



High Level Group on Non-discrimination, Equality and Diversity

Subgroup on Equality Data

Guidance note on the collection and use of data for LGBTIQ equality

EUROPEAN COMMISSION

Directorate-General for Justice and Consumers
Directorate D –Equality and Non-Discrimination
Unit D1 - Non-discrimination: LGBTIQ, Age, Horizontal Matters
European Commission
B-1049 Brussels

Guidance note on the collection and use of data for LGBTIQ equality

This document has been produced 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.

The Subgroup on Equality Data would like to express their gratitude to the EU Agency for Fundamental Rights (FRA) and two external experts, Joz Motmans and Aisa Burgwal, for their extensive input in the development of this guidance note, drawing on their wide-ranging knowledge and expertise.

Manuscript completed in March 2023

PDF ISBN 978-92-68-05054-5 doi:10.2838/398439 DS-03-23-327-EN-N

Luxembourg: Publications Office of the European Union, 2023

© European Union, 2023

The reuse policy of European Commission documents is implemented by the Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Except otherwise noted, the reuse of this document is authorised under a Creative Commons Attribution 4.0 International (CC-BY 4.0) licence (<https://creativecommons.org/licenses/by/4.0/>). This means that reuse is allowed provided appropriate credit is given and any changes are indicated.

Table of contents

Table of contents	5
1. Background to – and purpose of – this Guidance	7
1.1. Need for data for informed policy choices	8
1.2. What are ‘equality data’?	9
1.3. Purpose and structure of this guidance note.....	12
2. Challenges, data protection, and other ethical considerations in collecting and using equality data on SOGIESC	16
2.1. Challenges in collecting and using SOGIESC data.....	16
2.2. Data-protection rules do not prevent the collection of equality data on SOGIESC	21
2.3. Ethical considerations in relation to data on SOGIESC	23
3. General principles for collecting equality data on SOGIESC	27
3.1. Carry out needs assessments of users and potential users of data disaggregated by SOGIESC	27
3.1.1. Rationale	27
3.1.2. Guidance	27
3.2. Identify existing data sources that provide information on SOGIESC	28
3.2.1. Rationale	28
3.2.2. Guidance	29
3.3. Align definitions and categorisations related to SOGIESC.....	31
3.3.1. Rationale	31
3.3.2. Guidance	31
3.4. Mainstream information on SOGIESC in all national and EU surveys	33
3.4.1. Rationale	34
3.4.2. Guidance	41
3.5. Mainstream an intersectional approach in the collection and use of SOGIESC data	43
3.5.1. Rationale	45
3.5.2. Guidance	45
3.6. Provide sufficient budget for regular surveys to collect equality data on SOGIESC	46
3.6.1. Rationale	46
3.6.2. Guidance	47
3.7. Facilitate the effective use of SOGIESC equality data to strengthen and improve LGBTIQ equality	48
3.7.1. Rationale	48

3.7.2. Guidance	49
4. How to collect information on SOGIESC	52
4.1. How to collect data on sexual orientation	53
4.1.1. Rationale	53
4.1.2. Guidance	56
4.1.3. Examples for measuring sexual orientation.....	57
4.2. How to collect data on sex and gender (identity)	60
4.2.1. Rationale	60
4.2.2. Guidance	64
4.2.3. Examples on how to collect data on sex, gender, and gender identity	66
4.3. How to collect data on gender expression	71
4.3.1. Rationale	71
4.3.2. Guidance	72
4.3.3. Examples of measuring gender expression.....	73
4.4. How to collect data on sex characteristics	76
4.4.1. Rationale	76
4.4.2. Guidance	77
4.4.3. Examples of intersex measures.....	78
Annex: Glossary	82

1. Background to – and purpose of – this Guidance

Equality and non-discrimination are founding values of the European Union (EU). These values are enshrined in the EU's Treaties¹ and in the [Charter of Fundamental Rights of the European Union](#)². They are also an integral part of the [European Pillar of Social Rights](#)³.

In recent decades, legislative developments and policy initiatives on the national and EU level have helped build more inclusive societies, including for lesbian, gay, bisexual, asexual, trans, intersex and queer (LGBTIQ) people⁴. Notably, in November 2020 the European Commission published its first ever [LGBTIQ equality strategy 2020-2025](#)⁵. The strategy delivers on the European Commission's commitment to building a union of equality and marks a new era in efforts at the European level to promote equality for LGBTIQ people.

Nevertheless, a recent survey by the [European Union Agency for Fundamental Rights \(FRA\) on LGBTI people in the EU gives cause for concern](#). The survey was the largest of its kind providing information on the lived realities of LGBTI people in the EU. It points to persistent and concerning rates of discrimination, bias-motivated violence, and harassment against LGBTIQ people. There is an urgent need to collect robust and comparable data so we can measure progress in promoting equality and non-discrimination. However, information about the lived realities of LGBTIQ people across the EU is incomplete and fragmented. Where data are available, they are often not sufficiently comparable or of low quality.

Acknowledging the need for more and better information related to sexual orientation, gender identity, gender expression, and sex characteristics (SOGIESC), the European

¹ Article 10 [TFEU](#) and Article 2 [TEU](#).

² Article 21.

³ Principle 3.

⁴ LGBTIQ people are people: (i) who are attracted to others of their own gender (lesbian, gay) or multiple genders (bisexual, pansexual), or who experience little or no sexual and/or emotional attraction (asexual, aromantic); (ii) whose gender identity and/or expression does not correspond to the sex they were assigned at birth (trans, non-binary); (iii) who are born with sex characteristics that do not fit the typical definition of male or female (intersex); and (iv) whose identity does not fit into a binary classification of sexuality and/or gender (queer). See European Commission (2020), [LGBTIQ equality strategy 2020-2025](#), COM(2020) 698 final, Brussels, 12 November 2020, p. 3.

⁵ The strategy builds on the list of actions to advance LGBTI equality for 2015-2019, which was the first policy framework specifically combating discrimination against LGBTI people. See: European Commission (2020), [Final report 2015-2019 on the list of actions to advance LGBTI equality](#), Publications Office, Luxembourg of the European Union (Publications Office).

Commission in 2020 invited the FRA and the European Institute for Gender Equality (EIGE) ‘[...] to continue providing Member States with technical assistance and methodological support on the design and implementation of data-collection exercises on LGBTIQ people both on single and multiple grounds⁶’.

1.1. Need for data for informed policy choices

‘Reliable and comparable equality data will be crucial for assessing the situation of LGBTIQ people and effectively tackling inequalities’.

[EU LGBTIQ Equality Strategy \(2020-2025\)](#)

A key purpose of any data-collection process is to gather information that can help to: (i) drive change; (ii) design evidence-based, targeted policies and programmes; and (iii) ensure budgets that meet the needs of different groups of people. Data reflect the identities, experiences and needs of people and communities that have the right to be heard and respected. Everyone should be able to see themselves and their identities represented in surveys and other data-collection instruments⁷. However, the diversity of LGBTIQ people has traditionally not been visible in national surveys or administrative data sources⁸.

This general lack of – or incomplete – statistical evidence about LGBTIQ peoples’ identities and experiences makes it difficult to assess the true magnitude/prevalence of discrimination, structural inequalities, bias-motivated crime, and harassment suffered by LGBTIQ people. A lack of equality data also makes it difficult to monitor and evaluate the effectiveness of legislation and policies in tackling LGBTIQ inequality and discrimination. Many surveys and research exercises used to inform policy do not ask: (i) questions about SOGIESC; or (ii) questions on experiences of discrimination based on SOGIESC. This can make LGBTIQ people invisible within national and international datasets and for monitoring efforts in relation to equal treatment.

To make informed policy choices to counter discrimination and foster equal treatment, legislators and policymakers need data on people’s social positioning and on their

⁶ European Commission (2020), [LGBTIQ equality strategy 2020–2025](#), COM(2020) 698 final, Brussels, 12 November 2020, p. 24.

⁷ National Academies of Sciences, Engineering, and Medicine (2022), [Measuring Sex, Gender Identity, and Sexual Orientation](#), Washington, DC: The National Academies Press, p. 6.

⁸ ‘Administrative data’ refers to the data contained in such records that are collected and kept by a government agency.

experiences of discrimination based on SOGIESC. Thus far, there has been little systematic and recurrent data collection on LGBTIQ people across the EU, and where SOGIESC data exist, they vary greatly in scope and quality. Only a few Member States collect data on SOGIESC, while others deliberately refrain from doing so, as evidenced by the 2017 [Commission report on data collection in relation to LGBTIQ people](#). Information suggests that, although the availability of data on sexual orientation is generally low, it is still considerably greater than the availability of data on gender identity, gender expression, or sex characteristics. As the EU LGBTIQ equality strategy emphasises, this shows that there is a need to improve the collection and use of data on LGBTIQ equality⁹.

Box A 1 – Subgroup on Equality Data

In 2018, the [EU High Level Group on Non-Discrimination, Equality and Diversity](#) set up a Subgroup on Equality Data (the Subgroup) to support Member States in their efforts to improve the collection and use of equality data. It tasked the FRA with facilitating the work of the Subgroup, in line with the Agency’s mandate to develop methods and standards to improve the comparability, objectivity and reliability of equality data at the European level¹⁰.

To date, the Subgroup has developed:

- a set of [non-binding guidelines on collecting and using equality data](#);
- a [compendium of promising practices for the collection of equality data](#) implemented at national level;
- a [diagnostic mapping tool](#) to help Member States identify existing sources of equality data and identify gaps in their data;
- [practical guidance to Member States on improving the collection of data on racial or ethnic origin](#).

1.2. What are ‘equality data’?

The [European handbook on equality data](#) and the [Guidelines on improving the collection and use of equality data](#) define ‘equality data’ as any piece of information that is useful for the purposes of describing, analysing, reasoning about, and making decisions on the

⁹ European Commission (2020), [LGBTIQ equality strategy 2020-2025](#), COM(2020) 698 final, Brussels, 12 November 2020, p. 22.

¹⁰ [Council Regulation \(EC\) No 168/2007 of 15 February 2007 establishing a European Union Agency for Fundamental Rights, OJ L 53, 22.2.2007](#).

state of equality. This information may be quantitative or qualitative in nature. It can include aggregate data that reflect inequalities and their causes or effects in societies¹¹.

Equality statistics can be compiled from multiple data sources, such as population censuses; administrative registers; household and individual surveys; victimisation surveys; and attitudinal surveys. Other sources include: (i) complaints data (including aggregate profiles of complainants and offenders); (ii) criminal justice data (including court statistics and data on the outcomes of court cases, and compensation offered/sanctions applied); and (iii) other avenues of data collection (such as discrimination testing and diversity monitoring by employers, civil-society organisations, and service providers).

Data disaggregated by certain personal characteristics (including age, sex, racial or ethnic origin, religion or belief, disability, sexual orientation, gender identity, or being intersex/having a variation of sex characteristics) can be used to produce equality data, at an aggregated level for statistical purposes, if this is done voluntarily and in full compliance with legal provisions on data protection (see Section No 2.2 of this guidance document)

Box A 2 – Artificial intelligence

An important and relatively recent development concerns the use of data for creating algorithms to support decision-making and artificial intelligence systems. Such algorithms and artificial intelligence systems are frequently based on complex data processing, aiming at more efficient decision-making and sometimes at the full automation of processes.

These data-driven technologies need to make equality and non-discrimination a key consideration. As was demonstrated by the FRA's recent report¹², algorithms hold significant potential for discriminatory bias. Biases are often inherent in data that are frequently used to build algorithms and AI systems, the so-called training data. For instance, the FRA's simulation of algorithms to detect offensive speech used training data consisting of real social-media posts which had been labelled as offensive or not

¹¹ The United States National Science and Technology Council distinguishes between three types of data, all of which are essential for evidence-based policymaking:

'a) Data that helps us understand the quantity of people who have a particular experience or characteristic ("quantity" or "prevalence data"); b) Data that helps us understand how people experience systems differently ("quality" or "difference data"); and c) Data that helps us understand whether policies and interventions work for different populations ("understanding what works" or "assessment data").'

See Subcommittee on Sexual Orientation, Gender Identity, and Variations in Sex Characteristics (SOGI) Data and Subcommittee on Equitable Data of the National Science and Technology Council (2023), [Federal Evidence Agenda on LGBTQI+ Equity](#), *United States Government*, p. 8.

¹² FRA (2022), [Bias in algorithms – Artificial intelligence and discrimination](#), Publications Office, Luxembourg.

offensive¹³. The instances of online hatred in the data were often directed towards LGBTIQ, Muslim and Jewish people. As a result, in the FRA's simulation of algorithms to detect offensive speech, terms linked to sexual orientation such as 'gay' were more often incorrectly labelled as offensive, as were terms like 'Muslim' and 'Jew'¹⁴. Over-flagging may be helpful to identify and counter instances of hate speech towards these groups. However, it can have the adverse effect of members of these groups disproportionately having their content flagged as offensive¹⁵, which becomes particularly problematic if the flagged content is not subject to human review¹⁶.

It is necessary to analyse algorithms for bias, including by checking training data for sources of bias and by checking the outputs of the algorithms themselves. Even when there is no direct information on protected characteristics included in the data, the use of algorithms can still lead to discrimination due to data that are highly correlated with protected characteristics (proxies). A particularly problematic example is algorithmic profiling used by law enforcement, which uses different techniques to profile people based on correlations and patterns in data¹⁷. Profiling that results in discrimination based on protected characteristics is unlawful¹⁸, but can be hard to detect. Most algorithmic bias results in indirect discrimination through proxies (such as neighbourhoods or names as a proxy for ethnicity, or shoe size as a proxy for gender). Given the potentially limitless number of proxies, whose correlation with a protected characteristic may not be evident, this is more difficult to counteract¹⁹.

Bias can also be introduced if training data are unrepresentative. For instance, if a face-detection algorithm is mainly trained on male faces, its predictions may not be reliable when applied to female faces²⁰. Bias stemming from poor data quality can also be amplified by feedback loops, which occur when predictions of algorithms become the basis for future training datasets, for example in predictive policing²¹.

It is therefore important to also analyse the output of algorithms for bias. If the output of algorithms differs when only information on protected characteristics changes, this

¹³ FRA (2022), [Bias in algorithms – Artificial intelligence and discrimination](#), Publications Office, Luxembourg, pp. 11-12, pp. 55-56.

¹⁴ FRA (2022), [Bias in algorithms – Artificial intelligence and discrimination](#), Publications Office, Luxembourg, pp. 11-12, p. 62.

¹⁵ FRA (2022), [Bias in algorithms – Artificial intelligence and discrimination](#), Publications Office, Luxembourg, p. 11.

¹⁶ FRA (2022), [Bias in algorithms – Artificial intelligence and discrimination](#), Publications Office, Luxembourg, p. 77.

¹⁷ FRA (2018), [Preventing unlawful profiling today and in the future: a guide](#), Publications Office, Luxembourg, p. 19.

¹⁸ Article 11 (3) of the [Police Directive](#).

¹⁹ FRA (2022), [Bias in algorithms – Artificial intelligence and discrimination](#), Publications Office, Luxembourg, p. 24.

²⁰ FRA (2022), [Bias in algorithms – Artificial intelligence and discrimination](#), Publications Office, Luxembourg, p. 19.

²¹ FRA (2022), [Bias in algorithms – Artificial intelligence and discrimination](#), Publications Office, Luxembourg, p. 77.

indicates a risk of discrimination²². For this purpose, it is necessary to collect data on protected characteristics, including SOGIESC. Currently, the legal basis for this may be unclear for users of algorithms, since detecting algorithmic discrimination is not specified as a justification for processing sensitive personal data in the [General Data Protection Regulation \(GDPR\)](#)²³, although the EU's proposed Artificial Intelligence Act may provide some clarity²⁴.

1.3. Purpose and structure of this guidance note

Since SOGIESC categories can have different meanings in different societies and social contexts, there is no self-evident solution for measuring these categories. There is also no agreed-upon standard for how to best collect data on SOGIESC. This results in a lack of comparability between existing data-collection efforts at national level and across the EU. It also increases the time and cost of developing relevant categories and collecting data. Improving data-collection efforts and the quality of collected data will help Member States and the EU to gather more exact SOGIESC data to support their monitoring and reporting activities. It will also help to increase trust among LGBTIQ people in data-collection efforts by states and public institutions.

Although it supports awareness-raising on the opportunities and challenges of collecting SOGIESC data, this guidance note serves as a starting point to help Member States to effectively apply and monitor EU equality law. It sets out the challenges in collecting data on SOGIESC, and highlights promising paths to a more standardised and consistent approach, while reflecting the different national contexts and the varying requirements for the collection of equality data based on SOGIESC. Policymakers and other stakeholders may choose or prioritise actions set out in the guidelines that best meet the needs identified at the time in any given Member State.

EU Member States are encouraged to use these guidelines to: (i) inform the development and effective implementation of national action plans on LGBTIQ equality; (ii) inform their equality-data strategies; and (iii) monitor the outcomes of these strategies from an equality perspective.

²² FRA (2022), [Bias in algorithms – Artificial intelligence and discrimination](#), Publications Office, Luxembourg, p. 25.

²³ FRA (2022), [Bias in algorithms – Artificial intelligence and discrimination](#), Publications Office, Luxembourg, p. 24.

²⁴ FRA (2022), [Bias in algorithms – Artificial intelligence and discrimination](#), Publications Office, Luxembourg, p. 10.

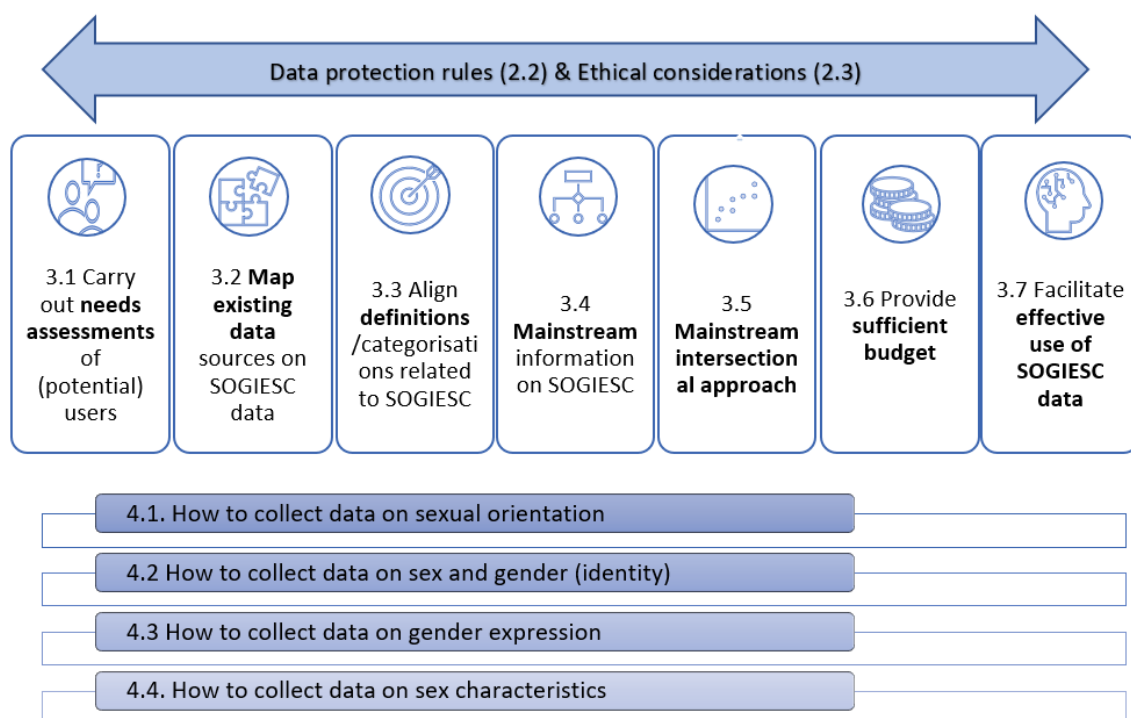
The target audience of these guidelines includes public and private entities involved in collecting and using data on equality for LGBTIQ people, such as:

- data producers in political and administrative bodies at national, regional, and local level;
- national statistical offices;
- private and public organisations;
- the justice system and the police;
- equality bodies and national human rights institutions;
- academia.

The guidance note begins by setting out the challenges, data-protection rules, and key ethical considerations when collecting and using SOGIESC data (Section 2). The main body of this guidance summarises two sets of principles for collecting and using equality data based on SOGIESC. The first set of 7 principles are general/cross-cutting in nature, and are discussed in Section 3. The second set of 4 principles provide practical guidance on how to collect data on SOGIESC, and are discussed in Section 4. The second set of practical principles includes several examples of past surveys to illustrate possible approaches. However, these are not intended as examples of best practice to be directly adopted. Where examples contain problematic elements, these are highlighted in the notes for consideration under each example.

Because the requirements for collecting equality data on SOGIESC may vary across different data sources (e.g. administrative, census, or survey data), the applicability of the following guiding principles might be easier for some data sources than others. The principles do not follow any particular order of priority. Depending on the user's needs, they can be read either together and applied as a whole, or they can be applied individually or combined.

Figure 1 – Guiding principles for the collection and use of equality data on SOGIESC



Alternative text: Figure laying out the structure of the guidance note. At the top is an arrow with the text ‘Data-protection rules (2.2) & Ethical considerations (2.3)’. Below are boxes with the general/cross-cutting principles and logos to illustrate them: ‘3.1 Carry out **needs assessments** of (potential) users’, ‘3.2 **Identify existing data** sources on SOGIESC data’, ‘3.3 Align **definitions/categorisations** related to SOGIESC’, ‘3.4 **Take** information on SOGIESC into account in all policies’, ‘3.5 **Take intersectional approach into account in all policies**’, ‘3.6 Provide **sufficient budget**’ and ‘3.7 Facilitate **effective use of SOGIESC data**’. Below this are four boxes with the practical principles: ‘4.1. How to collect data on sexual orientation’, ‘4.2 How to collect data on sex and gender (identity)’, ‘4.3 How to collect data on gender expression’, and ‘4.4. How to collect data on sex characteristics’.

Type your caption here.

© Type your Copyright information here.

Source: Type your source here.

The guidance note is complemented by a [compendium of promising practices for collecting equality data](#)²⁵, which compiles a range of practices on different

²⁵ FRA (2023), [Compendium of practices for equality-data collection](#).

discrimination grounds, including SOGIESC, and aims to provide practical examples on how to implement the principles on collecting and using equality data.

The Subgroup on Equality Data developed this guidance note with the support of the FRA and two leading experts in LGBTIQ equality. Before being finalised and endorsed by the High-Level Group on Non-Discrimination, Equality and Diversity, key European LGBTIQ representative²⁶ organisations and the Subgroup on LGBTIQ Equality were consulted.

²⁶ ILGA Europe, The Eurocentralasian Lesbian* Community (EL*C), OII Europe, Transgender Europe (TGEU), IGLYO.

2. Challenges, data protection, and other ethical considerations in collecting and using equality data on SOGIESC

2.1. Challenges in collecting and using SOGIESC data

In addition to the common challenges and gaps in the collection and use of equality data more broadly, which were explored in the first guidance note developed by the Subgroup²⁷, there are other and/or cumulative challenges to the collection and use of SOGIESC data. The 11 bullet points below discuss these additional challenges in more detail.

- **Limited reliable information on the composition of LGBTIQ people among the population in the EU and a lack of sampling frames.** So far, only Malta's census²⁸ includes questions on sexual orientation and gender identity, and no national census in the EU asks about gender expression or sex characteristics. There is therefore a lack of reliable estimates of the percentage of LGBTIQ people in the population, and this restricts quantitative research and makes it difficult to develop reliable sampling frames²⁹.
- **The need for valid and reliable measures which reflect contemporary terminology and concepts.** Definitions and categorisations need to be aligned across national data sources and between Member States to ensure comparability (see Guideline No **3.3** of this guidance document). Conceptions of LGBTIQ identities and

²⁷ Common challenges in collecting and using equality data include: the lack of a coordinated approach; the lack of sufficient resources; insufficient comparability across different data sources; insufficient consultation with relevant stakeholders; an imbalance of data collection on different grounds of discrimination and across different areas of life; an over-reliance on proxies; and inaccurate interpretation of the data-protection frameworks. See: European High Level Group on Non-Discrimination, Subgroup on Equality Data (2018), [Guidelines on improving the collection and use of equality data](#), Publications Office, Brussels, pp. 9-10.

²⁸ National Statistics Office – Malta, [Census](#).

²⁹ Daly, F., Crehan, P. R., Grzywnowicz, M. (2022), [The LGBTI Inclusion Index: An Innovative Tool to Incentivize Human Rights and Development Data](#). *Journal of Human Rights Practice*, 14(2), 600-621, p.614.

experiences are continuously evolving, so these concepts should be under periodic review while endeavouring to support comparability across time³⁰. Differences in language, culture and historical understandings should also be considered ([see Section 4 of this guidance document](#)). Moreover, terms commonly known within the LGBTIQ community may not be as well-known to the general public, so researchers should balance the use of precise terminology with the need for accessible questions, particularly when the survey is directed at the general population³¹. In doing so, researchers should avoid using terms which may be in common use but are considered outdated and/or derogatory (such as ‘hermaphrodite’ rather than ‘intersex’, or ‘transsexual’ as a term to refer to all trans people)³².

- **Small sample sizes that limit statistical power.** In most population-based surveys, LGBTIQ people represent less than 10% of the sample. This can be a challenge for data disaggregation and statistical power. It may result in LGBTIQ people being excluded entirely from the reported findings or being included with large confidence intervals. It can also limit any substantial intersectional analysis. This can be countered through increasing overall sample sizes, oversampling³³ LGBTIQ people, using booster samples³⁴, or conducting dedicated surveys, although these approaches all entail extra costs³⁵.
- **Methods of data collection.** There are specific challenges in surveying LGBTIQ people with traditional methods, such as personal interviews conducted either face-to-face or via telephone. Many LGBTIQ people are not ‘out’ in all aspects of their life and may not be willing to disclose aspects of their private identity or experiences, or may only be willing to do so anonymously. This can be countered through a self-completion section in a face-to-face interview (CASI – computer-assisted self-interviewing) ensuring confidentiality, or through an online survey

³⁰ National Academies of Sciences, Engineering, and Medicine (2022), [Measuring Sex, Gender Identity, and Sexual Orientation](#), Washington, DC: The National Academies Press, p. 124.

³¹ Center for American Progress (2022), [Collecting Data About LGBTQ+ and Other Sexual and Gender-Diverse Communities – Best Practices and Key Considerations](#).

³² Further information on terms to use and avoid can be found in European Parliament (2020), [Glossary of Sensitive Language for Internal and External Communications](#), and in Ginicola, M, Smith, C, and Filmore, J. (Eds.) (2017), [Affirmative Counseling with LGBTQ+ People](#), *American Counseling Association*, pp. 368-370.

³³ Oversampling is the practice of selecting respondents so that some groups make up a larger share of the survey sample than they do in the population.

³⁴ In the German Socio-Economic Panel (SOEP), a ‘boost sample’ of sexual and gender minority respondents was used to supplement the existing probability-based sample, see Fischer M., Kroh M., De Vries L., Kasproski D., Kühne S., Richter D. and Zindel Z. (2022), [Sexual and Gender Minority \(SGM\) Research Meets Household Panel Surveys: Research Potentials of the German Socio-Economic Panel and Its Boost Sample of SGM Households](#), *European Sociological Review*, (38(2)).

³⁵ Sell, R. L. (2017), [Challenges and Solutions to Collecting Sexual Orientation and Gender Identity Data](#). *American Journal of Public Health*, 107 (8), 1212-1214, pp. 1213-1214.

method (such as CAWI: computer-assisted web interviewing) ensuring anonymity. This might result in more respondents from different strata of the target population being willing to participate, including those who do not wish to disclose that they are LGBTIQ in a traditional survey conducted face-to-face or by telephone. Often, due to the lack of sampling frames,³⁶ the preferred method of data collection is a self-selection online survey. However, this also comes with its own challenges, as the sample will tend to be biased towards people more active within the LGBTIQ community, especially if it is not combined with a well-developed outreach campaign. Online methods in general pose the challenge of excluding people who lack access to technology or digital literacy.

- **Weighting methods depend on the method of data collection.** Where data are available for a representative sample (such as through the census in Malta³⁷), it is possible to weight these data based on the proportion of the population and their composition. When a non-probabilistic approach is used, post-stratification weighting can be applied to align the completed sample with estimated population totals. Weighting should also attempt to correct for potential exclusion, selection, and non-participation biases. This can include taking the channels used for survey promotion into account³⁸.
- **Lack of – or insufficient – participation/consultation of LGBTIQ people** and their representative organisations in the: (i) design and implementation of data collection; (ii) interpretation of results; and (iii) dissemination efforts. This lack of involvement might result in incorrect categorisations and interpretations of findings, as well as limited outreach. Considering the experience and expert knowledge of the communities is important in developing policy measures that reflect contemporary terminology and the situation at stake. Active consultation with LGBTIQ people/organisations at all stages of data collection, analysis, and use can be helpful for understanding the cultural context, the stigma LGBTIQ people might be facing, and other specific experiences.
- **The need to build trust.** Given the long history of persecution and criminalisation that LGBTIQ people have faced and continue to face, they may be reluctant to provide information on their sexual orientation, their gender identity, or whether they are intersex/have a variation of sex characteristics. This is especially the case for people who belong to multiple marginalised groups. Therefore, it is important

³⁶ A sampling frame is a list of units/people from which a sample is drawn.

³⁷ National Statistics Office – Malta, [Census](#).

³⁸ For instance, for the FRA LGBTI Survey 2019, affiliation weights were used to correct for over-representation of affiliated respondents and respondents who had been reached through campaigns by LGBTIQ organisations. See FRA (2020), [Technical report. A long way to go for LGBTI equality](#), Publications office of the European Union, Luxembourg. pp. 76-85.

to: (i) build trust by working together with civil-society groups; and (ii) clarify how the data are stored and used and for what purpose they are stored and used (see Section No 2.3 of this guidance document). Where possible, anonymised data should be made accessible to LGBTIQ communities and organisations for their own use. All public bodies should publish equality data that can readily be used by the public, civil-society organisations, and independent researchers. This should be in accessible formats for people who are not experts in the field. Special attention should be paid to making these data accessible for people with disabilities³⁹.

- **The difficult of reaching underrepresented and hard-to-reach groups.** Data-collection efforts need to be conscious of adopting sampling strategies that reach underrepresented LGBTIQ people. These efforts should consider, among other factors: gender; age; disability; racial and ethnic background; migration and refugee status; class; social origin; languages spoken by potential respondents; the level of respondents' involvement in the LGBTIQ community; access to housing (as LGBTIQ people disproportionately experience homelessness); access to technology; as well as the intersections between these factors. Survey methods should be adapted to the groups to be reached, and methodological limitations should be openly acknowledged. For example, a survey focused on LGBTIQ community centres would overwhelmingly reach people who are already actively engaged in the community and will be less likely to reach people who are 'closeted' (i.e. who do not publicly disclose that they are LGBTIQ)⁴⁰.
- **The need to improve outreach to LGBTIQ women.** Within the LGBTIQ community, cisgender⁴¹, endosex⁴² gay men tend to be overrepresented in data collection⁴³. A specific effort should therefore be made: (i) to ensure the participation of other members of the community, specifically targeting the women of the LGBTIQ community; and (ii) to ensure that civil-society organisations led by – and focused on – LGBTIQ women are involved in all parts of the dissemination process.
- **The need for explicit inclusion of trans and intersex people.** Often, research on LGBTIQ communities only considers sexual orientation (see Guideline No 3.4 of this guidance note), without asking about gender identity, gender expression, and sex

³⁹ Measures can include using 'simple English' (or equivalent), sign-language interpretation, large-print text, alt text for images, and colourblind-accessible colours in data visualisations.

⁴⁰ Human Rights Campaign Foundation (2019), [LGBTQ-Inclusive Data Collection: A Lifesaving Imperative](#). Washington, DC, Human Rights Campaign Foundation, p. 17.

⁴¹ Cisgender people are people whose gender identity corresponds to their sex assigned at birth.

⁴² An endosex person is someone whose sex characteristics fit the normative medical or social ideas for female and male bodies.

⁴³ For example, in the FRA LGBTI II Survey 2019, gay cisgender men represented 42% of the respondents. FRA (2020), [Technical report. A long way to go for LGBTI equality](#), Publications Office of the European Union, Luxembourg. p. 63.

characteristics. It is important to ensure that all parts of the community are made visible within the data, and to account for multiple affiliations within the LGBTIQ categories (for example, someone may be trans, intersex, lesbian and a woman).

- **Administrative data and the use of proxies** (see Guideline No 3.4 of this guidance document). Member States are increasingly moving towards using administrative data to produce both population-related statistical information and their censuses. Since administrative data sources rarely include information about sexual orientation, gender identity, or about being intersex/having a variation of sex characteristics, there is a risk that LGBTIQ people will increasingly be rendered invisible. Collecting SOGIESC information in administrative data without a legal ground or rationale being provided may also raise questions about data protection and the validity of the information provided⁴⁴. Administrative SOGIESC data collection will always be limited to: (i) the legal ground determining its purpose; and (ii) complying with data-protection rules. Collecting SOGIESC information in administrative data might also result in limiting comparability across data sources within and between countries. This is because the measurements in different data sources and countries might be based on differing concepts. Where proxy information is used, it might not reflect how the person in question would have self-identified, so it is likely to result in undercounts. Although they should be approached with caution, such data *can* be helpful, for instance where they provide the basis for further research that makes self-identification possible⁴⁵.

Box A 3 – Willingness to share SOGIESC information

Findings of the [Special Eurobarometer 437](#) point to broad support for providing sensitive personal information on an anonymous basis, especially among those who describe themselves as belonging to a minority group at risk of discrimination. Some 83% of respondents who self-identified as belonging to a sexual minority were in favour of providing, on an anonymous basis, information about their sexual orientation⁴⁶.

⁴⁴ Ashley, F. (2021), [Recommendations for Institutional and Governmental Management of Gender Information](#). *New York University Review of Law & Social Change*, 44(4):489-528, pp. 504, 514-515.

⁴⁵ Ashley, F. (2021), [Recommendations for Institutional and Governmental Management of Gender Information](#). *New York University Review of Law & Social Change*, 44(4):489-528, pp. 504, 514-515.

⁴⁶ European Commission (2015), [Special Eurobarometer 437: Discrimination in the EU in 2015](#), *European Union*, p. 99.

2.2. Data-protection rules do not prevent the collection of equality data on SOGIESC

‘Ensure that requirements for individuals to provide information on their sex or gender are relevant, reasonable and necessary as required by the law for a legitimate purpose in the circumstances where it is sought, and that such requirements respect all persons’ right to self-determination of gender’.

-[The Yogyakarta Principles plus 10](#)⁴⁷, Principle 6(G)

In the EU, the collection of personal data disaggregated by sensitive personal characteristics, such as SOGIESC, is protected by EU data-protection law and the [Charter of Fundamental Rights of the EU](#). However, data-protection rules do not stand in the way of collecting equality data for statistical purposes. On the contrary, the rules allow data to be processed while ensuring the respect of fundamental rights⁴⁸.

The [GDPR](#) provides for a single set of data-protection rules across the EU, meaning that all collection and processing of data on equality should be carried out in full compliance with the principles and safeguards set out under it. The GDPR requires that data collection should pursue a well-defined goal, such as measuring and documenting inequalities or meeting legal reporting requirements⁴⁹. More restrictive conditions apply to processing sensitive data, as mentioned in the [GDPR](#) on ‘special categories of personal data’⁵⁰.

Furthermore, the principle of ‘**data minimisation**’ lies at the heart of the EU’s general data-protection regime, requiring that personal data should be ‘adequate, relevant, and

⁴⁷ In 2006, international human-rights experts adopted a set of international principles relating to sexual orientation and gender identity serving as a universal guide to human rights. These Yogyakarta principles affirm binding international legal standards with which all States must comply. In 2017, [the Yogyakarta Principles plus 10 \(YP+10\)](#) were adopted to complement the original [Yogyakarta Principles](#).

⁴⁸ 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](#).

⁴⁹ Any organisation processing personal data must have a valid legal basis for that processing activity. The law provides six legal bases for processing: consent, performance of a contract, a legitimate interest, a vital interest, a legal requirement, and a public interest. See Article 6, [253 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, OJ 2016 L 119 \(General Data Protection Regulation\)](#).

⁵⁰ Under [Article 9 of the GDPR](#) special categories of personal data comprise ‘personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership’, as well as ‘genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person’s sex life or sexual orientation’.

limited to what is necessary in relation to the purposes for which they are processed'⁵¹. 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 stake⁵²;

Article 9(2) of the [GDPR](#) clearly stipulates that the general prohibition on processing special categories of personal data in line with the 'data minimisation' principle does not apply in several cases, including when:

- the data subject has given explicit consent;
- processing is necessary for reasons of substantial public interest, on the basis of EU or national law;
- processing is necessary for: (i) achieving purposes in the public interest; (ii) the purposes of scientific or historical research; or (iii) statistical purposes.

In addition, Recital 26 of the [GDPR](#) clarifies that the principles of data protection apply to special categories of personal data that concern an identified or identifiable natural person. By contrast, they do not apply to anonymous information or to personal data rendered anonymous in such a manner that the data subject is not identified or is no longer identifiable⁵³. Given the sensitivity of the information, data anonymity should be guaranteed even when linking data sets. Measures should be taken to prevent any abuse or data leaks following the collection of the data.

Data-collection efforts also need to ensure a balance between the benefits of the data collected and the risks associated with any potential misuse or unauthorised data disclosure⁵⁴. This ensures that data are only collected where the benefits of collecting the information outweigh the intrusion of privacy⁵⁵. Therefore, data should not be

⁵¹ Article 5(1)(c) and Recital (39) of the [General Data Protection Regulation](#).

⁵² FRA, Council of Europe (2018 edition), [Handbook on European data protection law](#), Section 3.3. Publications Office, Luxembourg.

⁵³ For more information, see: (i) Guiding Principle No 4 of the [Guidance Note on the collection and use of equality data based on racial and ethnic origin](#); (ii) the [European Commission's 2017 report on data collection in relation to LGBTIQ people](#); (iii) the EDPS preliminary opinion on data protection and scientific research from 6 January 2020; and (iv) the upcoming EDPB guidance on data protection and scientific research.

⁵⁴ National Academies of Sciences, Engineering, and Medicine (2022), [Measuring Sex, Gender Identity, and Sexual Orientation](#), The National Academies Press, Washington DC, p. 61.

⁵⁵ Conference of European Statisticians, 67th plenary session (2019). [In-Depth Review of Measuring Gender Identity](#). United Nations Economic and Social Council, Paris.

published or publicly accessible in a manner that allows individual data subjects to be identified, either directly or indirectly.

2.3. Ethical considerations in relation to data on SOGIESC

‘Recognise that the needs, characteristics and human-rights situations of populations of diverse sexual orientations, gender identities, gender expressions and sex characteristics are distinct from each other, and ensure that data on each population is collected and managed in a manner consistent with ethical, scientific and human-rights standards and made available in a disaggregated form’.

- [The Yogyakarta Principles plus 10](#), Principle 19.

The use of statistical and analytical categories for any data collection or for data disaggregation should always be guided by the overriding human-rights based principle of **doing no harm**. This principle was set out by the United Nations High Commissioner for Human Rights (OHCHR) in the [Human Rights-Based Approach to Data](#).

‘Doing no harm’ means: (i) that no data-collection activity should create or strengthen existing discrimination, bias, or stereotypes; and (ii) that the data collected should be used for the benefit of the groups they describe.

This guidance note recognises and reaffirms the OHCHR set of six principles of a human rights-based approach to data collection: participation, data disaggregation, self-identification, transparency, privacy, and accountability.

In addition to the set of human-rights based principles for data collection, the ethical principles set out in the bullet points below should underpin and guide all national efforts when collecting data on SOGIESC.

- Questions on SOGIESC should be **voluntary**: that is, nobody should be forced to disclose their sexual orientation, gender identity, or whether they have a variation of sex characteristics if they do not want to. Respondents should therefore have the opportunity to respond to SOGIESC questions with ‘prefer not to say/disclose’ and ‘don’t know’ or ‘questioning’.

- **Informed consent** to take part is crucial⁵⁶. This requires a clear affirmative act on the part of the data subject consenting to their data being processed. Consent must be freely given, specific, and unambiguous. It must cover all purposes of data processing, and all processing activities carried out for these purposes⁵⁷.
- When the purpose for collecting special categories of personal data is not clear, it is recommended to inform the subject as to: **(i) the legal ground for collecting the data; (ii) the purpose for which the data are collected; and (iii) how the data will be processed and protected.**
- **Civil-society organisations, including LGBTIQ organisations and survey experts, should be involved** as much as possible in all stages of developing a survey or data-collection instrument. Whenever possible, they should also be involved in interpreting and contextualising results in line with the underlying principle of doing no harm. During any consultation process, LGBTIQ representatives should not feel pressured to take part, and they should receive appropriate compensation for their contributions.
- **Inconsistencies in SOGIESC information provided by respondents to different administrative sources should not automatically be interpreted as attempts at falsification or as errors needing to be corrected or clarified.** This is particularly important for situations where SOGIESC information is collected as part of administrative information to provide permits, services, benefits, or employment. Inconsistencies may be due to various reasons, including: (i) changes in the respondents' identities; (ii) changes in their levels of comfort in disclosing such information; or (iii) limited response options (such as the use of binary gender or sex categories). Forms collecting SOGIESC information should clarify explicitly that respondents will not be penalised for any discrepancies⁵⁸.
- **Interviewers** may affect results since people might be more reluctant to reveal their sexual orientation, gender identity/expression, or the fact that they are intersex/have a variation of sex characteristics to an interviewer (for data-collection modes, see Section 2.1 of this guidance document).

⁵⁶ IE SOGI – UN Independent Expert on protection against violence and discrimination based on Sexual Orientation and Gender Identity (2019), [Data collection and management: An essential component in creating awareness and providing effective measures to address violence and discrimination based on SOGI](#), United Nations Human Rights Special Procedures.

⁵⁷ Recital (32) of the [General Data Protection Regulation](#) (GDPR).

⁵⁸ Subcommittee on Sexual Orientation, Gender Identity, and Variations in Sex Characteristics (SOGI) Data and Subcommittee on Equitable Data of the National Science and Technology Council (2023), [Federal Evidence Agenda on LGBTQI+ Equity](#), *United States Government*, pp. 39-40.

- For in-person interviews, it is important to **create a safe environment** where respondents can answer questions at their own pace, in privacy, and without fear of being overheard by other household members or passers-by⁵⁹. The interviewer should consider that answering sensitive questions may be emotionally taxing and could be 'triggering'⁶⁰ to some participants.
- The principles of avoiding harm should also be extended to cover the interviewers. Because **interviewers may visit the homes** of people unknown to them during research projects, they put themselves under a certain degree of psychological and physical risk. In such cases, interviewers should have access to psychological support during fieldwork⁶¹.
- When a survey aims to assess sensitive topics such as violence, discrimination, and sexual abuse, content warnings and information on **specialised support services** should be provided to respondents at the beginning and end of the survey.
- When the response category 'other' is included as a choice for a survey question, a **write-in possibility** should be available whenever possible. This allows respondents to elaborate in their own words on a particular issue/ground and reduces the risk of producing unreliable data, subsumed under a closed category of 'other'. Including the option 'other' can also inform possible changes to the answer categories in subsequent iterations of the survey. Where the question is about a respondent's identity, it might be preferable to use an alternative formulation such as 'Please describe' or 'Prefer to self-describe'.
- When conducting qualitative research, researchers should be aware of implicit bias, and they should therefore be **trained in diversity and anti-discrimination** to minimise potential discomfort and power imbalances when engaging with respondents. **Interviewer matching** (i.e. making sure the interviewer resembles the interviewee in some way such as in gender, language spoken, or migration-experience) can increase response rates and acceptance amongst respondents and thus improve the research results⁶². **Interviewee transcript review** (i.e. allowing interviewees to listen to a recording of the interview they have given so they can

⁵⁹ European Commission (2021), [Guidance note on the collection and use of equality data based on racial or ethnic origin](#). Publications Office, Luxembourg, p. 55.

⁶⁰ Triggering words can cause strong emotions (for example, frustration) because of previous experiences.

⁶¹ European Commission (2021), [Guidance note on the collection and use of equality data based on racial or ethnic origin](#). Publications Office, Luxembourg, p. 55.

⁶² European Commission (2021) [Guidance note on the collection and use of equality data based on racial or ethnic origin](#). Publications Office, Luxembourg, p. 55.

reflect on and correct the account they give) can be a useful tool to avoid misrepresentation or researcher bias when talking about SOGIESC topics⁶³.

Data-protection rules and other ethical considerations are general/cross-cutting in nature, and should be considered in the implementation of the guiding principles set out in this note.

⁶³ Hagens, V., Dobrow, M.J. and Chafe, R. (2009), [Interviewee Transcript Review: assessing the impact on qualitative research](#). *BMC Med Res Methodol* 9, 47; Rowlands, J. (2021), [Interviewee Transcript Review as a Tool to Improve Data Quality and Participant Confidence in Sensitive Research](#). *International Journal of Qualitative Methods*, 20.

3. General principles for collecting equality data on SOGIESC

3.1. Carry out needs assessments of users and potential users of data disaggregated by SOGIESC

3.1.1. Rationale

A needs assessment of users and potential users of data disaggregated by SOGIESC should help to identify the type of equality data needed for both evidence-based policy making and monitoring the state of equality and human rights.

The needs assessment should be based on: (i) relevant EU law; (ii) international and national normative standards to which Member States are subject (such as the [Charter of Fundamental Rights of the EU](#)); (iii) EU equality legislation; (iv) international human-rights law; and (v) national legislation. All of these standards and bodies of law may variously address equality and non-discrimination in relation to SOGIESC. The assessment should also reflect relevant policies and political commitments, including the [European Pillar of Social Rights](#), the [LGBTIQ equality strategy 2020-2025](#) and the [United Nations \(UN\) Agenda for Sustainable Development](#).

The needs assessment could be facilitated or coordinated by a national statistical institute or a relevant ministry/governmental department. Alternatively, a specific mandate and dedicated resources to carry out the assessment could be provided to a national human-rights institution, equality body, or research institute.

3.1.2. Guidance

Member States are encouraged to:

- undertake a comprehensive needs assessment for equality data in relation to SOGIESC, by considering all relevant policy and legislative developments, including international human-rights law and relevant policy and political commitments;

- conduct a user-needs assessment periodically, for example every 4-5 years, considering new legislative and policy developments or evaluations of existing laws and policies;
- identify data-collection priorities based on the urgency of the user needs;
- communicate the findings of the needs assessments through different channels and in different formats to all relevant institutions, bodies, statistical institutes, research institutes, civil-society organisations, etc.

3.2. Identify existing data sources that provide information on SOGIESC

3.2.1. Rationale

Identifying all available data (including data from administrative, governmental, academic and civil-society sources) related to SOGIESC is necessary to both: (i) obtain a coordinated approach to data collection; and (ii) create uniform categories, concepts, and definitions. Identifying all these data would make it possible to lay the foundation for a more systematic approach to SOGIESC data collection. It would also make it possible to provide a solid basis for necessary improvements by identifying the three important factors set out in the following bullet points.

- The first important factor is the different data sources and providers of equality data related to SOGIESC. In identifying these, it is important to also consider sources and providers that are not specifically designed to collect equality data but that still have variables that could be used to measure existing structural inequalities.
- The second important factor is the gaps and inconsistencies in the design of a study and in the methods, concepts, and measurements applied (these inconsistencies also include unnecessary duplicates in data collection).
- A distinction can be made between international/European data sources and national data sources. There are relatively few international/European studies that

collect international data on the experiences of LGBTIQ people⁶⁴. For national data sources, a further distinction can be made between the collection of data on a specific topic, such as discrimination, or the collection of data in general on population and/or household characteristics (e.g. national censuses). The latter can provide insight into the quantitative presence of LGBTIQ people in society and can possibly also provide details about the social position of these LGBTIQ people when comparable with other information (e.g. in a comparable work or living situation)⁶⁵.

3.2.2. Guidance

Member States and all other relevant institutions and actors involved in – or responsible for – the collection of SOGIESC data are encouraged to take the actions set out in the following bullet points.

- They should take due account of Guideline No 1 ‘Map existing sources of equality data and identify data gaps’ of the [Guidelines on improving the collection and use of equality data](#).
- They should identify and authorise an institution/body/organisation to initiate, coordinate, and execute the mapping exercise.
- They should undertake a comprehensive exercise to identify all existing data sources⁶⁶ that support the disaggregation of data by SOGIESC in statistical terms. They should also list all SOGIESC categories and proxy variables used in the different data sources, including their underlying concepts and definitions.
- They should use the results of the identification/mapping exercise to find similarities and discrepancies between data sources and over time on:
 - the policy area/legislative file for which the data have been used or are needed;
 - the data source (administrative register, survey data, complaint data, etc.);

⁶⁴ Some examples are: Out Now Global (2020), [Out Now Global LGBT 2020 Study: International LGBT2020 Homophobia Report](#), Out Now, Paris; FRA (2020), [A long way to go for LGBTI equality](#), Publications Office, Luxembourg; European Commission (2019), [Special Eurobarometer 493: Discrimination in the European Union](#), European Union; European Social Survey (2018), [ESS Round 9 Source Questionnaire](#), University of London, London; GESIS (2016), [European Values Study](#), GESIS, Cologne.

⁶⁵ For more information about international data sources and existing official and non-governmental data sources, see the [report on data collection in relation to LGBTIQ people](#).

⁶⁶ Data sources collected and/or used by national authorities in areas covered by EU law regarding equality and non-discrimination in relation to SOGIESC – these may also refer to proxies (such as sex of the partner along with the sex of the respondent, etc.).

- the categories/classifications used for the disaggregation, or the questions used in a survey to measure SOGIESC and the concepts and definitions underlying these questions (and relevant changes in categories/classifications over time);
 - the intersections between grounds of discrimination;
 - the areas of life for which the data are collected, such as: work, access to work, housing, access to housing, education, health, access to justice, access to goods and services, etc.
 - the methods of qualitative and quantitative data collection (e.g. face-to-face interview, online survey, mail survey, telephone interview, etc.), sample size, representativeness, and geographical coverage;
 - public accessibility of data and analyses;
 - the frequency with which the data were collected (i.e. the frequency of updates and existing time series, as well as any interruptions or variations in frequency).
- They should review the results of the data mapping in conjunction with the findings of the user-needs assessment (see Guideline No **3.1** of this guidance document) and assess the current needs to collect data broken down by sexual orientation, gender identity/expression, and sex characteristics.
 - They should ensure that aggregated statistical data based on SOGIESC enter the public domain in a transparent manner, especially in cases where such data can be used by the courts (Article 47 of the [Charter of Fundamental Rights of the EU](#)).
 - They should set up a data hub (with a dedicated web page) to collate and display SOGIESC data identified through the identification/mapping exercise in line with Guideline No 3 of the [Guidelines on improving the collection and use of equality data](#). This hub can also be hosted by the institution/body authorised to perform the needs assessment (see Guideline No **3.1** of this guidance document).
 - They should carry out a periodic exercise to identify LGBTIQ equality data considering both: (i) the frequency with which the identified data sources are updated, for example, every 4-5 years; and (ii) new potential data sources and suppliers.
 - Member States that do not collect SOGIESC data in official statistics or through large-scale surveys should consider a structured dialogue with different

stakeholders (e.g. decision-makers, public authorities, research institutes, civil-society organisations, etc.) to reach a consensus on what kind of data should and may be collected and under what conditions.

3.3. Align definitions and categorisations related to SOGIESC

3.3.1. Rationale

A uniform approach in data collection (including aligning the applied SOGIESC definitions) would support the harmonisation of the collected data: (i) over time; (ii) within different data sources; and (iii) across different data sources. Aligning definitions of sexual orientation, sex, gender, gender identity, gender expression, and sex characteristics would lead to better measurement and therefore maximise the comparability, validity, and reliability of the data collected. It would also minimise data-collection costs. Better measurement of sexual orientation, sex, gender, gender identity, gender expression, and sex characteristics will make it easier to identify sexual and gender minority populations and understand the challenges they face. The definitions from the Glossary of this guidance document may be taken as a starting point for such an alignment.

In summary, a uniform approach in this area would help to:

- maximise comparability (at the national level and across the EU);
- increase the validity and reliability of SOGIESC concepts;
- maximise the possibilities for linking different data sources for further analysis and equality reporting;
- minimise time, costs, and unnecessary duplication between data sources.

3.3.2. Guidance

Member States and relevant institutions/bodies involved in – or responsible for – collecting data based on SOGIESC are encouraged to take the 12 measures set out in the following bullet points.

- They should build on the most recent findings from academic research in sociology and social psychology on SOGIESC to increase the awareness of potential data collectors and data users about the social construction of the following terms: sex, gender, gender identity, gender expression, sexual orientation, and sex characteristics.
- They should engage in structured consultations with representatives of all relevant LGBTIQ communities to: (i) identify the most appropriate categories for disaggregation/questions on self-identification to be asked in a survey; and (ii) agree on operational definitions. In line with the [UN's human rights-based approach to data](#), community consultation and participation should take place as much as possible at all levels of research, including at the stage when questionnaires and measurements are being developed.
- Building on the results of the community consultations, they should carry out consultations with: (i) survey-methodology experts; (ii) survey-design experts; and (iii) data experts involved in collecting administrative data (e.g. national registers and censuses). In these consultations, they should discuss the challenges and problems of potential categorisations for disaggregation/survey questions (including proxy data on SOGIESC) in different data collections. These consultations should seek to reach agreement about the most effective, valid, and reliable approaches to the collection and use of data on LGBTIQ people. These consultations could also provide the basis for developing standards or promising practices in the measurement of relevant categorisations.
- They should improve response rates by adapting and revising questions/categories (in particular ambiguous questions and categories) where necessary. These adaptations and revisions could be based on insights from the community consultations.
- They should use precise and bias-free terminology that reflects the constructs of interest. The questions they use should clearly specify which components are being measured (e.g. sex, gender, gender identity, sex characteristics, and sexual orientation). They should make sure that one construct is not used as a proxy for another⁶⁷. In relation to this, they should duly consider [Guidelines No 4.1. to 4.4. of this guidance document](#).

⁶⁷ National Academies of Sciences, Engineering, and Medicine (2022), [Measuring Sex, Gender Identity, and Sexual Orientation](#), The National Academies Press, Washington DC, p. 6.

- They should duly consider differences in the requirements towards different data sources (administrative registers, censuses, survey data, etc.) and the feasibility of/leeway for applying uniform categories in different data sources.
- They should: (i) document the basic criteria and classification procedures in relation to SOGIESC; (ii) document their suggested combination of these criteria and procedures; and (iii) inform data users about the concepts these criteria and procedures are based on.
- They should take due account of Guideline No 9 ‘Enhance validity and reliability of equality data’ and Guideline No 11 ‘Improve comparability of equality data’ of the [guidelines on improving the collection and use of equality data](#).
- They should build on the findings from the identification exercise (Guideline No 3.2. of this guidance document) to identify commonalities and discrepancies between the categories/definitions applied across data sources and time.
- They should take due consideration of Guideline No 3.4 of this guidance document on taking information on SOGIESC into account in all policies to incorporate: (i) agreed-upon SOGIESC definitions/categories for disaggregation and the necessary combinations of these definitions/categories; and (ii) questions measuring discrimination/crime victimisation based on SOGIESC into EU and national surveys.
- They should strengthen and take into account in all policies the regular use of agreed-upon SOGIESC categories to disaggregate all data collected for the European Statistical System.
- They should strengthen cooperation between different national and international equality bodies to improve and coordinate their data collection based on SOGIESC.

3.4. Mainstream information on SOGIESC in all national and EU surveys

Data collectors and data users need multiple and complementary sources of equality data to enable evidence-based policy making and assess the state of equality for LGBTIQ people. A proper impact assessment of policies in the field of equality and non-discrimination requires regular data collection on all grounds of discrimination and for all areas of life covered by the relevant legislation. However, effort spent collecting more and better equality data should be balanced and guided by the EU’s data-protection rules, requiring states to collect personal data only for ‘specified, explicit and legitimate

purposes’ and only ‘relevant and limited to what is necessary in relation to the purposes for which [personal data] are processed’⁶⁸.

3.4.1. Rationale

3.4.1.1. Censuses

Censuses, large-scale national and international surveys, and administrative data can be valuable sources of equality data. However, they are often limited in the data they collect.

Censuses provide information on the size, composition, and spatial distribution of the population, as well as on its socioeconomic and demographic characteristics. Censuses collect data on each person of the general population based on registers, questionnaires, or mixed data sources.

The collection of personal data broken down by personal characteristics is protected by constitutional standards, EU data-protection law, and the [Charter of Fundamental Rights of the EU](#). Current EU legislation requires every census to collect and ask for information about **sex**. The question is mandatory and the response categories that are allowed are ‘male’ or ‘female’⁶⁹. This has created a problem for the LGBTIQ population, since some trans, gender-diverse, and intersex people do not fall within a binary framework of sex and/or gender. The limited options are thus exclusionary and will impact the validity of the data. Moreover, where countries legally provide for gender markers outside the binary on official documents (as is the case in Austria⁷⁰, Germany⁷¹, Denmark⁷², Italy⁷³, Malta⁷⁴, and the Netherlands⁷⁵), binary options may require respondents to make a false statement about their official legal status. In their censuses, Germany⁷⁶ includes the categories ‘diverse’ and ‘not stated in birth register’, and Malta⁷⁷ includes an ‘other’

⁶⁸ Article 5(1)(c) and Recital (39) of the [General Data Protection Regulation](#) (GDPR).

⁶⁹ Eurostat (2019), [EU legislation on the 2021 population and housing censuses: Explanatory notes](#), Publications Office, Luxembourg, p. 66.

⁷⁰ Ministry of Finance (2023), [Changing Assigned Sex](#). Note that these gender markers are limited to intersex people.

⁷¹ Federal Anti-Discrimination Agency (Antidiskriminierungsstelle des Bundes) (2023), [Male – Female – Diverse: The ‘third option’ and the General Act on Equal Treatment](#). Whether these gender markers are open to non-intersex non-binary people is unclear.

⁷² TGEU (2014), Denmark: [X in Passports and New Trans Law Works](#).

⁷³ Baska, M and Milton, J. (2022), [Italian court recognises non-binary person’s gender identity in historic first](#), Pink News.

⁷⁴ Human Rights Directorate (2020), [Legal Gender Recognition and Bodily Integrity](#).

⁷⁵ Government of the Netherlands (2022), [Changing the registration of gender on official documents](#).

⁷⁶ Statistische Ämter des Bundes und der Länder (2022), https://www.zensus2022.de/DE/Wer-wird-befragt/Musterfragebogen_Uebersicht/Fragebogen.html?nn=610864 \ "Haushaltebefragung

⁷⁷ National Statistics Office – Malta, [Census](#).

category for the question on sex, but only insofar as this corresponds to the record on the birth register. Austria, Denmark and the Netherlands rely on administrative data rather than conducting population censuses. In Italy, however, the 2022 census only provided for ‘male’ and ‘female’⁷⁸.

Overall, censuses rarely collect information about ‘special categories of personal data’ such as those related to SOGIESC⁷⁹. Not including questions on SOGIESC in censuses has wider implications, as censuses regularly build the sampling frame for different types of surveys. Therefore, it is important to reiterate that data-protection rules do not stand in the way of collecting equality data for statistical purposes, and that the principles of data protection do not apply to anonymous information or to personal data made anonymous in such a manner that the data subject is not identified or is no longer identifiable⁸⁰. So far, Malta is the only EU Member State to have included specific questions on sexual orientation and gender identity in its census⁸¹. Outside the EU, the United Kingdom, Canada and Australia have included a question on gender identity in their census, and Nepal, India and Pakistan have revised their censuses to include non-binary sex or gender categories⁸².

Where there are no specific SOGIESC questions, some idea of the lived realities of LGBTIQ people can be gained by using proxy information, for example asking questions about: (i) the census respondent’s partner; (ii) the census respondent’s relationship to other household members; (iii) the sex of the partner/household members in the census. Collecting such proxy information has increased in recent years, mainly because in an increasing number of countries same-sex couples can register their relationship or can legally marry. The EU regulations for the 2021 census⁸³ make it compulsory for Member States to provide data on the categories ‘in an opposite-sex marriage or registered partnership’ or ‘in a same-sex marriage or registered partnership’. However, the regulations also abolish the distinction between opposite-sex partners and same-sex partners for the category ‘partners in a consensual union⁸⁴’, making it more difficult

⁷⁸ ISTAT (2022), [Popolazione e Abitazioni: Documentazione](#).

⁷⁹ Under Article 9 of the [GDPR](#), special categories of personal data comprise ‘personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership’, as well as ‘genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person’s sex life or sexual orientation’.

⁸⁰ See European Commission (2021), [Round table on Equality Data in September 2021](#) and Principle 4 of the [Guidance note on the collection and use of equality data based on racial or ethnic origin](#).

⁸¹ National Statistics Office – Malta, [Census](#).

⁸² National Academies of Sciences, Engineering, and Medicine (2022), [Measuring Sex, Gender Identity, and Sexual Orientation](#), The National Academies Press, Washington DC, p. 114.

⁸³ [Regulation \(EU\) 2017/543](#).

⁸⁴ Eurostat (2019), EU legislation on the 2021 population and housing censuses – Explanatory Notes, Publications Office, Luxembourg, pp. 66-67.

to gather information on the number of same-sex couples who are not married or in a registered partnership.

Although data on same-sex partnerships are often used as a proxy for sexual orientation, the information they provide is limited. They do not capture people who are not currently in a relationship or who are not willing to disclose – or comfortable disclosing – their relationship. They also fail to capture respondents who are currently in an ‘opposite-sex’ relationship, which will particularly lead to an undercount of bisexual and pansexual people, as well as of trans people who have not had their gender legally registered.

Box A 4 – SOGIESC Questions in Malta’s Census

Q5

Indicate your registered sex (Mark one circle only)

- Male
- Female
- Other

Q7

Is the gender you identify with the same as your registered sex?

- Yes (Go to Q9)
- No

Q8

Specify the gender you identify with

(Open text field)

Q9

Which of the following best describes your sexual orientation? (Mark one circle only)

- Straight/Heterosexual
- Gay or Lesbian
- Bisexual
- Other (Specify)

Notes for consideration:

The census has some limitations. In Q5, the ‘other’ category refers specifically to registered sex, so this question should not be used to indicate the prevalence of non-

binary or intersex people. Q7 asks whether the respondent's gender differs from their registered sex instead of their sex assigned at birth. This question will therefore only capture trans respondents who have not had their gender legally recognised. Q9 is also limited in the options it presents, for instance by not capturing asexual people.

3.4.1.2. Large-scale surveys

Although SOGIESC data are still not collected through censuses in most countries, the situation appears to be changing for population-based surveys. **Large-scale surveys** are a powerful tool, which make it possible to collect a comprehensive set of equality data. General population surveys are based on a sample that is representative of the total population, for example the [Labour Force Survey](#), the [EU Survey on Living Conditions](#) (EU-SILC), or national health surveys.

In a recently published report, the Organisation for Economic Cooperation and Development (OECD) states that it is important to collect **representative** data on SOGIESC⁸⁵. Representativeness shows the extent to which the characteristics of a sample (such as its composition in relation to SOGIESC) corresponds to the characteristics of the population from which the sample has been drawn, following established rules. As stated in the [Guidelines on improving the collection and use of equality data](#), 'representativeness can be negatively affected if the sample size is too small – [that is], when it does not include a sufficient number of persons belonging to a targeted population group, or if specific subgroups of the population are systematically excluded from the data collection in the first place (because of language difficulties, impairments etc.)⁸⁶.'

Several EU and European Economic Area countries have already included questions on SOGIESC in population-based surveys⁸⁷. Belgium⁸⁸, Denmark⁸⁹, Finland⁹⁰, France⁹¹,

⁸⁵ OECD (2020), [Over the Rainbow? The Road to LGBTI Inclusion](#), OECD Publishing, Paris, p. 29.

⁸⁶ European Commission (2018), [Guidelines on improving the collection and use of equality data](#), Publications Office, Luxembourg, p. 21.

⁸⁷ A full list of nationally representative surveys asking about sexual orientation and gender identity conducted by public institutions in OECD countries as of 2018 is available in the OECD (2019), [Society at a Glance 2019: OECD Social Indicators](#), OECD Publishing, Paris, pp. 52-56. Further information on data collected by Member States is also detailed in European Commission (2017), [Data collection in relation to LGBTI People](#), Publications Office, Luxembourg.

⁸⁸ Institute for the Equality of Women and Men (Instituut voor de gelijkheid van vrouwen en mannen) (2022), [Enquête #YouToo?](#).

⁸⁹ Statens Serum Institut (2019), [Projekt SEXUS](#).

⁹⁰ Finnish Institute for Health and Welfare (2021), [School Health Promotion Study](#).

⁹¹ Institut National D'Études Démographiques (2020), [VIRAGE](#).

Ireland⁹², the Netherlands⁹³, Sweden⁹⁴ and Norway⁹⁵ have conducted surveys which ask about both sexual orientation and gender identity. Germany⁹⁶, Italy⁹⁷, Spain⁹⁸ and Iceland⁹⁹ have included questions on sexual orientation. Italy has implemented a pilot survey on discrimination which includes questions on sexual orientation and gender identity. Portugal also aims to include questions on sexual orientation, gender identity, and sex characteristics in an upcoming survey on gender-based violence¹⁰⁰. The inclusion of sex characteristics would be particularly important since no general population survey asks about whether respondents are intersex/have a variation of sex characteristics¹⁰¹.

3.4.1.3. Administrative data

Administrative data are distinct from data collected solely for statistical purposes (data collected solely for statistical purposes include census data and survey data). Administrative data are any data collected by public authorities. Administrative data primarily serve some purpose other than the production of statistics¹⁰², such as paying taxes, or registering births and deaths.

Unlike data from surveys, administrative data often cover entire populations (e.g. all taxpayers), and they might reveal information that otherwise would never have been collected. Administrative data may be collected more often than data collected for statistical purposes. For example, on the situation of LGBTIQ people in Europe, it is common to find administrative data from records on: (i) the number of people entering a legally recognised same-sex partnership (whether marriage or another status); and (ii) the number of individuals granted permission to change the record of their sex as assigned by birth.

While they can be useful, administrative data present many challenges and can be problematic as a source of SOGIESC data. Efforts to collect administrative data are not coordinated and tracked in the same way as efforts to collect statistical data. In some

⁹² Central Statistics Office (2019), [General Household Survey](#).

⁹³ Statistics Netherlands (2021), [Safety Monitor](#).

⁹⁴ Public Health Agency of Sweden (2022), [Public Health Survey](#).

⁹⁵ Statistics Norway (2021), [Quality of Life in Norway](#).

⁹⁶ German Institute for Economic Research (DIW Berlin) (2022), [Socio-Economic Panel](#).

⁹⁷ Istituto Nazionale di Statistica (2015), [Survey on discriminations by gender, sexual orientation and ethnic origin](#).

⁹⁸ Instituto Nacional de Estadística (2006), [Health and Sexual Behaviour Survey](#).

⁹⁹ Directorate of Health (2022), [Health and Wellbeing Study](#).

¹⁰⁰ OHCHR (2018), [Report of the Independent Expert on protection against violence and discrimination based on sexual orientation and gender identity: Responses by the Portuguese Authorities](#), p. 3.

¹⁰¹ OECD (2019), [Society at a Glance 2019: OECD Social Indicators](#), OECD Publishing, Paris, p. 15.

¹⁰² European Commission (2016), [European Handbook on Equality Data \(2016 revision\)](#), Publications Office, Luxembourg, Section 2.1.1.

cases, data at group level are not available at all¹⁰³. The variables may also be limited, for instance where data on legal gender are collected but data on gender identity are not. In many cases, SOGIESC data are being collected without there being a legal ground or rationale to do so (for example, when applying for a library card), which may be perceived as overly invasive and, in some cases, deter LGBTIQ people from accessing a particular service¹⁰⁴.

3.4.1.4. Use of proxy variables

A proxy variable is an indicator that is used to give an approximate representation of a phenomenon, related to the variable, but not directly measurable in a more objective way. For a variable to be a valid proxy, it must be closely related to the variable it replaces.

The [OHCHR's Human Rights-Based Approach to Data](#) states that populations must be self-identifying, meaning that population parameters cannot be imposed by an outside party or assigned through imputation or proxy. The primary concern with proxy reporting is that the resulting data will not reflect how a person would have identified if they could have responded for themselves¹⁰⁵. Therefore, collecting information on the self-identification of individuals in terms of SOGIESC takes precedence over proxy information.

Many Member States regularly collect population or census data that can be used as an imperfect proxy for SOGIESC data (for example in questions that ask about the sex of the respondent and their partner as a proxy for LGBTIQ relationships). This type of information is likely to be inaccurate for several reasons. For example, it fails to account for trans people who have not legally registered their gender, or for bisexual or pansexual people in a relationship with someone of a different sex. Such information should therefore only be used in the absence of any other means of collecting SOGIESC data.

The use of proxy reporting can be necessary in some cases (such as when collecting data on infants) but is likely to result in 'undercounts'. For example, if statistics are based only on prenatal screening (as a proxy for individuals with a variation of sex characteristics), they will miss those individuals whose variation of sex characteristics is revealed later in

¹⁰³ For instance, in health databases, intersex people are not specifically defined as such in health terms for statistical analysis, and are therefore not included in databases. See: Republic of Slovenia, Advocate of the Principle of Equality (2022), [Special report. The Situation of Intersex People in Medical Procedures](#), Advocate of the Principle of Equality, Ljubljana.

¹⁰⁴ Ashley, F. (2021), [Recommendations for Institutional and Governmental Management of Gender Information](#). *New York University Review of Law & Social Change*, 44(4):489-528. pp. 504, 514-515.

¹⁰⁵ National Academies of Sciences, Engineering, and Medicine (2022), [Measuring Sex, Gender Identity, and Sexual Orientation](#), The National Academies Press, Washington DC, p. 50.

life. Similarly, standard binary measures of sex are an inadequate proxy for the primary measurement of gender and sex characteristics, especially among sexually diverse and gender-diverse populations¹⁰⁶. If data on the number of changes in legal gender registration are used as a proxy for the size of the trans population (or for the proportion of the population that is trans), this can lead to undercounts since not all trans people can or wish to change their legal gender registration. It could also lead to inaccuracies, since it may include intersex people who do not identify as trans. Weighting based on estimates from other countries on the proportion of trans people seeking legal gender recognition (LGR) could improve the reliability of proxy data. However, this is only possible if the countries have comparable legislation, since legal hurdles can significantly impact the number of people seeking LGR.

Box A 5 – Use of proxies in the EU-SILC household survey

The **EU-SILC household survey** contains a question about sex and a question about the marital status of all household members¹⁶⁶. A note about same-sex couples is made to clarify that registered partnerships also count as a marital status. The combination of both questions could assess the prevalence of same-sex couples, although this information will only capture same-sex couples who are married or in a registered partnership.

[Sex] (*question repeats for all household members*)

- Male
- Female

[NOTE: All current household members aged 16 and over.]

[Marital Status] (*question repeats for all household members*)

- Never married
- Married
- Separated
- Widowed
- Divorced

[NOTE: All current household members aged 16 and over. Some countries have a legal framework for registering partnerships (in most countries these are same-sex partnerships, and they have a legal status parallel to married couples). This information must also be treated in a harmonised way and it is proposed to treat these legally recognised partners as married and to classify them under code 2 when the relationship still exists, or as 3-5 as appropriate (legal separation or death of one of the partners or so on).]

¹⁰⁶ National Academies of Sciences, Engineering, and Medicine (2022), [Measuring Sex, Gender Identity, and Sexual Orientation](#), The National Academies Press, Washington DC, p. 44.

Eurostat and Member States often do not disaggregate these data to estimate the prevalence of same-sex couples, even though the information is available. In the absence of better data being available, such disaggregation can be a useful tool.

3.4.2. Guidance

Member States and relevant institutions and bodies involved in – or responsible for – collecting data based on SOGIESC are encouraged to take information on SOGIESC into consideration in all policies. The bullet points below set out in more detail some considerations that Member States should bear in mind.

- If censuses do not provide information on SOGIESC, sample surveys should complement the censuses and include questions/modules capturing SOGIESC characteristics on a regular basis.
- When developing surveys or when collecting data as part of the [European Statistical System](#) through the [Labour Force Survey](#) and [EU-SILC](#), Member States should, where it is appropriate, systematically incorporate information based on SOGIESC. Adding optional questions on self-identification is also a possibility. Information based on SOGIESC could also be incorporated by developing and running thematic survey modules to cover, for example: (i) experiences of discrimination in different areas of life; (ii) bias-motivated harassment and violence; (iii) rights awareness; and (iv) reporting incidents of discrimination.
- To capture intersecting inequalities, Member States should introduce individual variables and survey questions that specifically relate to measuring inequalities and perceived discrimination on multiple grounds, including SOGIESC, in different areas of life.
- Where relevant and possible, Member States should avoid restricting their collection of SOGIESC data to adults only. Although minors are afforded greater protection under several of the major data-privacy laws (including the [GDPR](#)), data from European surveys prove the importance of also collecting data on people younger than 18. For example, the FRA’s LGBTIQ II survey collected data from young people aged 15-17¹⁰⁷. Although the [GDPR](#) recommendation for the age of consent to responding to surveys is 16, Member States may enact laws that lower the age of consent to 13 (but not under 13), according to Article [8\(1\)](#)¹⁰⁸.

¹⁰⁷ EU Agency for Fundamental Rights (2020), [A long way to go for LGBTI equality](#), Publications Office, Luxembourg.

¹⁰⁸ Article 8(1) of the [General Data Protection Regulation](#).

- Member States should consider using a question on minority-group affiliation when SOGIESC aspects cannot be measured directly in a survey. If it is impossible to add at least four additional questions (asking about sexual orientation, gender identity, gender expression, and/or variations of sex characteristics), an alternative but less exact approach can be found by asking the respondents to what extent they are part of other minority groups, providing a list of relevant examples which includes being LGBTIQ. An explanation or definition of what a minority group is (in terms of SOGIESC) should be provided. The definitions provided in [the glossary in the annex of this guidance document](#) can serve as a starting point for such an explanation.
- Member States should increase the number of people who otherwise would not be adequately represented in the survey samples due to their small numbers in the general population by using booster samples.
- If there are too few respondents in a category for meaningful analysis (such as a limited number of non-binary respondents compared to men and women), Member States should document the fact of having ‘unreliably’ small sample/cell size in technical reports and state why it is not possible to include certain categories in the main analysis. This increases visibility and prevents overly simplistic interpretations.
- Member States should promote the inclusion of SOGIESC in longitudinal surveys (surveys which repeatedly collect data on the same individuals over a longer period, e.g. panel surveys), to make it possible to assess individual change over time.
- When the purpose for collecting special categories of personal data is not clear, it is recommended that Member States state the legal ground and purpose for which the data are being collected. As stated in Article 6 of the General Data Protection Regulation (GDPR), the legal bases of data collection are consent; contract; legal obligation; vital interests; public task; or legitimate interest. Overly invasive questions that do not meet any of the above-mentioned purposes could increase LGBTIQ exclusion by dissuading LGBTIQ people from accessing particular services. The data collected should be necessary, proportionate and follow the overarching principle of doing no-harm. Therefore, if there is no legal ground to collect SOGIESC data, it is recommended to refrain from collecting data on SOGIESC in administrative data.
- Often the gender of a person is collected simply to make it possible to address this person in a correct manner. In these cases, ask the person for the correct salutation instead, including gender-neutral options. In this way, the person can inform you about their preferred salutations without sharing personal information. For

example, a question could be: ‘Which salutation do you prefer? Mrs, Mr, Mx or No salutation, just my first name’.

- Member States should use proxy variables/categories in existing data sources where the human rights-based principle of self-identification cannot be applied.
- For household surveys, in which one household member is providing information on other household members (a so-called proxy informant), the collection of SOGIESC information should be avoided to the extent possible¹⁰⁹, as it relates to sensitive data which requires the consent of the intended person. Therefore, SOGIESC questions should be asked only to the sampled target respondent. Non-statistical organisations and other research institutions may apply, as appropriate, their established practices for substituting households or respondents who refuse to take part in the interview or are unavailable.
- Data collection should enable respondents to self-identify whenever possible, and any proxy reporting should reflect what is known about how a person self-identifies.
- All data-collection activities should require well-informed consent from potential respondents, with no penalty for those who opt out of sharing personal information about themselves or other household members.

It should be clearly stated that data gathered through a proxy may not reflect how a person would have identified if they had been given the opportunity to self-identify and is likely to result in undercounts for marginalised populations. This limitation should be acknowledged in any analysis of such data¹¹⁰.

3.5. Mainstream an intersectional approach in the collection and use of SOGIESC data

‘Discrimination is often multidimensional and only an intersectional approach can pave the way to sustainable and respectful changes in society.’

¹⁰⁹ Survey module to measure experiences of discrimination developed by the Praia Group Task Team on Non-discrimination and Equality.

¹¹⁰ National Academies of Sciences, Engineering, and Medicine (2022), [Measuring Sex, Gender Identity, and Sexual Orientation](#), The National Academies Press, Washington DC, p. 50.

National equality laws in the EU rarely include provisions that relate to multiple or intersectional discrimination. Where they do include such provisions, there are discrepancies between countries, as the European network of legal experts on gender equality and non-discrimination notes¹¹¹. Regardless of legal arrangements and provisions, research data and evidence highlight that intersectional discrimination is a recurrent reality for many people across the EU.

For example, survey data from the FRA show that people experience discrimination based on a variety of grounds and as a combination of grounds, such as age and ethnicity, or religion and ethnicity (including migration background)¹¹². Findings from the FRA's LGBTI Survey II show that sexual orientation was named as an additional ground of discrimination experienced by 60% of intersex and 46% of trans people; while 46% of bisexual women and 29% of lesbian women mentioned sex as an additional ground of discrimination they had experienced. Out of those LGBTI respondents who perceived themselves as an ethnic minority (including migrants), 35% reported discrimination occurring due to ethnic origin or migration background. Of those who perceived themselves as belonging to a religious minority, 28% mentioned religion or belief. Of those who perceived themselves as a minority in terms of disability, 33% mentioned disability as an additional ground for the last incident of harassment they had experienced. Of intersex respondents, 41% also mentioned gender identity and expression as an additional ground of discrimination that they had experienced¹¹³.

All recent EU equality strategies and action plans address intersectionality by calling for actions to tackle intersecting and multiple forms of discrimination. In the EU anti-racism action plan, the European Commission encourages Member States to 'swiftly reach an agreement on the 2008 Commission proposal to implement equal treatment between persons irrespective of religion or belief, disability, age or sexual orientation'¹¹⁴. In addition, practitioners in the field recognise that addressing discrimination from the

¹¹¹ European network of legal experts in gender equality and non-discrimination (2020), [A comparative analysis of non-discrimination law 2019](#), Publications Office, Luxembourg, pp. 36-37.

¹¹² FRA (2021), [Equality in the EU 20 years on from the initial implementation of the equality directives](#), Publications Office, Brussels, p. 42.

¹¹³ FRA (2020), [A long way to go for LGBTI equality](#), Publications Office, Luxembourg, p. 45. For more intersectional analysis from findings of the FRA's second LGBTI survey on LGBTI people in the EU and North Macedonia and Serbia, see: Calderon-Cifuentes, PA. (2021), *Trans Discrimination in Europe. A TGEU analysis of the FRA LGBTI Survey 2019*; ILGA Europe, EuroCentralAsian Lesbian Community (EL*C) (2022), [Intersections, Lesbian Briefing](#), and ILGA-Europe and IGLYO (2022), [Intersections, Youth Briefing](#). Forthcoming briefings will address the intersectional experiences of intersex people, religious and ethnic minorities, older LGBTI people, trans people, and others.

¹¹⁴ European Commission (2020), [A Union of equality – EU anti-racism action plan 2020-2025, COM\(2020\) 565 final](#), Brussels, p. 6.

perspective of a single ground fails to adequately tackle the different ways in which unequal treatment can manifest itself¹¹⁵.

Box A 6 – Multiple and intersectional discrimination

‘Multiple discrimination’ describes discrimination that takes place on the basis of several grounds operating separately.

‘Intersectional discrimination’ describes a situation in which several grounds operate and interact with each other at the same time in such a way that they are inseparable and produce specific types of discrimination¹¹⁶.

3.5.1. Rationale

The complexity of LGBTIQ identities and experiences cannot be captured by a single ground, as they encompass multiple, intersecting characteristics that should be recognised¹¹⁷. Understanding intersectional experiences is therefore crucial to the design of effective measures and policies to adequately capture situations in which different grounds of discrimination intersect or act in combination with each another.

3.5.2. Guidance

Member States and relevant institutions/bodies involved in or responsible for collecting data based on SOGIESC are encouraged to take an intersectional approach in the collection and use of SOGIESC data. They are also encouraged to involve communities in the process. The bullet points below contain further suggestions for Member States.

- Member States and relevant institutions/bodies should develop strategies to: (i) adequately capture situations in which different grounds of discrimination intersect or act in combination with each another; and (ii) take into consideration multiple and intersectional discrimination in data analysis. To this end, EU Member States should use a comprehensive set of tools to collect equality data, including large-scale quantitative surveys covering different population groups and grounds of discrimination¹¹⁸.

¹¹⁵ FRA (2017), [Fundamental Rights Report 2017](#), Publications Office, Luxembourg, p. 58.

¹¹⁶ FRA and European Court of Human Rights (2018), [Handbook on European non-discrimination law – 2018 edition](#), Publications Office, Luxembourg, p. 59.

¹¹⁷ See also EIGE (2022), [Intersectionality](#).

¹¹⁸ FRA (2021), [Equality in the EU 20 years on from the initial implementation of the Equality Directives](#), Publications Office, Brussels, p. 20.

- Member States and relevant institutions/bodies should consider intersections both within the LGBTIQ community (e.g. a trans lesbian woman may experience specific discrimination separate from that faced by cisgender lesbians or trans people who are not lesbians), and with non-SOGIESC grounds of discrimination (age, ethnicity, disability, religion etc.).
- Member States and relevant institutions/bodies should use a comprehensive set of tools to collect equality data¹¹⁹, including large-scale quantitative surveys covering different population groups and different grounds of discrimination.
- Member States and relevant institutions/bodies should incorporate the measurement of multiple and intersectional discrimination into all existing systems for collecting data. For example, data collected on disability should ideally include information on SOGIESC and other potential characteristics that might lead to a greater risk of discrimination based on multiple factors or intersecting inequalities.
- For data collection through surveys, Member States and relevant institutions/bodies should ensure a sufficient net sample size to make possible the further disaggregation of data and intersectional analysis. To achieve this, they should consider using a suitable sampling method. Different sampling strategies can be applied to obtain an adequate sample size that fully represents the variations within the LGBTIQ target population.

3.6. Provide sufficient budget for regular surveys to collect equality data on SOGIESC

3.6.1. Rationale

Governments should ensure that data on equality is collected regularly to monitor compliance with commitments under EU law¹²⁰, even when it is not possible to collect

¹¹⁹ For more information, please consider the [European Handbook on Equality Data](#), as well as the [Guidelines on improving the collection and use of equality data](#).

¹²⁰ As mentioned in the [European Handbook on Equality Data](#), the Employment Equality Directive “expressly allow[s] – but do[es] not require – the Member States to maintain or introduce rules that allow discrimination to be established by ‘any means including on the basis of statistical evidence’”. Moreover, Governments themselves have on several occasions recognised the need to collect equality data (Copenhagen Declaration on Social Development and Programme of Action of the World Summit for

administrative data. In these cases, sufficient budgetary resources should be allocated at the national/regional level to collect equality data on SOGIESC through regular national/regional surveys conducted by a competent authority.

3.6.2. Guidance

In the absence of regular collection of SOGIESC data through administrative sources or surveys, Member States should provide **sufficient budgetary resources to collect such data every 4-5 years through a comprehensive national survey.**

The survey should achieve the four goals set out in the bullet points below.

- It should reach out to all LGBTIQ groups and actively involve groups that work specifically with – and for – specific groups that fall under the LGBTIQ category. It should also reach out to and involve multiply marginalised groups such as disabled LGBTIQ people, LGBTIQ people of colour, and LGBTIQ sex workers.
- It should aim to represent the diversity of this community and aim to cover populations and groups defined as ‘hard to reach’, including those living in poverty, homeless people, sex workers, migrants, asylum seekers and people who are not ‘out’ in their everyday life. This would also mean translating the survey tool (questionnaire and other fieldwork material) into languages relevant to these hard-to-reach groups, as well as involving civil society organisations¹²¹.
- It should be based on a sample size large enough to make it possible to both conduct more detailed data analysis and identify geographic differences/disparities including intersectional disparities. This means that the sample size should be large enough to make it possible to further analyse the data on: geographical distribution; age; racial or ethnic origin; disability; religion; sex assigned at birth; sexual orientation; gender identity/expression; sex characteristics; etc.

Social Development of 19 April 1995, A/CONF.166/9; Durban Declaration and Plan of Action; Vienna Declaration and Programme of Action of 12 July 1993, A/CONF.157/23; World Programme of Action concerning Disabled Persons; Standard Rules on the Equalisation of Opportunities for Persons with Disabilities).

¹²¹ UNECE (2020), [Poverty Measurement: Guide to Data Disaggregation](#), United Nations, New York and Geneva.

- It should be able to produce reliable and valid data for all LGBTIQ groups, i.e. by using the up-to-date survey measurement guidelines and choosing a sample design that also reaches hard-to-reach populations.

Regular surveys that collect equality data related to SOGIESC can be conducted as an independent survey with a specific focus on populations that self-identify as LGBTIQ, or as part of a general equality survey (i.e. a survey that also addresses other protected grounds). This can be done at the national (and, where relevant, regional) level by a competent authority, statistical institute, equality body or research organisation with the relevant capacity and expertise to conduct research.

3.7. Facilitate the effective use of SOGIESC equality data to strengthen and improve LGBTIQ equality

3.7.1. Rationale

Reliable, valid and comparable SOGIESC data are a prerequisite for evidence-based policies to foster equality and non-discrimination at the EU, national, and regional levels. SOGIESC data are seen as an essential element of all phases of a policy cycle to improve equality for LGBTIQ people. These SOGIESC data should therefore be systematically used for the development, implementation, and monitoring of both dedicated LGBTIQ policies/measures and wider equality frameworks at EU/national level.

The US National Science and Technology Council provides a detailed list of potential research questions that can be addressed with SOGIESC data in the areas of healthcare, housing, education, economic security, safety, and justice¹²². Some illustrative examples are set out in the three bullet points below.

- ‘What is the prevalence of homelessness among LGBTIQ+ adults? How does this compare to their non-LGBTIQ+ peers?’¹²³

¹²² Subcommittee on Sexual Orientation, Gender Identity, and Variations in Sex Characteristics (SOGI) Data and Subcommittee on Equitable Data of the National Science and Technology Council (2023), [Federal Evidence Agenda on LGBTIQ+ Equity](#), *United States Government*, pp. 18-33.

¹²³ Subcommittee on Sexual Orientation, Gender Identity, and Variations in Sex Characteristics (SOGI) Data and Subcommittee on Equitable Data of the National Science and Technology Council (2023), [Federal Evidence Agenda on LGBTIQ+ Equity](#), *United States Government*, p. 23.

- ‘How prevalent are various forms of job-related discrimination, harassment, or retaliation against LGBTQI+ people, such as discrimination in hiring, in wages, in equal employment opportunity, in fair treatment, in promotion or advancement, or in termination?’¹²⁴
- ‘What institutional contexts, policies, or practices promote a positive academic environment and contribute to higher rates of LGBTQI+ student retention and graduation? What individual-level, family-level, or community-level protective factors do LGBTQI+ people employ that help them to succeed in education and the workforce?’¹²⁵

The questions also provide examples of how to capture intersectional discrimination, such as the question ‘What barriers do LGBTQI+ minors and LGBTQI+ adults with disabilities face in accessing health services that require participation from guardians?’¹²⁶. By answering these questions, SOGIESC data can play a crucial role in policy development and evaluation.

3.7.2. Guidance

To facilitate the effective use of SOGIESC equality data, Member States and the relevant institutions and actors involved in or responsible for the collection of SOGIESC data are encouraged to take the actions set out in the bullet points below.

- They should use SOGIESC equality data to measure both: (i) the impact of LGBTQI-specific laws and policies; and (ii) the impact of wider equality and non-discrimination policies and measures on LGBTQI people.
- They should use SOGIESC equality data to: (i) report on progress towards LGBTQI equality at the national level; and (ii) support reporting efforts on the implementation of the EU’s [LGBTIQ equality strategy 2020-2025](#) and other relevant

¹²⁴ Subcommittee on Sexual Orientation, Gender Identity, and Variations in Sex Characteristics (SOGI) Data and Subcommittee on Equitable Data of the National Science and Technology Council (2023), [Federal Evidence Agenda on LGBTQI+ Equity](#), *United States Government*, p. 25.

¹²⁵ Subcommittee on Sexual Orientation, Gender Identity, and Variations in Sex Characteristics (SOGI) Data and Subcommittee on Equitable Data of the National Science and Technology Council (2023), [Federal Evidence Agenda on LGBTQI+ Equity](#), *United States Government*, p. 26.

¹²⁶ Subcommittee on Sexual Orientation, Gender Identity, and Variations in Sex Characteristics (SOGI) Data and Subcommittee on Equitable Data of the National Science and Technology Council (2023), [Federal Evidence Agenda on LGBTQI+ Equity](#), *United States Government*, p. 22.

instruments, such as other EU equality strategies¹²⁷ and the European Pillar of Social Rights.

- They should use SOGIESC equality data and indicators to inform processes and policies at the wider European and UN level. SOGIESC data could inform countries: (i) when reporting on relevant recommendations made by the Council of Europe¹²⁸; (ii) when reporting on progress towards achieving the 2030 Agenda with the UN's Sustainable Development Goals (SDGs)¹²⁹; (iii) when engaging with other international human-rights monitoring mechanisms, such as UN treaty bodies and special procedures (e.g. the independent expert on protection against violence and discrimination based on sexual orientation and gender identity); or (iv) when following up on recommendations made during the UN's Universal Periodic Reviews.
- They should use SOGIESC data to inform all stages of policy development. This could start with: (i) using SOGIESC data as a baseline for developing policies; (ii) designing indicators for monitoring purposes; and (iii) setting benchmarks. Later on in policy development they should use SOGIESC data to inform mid-term evaluations. And towards the later stages of policy implementation, they should use SOGIESC data to populate indicators, assess the effective implementation of measures, assess the outcomes of these measures, and monitor trends.
- They should ensure data can be disaggregated by SOGIESC and other relevant characteristics (age, disability, ethnic or racial origin, religion, and socioeconomic status) to reflect, analyse and report on differences in experiences in line with Guideline No 3.5 of this guidance document on taking an intersectional approach in the collection and use of data.
- They should use illustrative examples, such as anonymised quotes and personal stories collected during research to help contextualise research findings and increase the impact and outreach of the communication about the results.

¹²⁷ Promoting a Union of equality, the European Commission has adopted several policy instruments in addition to the [LGBTIQ equality strategy 2020–2025](#), including: the [EU anti-racism action plan 2020–2025](#); the [EU Roma strategic framework for equality, inclusion and participation 2020–2030](#); the [action plan on integration and inclusion 2021–2027](#); the [EU gender equality strategy 2020–2025](#); the [European Pillar of Social Rights action plan](#); and the [strategy for the rights of persons with disabilities 2021–2030](#).

¹²⁸ For example, on the Committee of Ministers' [Recommendation CM/Rec\(2010\)5 on measures to combat discrimination on grounds of sexual orientation or gender identity](#); or the Council of Europe's ECRI's [Country Monitoring](#). For more information, see the thematic page of the [Sexual Orientation and Gender Identity \(SOGI\) Unit](#).

¹²⁹ Target 17.18 of the [SDGs](#) requires States to 'increase significantly the availability of high-quality, timely and reliable data', disaggregated by both gender and 'and other characteristics relevant in national contexts'.

- They should ensure the timely dissemination of the collected data and make it freely accessible both online and offline to interested communities/researchers/policy makers.
- They should present the data in accessible and understandable language and formats, making it possible for different stakeholders and groups to use (for example, by LGBTIQ-representative organisations in their advocacy work, or by equality bodies in their monitoring and reporting on discrimination issues).
- They should tailor the presentation of the data depending on the audience (e.g. policy makers, experts, data specialists, the media) and ensure it is properly contextualised.
- They should ensure, where appropriate, that explanations and interpretations of the data are provided to avoid any misinterpretations and misuse of the data or of the conclusions drawn.
- They should document the data collection and all metadata according to high professional standards. They should draw up a technical report to document: (i) the survey methodology; (ii) the fieldwork (including the response rate achieved); (iii) the data-processing activities; and (iv) any relevant contextual information (such as population data disaggregated by relevant socio-demographic characteristics). The use and reporting of LGBTIQ data should always be contextualised with caveats and by providing documentation explaining how the data were collected.

4. How to collect information on SOGIESC

As stated in the [OHCHR's Human Rights-Based Approach to Data](#), any data-collection effort should include the means for free, active, and meaningful participation of relevant stakeholders, in particular the most marginalised population groups. When collecting data on SOGIESC, it is important to use community-appropriate terminology and ensure that data collection is culturally grounded¹³⁰. Data-collection efforts of SOGIESC aspects should be contextualised to explain cultural and historical factors. Definitions and translations should be developed in close consultation with LGBTIQ civil-society organisations, and with attention to local cultural terminology¹³¹. Wherever appropriate, civil society organisations, national human-rights institutions, equality bodies, and other relevant stakeholders should provide relevant perspectives and information (if they are qualified or in a position to represent the group's interests). It is also recommended to work closely with survey experts and civil-society organisations to develop the data-collection methodology. This is especially important: (i) in relation to the wording of the relevant questions and possible response categories; and (ii) when considering the degree of translatability of SOGIESC concepts in the respective Member State, while abiding by the principles of not doing harm.

One approach to achieve more clarity on underlying concepts in a survey across cultures and languages is to include 'info buttons'¹³². These could be helpful to give further information on the question or response categories. However, as many people will not open an info button, these should not be used for important information. Adding explanatory text via info buttons gives more insights and examples about the terms in question, ultimately ensuring a greater level of comprehension by the survey respondents, making the survey more reliable.

When collecting SOGIESC data, it is crucial **not to conflate sexual orientation, gender identity, gender expression, and sex characteristics**. When asked how they identify, respondents should always have the option to express various aspects of their identity and experience (for instance, an asexual, trans woman should have the ability to select

¹³⁰ National Academies of Sciences, Engineering, and Medicine (2022), [Measuring Sex, Gender Identity, and Sexual Orientation](#), The National Academies Press, Washington DC, p. 66.

¹³¹ For instance, the FRA's 2019 LGBTI Survey underwent a four-phase translation process, in which translators aimed for conceptual equivalence rather than word-for-word translation. The translations were checked by LGBTIQ country experts. See EU Agency for Fundamental Rights (2020), [A long way to go for LGBTI equality. Technical report](#), Publications Office, Luxembourg, pp. 18-19.

¹³² An 'info button' is a hover-over box in a web page providing extra information (for example, the definition of a term).

all of these in a questionnaire). Where routing¹³³ is applied based on identity categories, respondents with multiple identities should receive all relevant questions.

The following sections present practical guidance on how to collect data on sex, gender (identity), gender expression, sexual orientation and sex characteristics, including examples from past surveys to illustrate possible approaches. However, these examples are not intended to showcase best practice to be directly adopted, as is highlighted in the notes for consideration under each example.

4.1. How to collect data on sexual orientation

4.1.1. Rationale

Sexual orientation: ‘How one finds oneself feeling drawn (or not drawn) to another person in a sexual and/or romantic way¹³⁴. ‘This encompasses emotional, affectional, and sexual attraction to, and intimate and sexual relations with, individuals of a different gender or the same gender or more than one gender¹³⁵.

Research shows that people are willing to report their sexual orientation in survey research¹³⁶. There appear to be several factors that can positively influence this willingness and increase response rates. A key factor is respondents’ trust in the confidentiality of the responses¹³⁷. Avoiding stigmatising wording and providing clear definitions also has a positive impact on response rates^{138, 139}. Short, simple questions and fewer answer categories appear to reduce non-response rates, whereas complex questions, especially those seeking to combine several dimensions, are more likely to be

¹³³ A routing applies when a survey respondents choose to answer a survey with a particular response that the survey then asks them other questions pertaining to that response which it would not have asked other respondents who had chosen a different answer.

¹³⁴ European Parliament (2020), [Glossary of Sensitive Language for Internal and External Communications](#), p. 12.

¹³⁵ International Commission of Jurists (2007), [Yogyakarta principles: principles on the application of international human rights law in relation to sexual orientation and gender identity](#), p. 8.

¹³⁶ European Commission (2015), [Special Eurobarometer 437: Discrimination in the EU in 2015](#), *European Union*, p. 99.

¹³⁷ Saewyc, E. M. et al. (2004). [Measuring sexual orientation in adolescent health surveys: Evaluation of eight school-based surveys](#). *Journal of Adolescent Health*, 35(4), pp. 345.e.2-345.e15.

¹³⁸ Hayes, J. et al. (2012). [Prevalence of Same-Sex Behaviour and Orientation in England: Results from a National Survey](#). *Archives of Sexual Behaviour*, 41(3), pp. 631-639.

¹³⁹ Saewyc, E. M. et al. (2004). [Measuring sexual orientation in adolescent health surveys: Evaluation of eight school-based surveys](#). *Journal of Adolescent Health*, 35(4), pp. 345.e.2-345.e15.

skipped. Question placement will also have an impact. Generally, sensitive questions have higher non-response rates if they are placed directly at the beginning of a survey, and non-response rates tend to increase overall for questions the later they are placed in the survey. Questions placed directly after a sensitive question, such as on sexual or criminal behaviour, are also more likely to be skipped¹⁴⁰.

In different cultures and contexts, various concepts and labels exist for describing one's sexual and/or romantic orientation, but concepts such as straight, gay, lesbian, bisexual, pansexual and asexual are the most common in many languages. People who do not identify as heterosexual are often referred to as having 'non-heterosexual orientations'¹⁴¹.

In data-collection efforts, it is important to not focus only on minority groups such as LGBT+ people¹⁴². Sexual orientation is a part of *everyone's* human condition and can trigger positive and negative experiences. Therefore, an indicator of sexual orientation should make it possible to cover sexual-majority and sexual-minority populations.

Sexual orientation is a multidimensional construct encompassing (1) emotional, romantic, and sexual attraction, (2) identity, and (3) behaviour¹⁴³.

1. *Attraction* refers to the gender(s) to which a person is emotionally, romantically or sexually attracted and the strength of this attraction, including whether a person feels attracted at all.
2. *Identity* refers to a person's core internal sense of their sexuality, including how a person thinks of themselves and the way they want to be seen by others.
3. *Behaviour* refers to the gender(s) of sexual partners and how a person interacts romantically and/or sexually with others.

Sexual orientation is often defined based on the gender(s) of a person's desired or actual partners compared to their own gender. The three dimensions of sexuality – attraction,

¹⁴⁰ Saewyc, E. M. et al. (2004). [Measuring sexual orientation in adolescent health surveys: Evaluation of eight school-based surveys](#). *Journal of Adolescent Health*, 35(4), pp.345.e.2-345.e15.

¹⁴¹ Rieger, G., & Savin-Williams, R. C. (2012). [Gender Nonconformity, Sexual Orientation, and Psychological Well-Being](#). *Archives of Sexual Behaviour*, 41(3), 611-621; Coleman, E. (1990). Towards a synthetic understanding of sexual orientation. In D. P. McWhirther, S. A. Sanders, & R. J. Machover (Eds.), *Homosexuality/Heterosexuality*. Oxford University Press, Oxford.

¹⁴² Dean, J. J. (2011). [The Cultural Construction of Heterosexual Identities](#). *Sociology Compass*, 5(8), 679-687.

¹⁴³ National Academies of Sciences, Engineering, and Medicine (2022), [Measuring Sex, Gender Identity, and Sexual Orientation](#), The National Academies Press, Washington DC, p. 5.

identity, and behaviour – may not correspond to each other^{144,145}. For example, someone may be attracted to the same sex/gender but exclusively engage in sexual behaviour with people of the opposite sex/gender; or might be having sexual experiences with someone of the same sex/gender but still identify as straight.

Sexual-orientation identity is the cognitive (i.e. how a respondent sees or thinks of themselves) and social expression (i.e. it provides ‘label’ options that reflect a social status or community) of one’s sexual orientation. Thus, sexual-orientation identity is the dimension that is most consistently tied to experiences with material forms of discrimination, and is noted explicitly in laws and policies aimed at protecting or harming sexual minorities¹⁴⁶. It is also the dimension most frequently used in population-based data-collection settings to count and distinguish between sexual-minority and sexual-majority adult populations. When collecting data on young people, attraction may be a more reliable measure, since young people are more likely to still be coming to terms with their identity.

Sexual orientation identities can be categorised as set out in the bullet points below (see also the [glossary in the annex](#)).

- **Heterosexual, straight:** ‘A person whose romantic, emotional and/or physical attraction is to people of a different gender¹⁴⁷’.
- **Gay:** ‘A person who is sexually and/or emotionally attracted to people of the same gender. It traditionally refers to men, but other people who are attracted to the same gender or multiple genders may also define themselves as gay¹⁴⁸’.
- **Lesbian:** ‘A woman who has a romantic and/or sexual orientation towards women. Some non-binary people may also identify with this term¹⁴⁹’.
- **Bisexual:** ‘An umbrella term used to describe a romantic and/or sexual orientation towards more than one gender¹⁵⁰’.
- **Queer:** An umbrella term for belonging to the LGBTIQ community, which may refer to any or all SOGIESC elements. Previously used as a derogatory term to refer to

¹⁴⁴ Laumann, E., Gagnon, J. H., Michael, R. T., & Michaels, S. (1994), *The social organization of sexuality: sexual practices in the United States*, University of Chicago Press, Chicago.

¹⁴⁵ Richters, J. et al. (2014). [Sexual identity, sexual attraction and sexual experience: the Second Australian Study of Health and Relationships](#). *Sexual health*, 11(5), pp. 451-460.

¹⁴⁶ National Academies of Sciences, Engineering, and Medicine (2022), [Measuring Sex, Gender Identity, and Sexual Orientation](#), The National Academies Press, Washington DC, p. 75.

¹⁴⁷ IOM UN Migration (2021), [SOGIESC Glossary of Terms](#), UNHCR, p. 2.

¹⁴⁸ ILGA Europe (2022), [Our Glossary](#).

¹⁴⁹ Stonewall (2022), [List of LGBTQ+ terms](#).

¹⁵⁰ Stonewall (2022), [List of LGBTQ+ terms](#).

LGBTIQ individuals in the English language, queer has been reclaimed by people who identify beyond traditional gender categories and heteronormative social norms. However, depending on the context, some people may still find it offensive¹⁵¹.

- **Pansexual:** When a person is emotionally and/or sexually attracted to people regardless of their gender¹⁵².
- **Asexual:** Someone who experiences little to no sexual attraction, although they may experience romantic attraction¹⁵³.

4.1.2. Guidance

Member States are encouraged to do the following things.

- They should use precise terminology that reflects the concepts of interest. The term ‘sexual orientation’ should be clearly defined/explained. An info button with a definition could help respondents to surveys to understand the concept.
- They should avoid conflating attraction, identity and/or behaviour when measuring sexual orientation. As sexual orientation is a multidimensional construct, questions should clearly specify which component or components of sexual orientation are being measured. Depending on the dimension that is being measured (attraction, identity, and behaviour), the definition can vary.
- They should develop definitions in close consultation with LGBTIQ civil-society organisations, and with attention to local cultural terminology.
- When collecting data, they should allow respondents to self-identify whenever possible; any proxy reporting should reflect what is known about how a person self-identifies.
- When data collection is aimed at younger people, it is recommended to only use the scale of sexual attraction unless the aim of the survey is to measure sexual behaviour.
- They should use the dimension ‘sexual behaviour’ when sexual experiences/behaviours are the scope of the question. For instance, in HIV research it is more important to measure sexual behaviour than sexual attraction or

¹⁵¹ Adopted from ILGA Europe (2022), [Our Glossary](#).

¹⁵² ILGA Europe (2022), [Our Glossary](#).

¹⁵³ Adopted from IOM UN Migration (2021), [SOGIESC Glossary of Terms](#), UNHCR, p. 2.

identity¹⁵⁴. Sexual behaviour will also identify those people that do not identify as LGB+ but do interact romantically and/or sexually with people of the same sex/gender identity.

- They should use a clear time frame (over the respondent's lifetime, over the past 5 years of the respondent's life, last year, and currently) when asking about sexual orientation. Responses to a sexual orientation question are a subjective view of oneself and can change over the course of a person's lifetime and depending on the context.
- They should adopt terminology accepted by the relevant communities and make sure that answer options capture the different subgroups of interest.
- When asking a question about romantic and/or sexual partners, they should consider that respondents may have more than one partner.

4.1.3. Examples for measuring sexual orientation

4.1.3.1. Example of questions on sexual identity

The following example from **Belgium**¹⁵⁵ reflects the fact that one's sexual orientation might shift in a lifetime.

Q: How would you currently describe your sexual orientation?

- Straight
- Gay
- Lesbian
- Bisexual
- Asexual
- Otherwise:...
- I don't know
- I don't want to say

Notes for consideration

¹⁵⁴ Badgett, M. (2009). [Best Practices for Asking Questions about Sexual Orientation on Surveys](#). The Williams Institute, Los Angeles.

¹⁵⁵ Burgwal, A., & Motmans, J. (to be published). *Enough: research into the experiences with violence of LGBTI people in Flanders, Belgium*. The translation used in the article was 'heterosexual' and 'homosexual' rather than 'straight' and 'gay'.

By including a time frame, this question more accurately captures a respondent's experience. It is important, however, to ensure that it is possible for a respondent to select multiple options, as an asexual person may also identify with any of the other orientations.

4.1.3.2. Example of question on sexual attraction – Sexual health study in Belgium

An ongoing sexual health study in **Belgium** asked respondents to identify to whom they were sexually attracted¹⁵⁶. This example is problematic, since it groups all trans people together. The categories used by the same study for sexual behaviour ([below](#)) would be preferable.

Q: We would like to know who you are currently sexually/romantically attracted to. I am attracted to... (Check everything that applies to you).

- (Cisgender) men
- (Cisgender) women
- Transgender and/or non-binary persons
- No one
- I don't care about gender
- I don't want to say
- Other, namely: (text box)

Notes for consideration

Allowing for multiple answers makes this a more flexible question. However, grouping all trans people under one category does not capture the experience of a respondent who is, for instance, attracted to women (both cis and trans) but not to any other genders. Depending on the target audience, the use of the term 'cisgender' may also be confusing since the term is not widely known outside LGBTIQ communities. A brief clarification (Such as 'cisgender means "not transgender"') would be helpful. Moreover, the distinction between cisgender and transgender risks excluding intersex people.

¹⁵⁶ De Brier, N., et. al. (2022). [Prevalence and associated risk factors of HIV infections in a representative transgender and non-binary population in Flanders and Brussels \(Belgium\): Protocol for a community-based, cross-sectional study using time-location sampling](#). *PLoS ONE*, 17(4), pp. 1-22 De Brier, N., et. al. (2022). [Prevalence and associated risk factors of HIV infections in a representative transgender and non-binary population in Flanders and Brussels \(Belgium\): Protocol for a community-based, cross-sectional study using time-location sampling](#). *PLoS ONE*, 17(4), pp. 1-22 *PLoS ONE*, 17(4), pp. 1-22.

4.1.3.3. Example of question on sexual behaviour – Sexual health study in Belgium

The same study measured sexual behaviour in a way that would avoid binary gender answer options, and allow for multiple answers by giving a choice to distinguish between cisgender and transgender sexual partners. This was done with the question below¹⁵⁷.

Q: With whom have you had sexual contact in the last 12 months? (Multiple choice and multiple answers possible)

- Cisgender man or men
- Cisgender woman or women
- Transgender man or men
- Transgender woman or women
- Non-binary person(s)
- I don't know

Earlier in the questionnaire, respondents were asked whether they had sexual contact with at least 1 person in the last 12 months. To measure sexual experiences (sexual behaviour), the term 'sexual contact' needs a definition for the respondent to understand what the researchers meant. In this questionnaire, sexual contact was defined as 'all kinds of lovemaking involving genital contact, i.e. touching someone else's genitals, whether orally, vaginally or anally¹⁵⁸'.

Notes for consideration

This example uses inclusive response categories (preferable to those in the above example on sexual attraction), and the multiple-choice option allows respondents to describe their sexual behaviour more accurately. As with the example above, the use of the term 'cisgender' may be confusing depending on the target audience, since the term is not widely known outside LGBTIQ communities. The distinction between cisgender and transgender also risks excluding intersex people.

¹⁵⁷ De Brier, N., et. al. (2022). [Prevalence and associated risk factors of HIV infections in a representative transgender and non-binary population in Flanders and Brussels \(Belgium\): Protocol for a community-based, cross-sectional study using time-location sampling](#). *PLoS ONE*, 17(4), pp. 1-22.

¹⁵⁸ De Brier, N., et. al. (2022). [Prevalence and associated risk factors of HIV infections in a representative transgender and non-binary population in Flanders and Brussels \(Belgium\): Protocol for a community-based, cross-sectional study using time-location sampling](#). *PLoS ONE*, 17(4), pp. 1-22.

4.2. How to collect data on sex and gender (identity)

4.2.1. Rationale

Data-collection efforts often conflate sex, a biological variable, with gender or gender identity, a socially defined variable, or treat the respective concepts as interchangeable. Data are often only collected on legal sex/gender, without allowing for self-identification. Sometimes the goal of extrapolating data across surveys or the need for comparisons over time are used as arguments for not adapting already existing questionnaires/variables by adding new questions or revisions of already applied terms/categorisations. For example, for reasons of keeping historical continuity of information, some countries might retain a conflated category of sex and gender and not use a revised version (that differentiates between sex assigned at birth and gender identity).

However, a question on gender identity measures *different* information than a question on sex or a question on legal sex/gender¹⁵⁹, and the concepts do not always correspond to each other. A question on sex is often not straightforward for trans and intersex respondents, who may be unsure whether to respond with their sex assigned at birth or their gender identity. Legal sex/gender also fails to capture many trans people's experiences. Not all trans people wish or are able to obtain LGR. There can be many reasons for this. For example, some gender identities are not translatable into legal options (such as non-binary identities), and some people might not have access to legal sex/gender options depending on the legislation of a specific country (for instance restrictions due to medical criteria or there being a minimum age for LGR). This results in gender identity not being measured correctly when legal sex/gender is assessed.

As **sex**, **gender** and **gender identity** do not measure the same underlying concept, they should be disentangled and assessed separately^{160,161,162,163}. A difference can be made

¹⁵⁹ Legal gender refers to the gender currently recognised by law. Many Member States only recognise the options of male and female, which excludes other identities such as non-binary from being recorded on official documents like passports.

¹⁶⁰ Angus, J. (2012). [Gender, Sex, and Health Research: Developments and Challenges](#). *Canadian Journal of Nursing Research*, 44(3), pp. 3-5.

¹⁶¹ Lorber, J. (2006). [Shifting Paradigms and Challenging Categories](#). *Social Problems*, 53(4).

¹⁶² Ritz, S. A. et al. (2017). [First steps for integrating sex and gender considerations into basic experimental biomedical research](#). *the FASEB Journal*, 28(1), 4-13.

¹⁶³ Westbrook, L., & Saperstein, A. (2015). [New categories are not enough: rethinking the measurement of sex and gender in social surveys](#). *Gender & Society*, 29(4), 534-560.

between sex assigned at birth, gender, and gender identity. To do so, one needs to clarify and agree upon the underlying definitions of these different concepts.

According to a review of the United Nations Economic Commission for Europe (UNECE) on measuring gender identity, one of the major challenges to consider when collecting information on gender identity is **language**. The review states the following.

In some languages (Greek for example), the language is based on a binary sex framework where terms differ depending on whether a man or woman is speaking, or whether a man or woman is being spoken to. And some languages, [...], do not have different words for sex and gender (such as Romanian and Greek; possibly others). In Romanian, the word ‘gen’ for ‘gender’ also means ‘species’ and ‘sex’ — and so ‘identitate sexuala’ is used for both gender identity and sexual identity¹⁶⁴.

Sex is ‘a multidimensional construct based on a cluster of anatomical and physiological traits (**sex characteristics**), which include external genitalia, secondary sex characteristics, gonads, chromosomes, and hormones. It is usually assigned as female or male, most often defined at birth based on visual inspection of external genitalia¹⁶⁵. When constructed along with a binary model of male and female, the term carries the risk of neglecting the existence of intersex persons. For many trans people, sex is also not a straightforward concept.

Sex assigned at birth refers to ‘the sex that is assigned to a person at birth, typically based on the infant’s external anatomy¹⁶⁶’.

Gender is ‘a multidimensional construct that links gender identity (a core element of a person’s individual sense of self), gender expression, and social and cultural expectations about status, characteristics, and behaviour that are associated with sex traits¹⁶⁷. ‘Gender traditionally refers to a social and cultural construct of being a man or a woman. However, some people do not identify within the gender binary of man/woman. Gender exists independently of sex, and an individual’s gender does not always correspond with the sex assigned at birth^{168, 169}.’

¹⁶⁴ UNECE (2019), [In-Depth Review of Measuring Gender Identity](#). Conference of European Statisticians, Paris.

¹⁶⁵ National Academies of Sciences, Engineering, and Medicine (2022), [Measuring Sex, Gender Identity, and Sexual Orientation](#), The National Academies Press, Washington DC, p. 3.

¹⁶⁶ IOM UN Migration (2021), [SOGIESC Glossary of Terms](#), UNHCR, p. 7.

¹⁶⁷ National Academies of Sciences, Engineering, and Medicine (2022), [Measuring Sex, Gender Identity, and Sexual Orientation](#), The National Academies Press, Washington DC, p. 4.

¹⁶⁸ TGEU (2016), [Glossary](#).

¹⁶⁹ See also the [EU gender equality strategy 2020-2025, which defines gender as](#) ‘the socially constructed roles, behaviours, activities and attributes that a given society considers appropriate for women and men’ and is commonly used by EU institutions, including the European Institute for Gender Equality (EIGE).

Gender identity refers to ‘each person’s deeply felt internal and individual experience of gender, which may or may not correspond with the sex assigned at birth, including the personal sense of the body (which may involve, if freely chosen, modification of bodily appearance or function by medical, surgical or other means) and other expressions of gender, including dress, speech and mannerisms¹⁷⁰’.

Gender-diverse refers to the inherent diversity of possible gender experiences, which are therefore broader than cisgender or transgender. Gender-diverse is an umbrella term for many possible gender identifications, for example gender-fluid, gender-free, agender, bigender etc¹⁷¹.

Non-binary: ‘An adjective describing people whose gender identity falls outside the male-female binary. Non-binary is an umbrella term that encompasses a wide variety of gender experiences, including people with a specific gender identity other than man or woman, people who identify as two or more genders (bigender or pan/polygender) and people who don’t identify with any gender (agender)¹⁷²’.

Legal gender recognition (LGR) is a process in which a trans and/or intersex person’s gender is recognised in law, or the achievement of the process¹⁷³.

Cisgender: ‘A person whose current gender identity corresponds to the sex they were assigned at birth¹⁷⁴’.

Trans: ‘An umbrella term to describe people whose gender is not the same as, or does not sit comfortably with, the sex they were assigned at birth¹⁷⁵’.

Endosex: ‘A term describing a person who was born with sex characteristics that fit typical binary notions of male or female bodies. An endosex person may identify with any gender identity or sexual orientation¹⁷⁶’. See also **intersex**.

Intersex individuals, or persons with a variation of sex characteristics, are born with sex characteristics (sexual anatomy, reproductive organs, hormonal structure and/or

¹⁷⁰ International Commission of Jurists (2007), [Yogyakarta principles: principles on the application of international human rights law in relation to sexual orientation and gender identity](#), p. 8.

¹⁷¹ See for example the distribution of gender identities in [the EU Agency for Fundamental Rights’ LGBTI II survey](#), FRA (2020), [A long way to go for LGBTI equality](#), Publications Office, Luxembourg; European Commission (2019).

¹⁷² IOM UN Migration (2021), [SOGIESC Glossary of Terms](#), UNHCR, p. 5.

¹⁷³ ILGA Europe (2022), [Our Glossary](#).

¹⁷⁴ National Academies of Sciences, Engineering, and Medicine (2022), [Measuring Sex, Gender Identity, and Sexual Orientation](#), The National Academies Press, Washington DC, p. 4.

¹⁷⁵ Stonewall (2022), [List of LGBTQ+ terms](#).

¹⁷⁶ IOM UN Migration (2021), [SOGIESC Glossary of Terms](#), UNHCR, p. 7.

levels and/or chromosomal patterns) that do not match the typical societal definition of male or female (See Guideline No 4.4 of this guidance document).

On the topic of gender identities, it is important to recognise the diversity that exists. Many people identify outside the binary identities of men and women. People whose gender identity and sex assigned at birth do not match – or do not match completely – are usually grouped under the concept of **‘transgender and gender-diverse/non-binary’** (TGD)¹⁷⁷. This default grouping carries the risk of incorrectly including intersex people who may not identify with the sex assigned to them at birth but who would not identify as trans. Conversely, it also risks excluding those intersex people who *do* identify as trans but whose gender identity *does* align with the sex assigned to them at birth.

‘Gender-diverse’ is also an umbrella term for many possible gender identifications, for example non-binary, gender-fluid, gender-free, agender, bigender etc. (see, for example, the distribution of gender identities in [FRA’s LGBTI II survey](#) and in [the glossary in the annex of this guidance document](#)). Many gender-diverse people identify as trans, although not all do.

Questions about sex and sex assigned at birth can be very sensitive and considered invasive. Therefore, if data-collection efforts only seek to collect the gender of the respondents, a question on gender identity is sufficient. However, where data collection specifically aims to disaggregate trans respondents, a question on gender identity would not collect this information. Many trans people identify within the binary, so for instance both a trans and cis woman would select ‘woman’ as their gender identity, which does not allow for disaggregation¹⁷⁸.

To collect data on trans respondents, there are several possible approaches. Respondents can be asked directly whether they identify as trans. It is also possible to take a two-step approach asking first about gender identity and then about whether the respondent’s gender identity differs from their sex assigned at birth. An alternative two-step approach asks about gender identity and sex assigned at birth (unlike the first two-step question, this alternative does not make explicit reference to whether the respondent’s gender identity differs from their sex assigned at birth). The latter two approaches carry the above-mentioned risk of falsely classifying intersex respondents (note that these questions do not assess whether somebody is intersex/has a variation of sex characteristics, and this should be assessed in a separate question (See Guideline No 4.4 of this guidance document)).

¹⁷⁷ Motmans, J., Nieder, T. O., and Bouman, W. P. (Eds.). (2020). [Non-binary and genderqueer genders](#). Routledge, Oxon.

¹⁷⁸ Schilt, K., and Bratter, J. (2015). [From multiracial to transgender? Assessing attitudes toward expanding gender options on the US census](#). *Transgender Studies Quarterly*, 2(1), pp. 77-100.

4.2.2. Guidance

Member States and all other relevant institutions and actors involved in or responsible for the collection of SOGIESC data are encouraged to take the actions set out in the bullet points below.

- They should draw up questions that use precise terminology, and ensure the validity of the constructs in these questions, by clearly distinguishing measures of ‘gender’ from measures of ‘sex’. The term ‘gender identity’ should clearly be distinguished or understood as being distinct from the term ‘sex assigned at birth’ as well as from a (possible change in) legal gender registration or gender-affirming healthcare.
- The terms they use (such as ‘gender’, ‘gender identity’, ‘sex’ and ‘sex assigned at birth’) should be clearly defined/explained ([see the glossary in the annex of this guidance document](#)).
- They should develop definitions in close consultation with trans and intersex civil-society organisations, and with attention to local cultural terminology.
- To the extent possible, they should avoid collecting ‘sex assigned at birth’ as a standalone item in any data-collection context. ‘Sex’ as a separate criterion should only be measured when it is of analytical value.
- They should use inclusive measures of gender, which include the following.
 - One-step measures: attempts to identify trans respondents using a single question (e.g. ‘Do you identify as trans?’).
 - Two-step measures: a broader approach that tries to count persons who do not identify with the sex assigned to them at birth (e.g. trans, some intersex people) and those who do, consisting of a two-question sequence – commonly asking for (1) current gender and (2) whether the respondent’s gender differs from their sex assigned at birth; or (1) current gender and (2) sex assigned at birth. When cross-tabulated, the two-step measures can provide approximate counts of cisgender women and men, transgender women and men, and people who identify using terms outside the gender binary, such as non-binary or genderqueer.
 - Where possible, both questions should be presented on the same page. It may be preferable to ask about gender identity **before** asking about sex

assigned at birth. There is some evidence that being asked about sex assigned at birth first is uncomfortable for some trans respondents¹⁷⁹.

- If sex assigned at birth is asked first in a two-step approach, there should be a clarification stating that a further question on gender identity will follow, allowing respondents to self-identify.
- When aimed at younger people or children, they should adapt the wording of the questions accordingly to make it more understandable for the targeted population¹⁸⁰.
- As for the specific part on ‘sex assigned at birth’, and if a third category is legally allowed to be recorded in an official government certificate, a third free-text (write-in) response category should be added to the binary option (female/male) to provide for such a possibility. In countries where a third option is not possible in legal documents, this can possibly be omitted, although foreign-born people might need a third category.
- In situations in which the collection of transgender status is not a major objective, a single question asking for gender with categories of male, female, and a third category such as ‘Another gender’ would suffice. However, a question should not use ‘transgender’ as the third category. Since many trans people identify as male or female, such a measure may be invalidating and will underperform in identifying trans respondents¹⁸¹.
- They should periodically re-evaluate and expand answer options over time and across different settings, particularly for responses (and especially for write-in responses) by people who self-identify outside of binary gender norms.
- They should periodically assess item performance across different survey modes, including proxy reporting, in languages other than the national language¹⁸²: (i) for

¹⁷⁹ National Academies of Sciences, Engineering, and Medicine (2022), [Measuring Sex, Gender Identity, and Sexual Orientation](#), The National Academies Press, Washington DC, p. 119.

¹⁸⁰ Data on SOGIESC on young people are already being collected in some countries. However, questions might need to be rephrased. For example, using a simple phrase such as ‘to what extent do you feel like a boy/girl?’ is more understandable for adolescents. However, the use of terms such as ‘100%’ and ‘real boy’ is discouraged by current language guidelines in research. National Academies of Sciences, Engineering, and Medicine (2022), [Measuring Sex, Gender Identity, and Sexual Orientation](#), The National Academies Press, Washington DC, p. 132.

¹⁸¹ Schilt, K., and Bratter, J. (2015). [From multiracial to transgender? Assessing attitudes toward expanding gender options on the US census](#). *Transgender Studies Quarterly*, 2(1), pp. 77-100.

¹⁸² A commonly cited challenge for setting international standards for two-step data collection is the absence of words distinguishing between ‘sex’ and ‘gender’ in some languages (UNECE (2019), [In-Depth Review of Measuring Gender Identity](#). Conference of European Statisticians, Paris).

different social, demographic, and ethnic minority groups; (ii) for people with an immigrant background; and (iii) among young people.

- As recommended in the [2019 UNECE review on gender identity](#), in languages where the concept of gender is not well established, language issues might be rectified with definitions and guidance, online or via an interviewer. Different languages might require differently nuanced translations. Question testing is recommended to ensure respondents understand what is being asked. Such testing would need to be repeated on a recurring basis as public understandings change¹⁸³.

4.2.3. Examples on how to collect data on sex, gender, and gender identity

4.2.3.1. Example adopted from the FRA's LGBTIQ Survey:

Q1: How would you describe yourself today?

The response options 'Woman' and 'Man' include trans women and men.

Single response

- Woman
- Man
- Non-binary or Genderqueer or Agender or Polygender or Gender-fluid
- Do not identify as woman, man, or non-binary, please specify how would you identify ...
- Prefer not to say
- Don't understand the question
- Don't know

Q2: Are/were you a trans person?

(Ask if Q1='Woman', 'Man', or 'Do not identify as woman, man, or non-binary, please specify how would you identify')

The term 'trans' is used in this survey as a broad umbrella term that includes all those who are transgender, non-binary, gender variant, polygender, agender, gender-fluid, cross dressers, transsexual, or men and women with a transsexual past, and other terms

- Yes
- No
- Prefer not to say
- Don't understand the question
- Don't know

Q3: What sex were you assigned at birth?

¹⁸³ UNECE (2019), [In-Depth Review of Measuring Gender Identity](#). Conference of European Statisticians, Paris.

[INFO BUTTON: Sex assignment at birth is the classification of people as male, female, intersex, or another sex assigned at birth, often based on physical anatomy. The sex assigned at birth is recorded in your birth certificate when you were born.]

Single response

- Female
- Male
- Other (*please specify*)
- Prefer not to answer

How to count the collected data:

- Trans woman: Q1 = woman & Q2 = Yes
- Trans man: Q1 = man & Q2 = Yes
- Gender-diverse: Q1 = Non-binary or Genderqueer or Agender or Polygender or Gender-fluid

Notes for consideration

This approach is based on self-identification, allowing respondents to state for themselves whether they are trans. Since sex assigned at birth is not used as the basis for the respondent classification, the approach does not carry the risk discussed above of incorrectly classifying intersex respondents. Sex assigned at birth is instead collected as additional information that can inform the analysis and is asked about after the question on current gender. This is important, as qualitative research shows that asking about sex assigned at birth before asking about current gender may be uncomfortable for some trans respondents¹⁸⁴. This aligns with the experience of civil-society organisations consulted in the process of drafting this guidance document. One limitation of this question is the ‘Non-binary or Genderqueer or Agender or Polygender or Gender-fluid’ category, which does not capture all genders outside the binary. The approach to counting also assumes that all these respondents identify as trans, which may not be the case for some. The question does not capture information on intersex respondents, most of whom will have been assigned male or female at birth. To adequately capture this information, it is important to ask a separate question on whether respondents are intersex/have a variation of sex characteristics (see Guideline No 4.4 of this guidance document).

¹⁸⁴ National Academies of Sciences, Engineering, and Medicine (2022), [Measuring Sex, Gender Identity, and Sexual Orientation](#), The National Academies Press, Washington DC, p. 119.

4.2.3.2. Example asking about gender-matching sex assigned at birth – Australian Bureau of Statistics

When a question about sex assigned at birth is not necessarily needed, a two-step approach asking about whether a person’s gender identity matches their sex assigned at birth is recommended. This makes it possible for data on trans persons to be disaggregated, while having only one question on gender identity does not. However, and as mentioned above, this two-step grouping of questions carries the risk of incorrectly classifying intersex respondents. The Australian Bureau of Statistics takes this two-step approach, and additionally includes a question on legal gender registration.

Q1: How [do/does] [you/Person’s name/they] describe [your/their] gender?¹⁸⁵

[INFO BUTTON: Gender refers to current gender, which may be different to sex recorded at birth and may be different to what is indicated on legal documents.]

SINGLE RESPONSE

- Man or male
- Woman or female
- Non-binary
- [I/They] use a different term (please specify)
- Prefer not to answer

Q2: Is your gender the same as the sex you were registered at birth?

- Yes
- No (write in gender):

Q3: Have you had your gender registration adjusted on official documents?

- Yes
- No

How to count the collected data

In the guidance published online, details are only given on how the first question was counted¹⁸⁶. The box below sets out how the question was counted.

Preferred code	Alternative code	Label	Definition
----------------	------------------	-------	------------

¹⁸⁵ Australian Bureau of Statistics (2020), [Standard for Sex, Gender, Variations of Sex Characteristics and Sexual Orientation Variables](#), ABS.

¹⁸⁶ Australian Bureau of Statistics (2020), [Standard for Sex, Gender, Variations of Sex Characteristics and Sexual Orientation Variables](#), ABS.

1	M	Man or male	People who described their gender as man or male
2	F	Woman or female	People who described their gender as woman or female
3	X	Non-binary	People who described their gender as non-binary
4	T	Different term	People who described their gender as a term other than man/male, woman/female or non-binary*
5	Z	Prefer not to answer	People who preferred not to respond on how they describe their gender

* Except where the written response for 'Different term' indicates a variation of one of 'Man or male', 'Woman or female' or 'Non-binary', where that response will be coded to the associated label.

The count for trans and non-binary respondents can be derived as follows:

- Trans man: Q1= 'Man or male' AND Q2=No
- Trans woman: Q1= 'Woman or female' AND Q2=No
- Non-binary: Q1= 'Non-binary'

Notes for consideration

The three-step Australian-style approach provides an option for people who were registered other than male or female at birth, and its answers are mutually exclusive and mutually exhaustive. The question on gender registration is also an advantage, since low percentages on this question may indicate issues with the legal process (although not all trans people wish to recognise their gender legally). However, only giving respondents the options of 'yes' and 'no' for the question is overly simplistic. Requirements and cost can vary with different documents, so it is possible that a respondent will have, for instance, updated their passport without updating their birth certificate, which the question in its current form will not capture. The question does not explicitly mention gender identities beyond man, woman and non-binary, although these are collected in the '[I/They] use a different term' option. The approach carries the risk of incorrectly including intersex people who may not identify with the sex assigned to them at birth but who would not identify as trans. Conversely, it also risks excluding those intersex people who do identify as trans but whose gender identity does align with the sex assigned to them at birth. Finally, the write-in option in Q2 is unclear as to whether respondents should write in their current gender, or their sex assigned at birth.

4.2.3.3. Example of closed lists of gender categories – the ‘Being Transgender in Belgium’ study

In most cases, where measuring gender identity in its fullest is not the aim of the data collection, it is sufficient to offer a **closed list of gender categories** from which the respondent is asked to choose one option that matches best at this moment. This approach is suited to instances where neither the respondent’s sex assigned at birth nor their trans status is of relevance for the data collection.

One example of this approach is the question below.

Q1: How would you describe your gender identity? In your feelings and thoughts, you are (multiple options possible):

- A man
- A woman
- A man with a transgender past
- A woman with a transgender past
- A cross-dressing man
- A cross-dressing woman
- A trans man: a person who was born female but has a male gender identity
- A trans woman: a person who was born male but has a female gender identity
- Genderqueer
- Non-binary
- Polygender
- Gender-fluid
- Other: (complete) ...
- I don’t want to say

Q2: In order to be able to analyse the data from this study, we can only use a limited category of the above categories. We would like to ask you which group most closely matches your current gender identity? Tick the answer that best fits you at this moment. I prefer to be classified in the group of:

- (Trans) man
- (Trans) woman
- Cross-dresser
- Genderqueer/agender/polygender/genderfluid/non-binary
- I don’t know, I don’t have a preference

Notes for consideration

This approach gives respondents the agency to determine themselves which group they are classified in for the purpose of the study. However, the categories of man/woman with a transgender past may be confusing to respondents. Moreover, not all people who cross-dress would see themselves as trans, so including this as a category may also create confusion. In addition, the question may fail to capture all trans respondents, since some may only select the categories 'man' and 'woman' in Q1.

4.3. How to collect data on gender expression

4.3.1. Rationale

Gender expression refers to 'each person's presentation of the person's gender through physical appearance – including dress, hairstyles, accessories, cosmetics – and mannerisms, speech, behavioural patterns, names, and personal references. Gender expression may or may not conform to a person's gender identity¹⁸⁷.'

Gender non-conforming: 'A broad term referring to people who do not behave in a way that conforms to the traditional expectations of their gender, or whose gender expression does not fit neatly into a category. While many also identify as transgender, not all gender non-conforming people do¹⁸⁸.'

Gender expression is a useful variable in determining whether gender non-conforming people (independent of the person's sexual orientation, gender identity or sex characteristics) are treated differently, based on the degree that their gender expression aligns with the gender ascribed to them by others.

Asking about gender expression can involve: (i) continuum measures of femininity and masculinity; or (ii) questions on perceived gender conformity by others. Each approach can be useful depending on the aim of the research.

The continuum approach draws on an established line of research in psychology, which shows that femininity and masculinity should not be seen as polar 'opposites', as people can be low on both scales, high on both scales, somewhere in between, or outside of

¹⁸⁷ International Commission of Jurists (2017, [The Yogyakarta Principles Plus 10 - Additional Principles and State Obligation on the Application of International Human Rights Law in Relation to Sexual Orientation, Gender Expression and Sex Characteristics to Complement the Yogyakarta Principles](#), p. 6).

¹⁸⁸ Human Rights Campaign (2022), [Glossary of Terms](#).

the scales altogether^{189,190}. Cognitive interviews with queer women also indicate that a bipolar measure which asks respondents to rank themselves on a scale from very masculine to very feminine is conceptually and empirically problematic¹⁹¹. Unipolar measures asking about femininity and masculinity separately have also been shown to be more reliable in picking up nuanced responses for trans respondents in particular, compared to bipolar measures¹⁹².

Gender expression can be asked about by means of self-appraisal (how the respondent perceives their own gender expression) as well as by means of reflected appraisal (asking respondents how others perceive them). While responses to self and reflected appraisal tend to overlap, it is useful to measure the discrepancy between these, as this discrepancy has been associated with worse health outcomes¹⁹³. Qualitative research indicates that self-appraisal should be asked before reflected appraisal¹⁹⁴.

The second approach of asking about perceived conformity is easier to use in surveys and can be used in conjunction with questions about discrimination, harassment and violence, since people who are less conforming are more likely to experience these problems. This is more appropriate in an LGBTIQ-specific survey rather than a general population survey.

4.3.2. Guidance

Member States and all other relevant institutions and actors involved in or responsible for the collection of SOGIESC data are encouraged to take the actions set out in the bullet points below.

- They should use continuum scales (two Likert scales to measure a masculine gender expression and a feminine gender expression) as self-reported by the respondent.
- If possible, they should ask about both the respondents' subjective evaluation of their gender expression and how the respondent thinks others evaluate their

¹⁸⁹ Garbarski, D. (2023), [The measurement of gender expression in survey research](#), *Social Science Research* 110, p. 2.

¹⁹⁰ National Academies of Sciences, Engineering, and Medicine (2022), [Measuring Sex, Gender Identity, and Sexual Orientation](#), The National Academies Press, Washington DC, p. 104.

¹⁹¹ Garbarski, D., LaVergne, D. (2020). [The Measurement of Sexual Attraction and Gender Expression: Cognitive Interviews with Queer Women](#). In: Brenner, P.S. (eds) *Understanding Survey Methodology*. *Frontiers in Sociology and Social Research*, vol 4. Springer, Cham.

¹⁹² Garbarski, D. (2023), [The measurement of gender expression in survey research](#), *Social Science Research* 110, p. 9.

¹⁹³ Hart, CG, Saperstein, A, Magliozzi, D and Westbrook, L. (2019), [Gender and health: beyond binary categorical measurement](#), *Journal of Health and Social Behavior*, 60 (1), p. 109.

¹⁹⁴ Garbarski, D. (2023), [The measurement of gender expression in survey research](#), *Social Science Research* 110, p. 2.

gender expression. If there is limited space, they should ask about the latter (reflected appraisal)¹⁹⁵.

- They should consider that gender expression may vary across different contexts based on the respondents' comfort and safety. Respondents may not be able to express themselves in a manner that feels consistent with their wishes.
- In an LGBTIQ-specific survey, if the aim of the question is to evaluate how others perceive the respondent in order to identify how this would impact the respondent's experiences, asking about perceived conformity is most appropriate.

4.3.3. Examples of measuring gender expression

4.3.3.1. Example of a continuum approach¹⁹⁶

Q1. In general, how would you describe your appearance, style, and dress?

Not at all feminine	A little feminine	Somewhat feminine	Very feminine	Extremely feminine
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all masculine	A little masculine	Somewhat masculine	Very masculine	Extremely masculine
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q2. In general, how do you think people would describe your appearance, style, and dress?

Not at all feminine	A little feminine	Somewhat feminine	Very feminine	Extremely feminine
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all masculine	A little masculine	Somewhat masculine	Very masculine	Extremely masculine

¹⁹⁵ Garbarski, D. (2023), [The measurement of gender expression in survey research](#), *Social Science Research* 110, pp. 13-14.

¹⁹⁶ Garbarski, D. (2023), [The measurement of gender expression in survey research](#), *Social Science Research* 110.

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

Q3. In general, how would you describe how you walk, talk, sit, stand, and gesture?

Not at all feminine	A little feminine	Somewhat feminine	Very feminine	Extremely feminine
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all masculine	A little masculine	Somewhat masculine	Very masculine	Extremely masculine
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q4. In general, how do you think people would describe how you walk, talk, sit, stand, and gesture?

Not at all feminine	A little feminine	Somewhat feminine	Very feminine	Extremely feminine
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all masculine	A little masculine	Somewhat masculine	Very masculine	Extremely masculine
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Notes for consideration

This example uses unipolar measures of masculinity and femininity, and therefore avoids positioning these measures as polar opposites. The example captures both the dimensions of appearance and behaviour. More research is needed on whether the question order of appearance and behaviour has a significant impact, and on whether these survey questions could be combined¹⁹⁷. A drawback of this approach is that it assumes that respondents would understand their gender expression in terms of masculinity and femininity, which may not be the case, particularly for agender respondents. It would therefore be helpful to add a 'nowhere on this scale' option. Another limitation of the example is that it does not provide respondents with 'don't

¹⁹⁷ Garbarski, D. (2023), [The measurement of gender expression in survey research](#), *Social Science Research* 110, p. 14.

know’ or ‘prefer not to answer’ options, which are important from an ethical perspective (see Section No 2.3 of this guidance document).

4.3.3.2. Example of asking about perceived conformity¹⁹⁸

People can tell I’m transgender/gender non-conforming even if I don’t tell them.

- Always
- Most of the time
- Sometimes
- Occasionally
- Never

Notes for consideration

This is an example from a trans-specific survey. The question captures the extent of perceived visual conformity of the respondent with traditional gender identity expressions, which can be analysed in conjunction with rates of discrimination, violence, and harassment. Although this question is only directed at trans and gender non-conforming respondents, a similar question is possible for other LGBTIQ respondents. The Generations Survey, for instance, which focused on LGB respondents, asked: ‘How often, if ever, can people tell you are LGB even if you don’t tell them?’. The survey had the same response options of ‘always’, ‘most of the time’, ‘sometimes’, ‘occasionally’ and ‘never’¹⁹⁹. As mentioned above, the question also lacks ‘don’t know’ and ‘prefer not to answer’ options, which would be important to include from an ethical perspective (see Section No 2.3 of this guidance document).

¹⁹⁸Jaime, G., Mottet, L., Tanis, J., Harrison, J., Herman, J. and Keisling, M. (2011), [Injustice at Every Turn: A Report of the National Transgender Discrimination Survey](#), National Center for Transgender Equality and National Gay and Lesbian Task Force, Washington DC, p. 184.

¹⁹⁹ Krueger, E. A., Lin, A., Kittle, K. R., & Meyer, I. H. (2020), [Generations – A Study of the Life and Health of LGB People in a Changing Society \(Methodology and Technical Notes, Gallup Quantitative Survey\)](#), The Williams Institute, Los Angeles.

4.4. How to collect data on sex characteristics

4.4.1. Rationale

Sex characteristics ‘refer to each person’s physical features relating to sex, including genitalia and other sexual and reproductive anatomy, chromosomes, hormones, and secondary physical features emerging from puberty²⁰⁰’.

Intersex individuals, or persons with a variation of sex characteristics, are born with sex characteristics (sexual anatomy, reproductive organs, hormonal structure and/or levels and/or chromosomal patterns) that do not match the typical societal definition of male or female. Often, the variation of sex characteristics is not identified until later in life²⁰¹.

Being intersex/having a variation of sex characteristics is an important component of demographic status and private medical information. It can also be an aspect of identity, although not all people with a variation of sex characteristics will identify as intersex or see themselves as part of the LGBTIQ community²⁰².

Endosex: ‘A term describing a person who was born with sex characteristics that fit typical binary notions of male or female bodies. An endosex person may identify with any gender identity or sexual orientation²⁰³.’

While research on intersex people outside the clinical setting is still scarce, recent discussions about best-practice research methodologies have increased²⁰⁴. In academic literature, different terms are used to describe people with variations of sex characteristics. The terminology used to describe this population is shaped by social

²⁰⁰ International Commission of Jurists (2017, [The Yogyakarta Principles Plus 10 - Additional Principles and State Obligation on the Application of International Human Rights Law in Relation to Sexual Orientation, Gender Expression and Sex Characteristics to Complement the Yogyakarta Principles](#), p. 6).

²⁰¹ National Academies of Sciences, Engineering, and Medicine (2022), [Measuring Sex, Gender Identity, and Sexual Orientation](#), The National Academies Press, Washington DC, p. 139.

²⁰² National Academies of Sciences, Engineering, and Medicine (2022), [Measuring Sex, Gender Identity, and Sexual Orientation](#), The National Academies Press, Washington DC, p. 141.

²⁰³ IOM UN Migration (2021), [SOGIESC Glossary of Terms](#), UNHCR, p. 7.

²⁰⁴ Park, A. (2016), [Reachable: Data collection methods for sexual orientation and gender identity](#), The Williams Institute, Los Angeles.

views about sex and gender being binary concepts, which has also been largely adopted by medical professionals and policy makers.

By including a question about sex characteristics, intersex people/people with variations of sex characteristics are made visible in the collected data. It is important to not conflate having a variation of sex characteristics with having a certain (gender) identity, and to understand being intersex as an experience. Most people with a variation of sex characteristics identify as either a man or a woman, although the possibility to select/self-identify as intersex and/or gender-diverse (such as non-binary/genderqueer) should always be allowed when collecting data²⁰⁵.

Researchers should be aware that the concept of being intersex is not widely known in the intersex population itself, since many people with variations of sex characteristics are often given different names/diagnoses²⁰⁶, such as androgen insensitivity syndrome, congenital adrenal hyperplasia, Klinefelter syndrome or Turner syndrome²⁰⁷. Given the diversity of terms used, as well as the diversity in knowledge on terminology, it is very important to present the respondents with a clear and encompassing definition.

4.4.2. Guidance

Member States are encouraged to take the following actions.

- They should include at least one question to make it possible for people with a variation of sex characteristics to self-identify.
- They should use a standalone measure to ask respondents about whether they are intersex/have a variation of sex characteristics or are not (endosex). The question should not be added as a third response option to 'sex' or 'sex assigned at birth', as intersex persons are, as a general rule, assigned female or male. However, an open box should be provided to allow respondents to include national specificities, like the 'blank marker' in Germany²⁰⁸.
- Being intersex should not be conflated with being trans or with sexual orientation.
- They should consider that intersex people may have a gender identity different to the one assigned to them at birth. While some intersex people may identify as

²⁰⁵ Jones, T., Hart, B., Carpenter, M., Ansara, G., Leonard, W., & Lucke, J. (2016), [Intersex. Stories and Statistics from Australia](#), Open Book Publishers, Cambridge.

²⁰⁶ van Ditzhuijzen, J., & Motmans, J. (2020), [Kenni En Opvattingen over Intersekse: Een Nulmeting in Nederland En Vlaanderen](#), Rutgers and UZ Gent, Utrecht and Ghent.

²⁰⁷ Randjelovic, R. et. al. (2018), [Intersex Research Study](#), UNDP, pp. 46-48.

²⁰⁸ Statistische Ämter des Bundes und der Länder (2022), Zensus 2022.

trans, other may feel that the term does not encapsulate their specific situation as an intersex person in a mostly endosex society.

- They should add a clear definition of the concept of intersex or of what constitutes a variation of sex characteristics, with enough examples that are part of the umbrella term of intersex.
- They should develop definitions in close consultation with intersex civil-society organisations, and with attention to local cultural terminology.
- In addition to an explanation/definition of the concept, they should add an open follow-up question for those indicating they are intersex/have a variation of sex characteristics making it possible for respondents to describe, in the framework of the survey, experiences and needs connected to their variation to the extent they are willing to do so.

4.4.3. Examples of intersex measures

4.4.3.1. Knowledge of – and perceptions about – intersex people: a baseline measurement in the Netherlands and Flanders

Providing a definition of intersex is important and can significantly impact the response rate to surveys. A recent example comes from **a representative panel study from Belgium and the Netherlands**²⁰⁹. ‘Ad random’ selected participants were asked whether they were intersex in two different ways, which influenced the occurrence/visibility of intersex persons in the study results. The first time intersex was assessed in the panel study was before a definition was given, and respondents were able to tick the ‘intersex person’ option as applicable to them (among other options). In total, 9 people selected this option (0.4%). After the definition was given to the respondents, a different group of participants were asked whether they were an intersex person, or had some intersex condition. Now 28 people indicated that they identified themselves as an intersex person or that they had an intersex condition (1.6%). It therefore appears important to include a definition of intersex before assessing whether respondents are intersex or have a variation of sex characteristics.

It is also important to differentiate between questions which ask about whether the respondent has a variation of sex characteristics, and questions that ask whether they

²⁰⁹ van Ditzhuijzen, J., & Motmans, J. (2020), [Kennis En Opvattingen over Intersekse: Een Nulmeting in Nederland En Vlaanderen](#), Rutgers and UZ Gent, Utrecht and Ghent.

have been given a diagnosis related to such a variation, as the following two examples show. Unless the research is specifically on healthcare, the former approach is preferable.

4.4.3.2. Example asking whether respondents have a variation of sex characteristics: the FRA's LGBTI II survey (online)

The FRA's 2019 EU LGBTI II survey included the experiences of intersex people for the first time.

The question used by FRA was the following.

Q: Some persons are born with sex characteristics (like sexual anatomy, reproductive organs, and/or chromosome patterns) that do not belong strictly to male or female categories or belong to both at the same time. This is known as 'intersex'. Would you describe yourself as intersex?

- Yes
- No

[INFO BUTTON: *Variations of sex characteristics* can present themselves prenatally and at birth but also during childhood, in puberty or in adulthood. You might have noticed a variation in your sex characteristics at a very early age or later on in life, and you may have had surgical or medical treatment to modify them.]

Notes for consideration

The question provides a definition for clarity and allows respondents to self-identify. However, it does not allow respondents to describe their variations in their own words and does not include 'don't know' or 'prefer not to answer' options (see Section No 2.3 of this guidance document).

4.4.3.3. Example asking whether respondents have been diagnosed with a variation of sex characteristics – Research into the experiences with violence of LGBTI people in Flanders, Belgium

An ongoing LGBTI survey on violence in **Belgium** takes the following approach, which asks specifically about a diagnosis²¹⁰.

Q: Some persons are born with a body that does not fit within the normative definitions of man and woman. The experience of people born with such a body is known as 'intersex'. Have you ever been diagnosed with a form of intersex at birth or at a later date?

- Yes
- No
- I don't know
- I don't want to say

[POP-UP NOTE: Intersex refers to the experiences of people born with bodies that do not fit within the normative definitions of male and female. Intersex is very diverse and is not always about external sexual characteristics/genitals. Sexual characteristics are not only the genitals (penis/vagina), but also chromosomes, hormones, or internal sex organs (womb, hormone glands). Examples are being born with a female body but with XY chromosomes, or someone with a penis and a uterus. You may recognise yourself in this description but have never heard of the word intersex. Doctors and care providers may use medical terms such as 'pseudo-hermaphrodite', (C)AIS, Klinefelter, AGS/CAH, 17-, Alpha/5-, Beta/5-, Alpha Reductase and hypospadias, MRKH, Turner, or 'disorder/difference of sex development (DSD)'.]

Notes for consideration

The question provides an extensive definition in the pop-up note, including examples. However, it does not ask for the respondent's self-identification, but rather for a diagnosis. This medicalised approach may be difficult for intersex persons who were traumatised in medical settings. It might also exclude respondents who became aware of their variation of sex characteristics themselves and have not received a formal diagnosis. The term 'female body' in the explanatory text also reinforces normative assumptions of what is 'male' and 'female'. This could be rephrased, for instance, as 'external genitalia that are perceived as "female" (e.g. labia, vulva, clitoris)'. The text also refers to the diagnosis of 'pseudo-hermaphrodite'. While this term is still used in medicine, 'hermaphrodite' is considered a slur and should therefore either be avoided

²¹⁰ Burgwal, A., & Motmans, J. (to be published). Enough: research into the experiences with violence of LGBTI people in Flanders, Belgium.

or acknowledged to be a term that is often used derogatorily.²¹¹ Finally, 'sex characteristics' is a more appropriate term than 'sexual characteristics'.

²¹¹ Further information on terms to use and avoid can be found in European Parliament (2020), [Glossary of Sensitive Language for Internal and External Communications](#).

Annex: Glossary

Agender: ‘A term describing people who do not identify with any gender. This term may also encompass people who experience a lack of gender and people whose gender is neutral. Agender people have a range of sexual orientations, gender expressions and sex characteristics²¹².’

Allosexual/romantic: ‘Allo people experience sexual and romantic attraction, and do not identify as on the ace or aro spectrum. Allo is to ace [asexual] and aro [aromantic] spectrum identities, as straight is to LGB+ spectrum identities. It is important to use words that equalise experience, otherwise the opposite to ace and aro becomes ‘normal’ which is stigmatising²¹³.’

Asexual: ‘A person who may experience romantic or emotional attraction, but generally does not experience sexual attraction. Demisexual and greysexual/grey-asexual describe people with varying degrees of sexual attraction. Asexual may be used as an umbrella term encompassing demisexual, greysexual and other terms²¹⁴.’ Asexual people may identify as aromantic, gay, lesbian, bi, pan, etc.

Aromantic: ‘A person who does not experience romantic attraction. Some aromantic people experience sexual attraction, while others do not. Aromantic people who experience sexual attraction or occasional romantic attraction might also use terms such as gay, bi, lesbian, straight and queer in conjunction with asexual to describe the direction of their attraction²¹⁵.’

Bisexual (bi+): ‘Bi is an umbrella term used to describe a romantic and/or sexual orientation towards more than one gender. Bi people may describe themselves using one or more of a wide variety of terms, including, but not limited to, bisexual, pan, queer, and some other non-monosexual and non-monoromantic identities²¹⁶.’

Cisgender: ‘A person whose current gender identity corresponds to the sex they were assigned at birth²¹⁷.’

²¹² IOM UN Migration (2021), [SOGIESC Glossary of Terms](#), UNHCR, p. 5.

²¹³ Stonewall (2022), [List of LGBTQ+ terms](#).

²¹⁴ IOM UN Migration (2021), [SOGIESC Glossary of Terms](#), UNHCR, p. 2.

²¹⁵ Stonewall (2022), [List of LGBTQ+ terms](#).

²¹⁶ Stonewall (2022), [List of LGBTQ+ terms](#).

²¹⁷ National Academies of Sciences, Engineering, and Medicine (2022), [Measuring Sex, Gender Identity, and Sexual Orientation](#), The National Academies Press, Washington DC, p. 4.

Endosex: ‘A term describing a person who was born with sex characteristics that fit typical binary notions of male or female bodies. An endosex person may identify with any gender identity or sexual orientation²¹⁸.’ See also **intersex**.

Gay ‘refers to a person who is sexually and/or emotionally attracted to people of the same gender. It traditionally refers to men, but other people who are attracted to the same gender or multiple genders may also define themselves as gay²¹⁹.’

Gender is ‘a multidimensional construct that links gender identity (a core element of a person’s individual sense of self), gender expression, and social and cultural expectations about status, characteristics, and behaviour that are associated with sex traits²²⁰.’ ‘Gender traditionally refers to a social and cultural construct of being a man or a woman. However, some people do not identify within the gender binary of man/woman. Gender exists independently of sex, and an individual’s gender does not always correspond with the sex assigned at birth²²¹.’

Gender-diverse refers to the inherent diversity of possible gender experiences, which are therefore broader than cisgender or transgender. Gender-diverse is an umbrella term for many possible gender identifications, for example gender-fluid, gender-free, agender, bigender, etc²²².

Gender expression refers to ‘each person’s presentation of the person’s gender through physical appearance – including dress, hairstyles, accessories, cosmetics – and mannerisms, speech, behavioural patterns, names and personal references, and noting further that gender expression may or may not conform to a person’s gender identity²²³.’ ‘Gender expression is separated from gender identity as both cisgender or transgender men can for example deploy gender-stereotypically-masculine behaviour and expression, or not. Although there are variations in how gender expression (often defined as ranges of femininity and masculinity) is defined and expressed both across cultural groups and over time, at the core of the dominant gender-expression belief system is the expectation that people who are assigned female at birth should have a

²¹⁸ IOM UN Migration (2021), [SOGIESC Glossary of Terms](#), UNHCR, p. 7.

²¹⁹ ILGA Europe (2022), [Our Glossary](#).

²²⁰ National Academies of Sciences, Engineering, and Medicine (2022), [Measuring Sex, Gender Identity, and Sexual Orientation](#), The National Academies Press, Washington DC, p. 4.

²²¹ TGEU (2016), [Glossary](#).

²²² See for example the distribution of gender identities in [the EU Agency for Fundamental Rights’ LGBTI II survey](#), FRA (2020), [A long way to go for LGBTI equality](#), Publications Office, Luxembourg.

²²³ International Commission of Jurists (2017), [The Yogyakarta Principles Plus 10 - Additional Principles and State Obligation on the Application of International Human Rights Law in Relation to Sexual Orientation, Gender Expression and Sex Characteristics to Complement the Yogyakarta Principles](#), p. 6.

“feminine” gender expression, and those assigned male at birth should have a “masculine” gender expression²²⁴.’

Gender-fluid: ‘A person who does not identify with a single fixed gender or has a fluid or unfixed gender identity²²⁵.’

Gender identity refers to ‘each person’s deeply felt internal and individual experience of gender, which may or may not correspond with the sex assigned at birth, including the personal sense of the body (which may involve, if freely chosen, modification of bodily appearance or function by medical, surgical or other means) and other expressions of gender, including dress, speech and mannerisms²²⁶’.

Genderqueer: ‘A person who identifies as neither or both male or female, or a combination of genders²²⁷.’

Gender non-conforming: ‘A broad term referring to people who do not behave in a way that conforms to the traditional expectations of their gender, or whose gender expression does not fit neatly into a category. While many also identify as transgender, not all gender non-conforming people do²²⁸.’

Heterosexual, straight: ‘A person whose romantic, emotional and/or physical attraction is to people of a different gender²²⁹.’

Homosexual: ‘Sexually and/or emotionally attracted to people of the same, usually binary, gender. It is recommended to use the terms lesbian and gay instead of homosexual people. The terms lesbian and gay are being considered neutral and positive, and the focus is on the identity instead of being sexualised or pathologised²³⁰.’

Intersex: ‘Intersex individuals are born with sex characteristics (sexual anatomy, reproductive organs, hormonal structure and/or levels and/or chromosomal patterns) that do not fit the typical definition of male or female. The term “intersex” is an umbrella term for the spectrum of variations of sex characteristics that naturally occur within the human species. The term intersex acknowledges the fact that, physically, sex is a

²²⁴ West C, Zimmerman DH. (1987), [Doing Gender](#). *Gender and Society* 1(2), pp. 125-151.

²²⁵ Human Rights Campaign (2022), [Glossary of Terms](#).

²²⁶ International Commission of Jurists (2007), [Yogyakarta principles: principles on the application of international human rights law in relation to sexual orientation and gender identity](#), p. 8.

²²⁷ IOM UN Migration (2021), [SOGIESC Glossary of Terms](#), UNHCR, p. 5.

²²⁸ Human Rights Campaign (2022), [Glossary of Terms](#).

²²⁹ IOM UN Migration (2021), [SOGIESC Glossary of Terms](#), UNHCR, p. 2.

²³⁰ ILGA Europe (2022), [Our Glossary](#).

spectrum and that people with variations of sex characteristics other than male or female exist²³¹.’

Legal gender recognition: ‘A process whereby a trans and/or intersex person’s gender is recognised in law, or the achievement of the process²³².’

Lesbian: Refers to a woman who has a romantic and/or sexual orientation towards women. Some non-binary people may also identify with this term²³³.

LGBTIQ: Acronym for lesbian/gay/bisexual/trans/intersex and queer/questioning. Other popular variants add the letter A, where A stands for asexual, aromantic or agender and ‘+’. Adding the ‘+’ is generally considered more inclusive, encompassing other realities/identities such as (but not limited to) asexual, pansexual, gender-nonconforming²³⁴.

Non-binary: ‘An adjective describing people whose gender identity falls outside the male-female binary. Non-binary is an umbrella term that encompasses a wide variety of gender experiences, including people with a specific gender identity other than man or woman, people who identify as two or more genders (bigender or pan/polygender) and people who don’t identify with any gender (agender)²³⁵.’ Some non-binary people prefer a gender-neutral name or pronouns such as they/them rather than he/him or she/her (see personal pronouns).

Pansexual: ‘When a person is emotionally and/or sexually attracted to people regardless of their gender²³⁶.’

Personal pronouns: ‘Pronouns are words we use to talk about ourselves (“I”), to someone (“you”) or about other people (“she”, “he”, “they”, etc.) In English, “he” and “she”, called third-person pronouns, are gendered, while the singular “they” is gender-neutral. In other languages, such as Arabic and Korean, second-person pronouns (“you”) and first-person pronouns (“I”) may also be gendered. Some languages don’t require pronouns to indicate gender or have a gender-neutral option, including Japanese, Swahili, Swedish and Turkish. Sweden has added the gender-neutral pronoun “hen” to

²³¹ ILGA Europe (2022), [Our Glossary](#). This is a shortened version of the definition developed by the intersex community and OII Europe, see Dan Christian Ghattas (2015), [Standing up for the Human Rights of Intersex People – How Can You Help?](#), OII Europe and ILGA Europe, Brussels, p. 9.

²³² ILGA Europe (2022), [Our Glossary](#).

²³³ Stonewall (2022), [List of LGBTQ+ terms](#).

²³⁴ Adopted from European Parliament (2020), [Glossary of Sensitive Language for Internal and External Communications](#), p. 11. This guidance note uses the acronym ‘LGBTIQ’ to align with the terminology used in the Commission’s [LGBTIQ Equality Strategy 2020-2025](#).

²³⁵ IOM UN Migration (2021), [SOGIESC Glossary of Terms](#), UNHCR, p. 5.

²³⁶ ILGA Europe (2022), [Our Glossary](#).

their language to complement “han” (he) and “hon” (she). Other languages don’t mark gender at all, including Armenian, Finnish, Hungarian, Persian and Yoruba²³⁷.’

Queer: ‘Traditionally a negative term, queer has been reclaimed by some people and is considered inclusive of a wide range of diverse sexual orientations, gender identities and expressions. It may be used as an umbrella term for people with diverse SOGIESC, or as an alternative to the phrase “people with diverse SOGIESC” or the acronym LGBT. Queer is used by many people who feel they do not conform to a given society’s economic, social and political norms based on their sexual orientation, gender identity and gender expression²³⁸.’ ‘However, depending on the context, some people may still find the term offensive²³⁹.’

Sex is ‘a multidimensional construct based on a cluster of anatomical and physiological traits (**sex traits**), which include external genitalia, secondary sex characteristics, gonads, chromosomes, and hormones. It is usually assigned as female or male, most often defined at birth based on visual inspection of external genitalia²⁴⁰.’

Sex assigned at birth: ‘The sex that is assigned to a person at birth, typically based on the infant’s external anatomy [...]. The phrases “assigned female at birth” (AFAB) and “assigned male at birth” (AMAB) refer to people with typical male or female sex characteristics, regardless of their gender identity or gender expression. The phrase “coercively assigned female [male] at birth” (CAFAB and CAMAB) refer to intersex people assigned a binary sex, often via non-consensual surgeries²⁴¹.’

Sex characteristics ‘refer to each person’s physical features relating to sex, including genitalia and other sexual and reproductive anatomy, chromosomes, hormones, and secondary physical features emerging from puberty²⁴².’ People who have a variation of sex characteristics (VSC) that falls outside of the normative understanding of gendered bodies are often referred to as intersex.

Sexual orientation: refers to ‘how one finds oneself feeling drawn (or not drawn) to another person in a sexual and/or romantic way²⁴³.’ This encompasses emotional, affectional, and sexual attraction to, and intimate and sexual relations with, individuals

²³⁷ IOM UN Migration (2021), [SOGIESC Glossary of Terms](#), UNHCR, p. 5.

²³⁸ IOM UN Migration (2021), [SOGIESC Glossary of Terms](#), UNHCR, p. 2.

²³⁹ ILGA Europe (2022), [Our Glossary](#).

²⁴⁰ National Academies of Sciences, Engineering, and Medicine (2022), [Measuring Sex, Gender Identity, and Sexual Orientation](#), The National Academies Press, Washington DC, p. 3.

²⁴¹ IOM UN Migration (2021), [SOGIESC Glossary of Terms](#), UNHCR, p. 7.

²⁴² International Commission of Jurists (2017), [The Yogyakarta Principles Plus 10 - Additional Principles and State Obligation on the Application of International Human Rights Law in Relation to Sexual Orientation, Gender Expression and Sex Characteristics to Complement the Yogyakarta Principles](#), p. 6.

²⁴³ European Parliament (2020), [Glossary of Sensitive Language for Internal and External Communications](#), p. 12.

of a different gender or the same gender or more than one gender²⁴⁴.' The three aspects most commonly measured when identifying sexual orientation are sexual attraction, sexual behaviour, and sexual identity²⁴⁵.

SOGIESC: 'An acronym for sexual orientation, gender identity, gender expression and sex characteristics²⁴⁶.'

Trans/transgender: 'An umbrella term to describe people whose gender is not the same as, or does not sit comfortably with, the sex they were assigned at birth. Trans people may describe themselves using one or more of a wide variety of terms, including (but not limited to) transgender, transsexual, genderqueer (GQ), gender-fluid, non-binary, gender-variant, cross-dresser, genderless, agender, non-gender, third gender, bigender, trans man, trans woman, trans masculine, trans feminine and neutrois²⁴⁷.'

Variations of sex characteristics (VSC) refer to 'sex characteristics (sexual anatomy, reproductive organs, hormonal structure and/or levels and/or chromosomal patterns) that do not fit the typical definition of male or female. The term "intersex" is an umbrella term for the spectrum of variations of sex characteristics that naturally occur within the human species²⁴⁸.' 'Some of these variations may be apparent before or at birth, while others are not apparent until after puberty or later or may not be physically apparent at all. There are more than 40 intersex variations; experts estimate between 0.5% and 1.7% of the population is born with intersex traits²⁴⁹.'

²⁴⁴ International Commission of Jurists (2007), [Yogyakarta principles: principles on the application of international human rights law in relation to sexual orientation and gender identity](#), p. 8.

²⁴⁵ Adopted from National Academies of Sciences, Engineering, and Medicine (2022), [Measuring Sex, Gender Identity, and Sexual Orientation](#), The National Academies Press, Washington DC, p. 5.

²⁴⁶ IOM UN Migration (2021), [SOGIESC Glossary of Terms](#), UNHCR, p. 1.

²⁴⁷ Stonewall (2022), [List of LGBTQ+ terms](#).

²⁴⁸ ILGA Europe (2022), [Our Glossary](#).

²⁴⁹ IOM UN Migration (2021), [SOGIESC Glossary of Terms](#), UNHCR, p. 7.