European Commission

Collecting and using equality data in full compliance with EU General Data Protection Regulation and national data protection rules

Subgroup on Equality Data

High-Level Group on Non-discrimination,
Equality and Diversity

2024

This document has been prepared by the Subgroup on Equality Data of the High-Level Group on Non-Discrimination, Equality and Diversity.

The Subgroup on Equality Data is composed of representatives from EU Member States and Norway, the European Commission, Eurostat, Eurofound, EIGE and FRA. Member States representatives are experts from governmental departments dealing with non-discrimination policies, national statistical institutes and equality bodies. It provides a forum for relevant national authorities and EU institutions, to discuss challenges and opportunities in improving the collection and use of equality data.

Manuscript completed in 2024

Luxembourg: Publications Office of the European Union, 2024

PDF ISBN 978-92-68-22581-3 doi:10.2838/9633467 DS-01-24-013-EN-N

Collecting and using equality data in full compliance with EU General Data Protection Regulation and national data protection rules

In the EU, the processing of personal data is protected by EU data protection law and the Charter of Fundamental Rights of the EU– this includes processing of personal data disaggregated by sensitive personal characteristics. However, **data protection rules do not stand in the way of processing equality data for statistical purposes**. On the contrary, the rules allow data to be processed while ensuring the respect of fundamental rights[1].

According to the General Data Protection Regulation (GDPR)[2], the principles of data protection apply to the processing of special categories of personal data that concern an identified or identifiable natural person. By contrast, as detailed in Recital 26 of the Regulation, they **do not apply to the processing of anonymous information or of personal data rendered anonymous** in such a manner that the data subject is not or is no longer identifiable. This implies that it is possible to use information that is anonymous or has been rendered anonymous. Such data includes aggregate statistics produced to identify and record trends in equality.[3]

Personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, and the processing of genetic data, biometric data, data concerning health or data concerning a natural person's sex life or sexual orientation are considered sensitive data or, in the language of the Regulation, **'special categories of data'**[4] **and benefit from a higher level of protection**.

The purpose of this document is to summarize and reflect on the relevant aspects of the GDPR concerning the collection and use of equality data, providing guidance on ensuring compliance with EU data protection regulations and examples of National implementation. The Subgroup on Equality data developed this document based on the previous work produced by the Subgroup over the years. The document incorporates best practices from several Member States (Italy, Spain, Portugal, Croatia, and the Netherlands) as well as experiences shared by the Fundamental Rights Agency (FRA) and Subgroup members during their meetings.

---

[1] See FRA (2021), Equality in the EU 20 years on from the initial implementation of the equality directives, Publications Office, Brussels; European Commission (2021), Round table on Equality Data in September 2021.

[2] 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, 4.5.2016, p. 1–88.**

[3] FRA (2021), *Equality in the EU 20 years on from the initial implementation of the equality directives,* Luxembourg, Publications Office of the European Union (Publications Office), p.21.

[4] Article 9 of the General Data Protection Regulation.

Principles for the processing of personal data

The way personal data are processed has to comply with all **principles relating to the processing of personal data**, as set out in Article 5 of the GDPR: lawfulness, fairness, transparency, purpose limitation, data minimisation, accuracy, storage limitation, integrity, confidentiality and accountability.

The principle of **'data minimisation',** in particular, requires that personal data should be 'adequate, relevant, and limited to what is necessary in relation to the purposes for which they are processed'[5]. This means that:

- data processing should be limited to what is necessary to fulfil a legitimate purpose;
- personal data processing should only take place when its purpose cannot be reasonably fulfilled by other means;
- data processing may not disproportionately interfere with the interests, rights, and freedoms at stake[6].

If it is necessary to process sensitive data, as is often the case for the production of equality statistics, the **more stringent conditions laid down in Article 9** of the GDPR must be met as well. Article 9 of the GDPR has been formulated in such a way that the first paragraph contains a general prohibition on processing of sensitive data. However, this prohibition can be lifted if one of the derogations listed in the second paragraph of Article 9 applies[7], including:

- 9(2)(a): when the **data subject has given explicit consent to the processing** of that personal data for one or more specified purposes, except where EU or national law prevent the data subject from lifting the prohibition referred to in paragraph 1 of Article 9;
- 9(2)(g): when processing is necessary **for reasons of substantial public interest, on the basis of EU or national law** – providing the processing is proportionate to the aim pursued, respects the essence of the right to data protection and provides for suitable and specific measures to safeguard the data subject's fundamental rights and interests;
- 9(2)(j): when processing is necessary for **archiving purposes in the public interest, scientific or historical research purposes or statistical purposes**, in accordance with Article 89(1)[8] **based on EU or national law** - providing the processing is proportionate to the aim pursued, respects the essence of the right to data protection and provides for suitable and specific measures to safeguard the data subject's fundamental rights and the interests.

---

[5] Article 5(1)(c) and Recital (39) of the General Data Protection Regulation.

[6] FRA, Council of Europe (2018 edition), Handbook on European data protection law, Section 3.3. Publications Office, Luxembourg.

[7] For additional guidance on the interpretation of the GDPR see: Guidelines, Recommendations, Best Practices | European Data Protection Board (europa.eu).

[8] Article 89 of the General Data Protection Regulation sets out safeguards and the possibility of EU or national law to provide for derogations from certain data subject's rights, specifically regarding processing for archiving purposes in the public interest; scientific or historical research purposes; or statistical purposes.

Safeguards for data minimisation

It follows that when processing of equality data for scientific research purposes or statistical purposes is based on national law, such law must contain appropriate safeguards to ensure that technical and organisational measures are in place to **ensure respect for the principle of data minimisation**. Such measures may include pseudonymisation[9], provided that the specific purposes of the processing can be fulfilled in that manner. Furthermore, where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner.

However, as highlighted by the legal study commissioned by the European Data Protection Board (EDPB) on the appropriate safeguards under Article 89(1) GDPR for the processing of personal data for scientific research[10], the exact meaning of this article is debated in some Member States. While some argue that Member States shall adopt dedicated legislation specifying any own required safeguards, others claim that, according to this article, it is up to the researchers as controllers to implement safeguards. The study argues that Member States shall adopt national law, detailing the safeguards when personal data is processed for scientific research purposes to ensure the freedoms and the rights of data subjects.

In addition, in its preliminary opinion on data protection and scientific research of 6 January 2020, the EDPS clarifies that: “appropriate safeguards could include conducting a data protection impact assessment of likely risks for rights and freedoms of natural persons , appointing a data protection officer (mandatory in case of a public authority or body, regular and systematic monitoring of data subjects on a large scale, processing on a large scale of special categories of data), notifying a data breach, without undue delay and no later than 72 hours, when the breach is likely to pose a risk to the rights and freedoms of the data subjects, guaranteeing data security, and data minimisation through pseudonymisation or (unless it would impair the research) anonymisation.”

Examples of national practices

The Ombudsman in **Croatia** issued recommendations to the competent public administration bodies and authorities (especially in the fields of internal affairs, health, social welfare, labour, pension insurance and education) to start collecting and processing special categories of personal data, applying adequate protective measures to better understand structural discrimination.

The Croatian Public Health Institute established cooperation with the Ministry of Internal Affairs to improve monitoring of the health (and the social determinants for it) of members of the Roma national minority. This follows a positive opinion from the Personal Data

---

[9] ‘Pseudonymisation’ means processing personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure the personal data are not attributed to an identified or identifiable person.

Source: https://data.consilium.europa.eu/doc/document/ST-5419-2016-INIT/en/pdf p. 112.

[10] The study can be found here: Legal study on the appropriate safeguards under Article 89(1) GDPR for the processing of personal data for scientific research | European Data Protection Board (europa.eu)

Protection Agency on collecting data on national and ethnic affiliation in compliance with the GDPR. The Ministry of Interior shared the national affiliation data with the Croatian Public Health Institute, who then connected the personal identification number of the persons registered as Roma with the health data indicators in the database of a Croatian Public Health Institute. The measure was part of the activities undertaken in align with National Roma Inclusion Strategy 2013-2020.

The update of the measure is included in the Action Plan on implementing the National Plan for Roma Inclusion 2021-2027, for 2024-2025. In preparation for updated activity, the Croatian Public Health Institute together with Office for Human Rights and Rights of the National Minorities, are planning the following actions:

- informing the potential respondents of the second round of the "Baseline Data Study"[11] about potential usage of its data in the thematic research about its scope and on the necessity for Roma health;
- informing the respondents on importance of connecting the questionnaire data from "Baseline Data Study" with the personal identification number (that will allow the connection of the data from the existing database; and with the involvement of the Croatian Public Health Institute);
- ensuring the informed consent not only for participation in the fieldwork of both studies ("Baseline Data Study" and "Roma health" thematic research but also together with gathering the personal identification number of the respondents of the "Baseline Data Study";
- ensuring the informed consent for the usage of the individual data from the Croatian Public Health Institute connected to personal identification number, and combined with the personal data gathered through second round of "Baseline Data Study".

Statistics **Portugal** worked with the Data Protection Officer to ensure all GDPR requirements are met for the Survey on Living Conditions, Origins and Trajectories of the Resident Population in Portugal and a Data Protection Impact Assessment was carried out as well. For the questionnaire, adjustments have been made, such as:

- making the survey voluntary, unlike in all other household surveys, where it is mandatory;
- respondents and those who live in the household had the option not to give their names;
- interviewers submitted the completed interviews to the centralised data warehouse immediately, so that personal information from interviewees would be deleted from their work computers;
- names and addresses were excluded from the final database available for internal analysis purposes;
- data flow within Statistics Portugal computer systems was duly assessed, to address all possible security concerns.

---

[11] See: https://fra.europa.eu/en/promising-practices/roma-inclusion-croatian-society-baseline-data-study

In the **Netherlands,** Statistics Netherlands conducts a Barometer on Cultural Diversity within organisations. Because the privacy of individual employees is important, Statistics Netherlands analyses workforce data supplied by employers anonymously and only produces results at an aggregate level. There is no need to collect and present data on individual employees. All data are anonymous and aggregated at group level. If the size of the group is too small, the statistical office will not publish these data.

In **Spain,** a debate was launched between the National Statistics Institute, the Data Protection Agency, the Equality Body, the Roma Council, the EU Fundamental Rights Agency and more than 40 civil society organizations. This led to a concrete proposal[12] in 2023 to introduce a question on ethnic origin in statistics, studies and surveys in Spain, in full compliance with the GDPR.

The Spanish National Statistics Institute (INE) has announced it is going to include an ethnic origin question in the 2026 Survey of Essential Population and Housing Characteristics (ECEPOV), based on the principles of self-identification, no harm, anonymity and consent. The ECEPOV is designed with the main objective of providing detailed information on people, housing and buildings and would allow cross-referencing information on access to housing and its characteristics, nationality, place of birth, level of studies, knowledge of languages, employment situation, contact with new technologies and health. The survey will be carried out in full respect of the principles and safeguards of the GDPR, specifically:

- the data will not be published or made accessible to the public in a way that allows the data subjects to be identified, either directly or indirectly;
- each data subject will give his or her explicit consent in answering the survey;
- the data protection agency was involved in the whole process and all their recommendations were included.

In **Italy**, the National Statistical Programme (NSP) is the National Statistical System (SISTAN)'s tool for planning statistical activity of public interest. It defines statistical surveys, data processing and planning studies.

It requires a compilation of a document containing the objectives of the survey, data collection methods, main variables collected, legislative references, sensitive data and personal data, any starting lists used (e.g. full name, address, sex, date of birth of people in civil union), source and owner of the list, variables included, data dissemination. The Istat-Unar (Italian Office for the Promotion of Equal Treatment and the Elimination of Discrimination based on Racial or Ethnic Origin) project "Labour Discrimination against LGBT+ people and diversity policies in enterprises" (2018-2023) was inserted into the National Statistical Programme (NSP) and approved by the National Authority for Data Protection. The Italian National Statistical Institute (Istat) had a consultation with the Data Protection Authority to clarify some aspects of the project, after that the project was approved.

---

[12] https://urldefense.com/v3/__https://www.igualdad.gob.es/wp-content/uploads/EN_Proposal-consensus-introducing-an-ethnic-origin-question.pdf__;!!DOxrgLBm!FnTn9oNsrgECX_qx2iN73r0pocbwbDgZ8KrZWvkWimw37bMsfODYCn547WS2P5i8DWsXF6w5q34G7-mzMVlUym5n6_hotOtACcrKxA$

Three Computer Assisted Web Interviewing (CAWI) surveys were carried out targeted to LGBT+ people, based on respondents' self-identification.

A first survey of the project addressed all the people who chose Civil Union as an official recognition of their relationship and was carried out by Istat in 2020-2021[13]. Istat invited all people in civil union to complete the survey, using a list coming from administrative data because the civil status is public and it does not identify a specific sexual orientation. The explicit consent of the respondent to participate in the survey was asked.

A second survey - "Survey on Labour Discrimination against LGBT+ people (not in civil union)" - was carried out in 2022 (January-May)[14] and was addressed to LGB people who have never been in civil union. Istat implemented for the first time the snowball technique Respondent Driving Sample (RDS) in its web version (WebRD). It requires respondents to play an active role in recruiting new respondents who belong to their network of relationships. About fifty LGBT+ association agreed in facilitating the survey and, after signing an agreement with Istat on privacy protection, were tasked to identify the "seeds" of the network chain. Moreover, explicit consent to participate in the survey was needed from the respondent and a two-step model was adopted (in the first web-step the respondent is informed on the survey's goals and in the second, after giving the privacy consent, is directed to a web-page of the questionnaire). Participation was voluntary.

A data protection impact assessment (DPIA) on respondents was developed for this survey. It included: the risk analysis on the potential undermining of privacy and the identification of appropriate technical procedures to avoid such privacy breaches.

Istat took steps to implement the second edition of the Survey on Discrimination, with the Pilot survey planned for 2019. However, in 2018, the National Data Protection Authority pointed out some potential data protection risks for respondents and requested more information and guarantees from Istat, who suspended the survey. The drafting of the DPIA represents a standard practice for Istat nowadays but this was not the case in 2018. A privacy impact plan on respondents was developed by a dedicated task force for the Pilot survey, including the risk analysis on the potential undermining of privacy and the identification of appropriate technical and organisational measures for data security (based on the principles of integrity and confidentiality).

A favourable opinion on the implementation of the Pilot Survey on Discrimination was issued by the Data Protection Authority in December 2020.

The Pilot Survey on Discrimination was carried out in 2022-2023 to test the revised questionnaire and the CAWI technique (which had not been adopted in the first edition of the survey).

The Pilot sample survey on discrimination was designed to study the population's opinions and stereotypes on discrimination and to identify incidents of discrimination that they may have experienced. The main considered grounds of discrimination within the questionnaire are based on gender, health conditions, religious beliefs, being foreign or of foreign origin,

---

[13] https://www.istat.it/en/archivio/270626

[14] https://www.istat.it/it/files//2023/05/report-discriminazioni-15maggio.pdf

sexual orientation and gender identity. The survey covers many areas of daily life (looking for a job, working, studying, accessing various public services and other areas of daily life).

Getting in touch with the EU

In person

All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online (european-union.europa.eu/contact-eu/meet-us_en).

On the phone or in writing

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696,
- via the following form: european-union.europa.eu/contact-eu/write-us_en.

Finding information about the EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website (european-union.europa.eu).

EU publications

You can view or order EU publications at op.europa.eu/en/publications. Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre (european-union.europa.eu/contact-eu/meet-us_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex (eur-lex.europa.eu).

EU open data

The portal data.europa.eu provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.

Publications Office
of the European Union