, unlike the European Union, the Council of Europe takes a global perspective, which goes beyond the European Union’s borders and the European space, since Convention 108 has been ratified by 50 countries among them non-EU European countries, as well as African (Mauritius and Senegal) and South American countries (Uruguay). This different territorial scope of the Convention necessarily affects the manner in which common rules concerning data protection are defined and the content of these provisions.


35 See Greenleaf (n 7) 92 (“Europe’s new enthusiasm to project Convention 108 onto the world stage as a potential global treaty demonstrates the desire for a world order beyond the boundaries of Europe”). Many countries, also including non-members of Council of Europe, have not only ratified Convention 108, but also the Additional Protocol to the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data regarding supervisory authorities and trans-border data flows (2001).
In this sense, the main goal of the Convention is to set a minimum common standard among the Parties. The Convention is therefore a “simple, concise and technologically neutral instrument”. This simple nature and the principle-based approach adopted by the Council of Europe, represent the main distinction between Convention 108 and Regulation (EU) 2016/679, where the latter defines a long and detailed set of provisions.

In this regard, while the adoption of a principles-based approach is necessary for the Council of Europe, due to the lack of an adequate level of harmonization among the Parties to the Convention, for the European Union the adoption of a different model based on detailed provisions represents a choice. In this sense, other countries, such as the US, are proponents of a principles-based regulation.

From this perspective, the main question is whether one of these approaches is better suited than the other to regulating data protection, in a world where technology and society are continually evolving. While this is not the main goal of this analysis, the regulation of big data represents an interesting case study and suggests a number of considerations in this regard, set out in the final section.

Finally, Convention 108 and Regulation (EU) 2016/679 are based on two different social and technological scenarios, separated by more than thirty years, in which the forms of data processing evolved incredibly. In the light of this, a process of “modernisation” of the Convention is ongoing and is in its final stages. Nevertheless, age seems to have a limited impact on the

---

36 See also Explanatory Report to the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, Strasbourg, 28.1.1981, para 20.
38 See The White House (n 8) 47-48.
Convention, since its principles – which represent the backbone of the Directive 95/46/EC and, as a consequence of the new Regulation – are still effective.

3.1 The Guidelines on data protection in a Big Data world

The guidelines adopted in the context of Convention 108 are non-legally binding practical and operative instructions provided by the Council of Europe to the Parties to the Convention. They are primarily addressed to data controllers and data processors and their main purpose is to facilitate an effective application of the principles of the Convention.

Unlike the guidelines previously adopted by the Council of Europe, which deal with specific contexts or issues, these Guidelines focus on the use of a given technology (big data) and are not sector-specific. Moreover, since there are an increasing number of big data analytics applications, from predictive policing to credit scoring, which raise different questions and

---

40 See Wright and Raab (n 26).
41 See Guidelines, Section 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”).
42 See, e.g., Guidelines for prison and probation services regarding radicalisation and violent extremism (Adopted by the Committee of Ministers on 2 March 2016, at the 1249th meeting of the Ministers’ Deputies); Guidelines of the Committee of Ministers of the Council of Europe on child friendly justice (Adopted by the Committee of Ministers on 17 November 2010 at the 1098th meeting of the Ministers’ Deputies).
new issues in terms of data protection, the Guidelines cannot define a set of comprehensive provisions that take into account all the potential implications of big data uses.

For these reasons, the Guidelines are necessarily general, but may be complemented by further guidelines on the protection of individuals in specific fields of application of big data (e.g. healthcare, financial sector). This approach may not be completely satisfactory for operators, as they would like to have specific guidance regarding their field of interest (e.g. use of big data for humanitarian purposes, crime detection and prevention, etc.), but it is not possible to reach this level of detail focusing on big data in general. Detailed guidelines require the analysis of the peculiar uses of big data in a given field, taking into account the different related issues, which are affected by the manner in which analytics are applied in the specific context.

Despite this limit, the Guidelines represent an important step in regulating big data use, as the first international guidance in this field, where significant questions arise with regard to the traditional data protection regulatory paradigm. Nonetheless, the inadequacy of the existing data protection laws (including the recent Regulation (EU) 2016/679) has sometimes made it difficult for the Parties of the Council of Europe to accept the challenges of the novel approach of the Guidelines.

Moreover, the significant role played by the EU Commission and EU member states in the Council of Europe and the intention to ensure consistency between Convention 108, in view of its modernising process, and the new EU data protection framework have frequently led the Parties to a comparison between the solutions proposed in the drafts of the Guidelines and the rules provided by Regulation (EU) 2016/679 or national regulations. This EU-centered approach represents an issue that should be properly addressed, not only in the big data context, but in general with regard to the application of the principles of Convention 108.

Regarding the draft of the Guidelines, an exclusive focus on the EU model would have negatively affected their future application by all the Parties to Convention 108, and limited the effective adoption of the principles of the Convention in those countries where national data protection regulations have not reached the same standard of rigour as the EU Regulation.

45 See Guidelines, Section II (Scope).  
46 See above paras 2, 2.1 and 2.2. See also Gonçalves (n 18).  
47 See also above section 2.
3.2 The scope of the Guidelines

Awareness of the critical issues posed by the new forms of analytics-based data processing\(^{48}\) has characterised the entire drafting of the Guidelines. The starting point was the Preamble of the Draft Modernised Convention 108, which focuses on the protection of “personal autonomy based on a person’s right to control his or her personal data and the processing of such data”\(^ {49}\). This raises the following questions: how is it possible to safeguard this “right to control” in the big data context? Which forms of control are possible and adequate when data collection and analysis are often characterised by complexity and obscurity?

Due to the limits of the traditional approach to data protection in addressing these questions, the Guidelines suggest an application of the principles of Convention 108 that adequately takes into account “the given social and technological context” and “a lack of knowledge on the part of individuals”\(^ {50}\). To reach this goal, the Guidelines do not consider the notion of control as circumscribed by individual control (i.e. notice and consent), but 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”\(^ {51}\).

This points beyond the individual dimension of data protection to take into account its collective dimension\(^ {52}\). This is evident in some of the definitions


\(^{49}\) See Preamble of the Draft Modernised Convention 108 (“Considering that it is necessary to secure the human dignity and protection of the human rights and fundamental freedoms of every individual and [...] personal autonomy based on a person’s right to control of his or her personal data and the processing of such [personal] data”).

\(^{50}\) See Guidelines, Section I (Introduction). See also Section II (Scope) ("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").

\(^{51}\) See Guidelines, Section I (Introduction). See also below section 3.3.

\(^{52}\) See Alessandro Mantelero, ‘From Group Privacy to Collective Privacy: Towards a New Dimension of Privacy and Data Protection in the Big Data Era’ in Linnet Taylor, Luciano Floridi, Bart van der Sloot (eds), Group Privacy: New Challenges of Data Technologies (Springer International Publishing 2017) 139-158; Vedder (n 44).
used in the Guidelines. For instance, according to the Guidelines, personal data are any information relating to an identified or identifiable data individual, but they also encompass "any information used to single out people from data sets, to take decisions affecting them on the basis of group profiling information". This notion of personal data is consistent with the studies on group profiling and collective privacy and the more recent attitude of data gatherers, who are mainly interested in studying groups' behaviour and predicting this behaviour, rather than in profiling single users.

Another traditional aspect affected by the new big data paradigm is the distinction between sensitive and non-sensitive data, since analytics make it possible to infer sensitive data from non-sensitive data. The original draft of the Guidelines specified that "data that do not directly reveal sensitive information, but may provide such information when further processed or combined with other data, are considered sensitive data". The final version of the Guidelines amended this note in favour of a more detailed approach.
traditional comment, but paragraph 4.4 still reveals an awareness of the uncertain border that exists between these two categories of data.

Finally, the Guidelines abandon the traditional approach of data protection regulations, where the provisions are primarily addressed to data controllers. Today, although the processor “processes personal data on behalf of the controller”, there are contexts (e.g. cloud computing) in which users act as data controllers and service providers are processors, despite the fact that the entire architecture of data processing and data security is mainly defined in advance by service providers.

For this reason, even though several provisions of the Guidelines are directed at data controllers, effective adoption of these Guidelines often requires the active role of data processors. They are involved in the adoption of by-design solutions, cooperate with the users to facilitate the exercise of data subjects’ rights, and contribute to avoiding a discriminatory use of data.

This decision for some provisions to be directly addressed to data processors depends on the limitations of the pyramidal task distribution that characterises the data processing regulations based on the EU model. This model maintains the former data processing architecture, adopted in the ‘80s and ‘90s, where data were collected in an autonomous and centralised manner, largely processed in-house and rarely shared with third parties.

The adoption of horizontal and network-based models of organization, the outsourcing of part of the data processing and the sharing of information have increased the number and engagement of data processors in data management processes, with various interactions among different processors. Finally, as mentioned above, cloud computing has created the figure of a processor (i.e. the cloud provider) who has a critical control over the information provided by the data controller (i.e. the user of the cloud

---

60 See Guidelines, footnote no. 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”).

61 See Guidelines, Section IV, para 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”).


63 See Guidelines, Section IV, para 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”). On the anachronistic distinction between data controller and data processor in the Regulation (EU) 2106/679, see de Hert and Papakonstantinou (n 57) 184-185.

64 See Guidelines, Section IV, paras 4.1, 4.2, 4.3.

65 See Guidelines, Section IV, para 5.2.

66 See Guidelines, Section IV, para 7.5.
services). Today, then, both data controllers and data processors are required to put into practice the Guidelines.

3.3 The risk-based approach and the PESIA model

The risk-based approach is often represented as a new way to address the data protection issues concerning technological development. But this is only partially true. While, an express reference to risk management only appears for the first time in Regulation (EU) 2016/679, it should be stressed that risk analysis and risk management have been essential elements of data protection regulations since their origins.

The first data protection regulations were a response by legislators to the growing concern of citizens about the risk of computer-based social control, as the introduction of big mainframe computers gave governments and large corporations the opportunity to collect and manage large amounts of personal information.68

Over the years this risk has changed in nature, assuming new forms. The original concern about government surveillance has been complemented with new concerns regarding the economic exploitation of personal information (risk of unfair or unauthorised uses of personal information) and the information-based decision-making processes (risk of discrimination, large scale social surveillance, bias in predictive analyses).

Moreover, from a regulatory perspective, the interplay between legal scholars and computer scientists that characterised the origin of the first data protection regulations led to the adoption of a procedural approach. In this sense, the focus on data processing procedures represents a form of risk management, based on the regulation of the different stages of data processing (collection, analysis and communication) and the definition of the powers and tasks assigned to the different figures involved in the process.

The focus on the potential risks of data uses is therefore not only present in the provisions of Directive 95/46/EC which concern the security of processing69 or prior checking,70 but also represents a default approach of

68 See Colin J. Bennett, Regulating Privacy: Data Protection and Public Policy in Europe and the United States (Cornell University Press 1992) 29-33, 47; Myron Brenton, The Privacy Invaders (Coward-McCann 1964); Vance Packard, The Naked Society (David McKay 1964); Miller (n 67), chs. 1 and 2. See also Secretary’s Advisory Committee on Automated Personal Data Systems, ‘Records, Computers and the Rights of Citizens’ (1973) <http://epic.org/privacy/hew1973report/> accessed 27 September 2016 (“In more than one opinion survey, worries and anxieties about computers and personal privacy show up in the replies of about one third of those interviewed. More specific concerns are usually voiced by an even larger proportion”). See also Lee A. Bygrave, Data Protection Law. Approaching Its Rationale, Logic and Limits (Kluwer Law International 2002), 107-112.
69 See Directive 95/46/EC, Article 17.
70 See Directive 95/46/EC, Article 20.
data protection regulations. Nevertheless, the attention to these risks has led legislators to adopt provisions that are mainly focused on data security and data quality, without directly addressing the different social and ethical issues of data uses and defining a model to assess the various potential negative outcomes of these issues affecting individuals and society.71

The increasing use of big data analytics in decision-making processes has heightened the importance of investigating aspects regarding relations between individuals and society at large.72 Potential harms are not restricted to the well-known privacy-related risks (e.g. illegitimate use of personal information, data security), but should also include other prejudices, which mainly concern discriminatory or invasive forms of data processing.73 For these reasons, the existing Data Protection Impact Assessment should evolve into a broader and more complex Privacy, Ethical and Social Impact Assessment (PESIA).74

This assessment concerning the compliance of data use with ethical and social values is more complicated than the traditional data protection assessment. Whereas the driving values (e.g. integrity of data) in the context of data security and data management are technologically-based and can therefore be generalised across different social contexts, with social and ethical values the situation is different. These are necessarily context-based and change from one community to another, making it difficult to identify the benchmark that should be adopted in this kind of risk assessment.

This point is clearly addressed in the first section of the Guidelines, which sets out the principles underlying the entire document. Firstly, the section urges both data controllers and data processors to “adequately take into account the likely impact of the intended Big Data processing and its broader ethical and social implications”, in order to safeguard human rights and fundamental freedoms, in the light of Convention 108.75

71 See Guidelines, Section IV, para 2.3 (“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”).


74 Definition of the PESIA model is still in its infancy, see above fn. 31.

75 See Guidelines, Section IV, para 1.1.
Secondly, the Guidelines recognise the relative nature of social and ethical values\textsuperscript{76} and, in this sense, require that data uses are not in conflict with the “ethical values commonly accepted in the relevant community or communities and should not prejudice societal interests, values and norms”.\textsuperscript{77} Although the Guidelines recognise the difficulties of defining the values that should be taken into account in conducting the PESIA, they nevertheless do point out some practical steps to identify these values. They suggest that “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”, following the position of privacy scholars who have examined this issue.\textsuperscript{78}

Given the context-dependent nature of the social and ethical assessment and the fact that international charters may only provide high-level guidance, the Guidelines combine this general suggestion with a more tailored option, represented by “ad hoc ethics committees”.\textsuperscript{79} When the PESIA highlights “a high impact of the use of Big Data on ethical values”, these committees, which in some cases already exist in practice, should identify the specific ethical values to be safeguarded with regard to a given use of data, providing more detailed and context-based guidance for risk assessment.\textsuperscript{80}

In light of the above, the “architecture of values” defined by the Guidelines is based on three layers. The first general level is represented by the “common guiding ethical values” of international charters of human rights and fundamental freedoms. The second layer takes into account the context-dependent nature of the social and ethical assessment and focuses on the values and social interests of given communities. Finally, the third layer consists in a more specific set of ethical values provided by ethics committees and focused on a given use of data.


\textsuperscript{77} Guidelines, Section IV, para 1.2.


\textsuperscript{79} See Guidelines, Section IV, para 1.3 (“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”).

\textsuperscript{80} The same two-layer model, based on general guidelines and tailored guidance provided by \textit{ad hoc} committee, is already adopted in clinical trials. In this field, as the big data context, the specific application of technology poses context-related questions which must necessarily be addressed on the basis of the conflicting interests that characterise each case. This results in an “in the context” assessment of the conflicting interest; see Mantelero (21) 250.
3.4 The risk-assessment process

The Consultative Committee of Convention 108 is aware of the fact that current models of risk assessment in data protection can only partially address the potential outcomes of the use of data, mainly in the light of a broader assessment that encompasses ethical and social issues. In this sense, the Guidelines suggest a precautionary approach to regulating data protection in the field of big data.81

This approach should be adopted with regard to any new application of technology that may produce potential risks for individuals and society, which cannot be exactly calculated or quantified in advance.82 In this sense, the obscurity of big data uses, the uncertainty characterising the applications of data science in the field of analytics and their potentially high impact on certain essential aspects of society may warrant the adoption of this approach as the default setting.83

81 See Guidelines, Section IV, para 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”). On the distinction between the precautionary approach and the precautionary principle, see Jacqueline Peel, ‘Precaution - A Matter of Principle, Approach or Process? (2004) 5(2) Melb. J. Int. Law 483 <http://www.austlii.edu.au/au/journals/MelbJlIntLaw/2004/19.html> accessed 4 February 2017 (“One way of conceptualising what might be meant by precaution as an approach [...] is to say that it authorises or permits regulators to take precautionary measures in certain circumstances, without dictating a particular response in all cases. Rather than a principle creating an obligation to act to address potential harm whenever scientific uncertainty arises, an approach could give regulators greater flexibility to respond”).


These Guidelines require data controllers to adopt “preventive policies” to adequately address and mitigate the potential risks related to the use of big data analytics. This is consistent with the Draft Modernised Convention 108, which focuses on risk analysis and requires that data processing is designed “in such a manner as to prevent or minimise the risk of interference with [...] rights and fundamental freedoms.”

On the other hand, when the level of uncertainty is not high, the risk-assessment process prescribed by the Guidelines represents an important tool to tackle the risks stemming from the use of data. According to the general theory on the risk-based approach, this assessment process is divided into four separate stages: 1) identification of risks, 2) analysis of the potential impact of these risks, 3) selection and adoption of the measures to prevent or mitigate the risks, 4) regular review of the effectiveness of the measures. Moreover, to enable subsequent control of the effective level of compliance, data controllers should document both the risk assessment and the measures adopted.

This is the traditional circular scheme that characterises the risk-assessment, enriched by an explicit reference to by-design and by-default solutions. These two kinds of solutions, which are also mentioned in Regulation (EU) 2016/679, represent a key component of the modern risk-based approach to data protection.

According to the broader perspective adopted by these guidelines, the innovative element in risk-assessment concerns the range of the interests safeguarded and rights protected. In this sense, the assessment encompasses rights that go beyond the traditional sphere of data protection.


84 See Guidelines, Section IV, para 2.2.
85 See Article 8bis (2) of the Draft Modernised Convention 108.
88 See Guidelines, Section IV, para 2.10.
89 See, e.g., CNIL, ‘Privacy Impact Assessment (PIA). Methodology (how to carry out a PIA)’ (n 20).
90 See below para 3.6.
such as the right to non-discrimination,\textsuperscript{92} and it also takes into account the compliance of data processing with social and ethical values.

Given the complexity of this assessment and the various aspects that should be taken into account, it cannot be conducted only by data protection experts, but requires external auditors with specific and multi-disciplinary skills. For this reason, the Guidelines state that the risk 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”.\textsuperscript{93} Moreover, the collective dimension of the potential impact of the use of data leads the Consultative Committee to encourage the involvement of all the relevant stakeholders, giving voice to the different groups of persons potentially affected by the use of data.\textsuperscript{94}

As a result of the increasing use of big data in the decision-making processes and the extensive and pervasive nature of many forms of data processing based on analytics, the Guidelines do not leave the assessment exclusively in the hands of data controllers. In line with the approach adopted in Regulation (EU) 2016/679, when the use of big data “may significantly impact” the rights and fundamental freedoms of data subjects, controllers should consult supervisory authorities to seek advice and to mitigate the risks outlined in the impact assessment.\textsuperscript{95}

The Guidelines recognise the complexity of this assessment and the continuous evolution of both the potential risks and the measures to tackle them. In this regard, data protection authorities may play a significant role in supporting data controllers, informing them about the state-of-the-art of data security measures and providing detailed guidelines on the risk-assessment process.\textsuperscript{96}

Against this background, potential objections to this risk-based approach mainly concerns the heavy burden, in terms of costs and resources, it

\textsuperscript{92} See also Hildebrandt (n 48), 191-195; Solon Barocas, Andrew D. Selbsr, ‘Big Data’s Disparate Impact’ (2016) 104 (3) California Law Review 671-732. See also European Parliament (n 30).

\textsuperscript{93} See Guidelines, Section IV, para 2.6. See also, with regard to the Regulation (EU) 2016/679, Article 29 Data protection Working Party (n 20) (“where appropriate, it is recommended to seek the advice from independent experts of different professions (lawyers, technicians, security experts, sociologists, ethics, etc.)”).

\textsuperscript{94} See Guidelines, Section IV, para 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”). See also David Wright and Paul De Hert, ‘Findings and Recommendations’ in Wright and De Hert (n 72) 467. On ethical values and stakeholder analysis, see Clare Shelley-Egan et al., ‘SATORI Deliverable D2.2 (public version): Views of civil society organisations, government agencies/policymakers and media actors regarding ethics assessment of research and innovation’ (2015) <http://satoriproject.eu/work_packages/dialogue-and-participation/> accessed 13 February 2017; Clare Shelley-Egan et al., ‘Ethical Assessment of Research and Innovation: A Comparative Analysis of Practices and Institutions in the EU and selected other countries. Deliverable D1.1’ (2015) 28-29. <http://satoriproject.eu/work_packages/comparative-analysis-of-ethics-assessment-practices/> accessed 13 February 2017.

\textsuperscript{95} See Guidelines, Section IV, para 2.8.

\textsuperscript{96} See Guidelines, Section IV, para 2.8.
places on the shoulders of data controllers. In this sense the assessment of the impact of data processing with regard to ethical values and societal interests necessarily requires specific skills, the potential engagement of external experts and a significant amount of time for the analysis.

On the other hand, the complexity of risk assessment and risk management is strictly related to the nature of data processing. Thus, where the nature of the data and the type of processing imply low risks to individual and collective rights, risk assessment and risk management are neither time-consuming nor highly costly, while the burden increases proportionally with the severity of the risks.

These considerations suggest a wider observation about the costs of the risk-based approach. As in the field of environmental law, mitigating the impact of technology and business activities on fundamental rights and society has a significant global cost, which at the individual level differs from one entity to another, depending on the type of data processing. Nevertheless, according to a rights-based approach, the fundamental nature of the rights potentially affected by the use of data necessarily subordinate the cost-analysis to the prevailing safeguards of individuals' rights and fundamental freedoms.

This does not mean that costs cannot be an important factor in evaluating the proportionality of the measures taken and in defining risk management strategies, but the total amount of the costs is not per se a sufficient reason for reducing the level of protection afforded to fundamental rights. From this perspective, the protection of individuals and society against the risk related to the use of data should be effective and not affected by economic considerations.

Meanwhile, effective protection requires that data controllers are given sufficient incentive to comply with the legal provisions. To make sure that data protection practices are substantial and not merely formal, 98

---


99 See also Stefano Rodotà, ‘Data Protection as a Fundamental Right’ in Serge Gutwirth, Yves Poulet, Paul De Hert, Cécile de Terwangne and Sjaak Nouwt (eds), Reinventing Data Protection? (Springer 2009) 77-82. With regard to Big Data, see also, Bart van der Sloot. 2016. The individual in the big data era: moving towards an agent-based privacy paradigm. In Bart van der Sloot, Dennis Broeders & Erik Schrijvers (eds). Exploring the Boundaries of Big Data (Amsterdam: Amsterdam University Press) 182-5; Bart van der Sloot. 2016. Do Groups Have a Right to Protect Their Group Interest in Privacy and Should They? Peeling the Onion of Rights and Interests Protected Under Article 8 ECHR. In Linnet Taylor, Luciano Floridi, Bart van der Sloot (eds). Group Privacy: New Challenges of Data Technologies (Cham: Springer International Publishing) 197-224.

100 See also Kenneth A Bamberger and Deirdre K Mulligan, Privacy on the ground: driving corporate behavior in the United States and Europe (MIT Press, 2015).
inducements should be introduced to encourage companies to adopt the multiple risk assessment process required by the Guidelines.

This is why the original text of the Guidelines recommended that Parties “introduce some limitations to the liability of Data Controllers for damage caused by the risks referred to in paragraph 2.5 [i.e. the risk identified during the risk assessment process], when Data Controllers have processed Personal Data according to the provisions of this article”. The rationale of this provision was based on the trade-off between the costs of risk assessment and the liability costs, where the higher cost of the assessment could be compensated by a reduction in the potential costs in case of damages.

Although this approach is already adopted in other fields of law (e.g. corporate liability), the Parties correctly observed that the nature of the Guidelines, which are not a regulatory document, excludes the possibility of derogating to the liability rules laid down by the different national legislators. The final text thus retains the idea of providing incentives, but limits the initial proposal to administrative sanctions: the measures adopted by controllers to mitigate the risks “should be taken into account in the evaluation of possible administrative sanctions”. 101

3.5 The purpose specification and transparency

The most recent data protection regulations, as well as Regulation (EU) 2016/679, are still mainly focused on the traditional purpose specification principle. Nonetheless, in the big data context, the “transformative” use of data often makes it impossible to explain the description of all the possible purposes concerning the use of information at the time of data collection. Big data applications collect a large amount of information from different sources and analyse it to identify new trends and correlations in datasets. The analysis can therefore be conducted to pursue purposes not specified in advance – related to the correlations found in the datasets – and different from the purposes of the initial collection. 103 This often makes it impossible to use data analytics to process information that has been collected for “explicit” and “specified” purposes. 104

The traditional purpose-based approach of Convention 108 combined with the awareness of this limitation led the Consultative Committee of Convention 108 to consider the risk assessment and its transparency as possible ways to shed light on the specific uses of data, increasing the data subject’s awareness about their choices concerning personal information. 105

---

101 See Guidelines, Section IV, para 2.11.
103 See above section 2.1.
104 See article 5.4.b of the Draft Modernised Convention 108.
105 See Guidelines, Section IV, para 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,
In light of this and in line with the principle of transparency of data processing, the Guidelines recommend that the results of the risk-assessment be made publicly available.\textsuperscript{106} This provision is in line with Article 7bis (1) of the Draft Modernised Convention 108, according to which, the controller should inform data subjects of “any necessary additional information in order to ensure fair and transparent processing of the personal data”.

Communication of a summary of the potential effects of data uses, in terms of risk assessment, becomes “necessary additional information” in a context in which the traditional notice clearly shows its limits, due to its lack of clarity and the complexity and obscurity of data processing. This information on risk assessment ensures fairer and more transparent data processing, as well as providing data subjects with valuable input to help them to make their choices. Moreover, in many cases, the description of the risks surrounding data processing offers an indirect representation of the specific purpose of the use of data, since risk is necessarily purpose-related.

Regarding the level of disclosure of the results of the impact assessment, the Guidelines, in line with the suggestions of legal scholars,\textsuperscript{107} specify that the assessment results should be made publicly available “without prejudice to secrecy safeguarded by law”. Where there exits such secrecy, therefore, data controllers should “provide any confidential information in a separate annex to the assessment report”, which is not public, but can be accessed by the supervisory authorities.\textsuperscript{108}

Given the existing relationships between risk awareness and the data subject’s decisions, the Guidelines state that “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”.\textsuperscript{109} This provision takes into account the aforementioned limits in applying the purpose specification principle in the context of the current transformative use of data and recognises the central role played by risk in shedding light on data uses and in better defining the purposes of data processing.

\textsuperscript{106} See Guidelines, Section IV, para 3.3 (“the results of the assessment process described in Section IV.2 should be made publicly available, without prejudice to secrecy safeguarded by law”).

\textsuperscript{107} This is a critical aspect, due to the need to balance the transparency of data processing with security and the competitiveness of enterprises. It is possible to provide business-sensitive information in a separate annex to the impact assessment report, which is not publicly available, or publish a short version of the report without the sensitive contents. See Alessandro Mantelero, ‘Competitive value of data protection: the impact of data protection regulation on online behaviour’ (2013) 3(4) Int'l Data Privacy L. 234; Neil M. Richards and Jonathan H. King, ‘Three Paradoxes of Big Data’ (2013) 66 Stan. L. Rev. Online 41, 43; David Wright, ‘A framework for the ethical impact assessment of information technology’ (2011) 13(3) Ethics Inf Technol. 222. See also Recital no. 51 in the preamble to the Regulation (EU) 2016/679LIBE.

\textsuperscript{108} Guidelines, Section IV, para3.3.

\textsuperscript{109} Guidelines, Section IV, para3.1.
This is in line with the Draft Explanatory Report of the Draft Modernised Convention 108, which requires that the data subject should “be informed of the implications of his or her decision”.\textsuperscript{110} Moreover, voluntary risk-taking is an important aspect of individual autonomy, in terms of “ability to decide for one’s self what risks gives someone control over her own life and the ability to express her individuality through the chances she takes (and those she does not)”.\textsuperscript{111}

Finally, the original draft of the Guidelines provided specific guidance about the case in which collected data are further processed “for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes”,\textsuperscript{112} in order to narrow potential extensive interpretations of the exceptions concerning these purposes laid down by the Draft Modernised Convention 108.\textsuperscript{113}

With regard to data further processed for the aforementioned purposes, the draft of the Guidelines\textsuperscript{114} required that information “shall be stored in a form that permits identification of the data subjects for no longer than is necessary”.\textsuperscript{115} Article 5.4 (b) of the Draft Modernised Convention 108 also stated that appropriate safeguards for this further data processing could include “restriction to access and/or public availability of data”, in the absence of a public or individual legitimate interest to access such information. Nevertheless, these provisions on data re-use were thrown out during the final review of the Guidelines.

3.6 The by-design approach

In recent decades, the by-design approach has become increasingly popular among privacy scholars and rule-makers.\textsuperscript{116} Nevertheless, in order to embed

\begin{itemize}
  \item \textsuperscript{110} See Convention for the protection of individuals with regard to automatic processing of personal data [ETS No. 108], Draft Explanatory Report, prepared on the basis of the consolidated text of the draft modernised Convention 108 (document GR-J(2016)14), para 40.
  \item \textsuperscript{111} Koivisto and Douglas. 2015 (n 88), 10.
  \item \textsuperscript{112} See Draft Modernised Convention 108, Article 5 (4) (b) (“[Personal data undergoing processing shall be] collected for explicit, specified and legitimate purposes and not processed in a way incompatible with those purposes; further processing archiving purposes in the public interest, scientific or historical purposes or statistical purposes is, subject to appropriate safeguards, compatible with those purposes”).
  \item \textsuperscript{113} See Article 9 of the draft of the Modernised Convention and see also Article 9 of Convention 108.
  \item \textsuperscript{115} Guidelines, Section IV, para 3.3.
privacy and data protection values in the design of given devices or services, a preliminary analysis is necessary of the use of personal information and to assess the potential negative outcomes of data processing. On the basis of the risk assessment described above,\textsuperscript{117} suitable by-design measures can be adopted to reduce these outcomes. In this sense, the risk assessment process represents the first step in defining adequate by-design solutions.\textsuperscript{118}

Data protection regulations do not usually provide a specific description of the main aspects of the by-design approach.\textsuperscript{119} At the same time, this approach should be adequately tailored to the specific data processing context. The Guidelines therefore point out the key elements that should be taken into account by the developers of big data applications.

The strict relationship between risk assessment and by-design solutions implies that any change in the nature of the assessment affects the nature of the solutions adopted. Since the Guidelines use a broader notion of risk assessment, which encompasses the social and ethical consequences of the use of data, this multiple impact assessment necessarily leads data controllers and data processors to consider a broader range of by-design solutions to mitigate the additional ethical and social concerns.\textsuperscript{120}

More specifically, with regard to the implementation of the by-design approach in the big data context, the Guidelines focus on four main aspects: data minimisation, potential biases in the design of analytics, pseudonymisation and sensitive data.

The data should be collected and processed in such a way as to “minimise the presence of redundant or marginal data”.\textsuperscript{121} This is a contextual application of the minimization principle, which cannot be automatically applied in the field of big data, where analytics leads data scientists to maximise the collection of available information.

The transformative use of data makes it difficult to reduce the sources of information from the outset. Nonetheless, it is important to adopt a design paradigm that critically assesses the relevance of the collected data and, in light of this, exclude data that are “redundant or marginal”.

\textsuperscript{117} See above paras 3.3 and 3.4.

\textsuperscript{118} See Guidelines, Section IV, para 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”).

\textsuperscript{119} See, e.g., Article 25 of Regulation (EU) 2016/679.

\textsuperscript{120} See Guidelines, Section IV, para 4.1.

\textsuperscript{121} See Guidelines, Section IV, para 4.2.
Moreover, given the specific risks associated with big data use, the by-design approach should also shape data processing to 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”. Given the context-dependent nature of both risks and related by-design solutions, the Guidelines therefore suggest adopting an experienced-based approach.

The Guidelines suggest that “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”. Although analytics can provide different results on the basis of different amounts of data processed, this suggestion is designed to apply the precautionary approach and introduces an intermediate stage between the theoretical definition of the model and its application on a large scale, given the potential negative outcomes that might ensue for individuals and society.

Finally, the Guidelines highlight the role played by pseudonymisation in reducing the potential risks for data subjects and, with regard to the use of sensitive data, point out how by-design solutions can also be used to prevent non-sensitive data being used to infer sensitive information or, if this occurs, to extend the same safeguards to these data to sensitive data.

### 3.7 The data subject’s consent

Article 5.2 of the Draft Modernised Convention 108, in line with the vast majority of data protection regulations, states that data processing can be carried out “on the basis of the free, specific, informed and unambiguous consent of the data subject”. Although the consent-based model would appear to be outdated in the big data context, the Guidelines should be consistent with the framework provided by Convention 108 and its ongoing process of modernisation, which recognises the data subject’s consent as one of the legal bases for data processing.

---

122 See Guidelines, Section IV, para 4.2. See also The White House, Executive Office of the President, (n 29) 7 (which identifies the following technical themes that can cause discriminatory outputs: poorly selected data; incomplete, incorrect, or outdated data; selection bias; unintentional perpetuation and promotion of historical biases).

123 See Guidelines, Section IV, para 2.1.

124 This kind of test on a limited amount of data can offer insights to designers of big data applications with regard to the potential bias and risks of data uses, as well as shedding light on how to reach the same predictive results using a limited number of information sources.

125 See Guidelines, Section IV, para 4.5.

126 See Guidelines, Section IV, para 4.4.

127 See Mantelero (n 14).

128 It should be noted that, in a scenario characterized by complex data processing and concentration of control over information, the decision to retain a model mainly focused on “notice and consent” represents a risk. Indeed, companies can easily notify data
This is in line with the notion of individual’s self-determination, but the criticisms that arise in the big data context, with regard to the purpose specification principle and individual self-determination, led the Consultative Committee of Convention 108 to provide specific guidance regarding the data subject’s consent. In this vein, knowledge of the critical aspects concerning data uses and the related negative outcomes that may affect individuals represents an important factor in increasing data subjects’ awareness about data processing and the consequences of their choices.

For this reason, the Guidelines require that information provided to data subjects at the moment of data collection encompasses the results of the risk-assessment process, which shed light on the manner in which a given data processing affects individuals and society. In this sense, the Guidelines point out the existing relationship between risk assessment, public availability of the assessment results and data subject’s consent.

The Guidelines also add a further by-design suggestion, pointing out that the outcome of the assessment process “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”.

Moreover, since the data subject’s consent can be affected not only by a lack of awareness, but also by a lack of effective freedom of choice, the Guidelines state that 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”.

This condition of imbalance is a debated topic and the information asymmetry about data use that characterises big data applications may be considered per se a source of imbalance. Nonetheless, the Guidelines do not adopt a broad notion of imbalance, but do consider that the data subject’s consent is not freely given only when there is a “clear” imbalance of power. The data controller should demonstrate that this imbalance “does not exist or does not affect the consent given by the data subject”.

Finally, regarding the data subject’s self-determination, the nature of the right to the protection of personal data, which is part of the broader

subjects and require their consent without users having effective self-determination. See above sections 2 and 2.1; see also Solove (n 15), 1899.

129 See Guidelines, Section IV, para 5.1.
130 See above para 3.5.
132 See Guidelines, Section IV, para 5.3.
133 See also Recital no. 43 of the Regulation (EU) 2016/679 (“consent should not provide a valid legal ground for the processing of personal data in a specific case where there is a clear imbalance between the data subject and the controller”).
134 See Guidelines, Section IV, para 5.3.
category of personality rights.\textsuperscript{135} rules out the possibility that individuals may be perpetually affected by their previous decisions concerning personal information. The Guidelines therefore recognise data subjects’ right to withdraw their consent.\textsuperscript{136}

The complexity of data processing and unfriendly digital interfaces often make withdrawal of consent difficult. For this reason, the Guidelines ask controllers and, where applicable, processors to “provide easy and user-friendly technical ways” for data subjects to exercise their right to withdraw consent.\textsuperscript{137}

The same need also concerns the case in which data are processed in a manner incompatible with the initial purpose.\textsuperscript{138} In this case, data processing is unlawful and should be stopped, but data controllers do not necessarily comply with the law. Thus, when data subjects become aware of unlawful uses of data, they have to activate the remedies recognised by national data protection law to block data processing.

Unfortunately the main remedies, which consist in notifying the data controller and an appeal to the data protection authority, are often not easily available to ordinary users, since direct and quick communications with data controllers may be difficult and making a complaint to data protection authorities requires a minimum familiarity with the law.

For these reasons, which become more significant in the complex context of big data applications, the Guidelines require data controllers to actively facilitate the exercise of users’ rights in the case of infringements of the purpose limitation principle, providing them with “easy and user-friendly technical” solutions,\textsuperscript{139} such as a notification button on data controllers’ websites.

### 3.8 Data anonymization

The limits of the anonymization process in the context of big data have been acknowledged by much of the literature, which has highlighted the risks of re-identification by means of analytics.\textsuperscript{140} This suggests that the traditional

\begin{footnotesize}

\textsuperscript{136} See Guidelines, Section IV, para 5.2. See also Article 7(3) of Regulation (EU) 2016/679.

\textsuperscript{137} See Guidelines, Section IV, para 5.2.

\textsuperscript{138} See Guidelines, Section IV, para 5.2.

\textsuperscript{139} See Guidelines, Section IV, para 5.2.

dichotomy between personal data and anonymous data has been superseded.

This black and white distinction should be replaced by a different perspective, which considers personal data and anonymous data as the opposite ends of the same spectrum. The notion of anonymous data is not absolute, but relative and the focus should be on the risk of re-identification.

In light of the above, the Guidelines state that “as long as data enables the identification or re-identification of individuals, the principles of data protection are to be applied”. As a consequence, data controllers can process information as anonymous data (i.e. without applying the safeguards required for personal data) if the risk of re-identification is low, but must also take into account the specific nature of the data processing.

Against this background, the regulatory focus should be on the risk of re-identification rather than on the mere distinction between anonymous and personal data. In this sense, according to the Guidelines, controllers “should assess the risk of re-identification”.

As in all risk-assessment procedures, the assessment is necessarily context-based and technology-dependent. In this sense, the risk of re-identification should be analysed taking into account different factors, such as the time, effort or resources needed to re-identify data, as well as the nature of the data, the context of data use, the available re-identification technologies and their costs. Moreover, given the continual evolution of anonymizing and re-identification technologies, data controllers “shall regularly review the assessment of the risk of re-identification, in the light of the technological development”.


141 See Guidelines, Section IV, para 6.1.

142 See Guidelines, Section IV, para 6.2.

Since technology is not the only solution available to reduce the risk of re-identification, technical measures can be combined with legal or contractual obligations, as pointed out by the Guidelines.\textsuperscript{144} With regard to both these kinds of solutions, it is important to have an adequate level of accountability. Therefore, controllers “should demonstrate the adequacy of the measures adopted to anonymise data and to ensure the effectiveness of de-identification”.\textsuperscript{145}

### 3.9 The role of the human factor in big data-supported decisions

A set of provisions of the Guidelines concerns the role of the human intervention in big data-supported decisions. This is an important topic, which raises further concerns regarding individual freedom, decision-making processes and liability.

The presumed objective nature of algorithms – which clearly underestimates the role played by their human designers – combined with the fact that the decision-maker is often a subordinate of a given organisation raises critical issues with regard to role of the human decision-makers and their freedom of choice.

On the one hand, the supposedly reliable nature of these mathematics-based tools leads those taking decisions on the basis of the results of algorithms to believe the picture of individuals and society that analytics suggest. Moreover, this attitude may be reinforced by the risk of potential sanctions for taking a decision that ignores the results provided by analytics.

On the other hand, the decision maker may recognise the inadequacy of the results or strategies provided by algorithms and decide to act differently. In this case, the individuals affected by the decision should have the right to know the reasons that persuaded the decision maker to disregard the result of data processing.

Against this scenario, the Guidelines state that “the use of Big Data should preserve the autonomy of human intervention in the decision-making process”.\textsuperscript{146} This autonomy also encompasses the freedom of decision-makers not to rely on the recommendations provided by big data applications, but, in this case, “reasonable arguments” should be provided.\textsuperscript{147}

Further provisions concerning the decisions taken on the basis of the results of analytics regard potential bias in the use of data. The Guidelines stress

\textsuperscript{144} See Guidelines, Section IV, para 6.3.
\textsuperscript{146} See Guidelines, Section IV, para 7.1.
\textsuperscript{147} See Guidelines, Section IV, para 7.4.
the fact that the use of information cannot be de-contextualised and “the results provided by Big Data analytics should take into account all the circumstances concerning the data”.\textsuperscript{148} This is an important point in preventing the so-called dictatorship of data, where information is processed without taking into account the specific nature and characters of the original context where data were collected.

Moreover, the complexity and obscurity of many big data applications induced the Consultative Committee of Convention 108 to grant data subjects both the right to know the “reasoning underlying the processing, including the consequences for the data subject of this reasoning”\textsuperscript{149} and the right “to challenge this decision before a competent authority”, when they have been affected by a decision based on big data.\textsuperscript{150}

Finally, given the complex nature of big data processing and in order to reduce the potential bias in the use of big data, the Guidelines suggest a partial reversal of the burden of the proof, when “there are indications from which it may be presumed that there has been direct or indirect discrimination based on Big Data analysis”. In these cases, it is the controllers and processors who should demonstrate the absence of discrimination.\textsuperscript{151}

\section*{3.10 Open data}

There is a clear interplay between data protection and open data, as demonstrated by several opinions and guidelines provided by independent authorities.\textsuperscript{152} In line with the scope of providing practical guidance to data controllers and data processors, the Consultative Committee of Convention 108 directly addressed this issue, instead of merely affirming that open data policies should be consistent with data protection principles.

In line with the pedagogical approach which, in certain respects, should characterise guidelines on new and challenging topics, the Guidelines

\textsuperscript{148} See Guidelines, Section IV, para 7.2.

\textsuperscript{149} See Guidelines, Section IV, para 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”).

\textsuperscript{150} See Guidelines, Section IV, para 7.6.

\textsuperscript{151} See Guidelines, Section IV, para 7.5. This provision is in line with the regulations against discrimination. See, e.g., Directive EU 2000/43/EC, Article 8 (1), and Directive 2000/78/EC, Article 10.

encourage public and private entities to “carefully consider their open data policies concerning personal data”. Here, they should be aware of the fact that datasets made publicly available cannot be considered as isolated silos. On the contrary, big data analytics may be used primarily to merge and mine these open datasets to extract predictive knowledge.

Thus, when data controllers adopt open data policies, the risk-assessment process should take into account “the effects of merging and mining different data belonging to different open data sets”, with regard to the PESIA and the analysis of re-identification risk, which might increase with the merging of information belonging to different open datasets.

3.11 Education

The final provision of the Guidelines focuses on the role of information and digital literacy in the context of the present data-driven society. There is an extensive literature on the differing attitudes of various categories of individuals or segments of the population with regard to data protection. Although this is not the place to discuss the findings of these studies, the global picture that they provide is complex and presents many nuances, in terms of awareness of privacy implications.
In the light of the evolving technology, therefore, helping individuals to “understand the implications of the use of information and Personal Data”\(^{159}\) represents a fundamental task for governments, also with a view to stimulating participative digital citizenship. Moreover, a better understanding of data uses may help to reduce the digital divide affecting various countries or groups.

4. Conclusions

The advent of big data applications poses many different issues with regard to data protection which the Guidelines approved by the Consultative Committee of Convention 108 attempt to answer. The provisions discussed in the previous sections invite a number of brief remarks on the regulatory framework created by the combination of the principles of Convention 108 and the Guidelines. The central point concerns the adequacy of this principle-based approach in providing answers to the questions of evolving technology and a changing society.

To evaluate the strength of this approach, it is useful to compare the regulatory model adopted by the Council of Europe and the model of the European Union.\(^{160}\) The differences are evident: the European Union has adopted technologically neutral rules and an architecture built around detailed provisions, while the framework proposed by the Council of Europe maintains the original model adopted in the ‘70s and ‘80s (e.g. FIPPs, OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data), which is based on key principles, guidelines and \textit{ad hoc} frameworks.

On the one hand, Regulation (EU) 2016/679 sets out to define the regulatory paradigm in data protection for the coming decades and introduces various changes to the existing legal framework. However, the Regulation continues to be based mainly on the principles that inspired the previous Directive and national regulations in the ‘90s, without solving the problems concerning the application of these principles to the new big data scenario. Moreover, the specific nature of the legal instrument (a regulation rather than a directive) led the EU legislator to draft a long and detailed list of provisions which has replaced the principle-based approach of the Directive 95/46/EC.

Meanwhile, Convention 108 is engaged in a similar redrafting process,\(^{161}\) but the differing nature of the Council of Europe has necessarily meant that it has had to take a principle-based approach. But the broader nature of the principles makes it possible to give a specific interpretation of them with regard to given contexts.

\(^{159}\) See Guidelines, Section IV, para 9.
\(^{160}\) See also de Hert and Papakonstantinou. (n 39).
\(^{161}\) See the Draft Modernised Convention 108 (n 39).
In this sense the Guidelines not only apply the principles of Convention 108, but also complete the Convention and benefit from a greater margin of manoeuvre in implementing the provisions originally established in 1981. This has enabled the Council of Europe to adopt more innovative solutions in the Guidelines that Regulation (EU) 2016/679, which does not appear to take the new data processing paradigm and its consequences adequately into account.

Although innovative and complex uses of data necessarily require precise provisions, the legislative process seems unable to react quickly to socio-economic and technological change. The new EU regulation was adopted after lengthy negotiations which went on for four years after the presentation of the proposal of the European Commission and concluded a process started in 2009. Over this period, former “disruptive technologies”, such as big data, IoT, cloud computing, have become part of ordinary life, but do not find an adequate regulatory framework in Regulation (EU) 2016/679.

Such limits of the EU approach should lead us to carefully take into account the different regulatory model adopted by the Council of Europe, which combines a more flexible principle-based framework with specific guidelines. A principle-based regulation is influenced less by technological or social changes and ad hoc guidelines make it possible to contextualize the general provisions in specific technological and social scenarios, addressing new issues. Moreover, guidelines are more easily adopted and revised.

Although the debate on these two different ways of regulating technology is still open and it seems hard to reach a conclusive answer, the case of big data regulation raises many questions about whether Regulation 2016/679 represents the data protection framework for the next twenty years, as assumed by its proponents. This suggests that it may worth reconsidering the entire EU regulatory approach with a view to adopting a more flexible (but not weaker) data protection framework.

Acknowledgments

Part of this article was written during the period that I spent as visiting scholar at the Faculty of Law and Political Science at the Universitat Oberta de Catalunya (UOC) in Barcelona. I am indebted to all my colleagues at UOC and I wish to express my thanks to Ana María Delgado, Miquel Peguera, Raquel Xalabarder and Mònica Vilasau Solana for their kind hospitality. I am grateful to Charles Raab, Lee Bygrave and Viktor Mayer-Schönberger for the comments they provided to the draft version of the Guidelines, which provided me with various suggestions for these critical notes. I would also like to thank the anonymous reviewer for the constructive comments.
References


Bart van der Sloot. 2016. Do Groups Have a Right to Protect Their Group Interest in Privacy and Should They? Peeling the Onion of Rights and Interests Protected Under Article 8 ECHR. In Linnet Taylor, Luciano Floridi, Bart van der Sloot (eds). Group Privacy: New Challenges of Data Technologies (Cham: Springer International Publishing) 197-224.


Bennett, C.J. Regulating Privacy: Data Protection and Public Policy in Europe and the United States (Cornell University Press 1992) 29-33, 47.


Solove, D.J. Understanding Privacy (Harvard University Press 2008)


Mireille Hildebrandt, ‘Profiling: From Data to Knowledge. The challenges of a crucial technology’ (2006) 30(9) Datenschutz und Datensicherheit 550

Mireille Hildebrandt, Smart Technologies and the End(s) of Law : Novel Entanglements of Law and Technology (Edward Elgar Publishing, 2016) 21-61.

Myron Brenton, The Privacy Invaders (Coward-McCann 1964).


Stefano Rodotà, ‘Data Protection as a Fundamental Right’ in Serge Gutwirth, Yves Poullet, Paul De Hert, Cécile de Terwangne and Sjaak Nouwt (eds), Reinventing Data Protection? (Springer 2009) 77-82.


