OECD Statistics Working Papers 2018/09

Diversity statistics in the OECD: How do OECD countries collect data on ethnic, racial and indigenous identity?

Carlotta Balestra,

Lara Fleischer

https://dx.doi.org/10.1787/89bae654-en

OECD

Organisation for Economic Co-operation and Development

SDD/DOC(2018)9

Unclassified **English - Or. English**

8 November 2018

STATISTICS AND DATA DIRECTORATE

Diversity statistics in the OECD

How do OECD countries collect data on ethnic, racial and indigenous identity?

SDD WORKING PAPER No. 96

Carlotta Balestra, Statistics and Data Directorate, +33 (0)1 45 24 94 36, Carlotta.BALESTRA@oecd.org;
Lara Fleischer, Statistics and Data Directorate, +33 (0)1 45 24 17 65, Lara.Fleischer@oecd.org.

JT03439191

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

Diversity statistics in the OECD: How do OECD countries collect data on ethnic, racial and indigenous identity?

Carlotta Balestra and Lara Fleischer[1]

OECD Statistics and Data Directorate

1. This paper is part of a broader OECD-project on diversity, launched in early 2016. The authors thank Ann Morning (New York University), Kalinda Griffiths (UNSW Sydney), Claire Coleman (University of Sydney), Per Axelsson (Umeå University) and Andrew Sporle, (Univeristy of Auckland) for their advice on the challenges of measuring diversity; and Marco Mira D'Ercole (OECD Statistics and Data Directorate), Eva Degler, Cécile Thoreau, Marie-Anne Valfort (all OECD Directorate for Employment, Labour and Social Affairs), Marcos Diaz Ramirez, Chris McDonald and Laura-Sofia Springare (all OECD Centre for Entrepreneurship, SMEs, Regions and Cities) for their comments on previous versions of the document. Evidence on data collection is mainly based on information provided to the OECD by: Myles Burleigh, Australian Bureau of Statistics; Stephan Marik-Lebeck, Statistics Austria; Stephan Moens, Statistics Belgium; Magdalena Kostova, National Statistical Institute of Bulgaria; Hélène Maheux, Statistics Canada; Javiera Rauld and Ulla Suulamo, Instituto Nacional de Estadísticas of Chile; Eduardo Efrain Freire Delgado, National Administrative Department of Statistics of Colombia; Maria Elena Gonzalez Quesada and Karla Jinesta, Instituto Nacional de Estadística y Censos of Costa Rica, Robert Šanda and Gabriela Strašilová, the Czech Statistical Office; Annika Klintefelt, Statistics Denmark; Marja-Liisa Helminen, Statistics Finland; Gunter Brückner, Federal Statistical Office of Germany; Panagiota Poziou, Hellenic Statistical Authority; Marcell Kovacs, Hungarian Central Statistical Office; Anton Örn Karlsson, Statistics Iceland; Eoin Whelan, Central Statistics Office of Ireland; Yafit Alfandari, Central Bureau of Statistics of Israel; Valeria de Martino and Federico Polidoro, Italian National Institute of Statistics; Hideyuki Ishimaru, Ministry of Internal Affairs and Communications of Japan; Eonseong Kim, Statistics Korea; Sigita Šulca, Central Statistical Bureau of Latvia; Dalia Ambrozaitiene, Statistics Lithuania; Jérôme Hury, National Institute of Statistics and Economic studies of the Grand Duchy of Luxembourg; Pilar García Velázquez, National Institute of Statistics and Geography of Mexico; Han Nicolaas, Statistics Netherlands; Robert Didham, Statistics New Zealand; Paul Inge Severeide and Signe Vrålstad, Statistics Norway; Grzegorz Gudaszewski, Statistics Poland; Maria Veiga, Statistics Portugal; Andreea Cambir, National Institute of Statistics of Romania; Nikitina Svetlana, Federal State Statistics Service of the Russian Federation; Mária Katerinková, Statistical Office of the Slovak Republic; Danilo Dolenc, Statistical Office of the Republic of Slovenia; Ana Cánovas Zapata, National Statistical Institute of Spain; Mikael Schöllin, Statistics Sweden; Marcel Heiniger, Federal Statistical Office of Switzerland; Şebnem Beşe Canpolat, Turkish Statistical Institution; Pete Large and Pamela Spicer, Office for National Statistics of the United Kingdom; and Roberto Ramirez and Rachel Marks, US Census Bureau. Their time and knowledge, which helped build this review, is much appreciated.

OECD STATISTICS WORKING PAPER SERIES

The OECD Statistics Working Paper Series – managed by the OECD Statistics and Data Directorate – is designed to make available in a timely fashion and to a wider readership selected studies prepared by OECD staff or by outside consultants working on OECD projects. The papers included are of a technical, methodological or statistical policy nature and relate to statistical work relevant to the Organisation. The Working Papers are generally available only in their original language – English or French – with a summary in the other.

OECD Working Papers should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed are those of the authors.

Working Papers describe preliminary results or research in progress by the authors and are published to stimulate discussion on a broad range of issues on which the OECD works. Comments on Working Papers are welcomed, and may be sent to the Statistics and Data Directorate, OECD, 2 rue André Pascal, 75775 Paris Cedex 16, France.

This document, as well as any statistical data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

The release of this working paper has been authorised by Martine Durand, OECD Chief Statistician and Director of the OECD Statistics and Data Directorate.

www.oecd.org/sdd/publicationsdocuments/workingpapers/

Abstract / Résumé

Data on ethnic, racial and indigenous identity can help render certain minorities statistically visible, and expose potential discrimination and inequalities. This paper systematically reviews diversity data collection practices in OECD countries and selected key partners and identifies three common challenges: the legal treatment of 'sensitive' data and concerns around privacy; the use of different data sources for different policy purposes; and issues of comparability over time since identities are dynamic and multiple constructs. When relevant, recommendations and best practices to improve diversity data are put forward. These include: expanding the collection of data on ethnic and racial identities where legal frameworks permit; ensuring the representation of hard-to-reach populations such as indigenous communities; developing national diversity statistical standards to standardise information and allow linking data across sources; raising the timeliness and policy relevance of diversity data by including questions in both regular sample surveys and population censuses; and involving communities in the data collection process.

Keywords: Data collection, ethnicity, race, indigenous peoples, migration, well-being.
JEL Classification: C80, I3, J15.

****

Les données sur l'identité ethnique, raciale et autochtone peuvent aider à rendre certaines minorités visibles et à dévoiler des discriminations et des inégalités potentielles. Ce document examine systématiquement les pratiques de collecte de données sur la diversité dans les pays de l'OCDE et certains partenaires clés et identifie trois défis communs: le traitement juridique des données « sensibles » et les préoccupations relatives à la protection de la vie privée; l'utilisation de différentes sources de données à des fins politiques différentes; et les problèmes de comparabilité dans le temps puisque les identités sont des constructions dynamiques et multiples. Lorsque cela est pertinent, des recommandations et des meilleures pratiques pour améliorer les données sur la diversité sont avancées. Celles-ci comprennent: élargir la collecte de données sur les identités ethniques et raciales lorsque le cadre juridique le permet; assurer la représentation des populations difficiles à atteindre telles que les communautés autochtones; élaborer des normes statistiques nationales sur la diversité afin de standardiser les informations et de relier les données entre sources; assurer l'opportunité et la pertinence politique des données sur la diversité en incluant des questions dans les enquêtes par sondage régulières et les recensements de population; et impliquer les communautés dans le processus de collecte de données.

Mots clés : Collecte de données, appartenance ethnique, race, peuples autochtones, migration, bien-être.
Classification JEL : C80, I3, J15.

Table of contents

Diversity statistics in the OECD: How do OECD countries collect data on ethnic, racial and indigenous identity? 2

1. Introduction 6

2. Why measuring diversity is important 12

3. Overview on diversity data collection in the OECD 15

4. Common challenges in the collection of diversity data 20

5. Discussion and the statistical agenda ahead 38

References 40

Annex A. Diversity data collection in the OECD: Detailed results 48

Tables

Table 3.1. Estimated indigenous populations in OECD member and accession countries 18
Table 4.1. Frameworks regulating the collection of race-related data 23

Table A A.1. Collection of data on migrant status 48
Table A A.2. Collection of data on racial identity 50
Table A A.3. Collection of data on ethnic identity 52
Table A A.4. Collection of data on indigenous identity 54

Figures

Figure 3.1. Diversity collection practices 16

Boxes

Box 1.1. Glossary of key terms 7
Box 1.2. Diversity data and the 2030 Agenda 9
Box 2.1. Policy spotlight: The United Kingdom's Race Disparity Unit 13
Box 3.1. Is migrant status a good measure of diversity? 17
Box 3.2. OECD work with indigenous communities: recent estimates and lessons learned 17
Box 4.1. Community involvement in the collection and analysis of diversity data 24
Box 4.2. Identification methods beyond self-declaration 29
Box 4.3. Additional sources of information on diversity 31

1. Introduction

1. Diversity is part of our everyday reality. No matter where we live and work, we are in constant contact with people from a range of diverse backgrounds. The term "diversity" is used here as an umbrella concept that recognises that people, whilst similar in many ways, have different life experiences and characteristics, such as gender, age, race, ethnicity, physical abilities, religion and beliefs. Some diversity is primarily cultural (e.g. shared norms and behaviours), other may be biological (e.g. age, sex), and other yet is defined in personal terms (e.g. sexual orientation, religion).

2. Individuals can belong to many different groups and each of these shape a person's identity and how he or she is viewed by society.[2] Assessing differences among groups that share common traits is important to highlight disadvantages faced by, for example, migrants, youth or women (Steward, 2001[1]; OECD, 2017[2]). In the real world, however, various affiliations interact with each other and are tightly bound together, so that individuals may experience multiple advantages and disadvantages due to their belonging to different groups (OECD, 2017[2]).

3. This report focuses on specific aspects of diversity: racial, ethnic and indigenous backgrounds. These are crucial facets of self and group identity, and represent important lines along which our societies are becoming increasingly diverse. The racial and ethnic make-up of our societies is in flux: international migration has been adding to cultural and phenotypic diversity; while the boundaries between racial and ethnic groups are becoming blurred by high rates of intermarriage and a growing number of persons with mixed ancestry.

4. Terms such as "race", "ethnicity", and "indigenous identity" may carry different connotations depending on the context and country.[3] It would be hazardous to try to summarise them all; less ambitiously, this report provides a common language to understand how such concepts can be operationalised for statistical purposes (Box 1.1). When available, international and official standards are used as reference. This paper additionally considers statistics related to "migrant status" as a separate category, since in a number of countries information such as country of birth is de facto the only available information to make inferences about diversity.

2. In Sen's words: "*The same person can, for example, be a British citizen, of Malaysian origin, with Chinese racial characteristics, a stockbroker, a non-vegetarian, an asthmatic, a linguist, a bodybuilder, a poet, an opponent of abortion, a bird-watcher, an astrologer, and one who believes that God created Darwin to test the gullible*" (Sen, 2007[78]).

3. As with other social constructs, race and ethnicity may be given negative social meaning, when they are used to discriminate, to exclude, to exploit, or to profile people; but they can equally be sources of pride and form important parts of one's psychological identity (Sen, 2007[78]).

Box 1.1. Glossary of key terms

Diversity data – Refer to all types of disaggregated data used to inform about the presence and size of different groups within society (diversity data for enumeration), or to give evidence on the well-being outcomes of those groups and on the impacts of governmental policies and private sector practices on these groups (diversity data for assessing well-being outcomes and inequalities). The reader should keep in mind that no single definition or classification of diversity and its components is suitable to be recommended to all countries because ethno-cultural composition vary widely between them.

Collective identities – Refers to the sense of identity that is formed by belonging to a given group. An individual can form more than one identity. The term is used here to refer to identity shaped around racial, ethnic and indigenous backgrounds.

Race – In the absence of any internationally agreed definition, race is most often statistically characterised in terms of phenotype and appearance (e.g. skin colours), or with regard to ancestry. This should not be understood as an attempt to trace the definition of race to biological, anthropological or genetic factors but rather to (somewhat artificially) distinguish it from the concept of ethnicity.

Ethnicity – Describes a shared culture: the practices, values, and beliefs that characterise those belonging to a community. This multidimensional concept acts as an umbrella term encompassing language, religion traditions and other (United Nations, 2017[3]). A number of related concepts, including ancestry, citizenship and nationality, may overlap with ethnicity. However, ethnicity is not the same as nationality or citizenship, nor it is a measure of biology or genes.[4]

Indigenous identity – While no universal definition exists in international law, the term is used to refer to "tribal peoples whose social, cultural and economic conditions distinguish them from other sections of the national community, and whose status is regulated (wholly or partially) by their own customs or traditions or by special laws or regulations; and to peoples in independent countries who are regarded as indigenous on account of their descent from the populations which inhabited the country (or a geographical region thereof) at the time of conquest, colonisation or establishment of present state, and who, irrespective of their legal status, retain some or all of their own social, economic, cultural and political institutions" (ILO, 1989[4]).[5]

Migrant status –The *UN Recommendations on Statistics of International Migration* suggest that five basic criteria that might be used – either alone or in combination – to characterise an international migrant: residence; citizenship; time or duration of stay; purpose of stay, and place of birth. Generally, the status of migrant is defined in different ways in different countries, e.g.: *i)* someone whose country of birth differs from their country of usual residence; *ii)* someone whose nationality is not that of their country of usual residence; or *iii)* someone who has changed his/her country of usual residence for a period of at least a year, so that the country of destination becomes the country of usual residence (United Nations, 1998[5]). For the purposes of this paper, the term migrant status does not make any reference to a person's legal status, but rather refers to the catch-all category of "immigrant" or "foreign born" often used in demographic and social statistics and proxied by "country of birth", "citizenship" or "country of birth of one or both parent(s)" (European Commission, 2013[6]).

5. In many OECD countries, a vibrant equal rights movement, concerns about racism and discrimination, a call for the recognition of the rights of indigenous populations, and discussions on how to integrate increasingly diverse populations into society have made racial and ethnic background key issues in public debates. However, despite heightened awareness and policy interest, the availability of data on diversity in many countries have not always kept pace with emerging trends, partly due to the difficulties to operationalise complex and flexible constructs into valid and reliable statistical categories, and partly due to concerns about potential misuse of such data.

6. Whatever the reason, however, one fact is clear: without reliable data capturing population heterogeneity, it is difficult to get a sense of the presence and size of different groups within society (diversity data for enumeration) and to address discrimination and inequalities (diversity data for assessing outcomes and inequalities). Countries may be blinding themselves about some of the worse types of disadvantage by not collecting data on the experience of various minorities. The lack of relevant diversity data means that non-discrimination policies and laws are not implemented as effectively as they could if policies were better targeted, and that some communities remain statistically invisible within societies. The *UN Sustainable Development Goals* and other international initiatives demand

4. Race and ethnicity are terms that carry heavy intellectual and political baggage, and issues surrounding racial and ethnic identities are often contested within countries and across groups. Academic and popular understandings of racial and ethnic identities have changed dramatically over time. Prior to the 20th century, racial groups were generally perceived as permanent and distinct entities, with biology greatly responsible for differences in the cultures and the political and economic fortunes of these groups. This perspective, rooted in the experience of colonialism and slavery, established a mode of classification based on a rigid hierarchy of socially exclusive categories (especially in the United States). Today, social scientists generally agree that race is a socially constructed rather than a scientific category, and that most genetic markers do not differ sufficiently across "races" to be useful in biological or medical research. Studies have also documented the processes by which ethnic and racial boundaries have changed throughout history (Fujimura, Duster and Rajagopalan, 2008[93]; Morning, 2011[86]; Morning, 2014[87]; Roth, 2016[54]). This perspective implies that race is constructed by social, economic, political conditions that can change over time. For instance, past statistical practice in the United States included among "non-White" Irish, Italians, and eastern European Jews immigrants, groups that are now classified as "White" (Lee and Bean, 2004[89]) Nevertheless, although race is a social construction, perceptions may have real consequences – discrimination, attitudes by other people, inequalities, or as a positive source of identity. Disparities that result from racial discrimination can thus be seen as "biological expressions of race relations" (Krieger, 2000[91]).

5. Indigenous peoples usually have special status that distinguishes them from other groups. The crucial feature of indigenous people is their claim to a right of self-determination, by virtue of which they could determine their political status, pursue their economic, social and cultural development, and decide how to engage within their nations (United Nations, 2007[101]). Although indigenous peoples' relationship with the land and culture is similar across countries, their livelihoods, customs and effective rights vary widely between and within countries. For instance, in Canada, Australia, New Zealand and the Americas, indigenous communities occupied the land before the arrival of European settlers, while in other countries indigenous peoples may have endured domination by populations from neighbouring countries or from within. In Asia and Africa, most people consider themselves indigenous having achieved decolonisation and self-determination from European colonies (Bartlett et al., 2007[98]).

more robust data to monitor horizontal inequalities in their plea for "leaving no one behind" (Box 1.2; UN General Assembly (2015[7]).[6]

Box 1.2. Diversity data and the 2030 Agenda

On 25 September 2015, the UN General Assembly adopted the Sustainable Development Goals (SDGs), which aim to end poverty, tackle environmental change and fight injustice, as part of a new sustainable development agenda. Monitoring the implementation of the policy commitments should be based upon indicators "*disaggregated, where relevant, by income, sex, age, **race**, **ethnicity**, migratory status, disability and geographic location, or other characteristics*". Disaggregating SDGs by minority and vulnerable groups adds an important dimension to country-level reporting, highlighting horizontal inequalities between countries. At the same time, disaggregation helps identify policy priorities within countries: even when a country meets the individual targets on average, it could still improve its performance on the SDGs if a substantive share of its population falls behind on the respective indicators, as is often the case with minority groups.

Besides the inclusion of race and ethnicity as disaggregation variables of the 17 Sustainable Development Goals, four explicitly mention ethnicity, race or indigenous status:

- Goal 2: "*By 2030, double the agricultural productivity and incomes of small-scale food producers, in particular women, **Indigenous peoples**, […]*" (Target 2.3);
- Goal 4: "*By 2030, eliminate gender disparities in education and ensure equal access to all levels of education and vocational training for the vulnerable, including persons with disabilities, **Indigenous peoples** and children in vulnerable situations*" (Target 4.5);
- Goal 10: "*By 2030, empower and promote the social, economic and political inclusion of all, irrespective of age, sex, disability, **race, ethnicity**, origin, religion or economic or other status*" (Target 10.2);
- Goal 17: "*By 2020, enhance capacity-building support to developing countries ... to increase significantly the availability of high-quality, timely and reliable data disaggregated by income, gender, age, **race, ethnicity**, migratory status, disability, […]*" (Target 17.18).

Despite recognition of the important role that racial and ethnic backgrounds play in shaping people's well-being, data availability remains overall limited and further statistical effort is needed if countries are to live up to the 2030 SDG Agenda's aspiration of "leaving no one behind".

7. The time is thus right to assess the state of diversity statistics in OECD countries and in selected key partners to better understand the current challenges and the opportunities ahead. This assessment is based on results from a questionnaire sent (in November 2017) to national statistical offices (NSOs) in OECD and selected partner countries to assess current practices in the collection of diversity data. The analysis presented here is based primarily

6. Inequalities can be assessed by looking at vertical inequalities (i.e. among people ranked by whatever variable of interest) or at horizontal ones (i.e. differences in average outcomes between groups of people sharing some common trait such as gender, age or education; OECD (2017[2])). Inequalities by ethnic, racial or indigenous identity are part of horizontal inequalities.

on the feedback provided by NSOs, supplemented with publicly available information from survey descriptions or questionnaires.

8. The main purpose of this document is to provide data producers, policy makers, and communities with information on data collection practices in the field of racial, ethnic and indigenous identities. This effort represents a first, necessary step to develop evidence-based policies in the field. Given the diversity in countries' conditions (e.g. in terms of their demographic make-up and statistical infrastructure), the paper does not propose a uniform and standardised model of data collection and classification, but rather a toolbox that NSOs could use to inform the collection of valid and reliable diversity data going forward.

9. Key messages of this report are:

- Over the last decade, increasing awareness of societies' ethnic and racial diversity in some countries, and of the need for more effective inclusion policies, has created a momentum for collecting more and better diversity data. Such data can help expose discrimination and inequalities faced by certain groups and render them statistically visible.
- The measurement approaches and regulations that underpin what is often considered 'sensitive data' differ significantly across the OECD, with collection practices clustered around three broad categories: (1) all OECD countries collect information on diversity proxies such as country of birth (35 OECD members); (2) a small majority, mostly Eastern European countries as well as the United Kingdom and Ireland, gather additional information on race and ethnicity (16 OECD members); (3) only a handful of countries in the Americas and Oceania collect data on indigenous identity (6 OECD members).
- Collecting and analysing data on diversity is difficult but not impossible. Common challenges include (1) the treatment of 'sensitive' data and concerns around privacy as well as the reluctance of some groups to disclose their identity; (2) the use of different sources for different policy purposes, and the need to validate and link the collected data; and (3) issues of comparability over time, due to the flexible and contingent nature of collective identities.
- Potential improvements in diversity data include expanding data collection to variables beyond country of birth where legal frameworks permit it, ensuring the representation of hard-to-reach populations such as indigenous communities through non-standard sampling techniques; allowing respondents to report multiple identities; developing diversity statistical standards; and raising the timeliness and policy relevance of the data collected by including questions in regular sample surveys (rather than only in the population census); and linking data across sources. Involving communities in the data collection process is also important to improve trust and the quality of the information collected.

10. This report is part of an OECD-wide Diversity project that aims to address two main questions: 1) *How diverse and open to diversity are our societies?*; and 2) *How to make the most out of more diverse societies?* Related initiatives that point towards the importance of improving the quality of diversity data include recent OECD work on selected minority groups, such as indigenous people (OECD, 2017[8]; 2018[9]), and migrants (OECD, 2015[10]; 2017[2]; 2018[11]), as well as diversity data reviews by other international institutions (e.g. European Commission (2017[12]) or Open Society Foundations (2014[13])). The report also reflects a broader OECD endeavour that focuses on people's well-being, inclusiveness and equality. It is testimony of a growing awareness in countries that successful interventions to

tackle inequalities and discriminations need to build on better data on how well-being outcomes are distributed across population groups (OECD, 2017[2]).

11. This paper is organised as follows. Section 2 makes the case for collecting better diversity data and describes their possible policy use. Section 3 provides a broad overview of what type of diversity data OECD countries currently collect, and Section 4 describes common challenges in this endeavour. Section 5 concludes with some suggestions for advancing the statistical agenda on diversity.

2. Why measuring diversity is important

12. Provisions on equality and non-discrimination are enshrined in international human rights law and in the constitutions and legislations of most countries.[7] Despite international and domestic anti-discrimination laws, many people continue to face prejudice, harassment and hate crime because of their ethnic origins, which complicate efforts to improve their life chances and living standards (FRA, 2018[14]).[8] People from ethnic minority and indigenous communities show lower levels of well-being in a variety of key areas, such as education, employment, living standards, health, and housing (even after controlling for social and economic factors) (OECD, 2017[2]; FRA, 2017[15]; United Nations, 2009[16]). As race and ethnicity are beyond the control of each person, differences along these violate the notion of equality of opportunity.

13. Collection of accurate and comprehensive data on diversity is thus central to providing information on the size of different communities (**diversity data for enumeration**) and to implementing, monitoring, and evaluating civil rights laws and policies that aim to address disadvantages and promote equal opportunities in all sectors of society and across public services (**diversity data for assessing well-being outcomes and inequalities,** Aspinall (2012[17])).

14. There are three types of policy use cases for diversity data in the countries surveyed:

- **Distribution of resources.** Information on the size of recognised ethnic minorities and indigenous populations from census counts are used in a number of countries (e.g. Australia, New Zealand and Colombia) to allocate funds to these communities, to allocate parliamentary seats (e.g. Ireland) and to distribute federal resources (e.g. the United States).
- **Reveal inequalities between groups.** Diversity data can be used to shed light on how different population groups are doing, and identify areas where change is needed. For example, the United Kingdom's Cabinet Office conducted a Race Disparity Audit in 2017 to examine how people of different backgrounds are treated across public services including health-care, education, employment, home ownership and the criminal justice system (Cabinet Office (2017[18]); Box 2.1). The

7. The 1948 Universal Declaration of Human Rights proclaims that all human beings are born free and equal in dignity and rights, and that everyone is entitled to all the rights and freedoms set out in the Declaration, without distinction as to race, colour or national origin. Other UN Conventions committing countries to the elimination of racial discrimination and the promotion of understanding among all races include: the International Convention on the Elimination of All Forms of Racial Discrimination (ICERD); the International Covenant on Civil and Political Rights (ICCPR); the International Covenant on Economic, Social and Cultural Rights (ICESCR). In 2015, the UN General Assembly established a Forum for People of African Descent to serve as a consultation mechanism during the International Decade for People of African Descent (2015-2024). Moreover, Member States of the Council of Europe must abide by the European Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR) and the European Social Charter (ESC), alongside a range of recommendations and resolutions.

8. According to a 2017 Pew Research Center poll, 41% of Americans agree with the statement that racial discrimination is the main reason why many black people can't get ahead in life these days, a share which is up 9 points since previous year and is the highest level since 1994 (Pew Research Center, 2017[84]).

report yielded a complex picture: for instance, ethnic minorities are under-represented at senior levels in the public sector; Roma children are falling behind their peers at school; and black men face the highest likelihood of being found guilty in court. At the same time, among the poorest children, white British pupils do worst at school, with only 32% reaching the expected standard of reading, writing and maths at 11, and are more likely to smoke than their minority ethnic counterparts. This underscores the importance of diversity data for all groups of society, not only for minorities (DGME, 2014[19]).

Box 2.1. Policy spotlight: The United Kingdom's Race Disparity Unit

Beyond commissioning a Race Disparity Audit (to highlight the experiences and outcomes of people of all ethnicities across public services) the United Kingdom's Cabinet Office established a Race Disparity Unit (RDU) in 2016 with data, digital, policy and program expertise drawn from across Government, data partners, and the private sector.

The RDU tackled one of the key issues confronted by most OECD countries: transparency and accessibility of diversity data is often limited as public services do not systematically collect and release information on ethnic disparities in one place. In this context, the RDU developed an Ethnicity Facts and Figures data portal (www.ethnicity-facts-figures.service.gov.uk/) through extensive user-engagement, analysis, and data and digital standards to detail government information in one accessible point. The website, launched in October 2017, highlights "uncomfortable truths" and ethnic disparities; this information provides a basis for policy makers and society as a whole to discuss where action is needed.

Since its launch, the RDU has sought ideas and expertise from a range of stakeholders including academics, charities, community groups, local authorities and the private sector. Some of the recent policy responses as a result of the insights gained by the RDU's work include: targeted employment support in areas with high levels of unemployment, the establishment of a fund for youth unemployment, the release of more and better data on disparities in the criminal justice system, the introduction of measures to diversify the prison workforce, and the launch of an external review to share best practices on reducing exclusion.

- **Benchmark progress on diversity policies.** Diversity data can also be used to measure performance towards previously agreed equality targets, to see whether strategies are working. The Australian Prime Minister's Office releases an annual report on the *Closing the Gap* initiative, established in 2008 to reduce disadvantage among Aboriginal and Torres Strait Islander people across 7 targets related to education, employment and health. The latest report showed progress in some areas, but also that targets related to school attendance and achievement, employment, and life expectancy are not on track to be met (Australian Government, 2018[20]). In Canada, information on the visible minority population (i.e. persons, other than aboriginal peoples, who are non-Caucasian or non-white) is required under federal legislation for programs which promote equal opportunity and is used by governments, business, community groups, health care providers and researchers. For example, the census provides benchmark data that can be used by employers to compare the characteristics of their workforce with those of the population living in the same area (Statistics Canada, 2017[21]). Costa Rica's 2014

National Policy for a Society free of Racism, Racial Discrimination and Xenophobia includes an Action Plan which foresees the development of a system of indicators to monitor progress. Although so far, systematic follow-up of the Plan's implementation has been delayed, the Inter-Agency Commission for Monitoring and Implementing International Human Rights Obligations is currently working on a methodological guide to improve the accuracy of the indicators used (DGME, 2014[19]).

15. The need for better evidence on diversity has been widely acknowledged by international bodies over the past two decades. In 1996, the European Commission against Racism and Intolerance, in its first General Policy Recommendation, stated that "*it is difficult to develop and effectively implement policies […] without good data*" (ECRI, 1996[22]) and called for the collection of "*data which will assist in assessing and evaluating the situation and experiences of groups which are particularly vulnerable to racism, xenophobia, antisemitism, and intolerance*" (ECRI, 1998[23]). In more recent years, the UNECE and its advisory committee recognised that "*[i]n order to measure and combat ethnic and racial discrimination, ethnicity is an essential background characteristic in every survey, just as age and gender*" (CEIES, 2007[24]).

16. These international agreements set the framework within which Member States can develop, collect and disseminate official statistics in this field. The next section will show that some NSOs have made significant progress in describing the demographic, economic and social status of racial and ethnic minorities as well as indigenous populations, and disseminate statistics in this field. However, some OECD countries with well-established national statistical systems have only limited statistics on diversity, and some have none at all.

3. Overview on diversity data collection in the OECD

17. In November 2017, the OECD Statistics and Data Directorate conducted a survey of national statistical offices (NSOs) in OECD and selected partner countries to assess current diversity data collection practices, focusing on racial, ethnic and indigenous identity. Even though migrant status-related information such as country of birth is not a recommended proxy for ethnicity (Box 3.1), it has been included in the questionnaire as it is de-facto the only available information in many countries. All OECD Members, bar Estonia and France, have participated in this review, as well as the partner countries Bulgaria, Colombia, Costa Rica, Romania and the Russian Federation. The questionnaire covered issues such as legal frameworks that underpin diversity data collection, available data sources, and ways of identification (Annex B).[9]

18. This section presents the key findings from the stocktake of current practices in the collection of diversity data in OECD and selected partner countries. For a detailed breakdown by type of collective identity including survey vehicle, question wording and response options, see Annex A, and for an overview of data sources per country, see Annex C.[10]

19. The following considerations are useful when interpreting the results of this report. First, they primarily reflect the information provided by NSOs themselves. Second, most NSOs indicated only their most current data collection practices. Information on changes in question wording or methodology remains limited. Third, some additional desk research has been carried out for European countries by including national SILC, EU-SILC and EU-LSF surveys. This desk research was not systematic in all cases, and all possible available practices or data sources may not be reflected in this review. Lastly, in several instances, NSOs classified their responses differently from how these responses are presented here. These reclassifications are indicated as relevant.

20. NSO's collection practices on diversity data cluster around three broad categories (Figure 3.1):

- A large bulk of countries, mainly the older EU member states, only collect information on migrant status;
- Countries, mostly in Eastern Europe as well as the United Kingdom and Ireland, gather additional information on race and ethnicity;
- Countries in the Americas and Oceania collect data on racial/ethnic and indigenous identity.

9. Annex B is available at: www.oecd.org/sdd/diversity-statistics-in-the-OECD-annex-B.pdf.

10. Annex A is available at the back of this report and at: www.oecd.org/sdd/diversity-statistics-in-the-OECD-annex-A.pdf, Annex C is available at: www.oecd.org/sdd/diversity-statistics-in-the-OECD-annex-C.pdf.

Figure 3.1. Diversity collection practices

Note: A country is listed as collecting a type of diversity data if at least one data source (e.g. census, sample survey, population registry) includes relevant information. Categories for Estonia and France are based on online research, as these countries did not respond to the questionnaire used for this study. As not all OECD countries have indigenous populations, the denominator is set at 12 for this category (OECD, 2018[3]).

21. Some sort of information on **migrant status** is collected in all of the countries under study. The information available mostly relates to country of birth (41 countries), country of birth of parents (30) and year of arrival in the country (17) (Table A A.1).[11] Information on the parent's country of birth is especially relevant, as it can be used to identify natives with immigrant parents, who often continue to face challenges related to discrimination and lower performance on a range of well-being outcomes (OECD, 2017[25]). Information on migrant status is generally self-reported in sample surveys or censuses, although a number of countries use population registers (Austria, Denmark, Finland, Iceland, Israel, Latvia, Lithuania, Netherlands, Norway and Sweden). Although valuable estimates can be produced from most mainstreamed annual or quarterly surveys, producing data at a more granular level or for specific groups (e.g. by detailed country of birth or by individual socioeconomic characteristics) is far from straightforward. Moreover, as country of birth is only an imperfect proxy for ethnicity (Box 3.1), **there is a strong case to be made for collecting additional diversity data where legal frameworks permit it, beyond simply improving migration statistics.**

11. Annex A is available at the back of this report and at: www.oecd.org/sdd/diversity-statistics-in-the-OECD-annex-A.pdf.

Box 3.1. Is migrant status a good measure of diversity?

In general, collecting migration-related information on the foreign-born population and their children is a crude method for capturing diversity. Although such data are relatively readily available and often considered as 'objective', their use as proxy for ethnicity or race is problematic. The country of birth of a person neither takes account of the diversity of the country of origin of the individual or the parents (e.g. 'White' people in the United Kingdom that were born in former British colonies) nor does it capture cultural affiliation, the inherently self-perceived aspect of belonging to an ethnic group (Gill et al., 2005[26]). This view is also reflected in the *UN Principles and Recommendations for Population and Housing Censuses* (United Nations, 2017[3]), which state that country of birth or citizenship as well as questions on religion and language should not be taken as providing proper ethnic data.[12]

22. Seventeen countries, out of the ones covered by this review, collect official statistics on ethnicity, 8 on race, and 7 on their indigenous populations (Table A A.2; Table A A.3; Table A A.4).[13] Almost as many OECD countries, despite recognising the presence of indigenous populations on their territories do not gather any official information on them, a situation that limits policy options to improve well-being outcomes and grant rights for indigenous groups (Box 3.2).

Box 3.2. OECD work with indigenous communities: recent estimates and lessons learned

Although the "right to be counted" is fundamental in asserting the voice of indigenous populations, in many countries they remain invisible in the statistical systems due to the general absence of clear and consistent data (Axelsson, 2018[27]; OECD, 2018[9]). The most recent global estimates of indigenous peoples is at 302.45 million, and indigenous communities are found within over 90 different countries and seen in almost all regions in the world (Hall and Patrinos, 2012[28]; United Nations, 2009[16]).

Out of the OECD countries with indigenous communities, Denmark, Finland, France-New Caledonia, Japan, Norway, and Sweden do not currently collect official data, mainly because they are legally not allowed to collect information on the basis of ethnicity (Table 3.1).

In all cases the number of indigenous people is small relative to the national population and communities tend to be concentrated in specific locations. An OECD project on *Linking Indigenous Communities with Regional Development* was launched in 2017 to develop policy recommendations that improve economic development outcomes for indigenous people by better linking them with regional and rural development efforts. The project has involved participation from Australia, Canada, Colombia, the European Commission, Finland, New Zealand, Norway, Sweden and the United States, and focuses

12. The *UN Principles and Recommendations for Population and Housing Censuses* referenced throughout this paper refer to Revision 3 (United Nations, 2017[3]).

13. Annex A is available at the back of this report and at: www.oecd.org/sdd/diversity-statistics-in-the-OECD-annex-A.pdf.

on the four key themes of better statistics, land and economic development, business growth, and governance and capacity.

Table 3.1. Estimated indigenous populations in OECD member and accession countries

OECD member countries	Indigenous peoples	Population	% National population
Australia	Aboriginal	798 381	3.3 %
Canada	First Nation/Inuit/Metis	1 673 785	4.9 %
Chile	Various	2 185 729	9 %
Denmark (Greenland)	Inuit	50 220	(85%)
Finland	Sami	10 000	
France (New Caledonia)	Kanak	104 958	(39.1%)
Japan	Ainu	28 782	0.02%
Mexico	Various	12 250 947 - 25 699 111	10.1 % - 21.5 %
New Zealand	Maori	692 300	16.3 %
Norway	Sami	50 000 - 65 000	1% - 1.3 %
Sweden	Sami	20 000	0.20 %
United States	American Indian/ Alaskan native	6 706 210	2 %
Colombia	Various (65 Armerindin languages)	1 392 623	3.4 %
Costa Rica	Various inc. Bruca and Bribri	104143	2 %
Russian Federation	Various	257 895	0.2%

Note: Population data for Greenland refers to population born in Greenland. Greenland is defined as an autonomous country within Denmark, whilst New Caledonia is a special collectively of France. Data refer to 2017 for Chile; 2016 for Australia, Canada and the United States; 2015 for Mexico; 2014 for France; 2013 for New Zealand; 2011 for Costa Rica; 2005 for Colombia. Estimates for the Scandinavian countries are problematic because the identification of Sami peoples is based around the use of indigenous languages and the traditional practice of reindeer herding, even though this only holds true for the minority of these peoples today. The estimated size of the indigenous population in Mexico is 12.3 million (based on the spoken language of indigenous household), and 25.7 million if the definition is extended to those who self-identify as indigenous.
Source: OECD (2018), *Linking Indigenous Communities with Rural and Regional Development*, OECD publishing, forthcoming.

Initial work of the OECD project has revealed some lessons and insights about how to work with indigenous communities and the issues that are important to them. This includes engaging directly with indigenous leaders and institutions and involving them in the peer-review process, directly engaging with local indigenous communities, adapting views of development to incorporate indigenous values and perspectives, considering the fundamental legal rights frameworks that shape self-determination, and the need to address data gaps to support the development of better public policies (OECD, 2018[9]).

23. The three identified clusters of countries reflect different approaches to the recognition of diverse communities as well as actual levels of diversity. Those countries collecting information on all types of diversity tend to have very heterogeneous populations due to their status as former colonies (Chile, Mexico, Colombia, Costa Rica), settler colonies with pre-existing indigenous populations (Australia, Canada, New Zealand), or their legacy of the slave trade from the African continent (e.g. to the United States and other American countries). Several of these countries also have long employed active immigration policies as an element of nation building and fostering human capital.

24. In the United Kingdom and Ireland, as well as in many Eastern European countries, race and ethnicity are familiar topics in their national discourse due to either their status as

destination country for migrants from former colonies (United Kingdom) or due to the presence of significant national minorities (Eastern European countries and Ireland). On the contrary, most Western and Southern European countries only collect data on migrant status. This typically reflects explicit provisions against the collection of race and ethnicity data in their legal frameworks (European Commission, 2017[12]; Simon, 2007[29]; 2012[30]).

25. While this review is not specifically focusing on other types of diversity beyond race, ethnicity and indigenous identity, several countries provided additional information on their collection of religious data. Overall, 12 countries collect some information on religious affiliation: Australia, Canada, Germany, Hungary, Israel, Lithuania, Mexico, New Zealand, Portugal, Switzerland, the United Kingdom and Bulgaria include a question in their census. In Germany, information on religious affiliation is also included in its population statistics since the government collects taxes on behalf of predefined religious communities (e.g. the Catholic and Protestant Church of Germany). All these countries except Israel and Germany have explicit laws promoting religious data collection, but often specify that failure to answer is not a criminal offence. The Slovak Republic recently passed a law encouraging collection on religious affiliation for the upcoming 2021 Census round.

26. As for sexual orientation and/or gender identity, no census in OECD countries has ever included questions on in order to identify homosexual, bisexual, and transgender individuals. The bulk of population-based surveys apprehend the LGBT population in an indirect way, namely through the sex of the respondent's partner. This approach amounts to focusing on only a subset of the LGBT population, i.e. individuals living with a same-sex partner. That said, 15 OECD countries to date have already asked a question on sexual self-identification - whether the respondent thinks of herself as heterosexual, homosexual, bisexual or other- in one or several of their nationally representative surveys, noting that this trend is increasing. Additionally, three OECD countries have started collecting representative statistics on the transgender population: Chile, Denmark and the United States (OECD, 2019[31]).

4. Common challenges in the collection of diversity data

27. While the previous section argued that there is no one-size-fits-all way to collect diversity data, countries face common challenges in operationalising, collecting and processing information on racial and ethnic backgrounds. Some of the key issues include: (1) the treatment of what is often considered 'sensitive' data, concerns around privacy, reluctance of some groups to disclose their identity; (2) the use of different sources and the need to validate and link the collected data; and (3) issues of comparability relating to the flexible and contingent nature of collective identities. These issues shape the framework within which countries can develop, collect and disseminate diversity data, and the methodological and measurement options used. While recommended standards sometimes exist (e.g. some of the core principles of data collection put forward in the *UN Principles and Recommendations for Population and Housing Censuses*), in other instances NSOs need to develop internal guidance to ensure that the collection of diversity data across different data sources and over time rests on agreed concepts and common methodologies. When relevant, recommendations and best practices are put forward. Ongoing international initiatives, such as the *Subgroup on Equality Data* set up in 2018 by the EU High Level Group on Non-Discrimination, Equality and Diversity to draft non-binding guidelines on improving the collection and use of diversity data, should help provide further guidance to the challenges presented here.

4.1. Legal approaches to diversity data and privacy concerns

28. Diversity data are often qualified as "sensitive" or "special category" data; as such, their collection, dissemination and use is usually regulated by national and international legislation. The legal frameworks underpinning data collection can influence not only whether relevant information can be gathered, but in some cases also which groups are officially recognised.

29. Legal frameworks governing the collection of diversity data map on to the three country-clusters identified in Section 3 (Figure 3.1): the more encouraging the legal framework, the more data beyond migrant status is collected. However, specific approaches can differ significantly. In a few cases, sub-categories of people are explicitly defined, and a mandate for data collection is specified either for general statistical collection or only for a country's census in a specific round. In other instances, the legal basis does not extend to statistical activities per se but is grounded in human rights, anti-discrimination and/or privacy protection legislation. Even some of the countries specifically encouraging the collection of diversity data recognise it as 'sensitive', strongly emphasising that responses are voluntary, should be treated with special care and might not be published or used for discriminatory purposes.

30. Two main factors can account for the 'sensitivity' of such data. First, it is feared that it may be misused to maintain or deepen power relationships between majority and minority population groups (Simon and Piché, 2012[32]; Durante, Volpato and Fiske, 2010[33]). This concern is particularly heightened for groups who may have experienced ethnic profiling, segregation, genocide and violence (Chopin, Farkas and Germaine, 2014[13]), and especially so in countries where ethnicity-based data were used in the past to provide the basis for discriminatory policies. Second, several countries engage in a "colour-blind" approach that explicitly forbids the collection of any data on collective identities at

all. Ethnicity-blindness is supposed to ensure equality of all citizens in front of the State and, consequently, in social life more generally, regardless of ethnic or racial differences. Simon (2015[34]) refers to this colour-blind approach as "*equality through invisibility*", to emphasise that the liberties of a group are protected by eradicating diversity and forging sameness amongst all citizens.

31. In principle, the colour-blind approach upon which some countries (e.g. France, Denmark, Germany and Sweden) prohibit the collection of ethnic and racial data is well aligned with the democratic principle of impartiality. However, in practice, disadvantage groups can suffer from the lack of explicit information about their experiences and quality of life, and be disregarded by a democracy that has little knowledge about their demographic presence and the challenges they face. The absence of diversity data does not prevent resentments and differences from arising; and the experience of those societies that embrace ethnic diversity does not support the notion that systematic collection of data by ethnicity and race endangers the social cohesion of the country (Reitz, Zhang and Hawkins, 2011[35]).

32. Different minority communities often disagree on whether or not collecting data by racial and ethnic origin is desirable. In general, groups advocating against hate speech and crimes, or representing non-recognised communities (e.g. Afro-Europeans), are more vocal supporters of data collection, as are indigenous communities that feel invisible and deprived of their "right to be counted". Conversely, European anti-racist non-governmental organisations (NGOs), and Jewish and Roma communities often oppose personal data collection based on historical and current experience of data abuse, particularly by racist or xenophobic parties, to discriminate, stereotype, and misinterpret data (European Commission, 2017[12]).

33. Statistical legislation generally imposes few constraints in the case of data collection by **migrant status**. No country legally limits asking questions on migrant background, while 12 OECD countries have legislation encouraging such data collection (Australia, Austria, Canada, Chile, Germany, Latvia, Lithuania, Mexico, New Zealand, Slovak Republic, Turkey, United Kingdom and the Russian Federation).

34. Data collection on indigenous identity is generally rooted in law. The majority of countries collecting such information (5 out of 8 – Australia, Canada, New Zealand, Mexico, and Colombia) are officially obliged to do so. Chile and Costa Rica gather such data despite the absence of an explicit statistical obligation, but because of equality legislation (e.g. Costa Rica's National Policy for a Society free of Racism, Racial Discrimination and Xenophobia).[14]

35. The introduction of specific statistical legislation over the last decade, in particular in Latin American countries, reflects both the need to comply with the recommendations of

14. Switzerland, after acceding to the European Charter for Regional and Minority Languages in 1997, legally recognised linguistic groups as national minorities, and defined Romansh and Italian as regional or minority languages and Yenish as a non-territorial language. Switzerland is omitted in further analysis as this definition does not map onto the definition of indigenous populations used in this review. Similarly, the United Kingdom's data collection on national identity (referring to English, Welsh, Scottish, Northern Irish, or British) has been classified as ethnicity rather than indigenous identity. Some parts of Russian Federation's diverse indigenous population are legally recognised under the "The Unified List of Indigenous Peoples of the Russian Federation" (N 255, 2000) and protected under the Constitution (Article 69); this review treats this under the "ethnicity" category since official data collection uses the exact same identification question for ethnic and indigenous identity.

international human rights or statistical organisations and political movements for the recognition of indigenous populations and people of African descent (Simon, 2017[36]). For instance, Chile subscribed to the ILO's Indigenous and Tribal Peoples Convention in 2009. Colombia introduced mandates for the NSO to include ethnicity and territoriality variables for Afro-Colombians and to collect census data of the indigenous population in order to facilitate administration of funds to their respective territories.

36. In some cases, what constitutes state-recognised membership in specific indigenous communities is defined – this is the case, for instance, for Maori ethnicity and/or descent and iwi affiliation in New Zealand, or for Aboriginal and Torres Strait Islander identification in Australia. In Canada, the term "Aboriginal peoples" includes Indian, Inuit and Métis peoples ("First Nations people" and "North American Indian" are commonly used terms for "Indian"), with "Status Indians (Registered or Treaty Indians)" registered as Indians under the Indian Act of Canada. Many OECD countries with indigenous populations do not collect data pertaining to them (Denmark, Finland, France-New Caledonia, Japan, Norway, Sweden), even though some of these communities might be officially recognised as indigenous elsewhere in national or indigenous law. In Sweden, the Personal Data Act (1998:204) prevents official data collection on the Sami people (Axelsson, 2018[27]; OECD, 2018[9]).

37. Legislation on **race and ethnicity** is somewhat split between the two concepts. On the one hand, most countries collecting data on **ethnicity** operate under legislations encouraging such collection (Australia, Canada, Hungary, Latvia, Lithuania, Mexico, Poland, the Slovak Republic, the United Kingdom, the United States, Bulgaria, Romania and the Russian Federation). A minority of these countries have developed official standards that list major ethnic groups. These are often reviewed at each census round, thus emphasising the social construction and movable boundaries of the concept. Examples include the Australian Standard Classification of Cultural and Ethnic Groups (ASCCEG under Census and Statistics Regulation 2016 and the Census and Statistics Act 1905); and the US Office of Management and Budget 1997 Standards for Race and Ethnicity, which specify five minimum categories for data on race and two categories for data on ethnicity. 202 ethnicities are listed in Iceland's national legislation, and over 200 ethnic categories are included in the New Zealand Ethnicity Standard (Statistics Act 1975, Human Rights Act 1993, Bill of Rights Act 1990). The United Kingdom's list of ethnicities is conceived of as harmonised standards rather than official categories. Only the Czech Republic, Ireland, Israel and Slovenia gather ethnicity data despite the absence of a specific legislative framework.

38. **Race** as an ascribed attribute and skin colour, on the other hand, is a much more sensitive topic and generally, legal frameworks limit collection of this type of information. Only Canada, the United Kingdom, the United States and Colombia have official data collection mandates and legal definitions alluding to it (Table 4.1). Chile, Mexico and Costa Rica collect data on their Afro-descendant populations, but officially classify them as ethnic rather than racial group. The Czech Republic, Hungary, Latvia, New Zealand, Poland, Bulgaria and the Russian Federation, which all gather information on ethnicity, either forbid data collection by race and skin colour or stipulate that responses are voluntary and should be treated as sensitive data. For example, the Human Rights Act in New Zealand forbids the discriminatory use of information on skin colour. Hence, questions on this topic are asked only in specific circumstances, e.g. when of significance for epidemiological studies. Several older European Union member states, due to the legacy of genocide and categorisation of certain ethnic groups as inferior during World War II, either completely forbid or strongly limit data collection on both race and ethnicity (Austria, Belgium, Germany, Luxembourg, Portugal, Slovenia, Sweden and Turkey).

Table 4.1. Frameworks regulating the collection of race-related data

	Official definition	Legal encouragement for collection
Canada	Visible minorities are "persons, other than Aboriginal peoples, who are non-Caucasian in race or non-white in colour" (Employment Equity Act, S.C. 1995, c.44)	✓
United Kingdom	Race includes (a) colour; (b) nationality; (c) ethnic or national origins (Equality Act, 2010)	✓
United States	Official categories for race are White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or Pacific Islander	✓
Colombia	Physical characteristics (CONPES 3310, 2004)	✓

Note: Canada: The Canadian Multiculturalism Act, Canadian Human Rights Act and the Employment Equity Act are statutes which, in most cases, do not refer explicitly to the use of Census or survey data from Statistics Canada, but for which these data are used extensively in the design, implementation and evaluation of policies and programs under the legislation. It should be noted that while the Canadian legislation and the definition of visible minority allude to race and skin colour, they do not require the use of these terms in data collection and dissemination, nor in the design, implementation and evaluation of policies and programs. United Kingdom: The official definition of race is not part of statistical legislation. Obligation for data collection is provided in secondary legislation created for each census (e.g. Census Order 2010). Racial and ethnic response categories are not official but harmonised standards of groups with greater user need. United States: The 1997 Office of Management and Budget Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity recognise that these categories represent a social-political construct designed for collecting data on the race and ethnicity of broad population groups in this country, and are not anthropologically or scientifically based. Colombia: CONPES 3310 mandates to include ethnicity and territoriality variables for Afro-Colombians in statistical data collection.

39. In addition to national legislation, the 2018 EU General Data Protection Regulation (GDPR), which affects all EU countries, includes specific provisions on rights to privacy and diversity data. Several European countries have indicated that collection of data on race and skin colour, ethnicity, indigenous identity and religious beliefs will have to comply with the specifications of the GDPR guidelines, and the general interpretation of the regulation seems to be leaning towards refusing to collect sensitive data at all. However, this might be a too narrow interpretation of data protection laws, and other European Commission Directives highlight the importance of data processing rather than gathering.[15] Article 9 of the GDPR lists several exceptions to the prohibition of processing personal data including explicit consent of an individual or reasons of substantial public interest (European Commission, 2018[37]). Thus, while **specific safeguards to protect the fundamental rights and the privacy of individuals (e.g. voluntary consent) and appropriate data protection and disclosure control measures are considered essential** (e.g. the collected data must be aggregated or anonymised), data collection per se is possible under GDPR if national law allows it.

40. In addition to legal constraints, some respondents may feel a sense of unease when asked to disclose their ethnic identity, often linked to lack of trust in data collectors and previous negative discrimination experiences.[16] Such fear can lead to certain minority groups

15. There is a widely held belief that the law prohibits any collection of sensitive data pertaining to disability and ethnic origin. But equality data can be collected in compliance with the exemptions enumerated in Article 8 of Directive 95/46/EC (Chopin, Farkas and Germaine, 2014[13]). In fact, Article 8 requires that member states "*prohibit the processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, [or] trade-union membership, and the processing of data concerning health or sex life*", subject to certain exceptions, including for scientific, health, and medical research (European Commission, 1995[94]; Simon, 2007[29]).

16. A French survey conducted on a representative sample of students and employees to test their reactions to different methods for recording ethnic origins found that 12% of respondents felt very

such as Travellers and the Roma, or indigenous communities more generally, being considerably underestimated in surveys and censuses (FRA, 2009[38]; Font and Méndez, 2013[39]). Similar difficulties may impact migrant communities and undocumented individuals who might face denunciation and deportation (European Observatory on Health Systems and Policies, 2011[40]). **Training of enumerators and the involvement of minority communities in all stages of the data collection process including questions of data ownership are critical to mitigate these problems** (Box 4.1).

Box 4.1. Community involvement in the collection and analysis of diversity data

Active participation of minority groups in the collection and analysis of diversity data should be used to build trust that such data is used to promote equality rather than for wrongful purposes. Such participation is especially important so that members of these groups clearly have an opportunity to understand the rationale behind data collection, and how the data will be used and stored.

The quality of diversity data can also benefit from community involvement, e.g. by guaranteeing the validity and relevance of ethnic categories, increasing response rates by mobilising community members, supplying survey enumerators from these groups, or helping in the analysis and dissemination of data (European Commission, 2017[12]; European Network against Racism, 2015[41]).

In practice, the extent of communication with and involvement of community groups is often limited (OECD, 2018[9]; European Commission, 2017[12]). Examples of more inclusive data collection practices by official and non-official data producers include:

- **Ireland**: A combined racial and ethnic identity question was introduced in the 2005 census after consultations with the National Consultative Committee on Racism and Inter-culturalism, the equality body, an NGO for Travellers (Pavee Point) and relevant government departments. In previous census rounds, only Travellers were asked to state their ethnic status. Similarly, Traveller organisations worked in partnership with academic researchers at University College Dublin in the design, collection and analysis of the 2010 All Ireland Traveller Health Study. Despite initial reservations and distrust among Travellers, an unprecedented 80% response rate was achieved (All Ireland Traveller Health Study Team, 2010[42]).
- **Romania**: Ahead of the 2011 national census, the NGO Romani CRISS launched the "I am Roma" campaign to encourage Roma to self-identify and improve accuracy of census data. Roma enumerators (teaching assistants, health mediators, local workers) were recommended to municipalities and explained statistical language and procedures (European Network against Racism, 2015[41]).
- **Colombia**: Ahead of the 2018 Census, the national statistical office engaged in extensive consulting with different ethnic and indigenous communities to revise

uncomfortable about placing themselves in an "ethno-racial" category, with such reticence being strongest amongst immigrants and their descendants. Moreover, employees were wary of the inclusion of ethnic information in the personnel files of employers' (Simon and Clément, 2006[81]). On the other hand, in a 2009 representative survey of European ethnic minorities and people with a migration background, over 65% of respondents stated they were in favour of providing, on an anonymous basis, personal information about their ethnic origin, as part of a census, if that could help to combat discrimination their country (FRA, 2009[38]).

the categories used to identify them, resulting in slightly updated question wording and response options that better reflects the needs of the communities asked (DANE, 2018[43]).

- **Australia**: In various (sometimes one-off) research projects, local indigenous organisations have participated in or assisted with data collection. For instance, a collaborative project on culturally relevant well-being indicators between the Yawuru community, the Australian National University and the Kimberly Institute used measures identified by Yawuru themselves using their markers of living well (BCEC, 2016[44]). Consultations with Aboriginal and Torres Strait Islander people were also used in the context of the 2004 Western Australian Aboriginal Child Health Survey and the 2018 government report on Aboriginal and Torres Strait Islander Stolen Generations and descendants (part of the Action Plan for Healing funded by the Department of the Prime Minister and Cabinet), see Australian Institute of Health and Welfare (2018[45]).
- **Canada**: Recognising the importance of maintaining strong relationships with First Nations, Métis and Inuit communities, Statistics Canada has developed an Aboriginal Liaison Advisor Program (ALAP). Aboriginal Liaison Advisors work to build partnerships with First Nations, Métis and Inuit communities and organisations based on respect and trust. The program was founded in the 1980s and today has evolved to be a national program with 11 advisors located across the country. The program's objective is to increase knowledge of, and access to, Statistics Canada data, products and services for First Nations people, Métis, and Inuit across Canada, and to promote participation in Statistics Canada surveys. Another objective of the ALAP is to help build the statistical capacity of people and organizations. This is being achieved by delivering statistical workshops and webinars on different survey topics. In addition to these activities, the Aboriginal Liaison Advisors engage with communities and organisations on surveys and other initiatives. The success of the program can be seen in an increased participation of First Nations communities across the country. In 1986, roughly 130 communities did not participate in the Census of Population; however through the efforts of the ALAP there were only 14 communities that did not participate in the most recent Census conducted in 2016.
- **New Zealand**: Māori communities increasingly insist that, beyond simple consultation on data collected *about* them, ownership of their own personal and communal data is recognised. In New Zealand, the Te Mana Raraunga/ Māori Data Sovereignty Network was formed to protect Māori rights to data, based on the principle that Māori data should be subject to Māori governance. Various research and policy projects are underway, including a pilot to create an Iwi rohe geographic variable in the NSO's Integrated Data Infrastructure. Issues of indigenous data sovereignty are now being discussed also in Australia and Canada (Te Mana Raraunga, 2018[46]).

4.2. Different policy needs, different data sources

41. The various uses of diversity data outlined in Section 2 require different approaches to data collection. The type of information collected, and whether it is used alone or in combination with other data sources, can impact how well the information captures the

complexity of the real world and the potential use of diversity data. While enumeration typically is best achieved via population censuses and population registers, assessing well-being outcomes and inequalities is best achieved through more focused sample surveys. In an ideal case, data should be collected consistently across sources so that these records can be linked and/or compared.

42. Relevant sources indicated by the NSOs who answered the questionnaire are population censuses, sample surveys, and population registers, even if additional data sources exist and may be used to gather information on a country's race and ethnic makeup (Box 4.3). Each of these approaches comes with its own sets of measurement strengths and limitations.

43. **Population censuses** are the primary source for population statistics against which all others national sources are compared. By providing individual enumeration and recording of each person living in a defined territory over a well-defined reference period, censuses provide insight into the size and the diversity of minority groups that can be analysed over time, especially for small areas and communities small in size. The census provides national, regional and small area demographic information, which is essential for planning the provision of public services, such as health care, education and employment. In a number of countries (e.g. Ireland), the census also determines the apportionment of seats in national parliaments.

44. Participation in the census is often mandatory, although it may be optional to respond to questions relating to sensitive information such as ethnic origin, religion and language.

45. Among the major weaknesses of census enumerations is their limited focus. A census has to cover many different topics and racial, ethnic or indigenous identity is only one of them. Thus, the number of diversity-related questions is often limited, as is information on well-being outcomes (e.g. quality of housing, exposure to toxins, education, economic resources, criminal justice indicators, access to services, discrimination experiences). Also, in the process of processing and tabulating census data, low priority is often given to diversity-related information, and such data often only becomes available several years after being collected. Further, certain groups (e.g. undocumented migrants) may avoid being counted or may be excluded from the census count due to difficulties in accessing remote areas (e.g. some nomadic populations or indigenous people).[17] This may have consequences for other statistics: since the census is usually used to design the master sample of national surveys, those who are missed from census counts will be under-sampled in surveys. Finally, owing to the great expense and amount of labour involved, most censuses are carried out every 5-10 years; as such, they do fail to capture abrupt population shocks in a timely manner (e.g. rapid changes in migration inflows or changes in self-identification of the population in question, such as when receiving previously unknown information about ancestry). Moving towards digitalised data collection systems, rather than face-to-face interviews, might also impact upon response rates in communities with limited access to internet.

17. For instance, the 2013 New Zealand Census undercounted the indigenous population especially in remote areas, and estimates of under-counts were higher for the Maori ethnicity group than for any other group in the Census (Statistics New Zealand, 2014[80]). Similar issues affected the 2018 Census. Nevertheless, it is a misconception that all indigenous people in OECD countries live in remote areas – for example, the majority of the Aboriginal and Torres Strait Islander population in Australia now live in urban regions (OECD, 2018[9]).

46. **Sample surveys** are particularly useful to cover topics for which no administrative or census data are available, for example religion, sense of identity or experiences of discrimination. Sample surveys include both multi-purposes surveys (e.g. household and social surveys), which allow the analysis of multiple disadvantages, and specialised-topic surveys (e.g. labour force, health interviews, skill assessments, etc.) The surveys indicated by countries in the OECD questionnaire vary considerably by sample size, raining from just 1 000 up to 30 000 respondents, with the majority having sample sizes of 5 000-8 000. Frequency also varies by type of survey, with labour force surveys being collected most frequently (generally quarterly), income and general social surveys being fielded more or less annually, and specialised surveys every 2-5 years. In general, countries with well-established statistical systems (e.g. Australia, Canada, Chile, Denmark, Finland, Hungary, Israel, Norway, Mexico, New Zealand, and the United Kingdom) collect information on well-being outcomes of different groups, although most often limited to disaggregation by country of birth, as often the sample size of minority groups may limit the analysis and dissemination of survey information.

47. While sample surveys may be used to collect reliable information on broad diversity categories, very large samples and accurate sampling design are required to capture the whole spectrum of a population's collective identities, as some minority groups (i.e. indigenous people) may not be randomly spatially distributed throughout the population, hard to reach, or not consciously included in the sampling frames used by national statistics (OECD, 2017[2]). Small numbers of survey respondents mean that breakdowns of minority groups by regional areas (e.g. reservations in the case of indigenous people) or even age groups and gender can become meaningless for statistical purposes.

48. Beyond sample surveys at the national level, a variety of international or trans-national surveys contains some basic information on migrant status, thus allowing cross-country comparisons. Sometimes, more detailed information on ethnic and migrant backgrounds is provided in ad hoc modules that are carried out less frequently or as one-off. For instance, the 2008 and 2014 ad hoc modules of the *EU Labour Force Survey* covered migrants and their immediate descendants and their degree of integration in the labour market. The modules contained information on the respondent's country of birth, nationality, years of residence in the country, and country of residence one year before the survey. Nevertheless, the level of detail available in these ad-hoc modules remains limited: for instance, the 2011 ad hoc module of the *EU Statistics on Income and Living Conditions (EU-SILC)* on the inter-generational transmission of disadvantages included only four answer categories for questions on country of birth and citizenship of parents: present country of residence / another EU-27 country / another European country / outside Europe.[18]

49. Several methodological solutions exist or are under development to address some of the main short-comings that affect the quality of survey data. For instance, pooling across multiple years can increase sample size. For some variables that change slowly, this can be acceptable. However, for other outcomes, such as unemployment or poverty rates, pooling may not reflect developments in the field. Additionally, **new or enhanced survey methods using non-standard sampling techniques** such as time-location sampling or respondent driven sampling have been developed to deal with non-random non-response rates (Peress, 2010[47]). Several countries also conduct special surveys with enhanced sampling frames or

18. More detailed statistical information about migrant status should become available in the 2021 ad hoc module of the *EU Labour Force Survey* on the labour market situation of migrants.

oversampling strategies that focus on specific challenges of indigenous communities, for example with regard to health outcomes (see Annex C).[19] Lastly, methodological work is ongoing on how to improve coverage of hard-to-reach and disadvantaged population groups in data collection (e.g. *UNECE Task Force on Disaggregated Poverty Measures*).

50. Both in censuses and sample surveys, individuals are generally allowed to themselves declare which group they feel they are part of. Indeed, all countries rely on self-reporting of ethnic, racial and indigenous status, with the exception of Israel, which draws partly on nationality data from its population registry, and Latvia, which has a register-based system but did not provide further information. (For an overview of other identification methods, see Box 4.2).

51. Indeed, the *UN Principles and Recommendations for Population and Housing Censuses* state that information on ethnicity should be acquired through self-declaration (United Nations, 2017[3]). Similarly, according to the ILO's Indigenous and Tribal Peoples Convention apply (ILO, 1989[4]), "*self-identification as indigenous […is] regarded as a fundamental criterion for determining the groups to which the provisions*" of. Although from a theoretical and ethical viewpoint self-identification can be considered to be the optimal method, in practice it carries some measurement problems relating to the flexible and contingent nature of collective identities (see the discussion in the next sub-section).

52. **In an ideal case, and regardless of the type of information on diversity collected, consistency should be ensured across multiple data sources, so that these can be linked.** For instance, in countries that gather information on indigenous identity, both censuses and sample surveys include relevant identifiers (Table A A.4).[20] Information on race is also generally collected via both the population census (except Chile and Mexico) and sample surveys (except Ireland) (Table A A.2). The situation is somewhat different for ethnicity, where 6 out of 17 countries only include a question in the census (Table A A.3). Going forward, it will be essential to gather diversity information both in the population censuses and in sample surveys and to link such data, which calls for ethnic and racial categories to be defined in a comparable format across multiple sources. A number of countries are currently making efforts along these lines by exploring the possibility of record-linking of the same respondent and integrating information from censuses and administrative records (Bycroft et al., 2016[48]). When data for the same individual entity are not available from different data sources, or when the identifying information allowing records to be matched is insufficient, multiple frame methods and statistical modelling techniques can be used (see Groves and Harris-Kojetin (2017[49]) for a review). While these methods are not error-free, they have the potential to increase the breath of the information available for policy analysis.

53. Another way to reconcile the need for joined-up statistics across multiple outcomes with the need for in-depth measures on specific topics (or for certain population groups) is to link data covering outcomes for several life dimensions at a very broad level to more detailed sources providing specialised information on each aspect (Fleischer, Smith and Viac, 2016[50]).

19. Annex C is available at: www.oecd.org/sdd/diversity-statistics-in-the-OECD-annex-C.pdf.

20. Annex A is available at the back of this report and at: www.oecd.org/sdd/diversity-statistics-in-the-OECD-annex-A.pdf.

Box 4.2. Identification methods beyond self-declaration

Beyond self-declaration, other methods for ethnic identification exist. While observers' classifications may not match the self-identification of respondents, both approaches bring information about the way that individuals experience ethnicity and race in their daily lives. Indeed, a growing body of work shows that race and ethnicity influence the life chances of individuals in multiple ways, related not only to how people self-identify but also to how others perceive them (Saperstein, Kizer and Penner, 2016[51]).

- *Identification by community members*: individuals are considered as part of a group if they are recognised as such by members of the same group. This method could be used where acceptance by the community matters for shaping someone's ethnic identity, for example in the case of the indigenous or Aborigine peoples. In practice, however, indigenous identity is mostly only self-reported.
- *Identification by a third party (other than community members) based on visual observation*: an individual may be classified as member of a particular group if he or she is perceived as such, on the basis of his or her physical appearance (e.g. skin colour), by an external observer who is carrying out the classification. This is often presented as being the most controversial method for collecting data on ethnic and racial backgrounds, as it relies on the assumptions, evaluations or estimates of another party, whether a private person such as a teacher or head teacher, or of a public agent. Often, there is a mismatch between self or next-of-kin identification and third party identification. For example, a comparison between next-of-kin racial identifications and death certificates has shown that a large proportion of Black Hispanics in the United States are mis-identified on death certificates. This leads to a significant overestimate of their life expectancy because the race-specific mortality rates are inaccurate (Swallen and Guend, 2003[52]). Previous research also found that concordance between self-reported and third-party identification ranged between 33 and 77%, with the extent of overlap being highest for white Americans and lowest for Asians (Boehmer et al., 2002[53]). However, this approach may be appropriate, under certain circumstances, when the goal is to collect information to expose and tackle discrimination, which depends on how victims of discrimination are perceived by others.
- *Reflected identity by the individual:* This type of identification asks respondents not only about their identity but also about which ethnicity others might attribute to them (Roth, 2016[54]). This is an important aspect of perceived discrimination, but only employed by Mexico in official statistics.
- *Identification by a third party based on available knowledge*. This applies mainly in the context of censuses and sample surveys where a family member provides information on behalf of a person who is not available at the time of the interview. Children, in particular if below a certain age, are often identified by the parent or other responsible adult, who usually select their own ethnic group. This may, however, introduce a measurement bias especially for children of immigrants. For example, while immigrants in the United Kingdom tend to prefer national-origin identities (e.g. Indian), their children, when asked to identify themselves, tend to adopt hyphenated (e.g., Indian-British) or purely national (e.g. British) identities (Schimmele and Wu, 2016[55]).

54. **Population registers** exist in many OECD countries, although they are less documented than other standard sources of demographic information.[21] They are always run on a municipal basis but computerisation now makes it possible to centralise the data and establish national registers, thus enabling to regularly monitor the individual demographic trajectories of the entire *de jure* population at national level. Population registers record the movement into and out of a country, as well as births, deaths, internal movements and other information on residents. As such, they can provide timely information on international migration affecting a country; sometimes they also contain information on country of birth of parents, which is used as (imperfect) proxy for ethnic background. Indeed, in countries that solely rely on population registers, diversity data are usually limited to information on migrant status/country of birth (Table A A.1).[22] Unless linked with survey data, these data do not include any information on racial and ethnic identities, which require self-declaration.

55. Continuous updating is the key characteristic of a population register and its major advantage for producing demographic statistics, even if they are usually published annually. Since cancellation is less common than registration, mainly due to the fear of losing advantages linked to being registered in one's home country, population registers better capture immigrants than emigrants (Poulain and Herm, 2013[56]). In population registers, individuals are usually classified into pre-defined categories on the basis of indirect indicators, such as their country of birth, the nationality of their parents, or the language spoken at home, that are considered as "objective", in the sense that they are not based on feelings of affiliation or perceptions by others, but on factual information that can be assessed by a third party.

21. Austria, Denmark, Finland, Iceland, Israel, Latvia, Lithuania, the Netherlands, Norway, Slovenia and Sweden all use population registries.

22. Annex A is available at the back of this report and at: www.oecd.org/sdd/diversity-statistics-in-the-OECD-annex-A.pdf.

Box 4.3. Additional sources of information on diversity

Other statistical instruments that have not systematically scrutinised for the purposes of this review include:

- **Administrative sources,** which in some countries (e.g. the United Kingdom, New Zealand) may contain information on diversity. These are based on a wide array of recording systems, which are often linked in a systematic and secure way to produce an integrated data infrastructure. Although information generated by administrative records can be timelier than that generated by censuses, the underlying data need to be harmonised before being used in combination with other sources: ethnic statistics based on administrative sources are usually calculated using 'events' or numerators from one file, and denominators from the census. If the recording of ethnicity varies between the numerator and denominator, the resulting measures are affected by a numerator-denominator bias. Furthermore, the data result from administrative processes, and only a small part of the information recorded is processed for statistical purposes. Finally, administrative collections are generally understood by respondents to have a specific purpose, which may influence how people respond to ethnicity-related questions (Statistics New Zealand, 2005[57]).
- **Non-official international surveys**: Many international surveys include basic information on migrant background (e.g. country of birth) but almost never on race and ethnicity. For instance, the *OECD Programme for the International Assessment of Adult Competencies (PIAAC)* and the *OECD Programme for International Student Assessment (PISA)* have been extensively used to document the well-being of migrants and their offspring (OECD, 2017[2]; 2018[58]). Other international surveys, such as the European Social Survey or the World Value Survey, although containing proxies for ethnicity and migrant status, are characterised by small sample sizes, which may be insufficient to support analysis of outcomes for minorities.
- **Surveys on specific minorities** are conducted by a variety of stakeholders, including equality bodies, researchers, academics, demographers and NGOs. An example of an international survey targeting minority groups is the *European Union Minorities and Discrimination Survey (EU-MIDIS)*, conducted by European Union Agency for Fundamental Rights (FRA) in 2008 and 2016 and targeting selected ethnic minority and immigrant persons resident in the EU Member States (FRA, 2017[15]). The survey focuses on experiences of discriminatory treatment, racism, victimisation, awareness of rights, and reporting of complaints. Surveys on specific minority groups are often conducted also at national level. While these instruments have the potential of providing valuable and detailed information on a wide range of topics of great importance for minority groups and that are not often covered in other data sources (e.g. questions on perception of racial or ethnic origin by others, or on experiences of discrimination), they usually lack suitable sample frames for the target population, thus preventing direct comparison with official statistical sources.

4.3. Identities are dynamic and multiple

56. The construction of categories and the principles in accordance with which individuals can be categorised into different ethnic groups are among the most difficult issues involved in collecting diversity data. This section details which categories countries currently use for classification, including respondent options for selecting an identity (e.g. pre-coded vs open-ended responses, and the selection of multiple identities). Issues of comparability between sources and over time, as well as 'category jumping', i.e. people changing their ethnicity identifier, due to ethnic mobility and context effects, are also addressed.

57. **Statistical categories should reflect demographic changes as well as evolutions in the understanding of racial and ethnic identities, while remaining grounded in sufficient stability to allow comparability.** It is crucial that classification categories are validated by user groups and communities themselves to guarantee they reflect self-conceptions as much as possible and reduce instances of misreporting[23] (Brown, Hitlin and Elder, 2007[59]; Saperstein, 2006[60]). Basic criteria used to measure identities should be clearly explained to respondents and consistently used when disseminating the resulting data. Categories might also need to continually reviewed and updated to reflect changing societal trends such as immigration and inter-racial marriage: For instance, in the United States it is estimated that about 80% of African Americans have some White ancestry; and 50% of Mexican Americans have European or American Indian and Alaska Native ancestry. Over time, there have been a growing number of US respondents people who do not identify with any of the official race categories and whom have been racially classified as "Some Other Race", which was initially intended to be a small residual category (US Census Bureau, 2018[61]).

58. Countries that participated in this review currently use very different diversity categories, highlighting the fact that what might be called 'race' in one country might be seen as 'ethnicity' in another, and that boundaries are, to some extent, arbitrary (Morning, 2008[62]).

59. Existing **indigenous statistics** focus on a number of criteria, ranging from assessing indigenous status (self-identification of being indigenous) to tribal identification (which tribe/group a person belongs to) and language. Determination of indigenous status is frequently carried out in several steps, with an initial binary question asking about whether a person identifies as indigenous, and follow-up questions asking for the respondent's specific group/tribe and/or indigenous dialects and languages spoken (Table A A.4).[24]

60. Most Eastern Europe and Baltic countries share a view of **ethnicity** as relating to nationality and ancestry (rather than current citizenship) (Table A A.3). This is reflected either in the question wording (e.g. "*Which nationality do you feel you belong to?*" in Hungary; or *"What is your nationality? understood as the national or ethnical affiliation – do not confuse with citizenship"* in Poland) or in the response options provided (e.g. Lithuania asks about ethnicity but then lists Lithuanian, Polish, Russian as pre-coded categories). Poland and the Slovak Republic include religion-related groups such as

23. Misreporting occurs when respondents do not understand the intent or wording of the question, which may affect the reliability of the answers, especially for relatively small populations or geographic areas.

24. Annex A is available at the back of this report and at: www.oecd.org/sdd/diversity-statistics-in-the-OECD-annex-A.pdf.

Jewish/Yiddish as ethnic categories alongside nationalities, and the Israeli measure of ethnicity is a mix of information on country of birth from the population register and a self-reported question on religion. The Roma, as distinct ethnic group, are only explicitly listed among the response options in Hungary, Poland, the Slovak Republic and Bulgaria. Romania allows for self-identification of its considerable Roma community via an open-ended census question. Ireland is currently testing the addition of a Roma category in its Pilot Census 2018.

61. The majority of countries that collect data on both **race and ethnicity** use separate questions, although Ireland, the United Kingdom, Colombia and Costa Rica mix the two (Table A A.2).[25] Data collection in the United States relies on separate questions on race and ethnicity (with OMB standards directing that the latter should be asked first), conceptualising ethnicity solely as Hispanic origin (Office of Management and Budget, 1997[63]).[26] On the other hand, the United Kingdom and Ireland's questions refer to ethnicity, but the response categories include both ethnic groups and "racial" features such as skin colour (e.g. White, Black, mixed/multiple ethnic groups). Chile and Costa Rica consider their Afro-descendent populations as falling under ethnic rather than racial classifications.[27] However, since both refer to physical characteristics in the question and/or response options, for the purpose of this review they are considered under racial data collection. Some sort of conceptual overlap or even contradiction is not avoidable with mixed categories: Groups such as Irish, Italian, Russian, Jewish, and Serbian might all be included in the racial category "White". Conversely, the ethnic group "British" includes citizens from a multiplicity of racial backgrounds: Black, White, Asian, and more, plus a variety of race combinations.

62. Collective identities are socially constructed and situational rather than static. People may change how they identify themselves over time or they may identify themselves differently in different environments, which can be important for the interpretation of data and the dynamics of ethnicity. Category jumping, can reflect either *context effects* influencing responses or *ethnic mobility*.

63. Context effects refer to how (the mode) and where/why (the circumstances) the information is collected. A common cause of changes in identification relates to the perceived purpose of the data, e.g. a person's responses may differ where they understand the data in one collection relates to familial information and in another collection to social

25. Annex A is available at the back of this report and at: www.oecd.org/sdd/diversity-statistics-in-the-OECD-annex-A.pdf.

26. The identity category "Hispanic" was historically constructed and institutionalised in the United States by activists, officials, and media executives in the 1970s and 1980s. In the 1960 census, distinct groups of Latin American immigrants (e.g. Puerto Ricans, Mexicans, and Cubans) were classified as "white", and Latino activists claimed that this classification hindered their ability to portray their constituents as underrepresented minorities. A separate category "Hispanic" was first introduced in the 1980 census round. This led to other cultural changes, such as Spanish-language television expanding its reach to serve the now large, and newly unified, Hispanic community (Mora, 2014[100]).This shows that statistics do not only mirror the real world, but also influence how identities themselves are shaped in society.

27. For instance, although Afro-descendant groups claim phenotypic fasgos as part of their identity, the term "racial" is used by the National Institute of Statistics and Census of Costa Rica. This term is understood as a social construction of the "Afro-descendant" ethnic group that claims these aspects and not the term race as a biological category.

environments. This does not necessarily mean that people reply inaccurately, rather, they may be providing ethnicity responses that best reflect how they identify themselves relative to what they understand to be the purpose of the information. Since many collections are not primarily statistical in function, respondents' perceptions of the purpose of ethnicity questions may influence how they identify themselves, which can be different when completing, for instance, educational enrolments, benefit applications and census forms (Statistics New Zealand, 2005[64]).

64. Moreover, bias via context effects may be introduced through the mode of administration[28]: e.g. in interviewer-administered surveys actions on the part of the interviewer that deviates from survey protocols, either intentional or unintentional (i.e. *interviewer effect*), may cause a bias, especially for sensitive questions such as those on race and ethnicity. Interviewers may affect the responses that they obtain, not through any overt behaviour, but merely as a function of their observable characteristics (Brunton-Smith, Sturgis and Leckie, 2017[65]). For instance, African Americans who are interviewed by Whites are likely to give substantively different responses on some questions than their counterparts who are interviewed by another African American. The same is true of white respondents and black or white interviewers (*race-of-interviewer effect*). Similar effects were found with interviewer ethnicity (Weeks and Moore, 1981[66]; Reese et al., 1986[67]; Hurtado, 1994[68]).

65. Ethnic mobility as refers to people changing how they identify their ethnicity over time. For example, the social environment of people may change in ways that lead them to identify themselves with additional or different ethnicities. This may reflect both changes in social environment or living arrangements, such as partnership formation, change of job, moving to a different area, receiving new information (e.g. from DNA ancestry testing) or simply identifying oneself different at work and at home (Statistics New Zealand, 2005[64]).

66. Ethnic mobility can be quite significant and demographers have recently become very interested in this phenomena. For example, between the 2000 and 2010 Census in the United States, over 10 million people have been found to have changed their ethnic/racial category – a pattern that had not drawn attention before as diversity aggregates were stable (Liebler et al., 2017[69]).[29] Measures of reflected race, or one's perceptions of which ethnicity others believe one to be, have also been found to fluctuate over time (Saperstein, Kizer and Penner, 2016[51]). Ethnic mobility patterns will have to be monitored from a policy point of view, as they makes targeting of policies more difficult.

67. Especially for indigenous identification, ethnic mobility has emerged as a somewhat problematic issue. Australia, Canada, New Zealand and the United States experienced a large increase in the number of people self-reporting an indigenous ancestry in recent census waves that cannot be explained by population growth alone. On the one hand, it might be a sign of people feeling more comfortable to self-identity as indigenous. On the other hand, indigenous communities have expressed concern about this phenomenon,

28. The selection of the method has an impact on response rates and the reliability of the responses in general. On average, interviewer-administered surveys tend to yield higher response rates than self-administered surveys.

29. Similarly, in a longitudinal study in New Zealand, 8% of respondents changed ethnicity at least once during the three waves of the survey, with the strongest predictors of changing being having reported Pacific and Asian ethnicity at wave 1, as well as reporting more than one ethnic group (Carter et al., 2009[96]).

as from their perspective belonging to their community should reflect recognition by the tribes and clans, rather than solely individual self-reporting.[30] Statistics Canada, after extensive consultation with indigenous groups, is currently experimenting with changes to their indigenous identity question for their Census, with one variant including a follow-up question asking to which recognised indigenous organisation respondents actually belongs, using pre-coded and an "other" options. Other solutions suggested by research include combining questions on self-identified indigenous affiliation with those on descent (Kukutai, 2004[70]).

68. As general best practice principles for diversity data collection, the *UN Principles and Recommendations for Population and Housing Censuses* recommend that: *i)* free self-declarations / open questions should be used; *ii)* respondents should be able to indicate more than one ethnic affiliation; *iii)* categories for 'none' or 'not declared' should be allowed; *iv)* instructions should be provided on determining the ethnicity of children of mixed couples; and *v)* the basic criteria and classification procedures should be documented (United Nations, 2017[3]).

69. The response to a race and/or ethnicity question is obligatory rather than voluntary in most countries, even though regulations might differ between instruments (e.g. Canada allows for voluntary responses in its General Social Survey but not in the Census; see Annex A).[31] In reality, however, responses are generally not enforced, and not answering a question is very common.

70. In practice, the Czech Republic, Romania and the Russian Federation are the only countries that do not rely on pre-coded ethnicity categories (rather than open-ended questions as recommended by the UN). The Canadian ethnicity question is open-ended, but has very concrete examples of response categories in the explanatory text (Annex A).[32] An open-ended approach presupposes that the question itself has to be formulated in terms that are unequivocal for respondents.

71. For the majority of countries that use pre-defined response options instead, the number of pre-coded categories ranges from binary yes/no options (e.g. for Afro-descendants in Chile) to more detailed categories, e.g. up to 17 ethno-racial groups under 5 umbrella categories in the United Kingdom (Annex A). The *UN Principles and Recommendations for Population and Housing Censuses* warn that the pre-coding or the

30. Tribal membership for American Indians and Alaska Natives, for instance, is based on enrolment criteria set by individual tribes, often by degree of blood or blood quantum (Connolly, 2018[95]).

31. Annex A is available at the back of this report and at: www.oecd.org/sdd/diversity-statistics-in-the-OECD-annex-A.pdf.

32. The 2016 Canadian question and listing options read: "*What were the ethnic or cultural origins of this person's ancestors? An ancestor is usually more distant than a grandparent. For example, Canadian, English, Chinese, French, East Indian, Italian, German, Scottish, Cree, Mi'kmaq, Salish, Metis, Inuit, Filipino, Irish, Dutch, Ukrainian, Polish, Portuguese, Vietnamese, Korean, Jamaican, Greek, Iranian, Lebanese, Mexican, Somali, Colombian, etc.*". Changes in the list of examples provided may influence ethnicity counting. For instance, in the 2016 Canadian Census the list of examples of ethnic origins was updated to reflect the frequency of single responses reported in the 2011 National Household Survey. For 2016, "Iranian" and "Mexican" were added to the list of examples, while "Jewish" and "Salvadorean" were removed. Following these changes, the number of Canadians who reported their ancestors as Jewish in 2016 was about half of what was measured in 2011.

pre-classification of ethnic groups "*may have a tendency to lose detailed information on the diversity of a population*" (United Nations, 2017[3]). However, the use of a short list of categories is not generally intended to exclude groups, but rather to prevent the data collection process from becoming too complex and ensuring statistical significance for smaller groups. One way to mitigate this issue is to encourage self-disclosure by a blank column or an "other" option, allowing data subjects who do not "fit" into the listed ethnic categories to indicate different groups which are not on the list (Hasnain-Wynia and Baker, 2006[71]).

72. In the past, questions on race and ethnicity backgrounds have been based on asking people to tick one box only. More recently, however, a number of countries have moved towards multiple-response questions in a "mark all that apply" approach. 3 out 8 countries (Canada, the United Kingdom and the United States) enable individuals to report more than one race. Twelve out of 17 countries allow multiple responses for ethnicity, bar Israel, Latvia, the Slovak Republic and Slovenia. The vast majority of countries use a mix of pre-coded response options with an "other" category that usually allows respondents to specify the race/ethnic affiliation in a write-in box. Only Chile, Mexico, Colombia and Costa Rica do not make use of an open-ended option when collecting data on race (Annex A).[33]

A "mark all that apply" approach offers the opportunity to describe highly diversified societies more accurately and with more nuance, but also opens up new challenges about how to use, output and interpret the multitude of race and ethnicity categories (Perlmann and Waters, 2002[72]; Snipp, 2003[73]). Country practices and recommendations in this regard vary: The US Office of Management Budget recommends that an additional category – two or more races – should be reported with the results. This category should include only those who reported two or more races. A person who is counted in the "two or more races" category should be excluded from other race categories (Office of Management and Budget, 1997[63]). New Zealand recommends 'total responses' and 'single/combination outputs' as two approaches to ethnicity data output (Statistics New Zealand, 2005[64]). The (recommended) total response output method counts every ethnic group that a person identifies with, with people identifying themselves as belonging to two groups counted twice, implying that the sum of members of all ethnic groups will be greater than the total number of people. The advantage of this approach is that the relative size of the groups within the population is fairly represented (using as the denominator only the count of people for whom ethnicity is available). Conversely, the single/combination data provides useful information on the components of ethnic groups, but is more sensitive to ethnic mobility and contextual effects.

73. Given that virtually all countries have implemented changes in how diversity data is collected over time (e.g. changing question wording, updating response categories, or moving towards multiple response), comparability when mapping data across different sources or over time is a significant challenge (OECD, 2017[25]; Barnes et al., 2008[74]). Responses might also be processed differently, or different concepts be used between sources (i.e. *numerator-denominator bias* with e.g. administrative data as the former and survey data as the latter). **When data is compared across different collections, consideration needs to be given to how and when the data was collected** (especially in cases where the sources allowed different responses or where the data has been tabulated in

33. Annex A is available at the back of this report and at: www.oecd.org/sdd/diversity-statistics-in-the-OECD-annex-A.pdf.

different ways); also, **assumptions about uncertainties in the resulting data need to be made explicit** (Statistics New Zealand, 2005[64]).

74. **Bridging methods** that assign multiple-race responders to single-race categories can offer a solution to reconcile data collected under different standards by assigning some proportion of those reporting a specific combination of races to each of the races defining the group (Liebler and Halpern-Manners, 2008[75]; Perlmann and Waters, 2002[72]). The assignment is done so as to approximate in the aggregate how the individuals in this group would have responded had they been asked to report only a single-race. Allocating equal proportions of each multiple-race group to its component single races has often been proposed (Parker et al., 2004[76]). Going forward, countries should aim to develop **diversity statistical standards** and guidance to improve consistency and comparability across all data sources. The Ethnicity Standard in New Zealand or the harmonised groups in the United Kingdom represent good practice examples here.

5. Discussion and the statistical agenda ahead

75. Even though measurement approaches and regulations that underpin the collection of diversity data differ significantly across OECD countries, the need to collect more and better data to assess the well-being of minority population groups and to design more effective policies has been increasingly recognised. International agreements such as the 2015 Sustainable Development Goals emphasise that disaggregation by race, ethnicity and indigenous identity is essential to ensure that no one is left behind.

76. Current NSOs collection practices cluster around three broad categories: 1) all OECD countries collect information on some diversity proxies such as country of birth (36 OECD members); 2) a small majority, mostly Eastern European countries, the United Kingdom and Ireland, gather additional information on race and ethnicity (16 OECD members); and 3) only a handful of countries in the Americas and Oceania collect data on indigenous identity (6 OECD members). Diversity statistics are collected from the perspective of either enumerating the size of the relevant populations (typically in the census) or of comparing well-being outcomes across different population groups.

77. While privacy and human rights legislation sometimes prevents or discourages the routine collection of diversity data, the need to improve data availability and quality is being recognised in most countries. Many countries are piloting the addition of new ethnic response options to more accurately reflect the make-up of their societies (e.g. Ireland, the United States), while Belgium is considering allowing collection of race and ethnicity data within the restrictions imposed by the national legal framework. Within the European Statistical System, the inclusion of more detailed migration information is also being considered: The *Framework Regulation for Production of European Statistics on Persons and Households European* foresees the incorporation of questions on the country of birth of the respondent's parents in the *Labour Force Surveys* (from 2020), the *European Health Interview Survey*, the *European Union Statistics on Income and Living Conditions*, the *Household Budget Survey*s and the *Community surveys on ICT usage in households and by individuals*. The European Union Agency for Fundamental Rights is pursuing its *Roma and Travellers Survey* to collect comparable data in six selected Member States in 2018 (FRA, 2018[77]).

78. Overall, this review shows that collecting and analysing data on diversity is difficult but not impossible. Commonly identified challenges include: 1) the treatment of "sensitive" data and concerns around privacy legislation, as well as reluctance of some groups to disclose their identity; 2) the use of different sources for different policy purposes and the need to validate and link the collected data; and 3) issues of comparability over time relating to the flexible and contingent nature of collective identities.

79. While there is no one-size-fits all approach, for those countries where national law allows the gathering of diversity data, steps to improve diversity data collection include:

- Expand data collection to ethnicity/race/indigenous identity variables, while respecting the fundamental rights and privacy of individuals by ensuring appropriate data protection and disclosure control measures.
- Involve relevant communities into the process of survey development (including the wording of question and response categories), validation of the accuracy of self-

reported information, data collection efforts, and the dissemination of results. This will build trust and increase data quality.

- Ensure the representation of hard-to-reach populations, such as indigenous communities and the Roma population, through non-standard sampling techniques such as time-location sampling or respondent driven sampling, and include these communities among pre-coded response options where applicable.
- Gather information on diversity in both population censuses and sample surveys in order to provide robust demographic statistics and timely data that allows for the assessment of multiple well-being outcomes and discriminatory experiences. Where possible, link census, sample survey data and administrative records pertaining to these populations.
- When data is compared across two or more different collections, consideration needs to be given to how and when the data was collected. Also, assumptions about uncertainties in the resulting data need to be made explicit. NSOs should invest in developing diversity statistical standards and provide clear guidance to improve consistency and comparability across all data sources (censuses, sample surveys, administrative data).
- Allow respondents to declare more than one identity to better mirror the increasingly diverse make-up of our societies. Statistical categories should reflect demographic changes as well as evolutions in the understanding of racial and ethnic identities.
- Finally, diversity data should be used for policy purposes. While collection is the first step, it is important to ensure that data are used to inform strategies by government, businesses, and other actors in society, including minority communities themselves.

References

Alcoff, L. (2009), "Latinos beyond the Binary", *The Southern Journal of Philosophy*, Vol. 47/S1, pp. 112-128, http://dx.doi.org/10.1111/j.2041-6962.2009.tb00143.x. [99]

All Ireland Traveller Health Study Team (2010), *All Ireland Traveller Health Study: Summary of Findings*, School of Public Health, Physiotherapy and Population Science, University College Dublin, https://www.ucd.ie/t4cms/AITHS_SUMMARY.pdf. [42]

Aspinall, P. (2012), "Answer Formats in British Census and Survey Ethnicity Questions: Does Open Response Better Capture 'Superdiversity'?", *Sociology*, Vol. 46/2, pp. 354-364, http://dx.doi.org/10.1177/0038038511419195. [17]

Australian Government (2018), *Closing the Gap: Prime Minister's Report 2018*, Department of the Prime Minister and Cabinet, https://closingthegap.pmc.gov.au/sites/default/files/ctg-report-2018.pdf?a=1 (accessed on 25 September 2018). [20]

Australian Institute of Health and Welfare (2018), *Aboriginal and Torres Strait Islander Stolen Generations and descendants*, https://www.aihw.gov.au/getmedia/a6c077c3-e1af-40de-847f-e8a3e3456c44/aihw-ihw-195.pdf.aspx?inline=true. [45]

Axelsson, P. (2018), "Official statistics and Indigenous People: the state of play and recent developments", *Paper prepared for the 16th Conference of IAOS OECD Headquarters, Paris, France, 19-21 September 2018*, http://www.oecd.org/iaos2018/programme/IAOS-OECD2018_Axelsson.pdf (accessed on 25 September 2018). [27]

Barnes, T. et al. (2008), *A Comparative Analysis of Indirect Methodologies for Estimating Indigenous Life Expectancy*, Charles Darwin University. [74]

Bartlett, J. et al. (2007), "Identifying indigenous peoples for health research in a global context: a review of perspectives and challenges.", *International journal of circumpolar health*, Vol. 66/4, pp. 287-307, http://www.ncbi.nlm.nih.gov/pubmed/18018843. [98]

BCEC (2016), *Developing Culturally Relevant Indicators of Wellbeing: A Yawuru Example*, http://bcec.edu.au/assets/bcec-community-wellbeing-from-the-ground-up-a-yawuru-example.pdf. [44]

Boehmer, U. et al. (2002), "Self-reported vs administrative race/ethnicity data and study results.", *American journal of public health*, Vol. 92/9, pp. 1471-2, http://www.ncbi.nlm.nih.gov/pubmed/12197976. [53]

Brown, J., S. Hitlin and G. Elder (2007), "The importance of being "other": A natural experiment about lived race over time", *Social Science Research*, Vol. 36/1, pp. 159-174, http://dx.doi.org/10.1016/J.SSRESEARCH.2005.11.002. [59]

Brunton-Smith, I., P. Sturgis and G. Leckie (2017), "Detecting and understanding interviewer effects on survey data by using a cross-classified mixed effects location-scale model", *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, Vol. 180/2, pp. 551-568, http://dx.doi.org/10.1111/rssa.12205. [65]

Burhansstipanov, L. and D. Satter (2000), "Office of Management and Budget racial categories and implications for American Indians and Alaska Natives.", *American journal of public health*, Vol. 90/11, pp. 1720-3, http://www.ncbi.nlm.nih.gov/pubmed/11076238. [97]

Bycroft, C. et al. (2016), *Identifying Māori populations using administrative data: A comparison with the census*. [48]

Cabinet Office (2017), *Race Disparity Audit: Summary Findings from the Ethnicity Facts and Figures website*, https://www.ethnicity-facts-figures.service.gov.uk/static/race-disparity-audit-summary-findings.pdf (accessed on 25 September 2018). [18]

Carter, K. et al. (2009), "How Much and for whom does Self-Identified Ethnicity Change Over Time in New Zealand?Results from a Longitudinal Study", *Social Policy Journal of New Zealand* 36, pp. 32-45. [96]

CEIES (2007), *Final Statement of the 33rd CEIES Seminar Ethnic and Racial Discrimination on the Labour Market: Measurement, statistics and indicators*, Committee on Statistical Information in the Economic and Social Spheres, Valletta, https://ec.europa.eu/eurostat/web/conferences/conf-2007/33rd-ceies-seminar (accessed on 24 September 2018). [24]

Chopin, I., L. Farkas and C. Germaine (2014), *Ethnic Origin and Disability Data Collection in Europe: Measuring Inequality – Combating Discrimination*, Open Society, Brussels, https://www.opensocietyfoundations.org/sites/default/files/ethnic-origin-and-disability-data-collection-europe-20141126.pdf (accessed on 24 September 2018). [13]

Connolly, M. (2018), "Improving Lives of Indigenous People through Better Statistics: Meeting Policy and Program Needs", *Paper prepared for the 16th Conference of IAOS OECD Headquarters, Paris, France, 19-21 September 2018*, http://www.oecd.org/iaos2018/programme/IAOS-OECD2018_Connolly.pdf (accessed on 25 September 2018). [95]

DANE (2018), *Claves para entender el enfoque diferencial étnico del Censo 2018*, https://censo2018.dane.gov.co/claves-para-entender-el-enfoque-diferencial-etnico-del-censo-2018. [43]

DGME (2014), *Política Nacional para una Sociedad Libre de Racismo, Discriminación Racial y Xenofobia 2014-2025*, Dirección General de Migración y Extranjería, http://poblacion-indigena.iniciativa2025alc.org/politica-nacional-para-una-sociedad-libre-de-racismo-discriminacion-racial-y-xenofobia-2014-2025/ (accessed on 01 November 2018). [19]

Durante, F., C. Volpato and S. Fiske (2010), "Using the Stereotype Content Model to examine group depictions in Fascism: An Archival Approach", *European journal of social psychology*, Vol. 40/3, http://dx.doi.org/10.1002/ejsp.637. [33]

ECRI (1998), *General Policy Recommendation no. 4 on National surveys on the experience and perception of discrimination and racism from the point of view of potential victims*, European Commission against Racism and Intolerance, Council of Europe, Strasbourg. [23]

ECRI (1996), *General Policy Recommendation n. 1 on Combating Racism, Xenophobia, Antisemitism and Intolerance*, European Commission against Racism and Intolerance, Council of Europe, Strasbourg. [22]

European Commission (2018), *EUR-Lex - 32016R0679 - EN - EUR-Lex*, http://data.europa.eu/eli/reg/2016/679/oj (accessed on 24 September 2018). [37]

European Commission (2017), *Data collection in the field of ethnicity: Analysis and comparative review of equality data collection practices in the European Union*, http://ec.europa.eu/newsroom/just/document.cfm?action=display&doc_id=45791. [12]

European Commission (2013), *Using EU indicators of immigrant integration - EU Law and Publications*, Directorate-General for Migration and Home Affairs, European Commission, Brussels, https://publications.europa.eu/en/publication-detail/-/publication/0e4f1f56-14ad-446d-86e8-a360c40fc266 (accessed on 24 September 2018). [6]

European Commission (1995), *EU Directive 95-46-EC - Chapter 2 - Data Protection Commission - Ireland*, https://www.dataprotection.ie/docs/EU-Directive-95-46-EC-Chapter-2/93.htm (accessed on 24 September 2018). [94]

European Network against Racism (2015), *Equality data collection: Fact and principles*, http://enar-eu.org/IMG/pdf/edc-general_factsheet_final.pdf. [41]

European Observatory on Health Systems and Policies (2011), *Migration and health in the European Union*, Open University Press, New York, http://www.euro.who.int/__data/assets/pdf_file/0019/161560/e96458.pdf. [40]

Fleischer, L., C. Smith and C. Viac (2016), "A Review of General Social Surveys", *OECD Statistics Working Papers*, No. 2016/9, OECD Publishing, Paris, http://dx.doi.org/10.1787/bb54d16f-en. [50]

Font, J. and M. Méndez (eds.) (2013), *Surveying ethnic minorities and immigrant populations: methodological challenges and research strategies*, Amsterdam University Press. [39]

FRA (2018), *A persisting concern: anti-Gypsyism as a barrier to Roma inclusion*, European Union Agency for Fundamental Rights , Vienna, http://fra.europa.eu/en/publication/2018/roma-inclusion. [14]

FRA (2018), *Roma and Travellers Survey 2018*, European Union Agency for Fundamental Rights, http://fra.europa.eu/en/project/2018/roma-and-travellers-survey-2018 (accessed on 26 September 2018). [77]

FRA (2017), *Second European Union Minorities and Discrimination Survey - Main results*, European Union Agency for Fundamental Rights, Vienna, http://fra.europa.eu/en/publication/2017/eumidis-ii-main-results. [15]

FRA (2009), *EU-MIDIS Technical Report: Methodology, Sampling and Fieldwork*, European Union Agency for Fundamental Rights, Vienna, http://fra.europa.eu/eu-midis (accessed on 5 November 2018). [38]

Fujimura, J., T. Duster and R. Rajagopalan (2008), "Introduction: Race, Genetics, and Disease", *Social Studies of Science*, Vol. 38/5, pp. 643-656, http://dx.doi.org/10.1177/0306312708091926. [93]

Gill, P. et al. (2005), "Limitations and potential of country of birth as proxy for ethnic group.", *BMJ (Clinical research ed.)*, Vol. 330/7484, p. 196, http://dx.doi.org/10.1136/bmj.330.7484.196-a. [26]

Griffiths, K. (2018), "Statistics, rights and recognition: the identification of Indigenous peoples", *Paper prepared for the 16th Conference of IAOS OECD Headquarters, Paris, France, 19-21 September 2018*, http://www.oecd.org/iaos2018/programme/IAOS-OECD2018_Griffiths.pdf (accessed on 25 September 2018). [92]

Groves, R. and B. Harris-Kojetin (eds.) (2017), *Federal Statistics, Multiple Data Sources, and Privacy Protection*, National Academies Press, Washington, D.C., http://dx.doi.org/10.17226/24893. [49]

Hall, G. and H. Patrinos (2012), *Indigenous Peoples, Poverty, and Development*, Cambridge University Press. [28]

Hasnain-Wynia, R. and D. Baker (2006), "Obtaining Data on Patient Race, Ethnicity, and Primary Language in Health Care Organizations: Current Challenges and Proposed Solutions", *Health Services Research*, http://dx.doi.org/10.1111/j.1475-6773.2006.00552.x. [71]

Hurtado, A. (1994), "Does Similarity Breed Respect: Interviewer Evaluations of Mexican-Descent Respondents in a Bilingual Survey", *Public Opinion Quarterly*, Vol. 58/1, p. 77, http://dx.doi.org/10.1086/269409. [68]

ILO (1989), *Convention on indigenous and tribal peoples (No. 169)*, International Labour Organization, Geneva. [4]

Krieger, N. (2000), "Refiguring "Race": Epidemiology, Racialized Biology, and Biological Expressions of Race Relations", *International Journal of Health Services*, Vol. 30/1, pp. 211-216, http://dx.doi.org/10.2190/672J-1PPF-K6QT-9N7U. [91]

Kukutai, T. (2004), "The Problem of Defining an Ethnic Group for Public Policy: Who is Maori and Why Does it Matter? - Ministry of Social Development", *Social Policy Journal of New Zealand* 23, pp. 83-106. [70]

Kukutai, T. and R. Didham (2009), "In Search of Ethnic New Zealanders: National Naming in the 2006 Census", *Social Policy Journal of New Zealand* 36. [90]

Lee, J. and F. Bean (2004), "America's Changing Color Lines: Immigration, Race/Ethnicity, and Multiracial Identification", *Annual Review of Sociology*, Vol. 30/1, pp. 221-242, http://dx.doi.org/10.1146/annurev.soc.30.012703.110519. [89]

Liebler, C. and A. Halpern-Manners (2008), "A practical approach to using multiple-race response data: a bridging method for public-use microdata", *Demography*, Vol. 45/1, pp. 143-55, http://www.ncbi.nlm.nih.gov/pubmed/18390296. [75]

Liebler, C. et al. (2017), "America's Churning Races: Race and Ethnic Response Changes between Census 2000 and the 2010 Census", *CARRA Working Paper Series* Working Paper #2014-09, https://osf.io/preprints/socarxiv/d837h. [69]

Montagu, A. (1942), *Man's most dangerous myth : the fallacy of race*, AltaMira Press. [88]

Mora, C. (2014), *Making Hispanics - How activists, bureaucrats and media created a new American*, The University of Chicago Press. [100]

Morning, A. (2014), "And you thought we had moved beyond all that: biological race returns to the social sciences", *Ethnic and Racial Studies*, Vol. 37/20, pp. 1676-1685, http://dx.doi.org/10.1080/01419870.2014.931992. [87]

Morning, A. (2011), *The nature of race : how scientists think and teach about human difference*, University of California Press. [86]

Morning, A. (2008), "Ethnic Classification in Global Perspective: A Cross-National Survey of the 2000 Census Round", *Population Research and Policy Review*, Vol. 27/2, pp. 239-272, http://dx.doi.org/10.1007/s11113-007-9062-5. [62]

OECD (2019), *Society at a Glance 2019: OECD Social Indicators*, OECD publishing, Paris, forthcoming. [31]

OECD (2018), *Indicators of Immigrant Integration 2018: Settling In*, OECD Publishing, Paris, forthcoming. [11]

OECD (2018), *Linking Indigenous Communities with Rural and Regional Development*, OECD publishing,Paris, forthcoming. [9]

OECD (2018), *The Resilience of Students with an Immigrant Background: Factors that Shape Well-being*, OECD Reviews of Migrant Education, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264292093-en. [58]

OECD (2017), "How does having immigrant parents affect the outcomes of children in Europe?", *Migration Policy Debates*, OECD Publishing, Paris, http://www.oecd.org/els/mig/How-does-having-immigrant-parents-affect-the-outcomes-of-children-in-Europe.pdf (accessed on 24 September 2018). [25]

OECD (2017), *How's Life? 2017: Measuring Well-being*, OECD Publishing, Paris, http://dx.doi.org/10.1787/how_life-2017-en. [2]

OECD (2017), "Linking Indigenous Communities with Rural and Regional Development", *Paper presented at the 19th session of the Working Party on Rural Policy*, OECD Publishing, Paris, http://www.oecd.org/cfe/regional-policy/Indigenous-Commuities-project.pdf (accessed on 24 September 2018). [8]

OECD (2015), *Indicators of Immigrant Integration 2015: Settling In*, OECD Publishing, Paris. [10]

OECD (2013), *How's Life? 2013: Measuring Well-being*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264201392-en. [79]

Office of Management and Budget (1997), *Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity*, https://obamawhitehouse.archives.gov/omb/fedreg_1997standards (accessed on 24 September 2018). [63]

Ogunwole, S. (2002), *The American Indian and Alaska*, https://www.census.gov/prod/2002pubs/c2kbr01-15.pdf. [85]

Parker, J. et al. (2004), "Bridging between two standards for collecting information on race and ethnicity: an application to Census 2000 and vital rates", *Public health reports (Washington, D.C. : 1974)*, Vol. 119/2, pp. 192-205, http://dx.doi.org/10.1177/003335490411900213. [76]

Peress, M. (2010), "Correcting for Survey Nonresponse Using Variable Response Propensity", *Journal of the American Statistical Association*, Vol. 105/492, pp. 1418-1430, http://dx.doi.org/10.1198/jasa.2010.ap09485. [47]

Perlmann, J. and M. Waters (2002), *The New Race Question: How the Census Counts Multiracial Individuals*, Russell Sage Foundation. [72]

Pew Research Center (2017), *Views on race, immigration and discrimination*, http://www.people-press.org/2017/10/05/4-race-immigration-and-discrimination/ (accessed on 25 September 2018). [84]

Poulain, M. and A. Herm (2013), "Population Central Population Registers as a Source of Demographic Statistics in Europe", *Population*, Vol. 68/2, pp. 183-212. [56]

Reese, S. et al. (1986), "Ethnicity-of-Interviewer Effects Among Mexican-Americans and Anglos", *The Public Opinion Quarterly*, Vol. 50/4, pp. 563-572, http://dx.doi.org/10.2307/2748759. [67]

Reitz, J., H. Zhang and N. Hawkins (2011), "Comparisons of the success of racial minority immigrant offspring in the United States, Canada and Australia", *Social Science Research*, Vol. 40/4, pp. 1051-1066, http://dx.doi.org/10.1016/J.SSRESEARCH.2011.03.009. [35]

Ringelheim, J. (2006), "Processing Data on Racial or Ethnic Origin for Antidiscrimination Policies: How to Reconcile the Promotion of Equality with the Right to Privacy?", *Jean Monnet Working Paper Series*, No. 13, Center for Human Rights and Global Justice. [83]

Rockquemore, K., D. Brunsma and J. Feagin (2008), *Beyond Black : biracial identity in America*, Rowman & Littlefield Publishers. [82]

Roth, W. (2016), "The multiple dimensions of race", *Ethnic and Racial Studies*, Vol. 39/8, pp. 1310-1338, http://dx.doi.org/10.1080/01419870.2016.1140793. [54]

Saperstein, A. (2006), "Double-Checking the Race Box: Examining Inconsistency between Survey Measures of Observed and Self-Reported Race", *Social Forces*, Vol. 85/1, pp. 57-74, http://dx.doi.org/10.1353/sof.2006.0141. [60]

Saperstein, A., J. Kizer and A. Penner (2016), "Making the Most of Multiple Measures: Disentangling the Effects of Different Dimensions of Race in Survey Research", *American Behavioral Scientist*, Vol. 60/4, pp. 519-537, https://sociology.stanford.edu/publications/making-most-multiple-measures-disentangling-effects-different-dimensions-race-survey. [51]

Schimmele, C. and Z. Wu (2016), "Policy Brief No. 22 - The New Immigration and Ethnic Identity", *Population Change and Lifecourse Strategic Knowledge Cluster Research/Policy Brief*, Vol. 1/6, https://ir.lib.uwo.ca/pclc_rpb/vol1/iss6/7 (accessed on 24 September 2018). [55]

Sen, A. (2007), *Identity and Violence: The Illusion of Destiny*, Penguin. [78]

Simon, P. (2017), "The failure of the importation of ethno-racial statistics in Europe: debates and controversies", *Ethnic and Racial Studies*, Vol. 40/13, pp. 2326-2332, http://dx.doi.org/10.1080/01419870.2017.1344278. [36]

Simon, P. (2015), "The Choice of Ignorance: The Debate on Ethnic and Racial Statistics in France", Springer, Cham, http://dx.doi.org/10.1007/978-3-319-20095-8_4. [34]

Simon, P. (2012), "Collecting ethnic statistics in Europe: a review", *Ethnic and Racial Studies*, Vol. 35/8, pp. 1366-1391, http://dx.doi.org/10.1080/01419870.2011.607507. [30]

Simon, P. (2007), *Ethnic statistics and data protection in the Council of Europe countries*. [29]

Simon, P. and M. Clément (2006), "How should the diverse origins of people living in France be described? An exploratory survey of employees' and students' perceptions", *Population & Societies* 425. [81]

Simon, P. and V. Piché (2012), "Accounting for ethnic and racial diversity: the challenge of enumeration", *Ethnic and Racial Studies*, Vol. 35/8, pp. 1357-1365, http://dx.doi.org/10.1080/01419870.2011.634508. [32]

Snipp, C. (2003), "Racial Measurement in the American Census: Past Practices and Implications for the Future", *Annual Review of Sociology*, Vol. 29/1, pp. 563-588, http://dx.doi.org/10.1146/annurev.soc.29.010202.100006. [73]

Statistics Canada (2017), *Visible Minority and Population Group Reference Guide, Census of Population 2016*, https://www12.statcan.gc.ca/census-recensement/2016/ref/guides/006/98-500-x2016006-eng.cfm. [21]

Statistics New Zealand (2014), *2013 Census counted over 97 percent of population*, https://www.stats.govt.nz/news/2013-census-counted-over-97-percent-of-population (accessed on 24 September 2018). [80]

Statistics New Zealand (2005), *Understanding and Working with Ethnicity Data*, http://archive.stats.govt.nz/~/media/Statistics/surveys-and-methods/methods/class-stnd/ethnicity/Understanding%20and%20Working%20with%20Ethnicity%20Data.pdf. [57]

Statistics New Zealand (2005), *Understanding and Working with Ethnicity Data - A Technical Paper*, http://archive.stats.govt.nz/~/media/Statistics/browse-categories/population/census-counts/review-measurement-ethnicity/understanding-working-ethnicity-data.pdf. [64]

Steward, F. (2001), *Horizontal inequalities: A neglected dimension of development*, United Nations University, World Institute for Development Economics Research. [1]

Swallen, K. and A. Guend (2003), "Data Quality and Adjusted Hispanic Mortality in the United States, 1989–1991", *Ethnicity & Disease*, Vol. 13, pp. 126-133. [52]

Te Mana Raraunga (2018), *Te Mana Raraunga / Maori Data Sovereignty Network*, https://www.temanararaunga.maori.nz/. [46]

UN General Assembly (2015), *Transforming our world: the 2030 Agenda for Sustainable Development*, https://sustainabledevelopment.un.org/post2015/transformingourworld (accessed on 24 September 2018). [7]

United Nations (2017), *Principles and Recommendations for Population and Housing Censuses Revision 3*, Department of Economic and Social Affairs, New York, http://unstats.un.org/unsd (accessed on 24 September 2018). [3]

United Nations (2009), *State of the World's Indigenous Peoples, First Volume*, UN DESA, United Nations Department of Economic and Social Affairs, https://www.un.org/development/desa/indigenouspeoples/publications/2009/09/state-of-the-worlds-indigenous-peoples-first-volume/ (accessed on 24 September 2018). [16]

United Nations (2007), *Declaration on the Rights of Indigenous Peoples*, https://www.un.org/development/desa/indigenouspeoples/declaration-on-the-rights-of-indigenous-peoples.html (accessed on 24 September 2018). [101]

United Nations (1998), *Recommendations on Statistics of International Migration*, United Nations, New York, https://unstats.un.org/unsd/publication/seriesm/seriesm_58rev1e.pdf (accessed on 24 September 2018). [5]

US Census Bureau (2018), *Research to Improve Data on Race and Ethnicity*, https://www.census.gov/about/our-research/race-ethnicity.html (accessed on 26 September 2018). [61]

Waters, M. (1999), *Black identities: West Indian immigrant dreams and American realities*, Russell Sage Foundation, http://www.hup.harvard.edu/catalog.php?isbn=9780674007246 (accessed on 24 September 2018). [102]

Weeks, M. and R. Moore (1981), "Ethnicity-of-Interviewer Effects on Ethnic Respondents", *Public Opinion Quarterly*, Vol. 45/2, p. 245, http://dx.doi.org/10.1086/268655. [66]

Annex A. Diversity data collection in the OECD: Detailed results

Table A A.1. Collection of data on migrant status

Country	Characteristic available			
	Country of birth (41)	Year of arrival in country (17)	Parent's country of birth (30)	Other
Australia	✓	✓		Immigration status, reason for immigration
Austria (register-based)	✓	✓	✓	
Belgium	✓		✓	
Canada	✓	✓	✓	Immigration status, immigrant admission (categories), citizenship, languages spoken
Chile	✓	✓		
Czech Republic	✓		✓	Previous place of residence
Denmark (register-based)	✓	✓	✓	Previous place of residence
Estonia	✓	N/A	✓	
Finland (register-based)	✓	✓	✓	Languages spoken
France	✓	N/A	✓	
Germany	✓	✓	✓	Type of citizenship acquisition
Greece	✓	✓		Previous place of residence, reason for immigration
Hungary	✓		✓	Languages spoken
Iceland (register-based)	✓		✓	
Ireland	✓		✓	
Israel (partly register-based)	✓		✓	

Italy	✓		✓	
Japan	✓			
Korea	✓	✓		
Latvia (partly register-based)	✓		✓	
Lithuania (partly register-based)	✓		✓	
Luxembourg	✓	✓	✓	
Mexico	✓	✓		Previous place of residence, person in household emigrating, date of last emigration and return
Netherlands (register-based)	✓	✓	✓	Previous place of residence, age at arrival, reason for immigration (non EU-EFTA)
New Zealand	✓	✓		Migrants' perceptions about prejudice, discrimination, satisfaction, attitudes to migrants and migration flows, languages spoken
Norway (register-based)	✓		✓	
Poland	✓		✓	Languages spoken
Portugal	✓	✓	✓	Previous place of residence
Slovak Republic	✓		✓	
Slovenia	✓	✓	✓	
Spain	✓		✓	
Sweden (register-based)	✓	✓	✓	Reason for immigration
Switzerland	✓		✓	Languages spoken
Turkey	✓	✓	✓	Previous place of residence
United Kingdom	✓	✓	✓	History of residence in the UK, languages spoken
United States	✓			Citizenship
Colombia	✓			
Costa Rica	✓			
Romania	✓		✓	
Bulgaria	✓		✓	Languages spoken
Russian Federation	✓			Languages spoken

Note: A tick indicates that a variable is collected in at least one data source (e.g. census, household survey, population registry). Information from SILC (core survey) and the EU-SILC 2011 ad-hoc module on Transmission of Intergenerational Disadvantages has also been considered for all European countries, including for Estonia and France.

Table A A.2. Collection of data on racial identity

Country (8)	Source	Question	Response categories		Multiple responses possible	Voluntary response	Internal comparability across current instruments
			Pre-coded	Open-ended			
Canada	• Census • Sample survey	Is this person…?	White, South Asian, Chinese, Black, Filipino, Latin American, Arab, Southeast Asian, West Asian, Korean, Japanese	Other	✓	Only for social surveys	✓
Chile	• Sample survey	Do you consider yourself Afro-descendant? (Moreno de Azapa, Descendant of Morena family, Black, Zambo, Mulatto)	Yes, No				N/A
Ireland	• Census	What is your ethnic or cultural background?	**White** Irish, Irish Traveller, Other White **Black or Black Irish** African, other Black **Asian or Asian Irish** Chinese, other **Other, incl. mixed**	Other		✓	N/A
Mexico	• Sample survey	According to his culture, history and traditions, is (name) Black, this is, Afro-Mexican or Afro-Descendant?	Yes, Yes in part, No				✓

United Kingdom	• Census • Sample survey	What is your ethnic group..?	**White:** English/Welsh/Scottish/Northern Irish/British, Irish, Gypsy or Irish Traveller, Any other White background **Mixed/multiple ethnic groups**: White and Black Caribbean, White and Black African, White and Asian, Any other Mixed / Multiple ethnic background **Asian/ Asian British:** Indian, Pakistani, Bangladeshi, Chinese, Any other Asian background **Black/ African/ Caribbean/ Black British:** African, Caribbean, Any other Black/ African/Caribbean background **Other ethnic group:** Arab	Any other ethnic group	✓		✓
United States	• Census • Sample survey	What is this person's race?	White, Black/ African Am or Negro, American Indian or Alaska Native, Asian Indian, Chinese, Filipino, Other Asian, Japanese, Korean, Vietnamese, Native Hawaiian, Guamanian or Chamorro, Samoan, Other Pacific Islander, Some other race	Print race for American Indian or Alaska Native, Other Asian, Other Pacific Islander, Some other race options	✓	N/A	✓
Colombia	• Census • Sample survey	According to your culture, people or physical characteristics, do you recognize yourself as being…?	Indigenous, Rom, Raizal, Palenquero, Black/ Mulatto/ Afro-Colombian/ Afro-descendant, None of the above				✓
Costa Rica	• Census • Sample survey	(Name) considers him/herself	Black or Afro-descendent, Mulatto, Chinese, White or Mestizo, Other, None				Collapse of categories for some surveys

Note: Don't know/refused to answer response options have not been noted. Some answers have been reclassified by the authors as racial identity due to the content of the question, even though national statistical offices themselves reported it as ethnicity. Internal comparability across current instruments only refers to the most recent wave of censuses and sample surveys considered in this paper, and does not imply comparability over time or across other data sources such as administrative records.
Ireland: The Pilot Census 2018 has included additional response options of Roma, Indian/Pakistani/Bangladeshi, Arabic for ethnic background, United Kingdom: There are minor differences in the response categories for the Scottish census questionnaire. Colombia: For the upcoming 2018 National Population and Housing Census, a consultation and agreement process was held with the ethnic groups of the country (Afro-Colombian, Indigenous and Gypsies), and the response categories to the race question are planned to be expanded as follows: "Indigenous, to which indigenous group do you belong, to which clan do you belong? Gypsy, to which Vitsa do you belong, to which Kumpania do you belong? Raizal of San Andrés, Providencia and Santa Catalina Archipelago? Palenquero of San Basilio? Black, Mulatto, Afro-descendant, Afro-Colombian? No ethnic group." United States: The proposed design of the race question for the 2020 Census will include several significant changes: examples and instructions to print origins for the White and Black or African Am. categories, removal of the term "Negro", examples for the Indian or Alaska Native Category, and change of "Guamanian or Chamorro" to "Chamorro".

Table A A.3. Collection of data on ethnic identity

Country (17)	Source	Question	Response categories		Multiple responses possible	Voluntary response	Internal comparability across current instruments
			Pre-coded	Open-ended			
Australia	• Census	What is the person's ancestry?	English, Irish, Scottish, Italian, German, Chinese, Australian	Other ancestry 1 +2	✓		N/A
Canada	• Census • Sample survey	What were the ethnic or cultural origins of this person's ancestors? An ancestor is usually more distant than a grandparent. For example, Canadian, English, Chinese, French, East Indian, Italian, German, Scottish, Cree, Mi'kmaq, Salish, Metis, Inuit, Filipino, Irish, Dutch, Ukrainian, Polish, Portuguese, Vietnamese, Korean, Jamaican, Greek, Iranian, Lebanese, Mexican, Somali, Colombian, etc.		Specify as many origins as applicable	✓	Only for social surveys	✓
Czech Republic	• Census	Ethnicity:		✓	✓	✓	N/A
Hungary	• Census • Sample survey	Which nationality do you feel you belong to?	Hungarian, Bulgarian, Roma, Greek, Croatian, Polish, German, Armenian, Romanian, Ruthenian, Serbian, Slovak, Slovenian, Ukrainian, Arabian, Chinese, Russian, Vietnamese	Other	✓	✓	Collapse of categories for LFS
Israel	• Census • Sample survey • Register	a) Nationality (from register) + b) What is your religion?	Jewish, Moslem, Christian, Druze	Other			✓
Ireland	• Census	What is your nationality?	Irish, Other, No nationality	Other	✓	✓	N/A
Latvia	• Register	N/A	N/A	N/A			✓
Lithuania	• Census	What is your ethnicity?	Lithuanian, Polish, Russian	Other	N/A	✓	N/A

Poland	•	Census	What is your nationality? (understood as the national or ethnical affiliation – do not confuse with citizenship)	Polish, Belarusian, Czech, Karaitic, Lithuanian, Lemko, German, Armenian, Romany, Russian, Slovakian, Tatar, Ukrainian, Jewish	Other	✓	N/A	N/A
Slovak Republic	•	Census Sample survey	Nationality:	Slovak, Hungarian, Roma, Ruthenian, Ukrainian, Czech, German, Polish, Croatian, Yiddish, Bulgarian	Others			Different response categories for census and LFS
Slovenia	•	Census	Nationality/ethnicity:	Slovenian, Italian, Hungarian, Nationally/ethnically indeterminate	Other nationality/ ethnicity		✓	N/A
New Zealand	•	Census Sample survey	Which ethnic group do you belong to?	New Zealand European, Maori, Samoan, Cook Island Maori, Tongan, Niuean, Chinese, Indian	Other such as Dutch, Japanese, Tokelauan. Please state	✓		✓
United Kingdom	•	Census Sample survey	How would you describe your national identity?	English, Welsh, Scottish, Northern Irish, British	Other	✓		✓
United States	•	Census Sample survey	Is this person of Hispanic, Latino or Spanish origin? For this census, Hispanic origins are not races.	No, not of Hispanic, Latino or Spanish origin; Yes, Mexican, Mexican Am., Chicano; Yes, Puerto Rican; Yes, Cuban	Yes, another Hispanic, Latino or Spanish origin	✓	N/A	✓
Bulgaria	•	Census Sample survey	Ethnic group?	Bulgarian, Turkish, Roma	Other	✓	N/A	✓
Romania	•	Census	Which ethnic group does the person consider belonging to?		✓	✓	✓	✓
Russian Federation	•	Census Sample survey	Your nationality:		✓	✓	✓	✓

Note: Don't know/refused to answer response options have not been noted. Ireland: The Pilot Census 2018 has included the change of wording from "nationality" to "citizenship". Internal comparability across current instruments only refers to the most recent wave of censuses and sample surveys considered in this paper, and does not imply comparability over time or across other data sources such as administrative records.

Table A A.4. Collection of data on indigenous identity

Country (7)	Source	Question	Response categories		Multiple responses possible	Voluntary response	Internal comparability across current instruments
			Pre-coded	Open			
Australia	• Census • Sample survey	Is (person) of Aboriginal origin, Torres Strait Islander origin or both?	Neither Aboriginal nor Torres Strait Islander origin, Aboriginal but not Torres Strait Islander origin, Torres Strait Islander but not Aboriginal origin, Both Aboriginal and Torres Strait Islander origin		✓		✓
Canada	• Census • Sample survey	a) Is this person an Aboriginal person, that is, First Nations (North American Indian), Metis or Iunuk (Inuit)? b) Is this person a Status Indian (Registered or Treaty Indian as defined by the Indian Act of Canada)? c) Is this person a member of a First Nation/Indian band?	a) No, First Nations (North American Indian), Metis, Inuk (Inuit) b and c) Yes, No	c) Specify name of First Nation/Indian band	✓	✓ (except census)	✓
Chile	• Census • Sample survey	a) Do you consider yourself as belonging to an indigenous or native people? b) To which one?	a) Yes, No b) Mapuche, Aymara, Rapa Nui, Lican Antai, Quechua, Colla, Diaguita, Kawésqar, Yagán or Yámana	Other			N/A
Mexico	• Census • Sample survey	a) According to (name)'s culture, does she(he) consider herself(himself) indigenous? b) Does (name) speak an indigenous dialect or language? c) Which dialect or indigenous language does (name) speak?	a and b) Yes, No	c) Specify			✓
New Zealand	• Census • Sample survey	a) Are you descended from a Māori (that is, did you have a Māori birth parent, grandparent or great-grandparent, etc)? b) Do you know the name(s) of your iwi (tribe or tribes)? c) In which language(s) could you have a conversation about a lot of everyday things?	a) Yes, No c) English, Māori, Samoan, New Zealand Sign Language	b) Write in iwi and rohe (iwi area) other for the language c) Other	✓		✓

United States	• Census • Sample survey	What is this person's race?	**White** (e.g. German, Irish, English, Italian, Lebanese, Egyption, etc) **Black or African Am,** (e.g. African American, Jamaican, Haitian, Nigerian, Ethiopian, Somali, etc) **American Indian or Alaska Native** (e.g. Navajo Nation, Blackfeet Tribe, Mayan, Aztec, Native Village of Barrow Inupiat Traditional Government, Nome Eskimo Community, etc) **Chinese** **Filipino** **Asian Indian** **Vietnamese** **Korean** **Japanese** **Other Asian** (e.g. Pakistani, Cambodian, Hmmong, etc) **Native Hawaiian** **Samoan** **Chamorro** **Other Pacific Islander** (e.g. Tongan, Fijian, Marshallese, etc) **Some other race**	Print name of enrolled or principal tribe(s) for American Indian or Alaska Native option	✓	N/A	Reporting of multiple races allowed since 2000
Colombia	• Census • Sample survey	Which of the following ethnic groups do you consider yourself belonging to?	Indigenous, Gypsy, Raizal of the archipelago, Palenquero, Black/ Mulatto (Afro-descendant), None of the above				✓
Costa Rica	• Census • Sample survey	a) (Name) considers himself/herself indigenous? B) What is the indigenous group (name) belongs to? C) Does (name) speak any indigenous language?	a and c) Yes, No b) Bribri, Brunca or Boruca, Cabécar, Chorotega, Huetar, Maleku or Guatuso, Ngöbe or Guaymí, Teribe or Térraba, From other country, No indigenous group				✓

Note: Don't know/refused to answer response options have not been noted. Internal comparability across current instruments only refers to the most recent wave of censuses and sample surveys considered in this paper, and does not imply comparability over time or across other data sources such as administrative records.
Australia: The possible inclusion of an additional question on which people/clan/tribe/nation/language group the Aboriginal and/or Torres Strait Islander person is from.is currently under discussion for the 2021 Census of Population and Housing. Mexico: Just as with race and ethnicity, additional information on reflected indigenous status is collected.