 DOI: https://doi.org/10.15388/OS.LAW.2020.11

COMBATING DISCRIMINATION THROUGH BIG DATA – FUTURE OF EQUALITY?

Sandra Sakolciová[1], Adam Máčaj[2]

Abstract. The paper aims firstly to assess the future of anti-discrimination measures and policies, above all through the lens of ethnic data utilization. The question posed is not only whether massive collection and usage of such data is viable in relation to its result, but also whether such ethnic data collection is an obligation incumbent upon state authorities, in international and European human rights protection systems in particular. On the other hand, this article aims to compare existence of such obligation with the existing standards on right to privacy and implications for this right stemming from such use of Big Data. The negative impact resulting from such obligation in this regard could weigh heavily on protection of personal data, currently one of the main concerns throughout Europe and the EU.
Keywords: discrimination, ethnicity, Big Data, data protection, positive obligations, privacy.

1. INTRODUCTION

Personal data are at the crossroads nowadays, with data protection, technological development, and right to privacy all playing an essential role in current regulation of the matter. On the other hand, broad data collection can provide society with great insight into patterns of inequalities and direct public authorities to potential root causes and solutions to such persisting schemes. The problem then is whether the stricter approach to regulation of personal data and privacy protection collection should prevail, despite the potential benefits the opposite approach could bring to tackle the endemic discrimination against vulnerable communities.

The aim of this paper is to establish whether collection of sensitive data can be established as an obligation stemming from responsibility of states to protect human rights. The potential conflict between such obligation and protection of privacy is then analysed as well. If the results cannot conclusively support the equality data collection as an obligation of states, the paper seeks to outline the dangers of such practice, and consider the necessary safeguards, should the collection and utilization of ethnic data be contemplated as a good practice.

1 *Master of Law, PhD. candidate at Comenius University in Bratislava, Faculty of Law, Department of International Law and International Relations, sandra.sakolciova@flaw.uniba.sk.*

2 *Master of Law, PhD. candidate at Comenius University in Bratislava, Faculty of Law, Institute of European Law, adam.macaj@flaw.uniba.sk.*

2. COLLECTION OF ETHNIC DATA AS AN ANTI-DISCRIMINATION POLICY

It must be recalled that ethnic data fall under the category of so called „sensitive data". Both the European Union's General Data Protection Regulation,[3] as well as the Council of Europe's Convention 108[4] generally prohibit the collection of sensitive data, including data on ethnicity, race, religion, sexual orientation and similar. Some countries still rely on the assumption that EU law prevents them from collecting such data whatsoever. Both mentioned legal instruments, however, include a list of exceptions which enable collection of sensitive data under specific conditions. One of them is for example processing for reasons of substantial public interest, which could well be tackling discrimination against ethnic groups.[5] It is therefore clear that the prohibition is not absolute and the question we focus on here is rather the existence or non-existence of an obligation to collect ethnic data and its implications or risks, not the possibility or impossibility to collect them.

To properly assess the feasibility, or indeed necessity, of collecting ethnic data for the purpose of tackling discrimination, it is necessary to assess firstly the status of such practice. The two alternatives presented, having regard to effect of such data collection, are either that states have a binding obligation to collect, assess, and utilize ethnic data for anti-discriminatory policies, or such collection is a desirable measure, absence of which however does not infringe positive obligations of state. Such positive obligation would serve to prevent and investigate offences against human rights, in this case, unlawful discrimination. Prevention and investigation would accordingly form a component part of duty to ensure human rights (Kamber, 2017, p. 47). On the other hand, if collection of such data is not considered to be binding, and it did not attain the status of legal obligation, states cannot *de lege lata* be held responsible for omissions or deficiencies in implementation of the ethnic data collection. Therefore, it is essential to assess the abovementioned status, in order to fully appreciate the implications in this area.

2.1. The obligation approach

One of the most prominent actors amongst human rights bodies across Europe, the European Committee of Social Rights ("ECSR"), also offers one of the starkest impetus for assessing collection of ethnic data as an obligation incumbent upon State Parties to the European Social Charter ("ESC"). It required states to produce a specific data on disadvantaged groups, such as children that dropped out of schools, and disaggregate the data on the basis of ethnicity.[6] Drawing in part from lack of evidence on effectiveness of state action supported by reliable statistical data, it considered situa-

3 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

4 Council of Europe, Convention for the Protection of Individuals with Regard to the Automatic Processing of Individual Data, 28 January 1981, ETS 108.

5 Article 9 (2) g) of the GDPR.

6 Such specific request was made by the ECSR regarding Bulgarian school system (see ECSR, Conclusions 2005 – Bulgaria – Article 17(2), 2005/def/BGR/17/2/EN, 30 June 2005).

tion in Bulgaria to be incompatible with right to education, as guaranteed by the ESC. Therefore, the ECSR considered collection of data on basis of ethnicity not as an obligation per se, yet it viewed the practice in conjunction with social rights guaranteed by the ESC as a necessary piece of evidence to further assess anti-discrimination policies. Absent such proof, the ECSR indicated it would not shy from finding a violation of state obligations under international human rights.

Similarly, the ECSR established that combating discrimination through official collection of sensitive ethnic data is not the only measure available to states, and even if official prohibition on such data collection is in effect under national law, the states must find alternative means of assessing the extent of the problem, even if such assessment should be based only on unofficial supporting estimates.[7] Therefore, the ECSR apparently does not merely require states to collect the relevant data to assess effectiveness of its policies. Additionally, it requires states to consider data collected by third parties, non-state bodies, and consider them at least *prima facie* reliable source of evidence to be considered in policymaking.

The rationale behind this approach seems to be one of necessity. It has been argued that such data collection must be produced to picture extent of discrimination, as well as its tackling, specifically in areas where concepts of indirect discrimination are recognized (Simon, 2007, p. 69). Indeed, even the proper implementation of decisions of most prominent human rights bodies, such as the European Court of Human Rights ("ECtHR"), may require states to collect such data, even if they were considered earlier only as *prima facie* evidence of unlawful discrimination.[8] However, the finding of such discrimination by the ECtHR is subject to execution under supervision of CoE's Committee of Ministers ("CM"), under Art. 46(2) of the European Convention on Human Rights ("ECHR"). In the course of supervision in executing a judgment in which discrimination has been already found, the CM strongly urges states condemned to provide additional statistics, over a longer period of time, so that it can properly assess overrepresentation of members belonging to vulnerable communities that have been found to be subjected to discrimination.[9]

The approach adopted in the CM therefore unequivocally requires states to collect pertinent data on ethnicity once it supervises their execution of ECtHR judgments. This obligation therefore crystallizes before the CM only at a later stage of the proceedings, it does not however conclusively contradict the findings of ECSR that such obligation exists even before the discriminatory pattern is established by any ruling. As mentioned by Simon (2007, p. 69), the purpose of collecting these data on ethnicity is, *inter alia*, also to facilitate possible legal proceedings. Therefore, to allow aggrieved parties to lodge their claims, such data need to be available even without intervention by the CM. Accordingly, the view adopted by the ECSR has merit, and the teleological approach indicates collection of data must be approached as a means to prevent and suppress discrimination, not remedy it.

7 See in this regard ECSR, Conclusions XVIII-1 – Czech Republic – Article 1(1), XVIII-1/def/CZE/1/1/EN, 30 October 2006; ECSR, ERRC v. Greece, Decision on the merits, Collective Complaint No. 15/2003, 8 December 2004, para. 28.

8 The appreciation of available data is present e. g. in ECtHR, D. H. and others v. the Czech Republic, judgment [GC], application no. 57325/00, 13 November 2007, para. 191; ECtHR, Horváth and Kiss v. Hungary, judgment, application no. 11146/11, 29 January 2013, para. 128.

9 CM, Notes on the agenda, 1348th meeting, CM/Notes/1348/H46-11, 6 June 2019.

2.2. The good practice approach

This approach is similar to the first approach to data collection as an obligation, in that it views collecting ethnic data as a practice enabling states to properly understand the depths of discriminatory schemes within its territory, and assess effectiveness of policies adopted to eliminate it. Similar recommendation was made e. g. by Parliamentary Assembly of Council of Europe ("PACE"), or European Commission against Racism and Intolerance ("ECRI"), which outlined several areas where data collection is to be utilized, and reiterated necessity of sufficient safeguards against abuse.[10]

Similarly, the collection of disaggregated data is viewed as essential in pursuit of sustainable development, such as 2030 Sustainable Development Goals ("2030 SDGs"), which require accessible, timely and reliable disaggregated data to monitor progress.[11] In fact, states pledged to provide such data already by 2020,[12] and should therefore follow the proposed course of action. Although the 2030 SDGs were adopted by the United Nations General Assembly ("UNGA") and establish therefore only recommendations for states (Öberg, 2006, p. 883), such approach establishes a good practice recognized in the international community.

Under the EU law, status of obligation to collect ethnic data as an anti-discrimination measure is ambiguous. On the one hand, it is argued that collection and analysis of such data is "[t]he most effective and economically viable way to assessing the impact and enforcement of anti-discrimination law and policy" (Farkas, 2017, p. 45). The research indicates that such data collection is at least implied in EU anti-discrimination law and the Racial Equality Directive (Farkas, 2017, p. 46). On the other hand, some voices call for active enforcement of this obligation, even calling for the European Commission to launch infringement proceedings under Art. 258 of the Treaty on the Functioning of the EU ("TFEU") against those states that argue data protection legislation does not permit data collection on the basis of racial and ethnic origin.[13]

This however does little to resolve the dilemma whether EU law follows the principle establishing sensitive data collection as obligation, or it is merely one of possible ways to implement Racial Equality Directive. Arguing that it is an infringement of EU law to misrepresent data protection legislation as prohibiting such data collection does little to clarify whether the actual omission to collect the data violates EU law of itself, or it only arises to the level of infringement under Art. 258 of the TFEU once such omission is complemented by wilful misrepresentation of EU law. In the absence of authori-

10 For example, PACE asked inclusion of Roma and Travelers to be appropriately designed and its impact effectively monitored, in conformity with data protection requirements (PACE, Promoting the Inclusion of Roma and Travellers, Resolution 2153 (2017), 27 January 2017; ECRI, General Policy Recommendation No. 13 on Combating Anti-Gypsyism and Discrimination against Roma, 24 June 2011. See below for a more detailed discussion of the required safeguards.

11 This is outlined in para. 48 of Resolution of the United Nations General Assembly no. A/RES/70/1, adopted on 25 September 2015.

12 Goal 17.18 of the 2030 SDGs.

13 Such approach was proposed by Philip Alston, acting United Nations Human Rights Council Special Rapporteur on extreme poverty and human rights, End-of-mission statement on Romania, 11 November 2015, available online, URL: https://www.ohchr.org/en/newsevents/pages/displaynews.aspx?newsid=16737&langid=e%252523sthash.42v5aeft.dpuf (last accessed 20.5.2020).

tative interpretation, it must be concluded that it yet remains to be seen which of the two outlined approaches shall prevail within the EU. Nevertheless, even if approach takes the view that a binding obligation is indeed established, it remains necessary to analyse the impact on protection of privacy and implications such obligation may bear on the communities concerned.

3. THE POTENTIAL FOR ABUSE AND PROTECTION AFFORDED

Despite the call of prominent human rights bodies including CM, ECSR, ECRI, and NGOs such as ERRC, for collecting ethnic data, many countries still remain reluctant to do so. What are their arguments and grounds for not complying with those recommendations? Are they relevant? In our opinion, the counter-voices shall never be ignored as any large-scale measures usually do carry side-effects and tend to impact the whole or big parts of the population. This part of the paper will therefore elaborate on the potential risks connected with collection of ethnic data on a massive scale. These include potential threats to privacy, data protection and last but not least, risk of abuse of information concerning ethnicity either by private actors or by the states themselves.

3.1. Privacy at risk?

Those calling for collection of ethnic data, as well as those opposing it, refer to potential human rights abuse. The main reason not to collect ethnic data put forward by the opposing states or scholars is usually to prevent violation of the right to privacy and the right to data protection.

Both rights are at stake when it comes to determining people's membership in certain groups or categories according to criteria such as race or ethnicity. These are contested concepts and often it is not distinguished between them at all. In the 19th century, they were understood as biological categories determining individuals' abilities and intelligence. This understanding often served as justification of domination, exploitation and even extermination. Nowadays, these theories are not supported anymore by the majority of scientists and they are considered to be a social construct (Ringelheim, 2006, p. 32-33), rather than an objective fact.[14] Understanding ethnicity as a social construct means that its definition and categories will depend on the society itself. In practice, this leads to use of different classification criteria in different countries. The various methods include self-identification (a method used for census)[15] and classification by a third party according to objective criteria (usually a state on the basis of language, parents' origin place of birth or similar) or even subjective criteria and feelings (classification by members of a community) as is the case of

14 Unfortunately, even this perception has not totally eliminated inequality and exclusion. Even today, discrimination takes either a hidden, indirect form (where data and technology are called for help) or explicit and direct discrimination. The latter just culminated in the United States after the police killed George Floyd and is strongly protested against at the time of writing this article.

15 The level of individual autonomy will depend on whether there is a pre-established list of ethnic groups which can be ticked or a space for writing down the perceived ethnicity freely. Usually there is already a list of options (such data is easier to analyse), however, in a few countries such as Slovakia, Romania or Estonia, only one ethnic origin can be selected in the census (see Farkas, 2017, p. 16).

Indians in the United States who are members of tribes recognized under federal jurisdiction (Ford, 1994, p. 1263).

The question is whether classification of individuals as members of certain groups by third parties is in compliance with the right to privacy and the right to data protection. For example, if someone is identified as Roma based on certain criteria but does not feel like Roma at all, is it alright to treat him as a member of this ethnic group? Being a member of a protected group may influence the way majority sees them and those memberships may carry also burdens, such as discrimination, which we focus on in this article.

Self-determination, individual identity and autonomy are notions which are considered to be at the heart of the right to privacy. They are all interconnected and aim to provide an individual with a power to define their own concept of existence.[16] The ECtHR emphasizes these principles in its established case-law under the Article 8 (Right to privacy) of the ECHR. The notion of personal autonomy is considered to be "an important principle underlying the interpretation of its guarantees, protection is given to the personal sphere of each individual, including the right to establish details of their identity as individual human beings."[17]

Individual autonomy is also an important principle of the right to data protection. Right to data protection was introduced later than the right to privacy and even though they are not the same, they are closely connected. While the European Union introduced the right to data protection in the Charter of Fundamental Rights of the European Union as a separate right in a separate article, the ECtHR considers it to fall under the Article 8 of the ECHR and the right to private and family life. Their scope differs and the ECtHR's approach suggests that the scope of the right to data protection is narrower and could be even treated as a subcategory of the right to privacy. Notwithstanding the exact scope and relationship of those rights, it is important for this paper that both of them are considered to be at danger by collecting ethnic data.

The close connection of the right to privacy and data protection is manifested especially by the common value which they protect and that is the privacy[18] understood as people's privilege to influence the way they are seen and to be able to protect their personal life. In order to achieve this in the era of data and technology boom, they need to possess certain control when it comes to information concerning them. This is understood as an (individual) informational self-determination This concept of "informationelle Selbstbestimmung" was developed by the German Federal Constitutional Court in its landmark Census case[19]. Relying on the principle of human dignity and free development of personality, this court stressed out the importance of individual autonomy which must be safeguarded by the states through enabling individuals to participate in the processing of their data. Despite not being an absolute right, every individual shall be protected against any disproportionate or unreasonable interference with their right to informational self-determination.

16 US Supreme Court, Planned Parenthood of Southeastern Pa. v. Casey, 505 U.S. 833 (1992).

17 ECtHR, Christine Goodwin v. the United Kingdom, judgment, application no. 28957/95), 11 July 2002.

18 Recital 4 of the GDPR.

19 German Federal Constitutional Court, judgment, 15 December 1983, 1 BvR 209, 269, 362, 420, 440, 484/83.

If we look at the methods of determining people's membership in ethnic groups, the self-identification method seems to be the one which is in compliance with these values. On the other hand, classification by third parties does appear to be a rather disproportionate interference with the self-determination principle and individual autonomy as the pillars of both, the right to privacy as well as the right to data protection. ECRI, for example, reiterates in its recommendation documents that classification shall be, if no justification exists to the contrary, always based on voluntary self-identification.[20] Article 3 of the CoE's Framework Convention for the Protection of National Minorities[21] stipulates that every person belonging to a national minority shall have the right freely to choose to be treated or not to be treated as such. It follows that the states cannot determine the membership of individuals in any national minority and it would be contrary to the principle of self-determination, if ethnicity was not to be determined by the individuals themselves.

Even though many countries adopted the self-identification approach when collecting data on race or ethnicity in order to comply with privacy standards (Farkas, 2017, p. 4), this approach is not flawless. Paradoxically – data adopted on such basis may be less workable than other. There are three practical reasons therefor. Firstly, if individuals can identify their ethnicity freely, there can be basically an indefinite number of possibilities and it could be difficult to provide useful statistics (Skerry, 2000, p. 49-55) However, the compromise would be to provide a list of options which would include also a possibility to indicate "Other" if nothing in the list fits. Secondly, if the declaration of ethnicity is not compulsory ("voluntary self-identification" as suggested by ECRI), then the results of data analysis may not be in line with the reality, should people be reluctant to provide this information. This could concern especially groups which are stigmatized in the society (Rallu et al., 2004, p. 505). Again, the reasons for not revealing information, such as ethnicity, is connected with its true sensitivity, especially when there is a fear of abuse of such data. The third reason, why self-identification may challenge usefulness of such data is similar to the previous one. It is less probable, but not excluded, that people may indicate ethnicity which does not correspond with the real life conditions. Let us take an example of someone who lives in a Roma community, speaks the language, his or her parents are Roma, but indicates in the official census that he or she belongs to some other ethnic group or does not belong to any ethnic group. If this person does not get admitted to a school due to an indirectly discriminating measure (against Roma) then this would not be reflected in the statistics. If many people would act in the same way, then the statistics, which could have discovered discrimination, would be spoilt. Similarly, as in the previous case, the motivation not to indicate Roma ethnicity could be fear of abuse or fear of being labelled as Roma and face potential direct discrimination which stems from either individual or general experience.

20 See ECRI, General Recommendation VIII, 1990, A/45/18; ECRI, General Policy Recommendation No. 13 on Combating Anti-Gypsyism and Discrimination against Roma, 24 June 2011

21 Council of Europe, Framework Convention for the Protection of National Minorities, 1 February 1995, ETS 157.

3.2. The "Equality paradox" and other concerns

There are two hypothetical scenarios which may happen if a state starts collecting equality data. If any of those occurs, we can speak of an "equality paradox":

i) People may not be willing to provide information on their ethnicity (due to any of the above mentioned reasons, such as lack of trust and fear) and/or
ii) The provided ethnic data will be misused and will negatively affect members of the ethnic groups (e.g. will be used to directly discriminate against someone).

Equality paradox is a term which may be used to name an event, when an effort to combat discrimination leads to discrimination or at least to fear of discrimination and further stigmatization. There is one factor which always needs to be taken into account and that is the "human factor". Prejudice, dishonesty, misuse, incorrect interpretation, misunderstanding or even malice has always been present in the human race and it will probably always be. A state collects data, but it is people who work with them. And it is not just the human factor in a true sense that could play a role. Algorithms, which are often involved in a decision-making process, could produce biased results, too. Discrimination can occur in their design as well as implementation if the input information is biased.

There is another paradox concerning fighting discrimination through big data. One of the main safeguards of data protection is anonymization. It is a technique which removes personal identifiers from the data. Anonymized data are considered to be safe data and therefore they are not covered by the GDPR as they do not relate to an identified or identifiable natural person.[22] Anonymization could thus present a practical realization of another important equality data safeguard recommended by ECRI, namely confidentiality.[23] However, if a person is not identifiable, it does not mean he or she is not reachable (Barocas & Nissenbaum, 2014, p. 45). With big data, it is possible to infer certain information from a combination of other information. Sensitive data could be often inferred from other, non-sensitive data. The more data is available, the easier it is to link them together. It is even possible to infer sensitive data, such as ethnicity or religion from anonymous data. One real example comes from the US. The New York taxi commission released all taxi trip data, such as location coordinates or number of passengers. Although taxi licenses and medallion numbers had been anonymized, it was still possible to infer personally identifiable information by linking other data. As a result, it was possible (with high accuracy) to infer whether a driver was a Muslim or not, e.g. by finding out if he made a break during time of regular prayers (Kammourieh et al., 2017, p. 61). Anonymization is considered to be a failure also due to successful re-identification attacks. Theoretical and practical studies,[24] as well as real cases show that it is possible to re-identify data subjects by using various de-anonymization methods. One of the well known examples is the Netflix Prize Data Study. The Netflix company wanted to improve its recommendation algorithm and declare a public competi-

22 Recital 26 of the GDPR.

23 ECRI, General Policy Recommendation No. 13 on Combating Anti-Gypsyism and Discrimination against Roma, 24 June 2011.

24 See e.g. Wondracek, 2010.

tion with a reward for the best team. For this purpose, it published millions of records on how its users previously rated movies. University researchers found out that potential attackers would be able to identify the concrete users with a very little information about them available. Specifically, if an attacker knew when approximately a user rated six movies, he would be able to disclose his or her identity in 99% of the cases (Ohm, 2010, pp. 1720-1721). It should be also emphasized that this study is from 2006 and since then, there has been a huge progress in technology.

When it comes to misuse of data, the risk lies also with the states themselves who do or will collect ethnic data. History shows that states are the biggest human rights abusers. During the World War II, population registers played a key role in persecuting Jews and Roma people (Seltzer, 1998, pp. 511-552). Even nowadays, there are concerning incidents and treatment of minorities such as reported acts of police brutality against ethnic communities[25] or pervasive use of technology which includes methods such as profiling (e.g. criminal profiling).[26] The result of such profiling may be that a certain ethnic group is more prone to commit crime in general (or to commit specific crimes) than other group (Gross & Livingston, 2002, p. 1415). Ethnic profiling becomes an incisive technique of control and can be perceived as humiliating by members of a targeted group because they have been scrutinized on a basis of their identity and not on the basis of their individual behaviour (Goldini, 2013, p. 24). It is difficult to foresee or even take precautions against misuse of data by the sovereign states and the strongest safeguard is always the international community itself.

Fortunately, most of the times the ethnic data and population registers are not misused and, as a reliable source of data, serve a good purpose in many fields (Goldini, 2013, p. 30). The states however need to solve many problems connected with collection and meaningful use of equality data. In order for a state to develop a workable monitoring system, data need to be collected not only at the national level, but also on the lower, institutional level. While data collection on the national level provides for information on the percentage of minorities in the general population (obtained for example through population registers), data provided by relevant public or private entities (such as employers or schools) allow for comparing the proportion of the minorities' members present in those entities in order to identify potential underrepresentation, which may indicate discrimination (Ringelheim, 2007, pp. 19-20). What information do the states need from the institutions? Is it safe to store all the collected information about individuals, including the ethnic data, in one place? Should there be more state bodies controlling the whole process or some independent third parties? Does the state have financial sources for such a big scale monitoring? Is the monitoring going to be transparent and would the transparency undermine confidentiality of the data? Is it possible that the routine classifying according to ethnicity will cause even stronger polarisation in the society? These and many more questions will need to be taken into account by the states.

25 Note e.g. ECtHR, Linguar v. Romania, judgment, application no. 48474/14, 16 April 2019.

26 Council of Europe Commissioner for Human Rights, Ethnic Profiling: A persisting practice in Europe, 9 May 2019, available online, URL: https://www.coe.int/en/web/commissioner/-/ethnic-profiling-a-persisting-practice-in-europe (last accessed 25.6.2020).

4. CONCLUSION

Discrimination is still prevalent in our society. In order to fight against it, the most prominent human rights bodies such as ECSR or ECRI call for collection and evaluation of relevant data, including sensitive data. The so called "equality data" should provide useful statistics and enable identification of potential underrepresentation and other forms of discrimination against members of protected groups. Even though ECSR does not consider such collection to be an obligation *per se*, the absence of equality data may lead to violation of international human rights norms. However, such clear approach is not shared by all relevant international bodies, and it is often ambiguous if the states are actually obliged to collect ethnic data or it is only recommended as a good practice.

Despite the obvious benefits that ethnic data may bring to combat indirect discrimination of minorities, there are still concerns of their impact on human rights. The potential for abuse arising from processing of sensitive data has been recognized even by bodies arguing such data collection is mandatory. This article discussed the impact equality data may have on the right to privacy and the right to data protection. Two main problems were identified. The first problem is connected to the notions of individual autonomy and self-determination, which are at the core of both rights and which could be undermined if people's ethnicity would be defined by a third party based on objective criteria. A feasible solution seems to be to apply the method of voluntary self-identification. However, this method has some practical flaws which may spoil the statistics. Secondly, there is a risk of misuse of ethnic data, either by those who have access to the data (especially the state itself), or those who manage to gain access to the data (hacking, re-identification attacks).

Even if the risks do not materialize, people may still fear to provide information on their ethnicity, especially members of ethnic groups which are stigmatized in the society. The term "equality paradox" was introduced to name the situations when the effort of combating discrimination actually leads to discrimination or violation of other human rights. The same applies to cases when there is only a fear of discrimination or persecution, which may have a very bad impact on the ethnic groups, as well as on the whole society due to risk of increased polarisation.

To conclude, it is, without a doubt, important to take positive action and combat discrimination using the potential of the new technologies. Nevertheless, careful scrutiny when it comes to collecting ethnic data, which has already culminated into one of the biggest massacres in history, shall never be underestimated. Any potential risks of harm, even the less serious ones, shall always be carefully analysed. This article highlighted some of them and recommends considering them *a priori* if any ethnic data collection is initiated.

Bibliography

Legal sources

1. Council of Europe, Convention for the Protection of Human Rights and Fundamental Freedoms, 4 November 1950, ETS 005.

2. Council of Europe, Convention for the Protection of Individuals with Regard to the Automatic Processing of Individual Data, 28 January 1981, ETS 108.

3. Council of Europe, Framework Convention for the Protection of National Minorities, 1 February 1995, ETS 157.

4. Council of Europe, European Social Charter (revised), 3 May 1996, ETS 163.

5. European Union, Charter of Fundamental Rights of the European Union, 26 October 2012, 2012/C 326/02.

6. European Union, Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin, OJ L 180.

7. European Union, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ L 119.

8. European Union, Treaty on the Functioning of the European Union (consolidated version), 26 October 2012, OJ C 326.

Literature

9. Barocas, S. et Nissenbaum, H. (2014) 'Big Data's End Run Around Anonymity and Consent', Privacy, Big Data, and the Public Good, Cambridge University Press, pp. 44-75. doi: 10.1017/CBO9781107590205.

10. Farkas, L. (2017) Data collection in the field of ethnicity. Luxembourg: Publications Office of the European Union. doi: 10.2838/447194.

11. Ford, Ch. A. (1994) Administering Identity: The Determination of 'Race' in Race-Conscious Law. California Law Review, Vol. 82, No. 5, pp. 1231-1285. doi: 10.2307/3480910.

12. Goldini, M. (2013) Profiles of Discrimination: A Critical Argument against Racial Profiling. Sortuz. Onati Journal of Emergent Socio-legal Studies, Vol. 5, Issue 1, pp. 19-35. ISSN 1988-0847.

13. Gross, S. et Livingston, D. (2002) Racial Profiling Under Attack. Columbia Law Review, 1413-38. doi: 10.2307/1123676.

14. Kamber, K. (2017) Prosecuting Human Rights Offences: Rethinking the Sword Function of Human Rights Law. Boston: Brill.

15. Kammourieh, L. et al. (2017) Group privacy in the age of big data. Taylor, L. et al. (2017) Group Privacy: New Challenges of Data Technologies. Dordrecht: Springer. ISBN 978-3-319-46608-8.

16. Ohm, P. (2010) Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. UCLA Law Review, Vol. 57, pp. 1701-1776.

17. Öberg, M. D. (2006) 'The legal effects of resolutions of the UN Security Council and General Assembly in the jurisprudence of the ICJ', European Journal of International Law, 16(5), pp. 879–906. doi: 10.1093/ejil/chi151.

18. Rallu, J.-L. et al. (2004) Démographie et Ethnicité : Une Relation Ambigué. Caselli, G. et al. (2004) Démographie : analyse et synthèse. Paris: Institut national d'études démographiques. ISBN 9782733220153.

19. Ringelheim, J. (2007) Processing Data on Racial or Ethnic Origin for Antidiscrimination Policies: How to Reconcile the Promotion of Equality with the Right to Privacy? Center for Human Rights and

20. Global Justice Working Paper No. 13, 2007/Jean Monnet Working Paper 08/06.

21. Seltzer, W. (1998) Population Statistics the Holocaust, and the Nuremberg Trials. Population and Development Review, Vol. 24, No. 3, pp. 511-552.

22. Simon, P. (2007) „Ethnic" statistics and data protection in the Council of Europe countries. Study Report. Strasbourg: Council of Europe.

23. Skerry, P. (2000) Counting on the Census? Race, Group Identity, and the Evasion of Politics. Washington D.C.: Brookings Institution Press. ISBN 978-0815779643.

24. Wondracek, G. et al. (2010) A Practical Attack to De-Anonymize Social Network Users. IEEE Symposium on Security and Privacy, pp. 223-238. doi: 10.1109/SP.2010.21.

Jurisprudence

25. ECSR, ERRC v. Greece, Decision on the merits, Collective Complaint No. 15/2003, 8 December 2004.

26. ECSR, Conclusions 2005 – Bulgaria – Article 17(2), 2005/def/BGR/17/2/EN, 30 June 2005.

27. ECSR, Conclusions XVIII-1 – Czech Republic – Article 1(1), XVIII-1/def/CZE/1/1/EN, 30 October 2006.

28. ECSR, COHRE v. Italy, Decision on the merits, Collective Complaint No. 58/2009, 25 June 2010.

29. ECtHR, Christine Goodwin v. the United Kingdom, judgment, application no. 28957/95), 11 July 2002.

30. ECtHR, D. H. and others v. the Czech Republic, judgment [GC], application no. 57325/00, 13 November 2007.

31. ECtHR, Horváth and Kiss v. Hungary, judgment, application no. 11146/11, 29 January 2013.

32. ECtHR, Linguar v. Romania, judgment, application no. 48474/14, 16 April 2019.

33. German Federal Constitutional Court, judgment, 15 December 1983, 1 BvR 209, 269, 362, 420, 440, 484/83.

34. US Supreme Court, Planned Parenthood of Southeastern Pa. v. Casey, 505 U.S. 833 (1992).

Other sources

35. Council of Europe Commissioner for Human Rights, Ethnic Profiling: A persisting practice in Europe, 9 May 2019, available online, URL: https://www.coe.int/en/web/commissioner/-/ethnic-profiling-a-persisting-practice-in-europe (last accessed 25.6.2020).

36. Council of Europe Committee of Ministers, Notes on the agenda, 1348th meeting, CM/Notes/1348/H46-11, 6 June 2019.European Commission against Racism and Intolerance, General Policy Recommendation No. 13 on Combating Anti-Gypsyism and Discrimination against Roma, 24 June 2011.

37. Parliamentary Assembly of the Council of Europe, Promoting the Inclusion of Roma and Travellers, Resolution 2153 (2017), 27 January 2017.

38. United Nations General Assembly, Resolution no. A/RES/70/1, adopted on 25 September 2015.

39. United Nations Human Rights Council Special Rapporteur on extreme poverty and human rights, End-of-mission statement on Romania, 11 November 2015, available online, URL: https://www.ohchr.org/en/newsevents/pages/displaynews.aspx?newsid=16737&langid=e%252523sthash.42v5aeft.dpuf (last accessed 20.5.2020).