



Equality Data & Privacy

How do we collect diversity data responsibly and lawfully?



What is equality data?

Definition

“All information that is useful for the purpose of describing and analysing the state of equality. It may be qualitative or quantitative.”

Source: European Commission (2016): European handbook on equality data. 2016 revision. Directorate-General for Justice and Consumers. Directorate D — Equality.

Why is this important?

28.7%

of all professors in the Netherlands are women



Under-recommendation in school advice for children from a migrant background



Higher risk of discrimination for students of colour applying for internships

No data, no policy.

To tackle discrimination and inequality, organisations need to understand the situation.
Equality data enables them to implement evidence-based policies.

Reversal of the burden of proof (Article 10 Awgb)

When discrimination is presumed, the employer must prove that it has not breached the principle of equal treatment.

CSRD compliance

Large organisations (>250 employees) must publicly report on equal treatment. Without equality data, this is not possible.

Common objections – and how to address them

“We are not allowed to collect diversity data (GDPR).”



It is permitted – provided there is a clear purpose, this is communicated, and the GDPR conditions are met.

“Categories used are inappropriate or offensive.”



Involve the target audience in choosing the categories. Always provide a free-text field and acknowledge complexity.

“Will my data remain anonymous?”



Traceability is prevented by minimum cell size, aggregation, access restrictions, and dataset separation.

“Putting people into boxes stigmatises them.”



Always provide context for data. Qualitative follow-up research prevents the reinforcement of stereotypes.

Legal framework: is it permitted and how?

1 Explicit consent

Voluntary

- no negative consequences if consent is refused

Informed

- the purpose, use and retention period must be clear

Unambiguous

- active tick, no pre-ticked box

Active

- active step to give consent

Specific

- separate consent for each purpose

Demonstrable

- document each instance of consent

Explicit

- confirmed in writing or digitally (for special categories of personal data*)

2 Substantial public interest (Article 25 UAVG)

- Processing is aimed at eliminating disadvantages faced by a minority group
- **Limited to:** country of birth, country of birth of parents/grandparents
- The person concerned has not filed a written objection
- **Please note:** in the case of equality data (broader scope), consent is the appropriate legal basis

** Special categories of personal data (race, religion, health, sexual orientation) always require an exception under Article 9 of the GDPR, in addition to a basis under Article 6.*

Basic GDPR principles in practice

Transparency

Explain in clear language why data is collected, how long it is retained and who has access to it.

Purpose limitation

Use data only for the purpose for which consent has been given; do not use it for any other processing operations.

Data minimisation

Do not collect more information than is strictly necessary for the purpose.

Accuracy

Keep data up to date and correct errors – incorrect data leads to flawed policies.

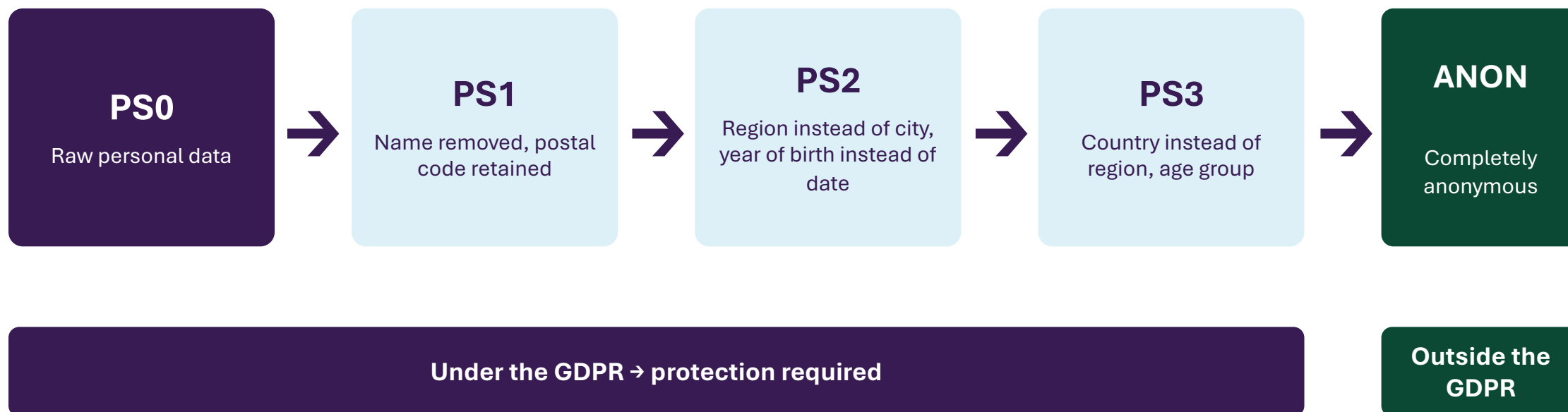
Storage limitation

Set a retention period. Delete raw data once this period has expired.

Security

Encryption, access restrictions, logging, rapid aggregation, and dataset separation.

From personal data to anonymised data



Six human rights principles for responsible data processing

1 Participation

Involve staff and students in all stages: from defining the problem to interpretation and policy-making.

2 Data disaggregation

Break down the data by diversity dimensions to reveal inequalities within each group.

3 Self-identification

Data subjects determine how they wish to categorise themselves. Third parties may never categorise individuals.

4 Transparency

Provide full information about the purpose, use, retention period and recipients – in plain language.

5 Privacy

Appropriate technical and organisational measures: encryption, aggregation, access restriction.

6 Accountability

Account for how data has been used and what policies have resulted from this.

Good questionnaires: practical choices

Self-categorisation

Let respondents categorise themselves – not based on administrative data. Always offer the options to tick multiple boxes, skip questions, and use a free-text field.

Geographical origin

Use regions (Europe / outside Europe) rather than 'Western/non-Western'. This is consistent with the Statistics Netherlands classification and reflects the perceptions of those involved.

Disclaimer accompanying the questionnaire

Acknowledge that categories do not fully capture the complexity of identity. Explain why certain categories have been chosen.

Intersectionality

When possible, consider several dimensions at once. A white woman faces different barriers to a woman of colour.

Care for respondents

Questions about ethnicity and racialisation can be challenging. Communicate in advance and offer follow-up support and feedback.

Organisation-wide survey

Line managers should not collect equality data; a DE&I office should. This reduces pressure and increases reliability.

DiMIS – Diversity Minimal Item Set

What is DiMIS? A validated minimum set of questions (items), designed to identify dimensions of diversity amongst respondents, developed for use in the health sciences and broader research practice. The set is deliberately brief so it can be widely applied, including in existing surveys such as the Employee Satisfaction Survey.

Source: Gültzow e.a. (2026). The Dutch version of the Diversity Minimal Item Set (DiMIS-NL). OSF. <https://doi.org/10.17605/OSF.IO/HWDPI>

Gender & sex

Including non-binary and gender-diverse identities. Not the same as sex (biological).

Age

Preferably report using age groups (not dates of birth) to limit traceability.

Socio-economic status

E.g. parents' educational attainment, income bracket or first-generation student/staff member.

Care work

Informal care and/or childcare responsibilities that affect availability and participation.

Sexual identity

LGBTIQA+ identities. Special category of personal data: explicit consent required.

Ethnic-racial identity

Based on self-identification. Use geographical origin, not 'Western/non-Western'.

Religion & world view

Special category of personal data. Ask a wide range of questions, including philosophical beliefs and atheism.

Mental health

Self-reported mental health symptoms or diagnoses. Stigma-sensitive.

Physical health & disability

Including invisible disabilities. Ask about 'disabilities affecting daily functioning'.

Equality Data and AI: the paradox of algorithmic bias

The paradox

Algorithms do not discriminate because of errors in the code, but because the training data already gives a distorted view.

Without equality data, you cannot detect bias in algorithms. So having less data does not mean that there is no protection.

“Race blind is not race neutral, and unawareness does not equal fairness.”

EU AI Act

Regulation 2024/1689

Providers of high-risk AI systems must:

- ✓ **Examine their training data for potential biases**
- ✓ **Take corrective measures**
- ✓ **Special categories of personal data may be processed**

Conclusion

Institutions that use automated systems for recruitment, assessment or monitoring require equality data to ensure compliance with both the GDPR and the AI Act.

Conclusions & recommendations

Conclusions

- ☐ Privacy and inclusivity are complementary – not contradictory
- ☐ Equality data is crucial to reveal structural inequality
- ☐ The collection of equality data can be carried out lawfully within the framework of the GDPR
- ☐ Transparency, voluntariness and self-identification are core principles
- ☐ Absence of data increases the risk of discrimination
- ☐ Need for continuous monitoring, evaluation and transparency

Recommendations

- ☐ Privacy by design & privacy by default
- ☐ Use clear and understandable communication about purpose and use
- ☐ Request explicit, voluntary and informed consent
- ☐ Enable self-identification and multiple identities
- ☐ Carry out regular audits and evaluations
- ☐ Promote training and awareness amongst staff
- ☐ Develop a feedback mechanism
- ☐ Foster a culture of accountability, openness and transparency

Key references: European Commission & HLEG

2016	<u>European Commission – European Handbook on Equality Data (revised edition)</u>	Reference work on the definition, collection and use of equality data. Introduces the concept and provides frameworks for Member States and institutions.
2017	<u>European Commission – DG Justice – Data Collection in the Field of Ethnicity</u>	A comparative analysis of data collection practices in the EU relating to ethnicity. Discusses issues of definition, categorisation and methodology.
2018	<u>EC – High Level Group on Non-Discrimination (HLEG) – Guidelines on Improving the Collection and Use of Equality Data</u>	Twelve specific recommendations for policymakers and institutions, including the harmonisation of categories, self-identification and data quality. Key reference for this Guidance.
2021	<u>HLEG – Subgroup on Equality Data – Guidance Note on the Collection and Use of Equality Data Based on Racial or Ethnic Origin</u>	Operational implementation of the 2018 Guidelines specifically relating to race and ethnicity. Reaffirms the six human rights principles (OHCHR) and provides practical guidance.
2000	<u>Racial Equality Directive (Directive 2000/43/EC)</u> European Union	Legal basis for equal treatment regardless of race or ethnic origin. Also provides for the reversal of the burden of proof in cases of suspected discrimination (Article 8).



Working together towards inclusive higher education and research

*“Privacy and inclusivity are not opposites — they
reinforce each other.”*

This guidance was developed with contributions from members of an ad hoc working group chaired by Marlon Domingus (Erasmus University Rotterdam). The working group also included: Aya Ezawa (Leiden University), Jule van den Berg (Erasmus University Rotterdam), Helen Tibboel (Erasmus University Rotterdam), Nellie Konijnendijk (Royal Netherlands Academy of Arts and Sciences, National Network of Diversity Officers), Jan van den Berg (Amsterdam University of Applied Sciences), Floor May-Smit (NHL Stenden), Dick Houtzager (Office of the National Coordinator against Discrimination and Racism), Onur Şahin (State Commission against Discrimination and Racism), Danique Knibbeler (NautaDutilh), and representatives from the Advisory Committee Diverse and Inclusive Higher Education and Research.
