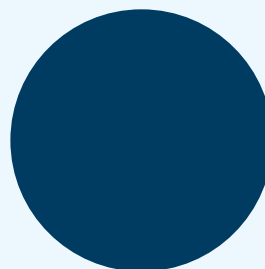
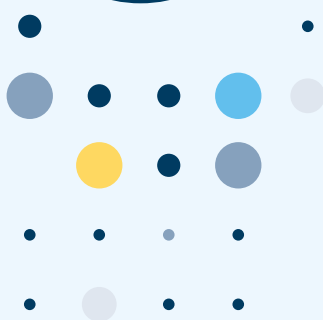
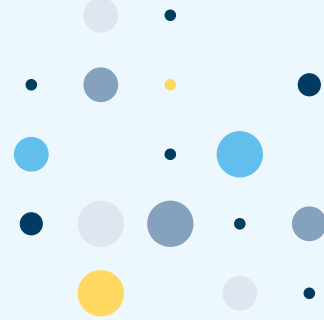
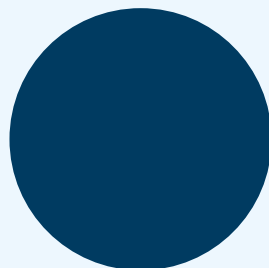
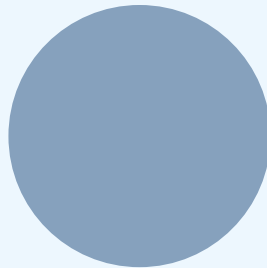
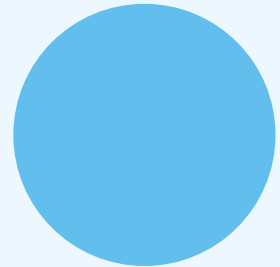
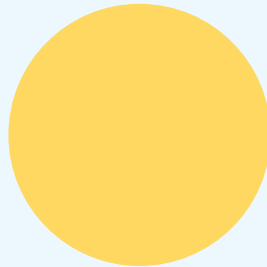
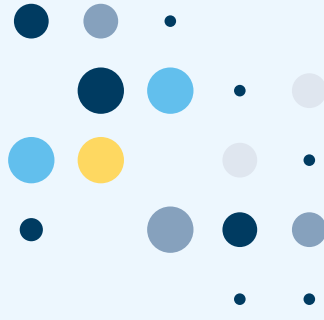
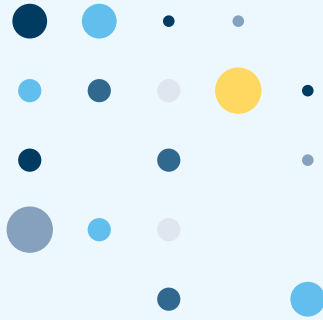
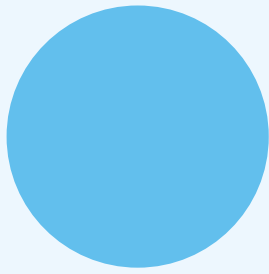


**Inclusive data taskforce**

**Recommendations report**

**Leaving no one behind. How can  
we be more inclusive in our data?**





# Table of Contents

<b>Chair's foreword</b>	<b>1</b>
<b>Introduction</b>	<b>3</b>
Why was the Inclusive Data Taskforce established?	3
What did we aim to achieve?	3
What did we do?	4
Our report and recommendations	5
<b>What are the critical data gaps?</b>	<b>7</b>
Groups or characteristics missing from the data	8
Groups for whom there are insufficient data	9
Groups for whom data are of insufficient quality	9
<b>How inclusive are the UK's current approaches to data collection?</b>	<b>11</b>
Addressing participation in data collection activities	11
Ensuring that the data collected meet respondent and user needs	13
Ensuring data collected are of sufficient quality to accurately count everyone in society and monitor their outcomes	14
<b>How can we ensure everyone in society is represented in data analysis and outputs?</b>	<b>16</b>
Lack of granularity in data to enable an understanding of specific subgroups and the intersectionality of personal characteristics	17
Lack of resources and/or skills to find and analyse data	19
Lack of suitable and accessible outputs	20

<b>What does good practice look like and what are the conditions that enable inclusivity to thrive?</b>	<b>22</b>
Building trust through engagement	22
Answering the right questions	23
Engaging all communities	23
Partnerships	24
Inclusivity in routine data collection	24
Developing new approaches to inclusive data collection	24
Making the best use of existing data to improve inclusiveness	25
Inclusive dissemination and communication	26
<b>Our recommendations</b>	<b>27</b>
How can we be more inclusive in our data?	27
<b>Supporting evidence</b>	<b>37</b>

# Chair's foreword

I am pleased to present the report of the Inclusive Data Taskforce and the recommendations we have developed to help achieve a more inclusive data system.

The Taskforce was established in October 2020, with the goal of ensuring that data and evidence across the UK is reflective and inclusive of all, so 'that everyone in society counts and is counted and no one is left behind.'

The COVID-19 pandemic, which has provided the context for all of our work, has highlighted just how important good and timely data is for public policy and society. Inclusive data helps us understand how events impact differentially on individuals, groups and communities. In turn this must enable those responsible in government, local authorities and wider society, and all of us as individuals and communities, to address the disparities and inequalities which exist in the UK in 2021.

The Taskforce has brought together a range of expertise, including senior academics and civil society leaders; on subjects such as social and economic inequalities, ethnicity and identity; qualitative and quantitative methodologies, and research ethics. I am grateful to them for their hard work and their willingness to meet the tight deadlines to which we had to work. Like me, they deeply appreciate the support and hard work of the ONS team who have supported us so ably throughout. We have looked at how inclusivity can be improved across the research process from study design, data mapping, data collection, data analysis through to the presentation of findings.

As part of our process, we conducted extensive consultation and engagement. We gathered a range of insights to examine the current comprehensiveness of UK data and evidence.

**We sought to understand what works well and what could be improved, the gaps and barriers to achieving more inclusive data, the impacts of data not being fully inclusive, and how inclusivity can be improved.**

I am grateful to everyone who took part and contributed their time, views, experiences and ideas for solutions and ways forward. The Taskforce has been struck by the willingness of people to engage with the consultation activities, to organise events themselves and to send their work to us. The contributions of so many different individuals and groups have been integral to our deliberations and recommendations. They have helped us better understand how data inclusivity might be achieved in practice.

Consultation of the depth and breadth carried out was only possible given our timescale because of the team at the ONS Centre for Equalities and Inclusion, who worked tirelessly on our behalf on engagement, consultation and research, gathering and building the evidence base for this report.

I believe our recommendations will land on fertile ground, given the contributions we have received from so many quarters, which is reassuring as there is much to be done.

We hope that the recommendations outlined in this report, which are extensive and comprehensive, will help chart a way forward in improving the representativeness of UK data and evidence, so that the voices and experiences of all are better reflected in the future.

**Dame Moira Gibb**

Chair of the Inclusive Data Taskforce  
July 2021



## Introduction

### Why was the Inclusive Data Taskforce established?

A year ago, the National Statistician convened an independent Taskforce to recommend how best to make a step-change in the inclusivity of UK data and evidence. This followed publication of *Statistics for the Public Good*, the strategic vision of the UK Statistics Authority which aims to ensure that everyone in the UK counts and is counted, and no one is left behind.

Taskforce members, chaired by Dame Moira Gibb, comprise a diverse group of senior academics and civil society leaders with wide ranging expertise across equalities topics, methodologies, geographies, and data ethics.

### What did we aim to achieve?

We were asked to consider four important questions, the answers to which form the basis of our recommendations to the National Statistician:

- how can we improve inclusiveness in our approach to the collection, analysis and reporting of data and evidence?
- how can we make most effective use of existing data, such as administrative, census and survey data to understand equalities and inclusion?
- what are the critical data gaps that hinder our understanding of equalities and inclusion and how can we address them?
- how can we build on our own and others' experiences in improving our approach to equalities and inclusion going forward?

## What did we do?

Over a 9-month period, we set ourselves an ambitious programme of engagement activities to listen and learn. This included conversations with those who may be asked to participate in research and share their data with researchers and with those who collect or use data and evidence. We heard from people across all four countries of the UK through a range of consultation activities carried out between January and May of 2021 including:

- a 12-week online open consultation on CitizenSpace
- seven roundtable discussions and six in-depth interviews with central and local government representatives, and those in the devolved nations
- four roundtable discussions and two in-depth interviews with academics and representatives of learned societies
- discussions with over 80 civil society leaders working in 15 different equalities areas
- discussions with over 90 members of the public with lived experience of equalities issues

Participants were drawn from a range of backgrounds and were selected based on the equalities work that currently takes place. Consultation events were held online, as they took place during the pandemic, when face-to-face meetings were restricted. To ensure we heard from people who may be less able to access the internet, we also did a paper-based consultation by post with those at risk of digital exclusion. Other groups and organisations invited us to events that they organised to contribute their perspectives to the consultation. These included:

- a round table discussion organised by people identifying as Sikh
- a round table discussion organised by people identifying as Muslim
- an event organised by the Open Data Institute
- an event organised by the Foreign, Commonwealth and Development Office

During our monthly meetings, we considered papers and presentations on a wide range of topics relating to inclusive data and evidence. Some of these we commissioned ourselves on topics we felt were important to learn more about, while others were contributed by people wanting to bring their work to our attention.

We are very grateful to all those who shared their views and experiences with us and have tried to do justice to the wealth of information gathered in our recommendations for the way forward. For those interested in having a more detailed look at the findings from each of the consultation activities, they have been published separately and are available to view online. Links to each report can be found in the supporting evidence section.



This report is also available in Welsh. A summary version of this report is available in English, Welsh and easy read format. Easy read versions of the summary report are available in English, Welsh, Polish, Romanian, Punjabi, Mandarin, Cantonese, Arabic and Farsi. If you require another format, please email us at [equalities@ons.gov.uk](mailto:equalities@ons.gov.uk) or call **0800 298 5313**.

## Our report and recommendations

We have carefully considered all the evidence gathered in developing our recommendations and this report summarises many of the issues people raised with us. Throughout the consultation activities and the evidence received during our meetings, it became clear that the step-change we need towards more inclusive data and evidence will require:

- development of a strong culture of trust and trustworthiness, recognising the need to ensure confidence in data collection both of those sharing their data and those using data for the public good
- a system-wide approach and real commitment to being more inclusive in what is measured and how, who is engaged with, widening access to data and evidence, and improving the accuracy of the insights produced to better reflect people's lives
- clarity about what inclusivity means in practice, to ensure data and evidence comprehensively and reliably reflect our society, now and as it evolves

The Taskforce aimed to be as inclusive as possible in considering needs across the four countries of the UK. There are different legislative frameworks in place across the UK and we have been mindful of this in how we refer to populations and groups of interest in relation to equalities and inclusion data and evidence. The Taskforce has focused on a wide range of populations and groups that may be at greater risk of disadvantage, discrimination, or marginalisation, both generally and in our statistical picture of the UK. This has included:

- those with characteristics protected under the Equality Act (2010) which applies in England, Scotland and Wales
- groups which feature in the Sustainable Development Goals as those are particularly important for disaggregation to ensure we leave no one behind
- others not explicitly part of these frameworks, but whom we know to be under-represented in UK statistics and evidence such as residents of some types of communal establishments, non-private household populations more generally, and undocumented migrants



Throughout our report and recommendations, we have used **'relevant groups and populations'** to refer to the social groups and populations of participants who are the specific focus of our recommendations in relation to greater inclusivity of data collection, recording and reporting. This includes those with protected characteristics (in England, Wales and Scotland) and also refers to those groups completely missing from the data or for whom data are lacking or are of poor quality.

We also refer to **'racialised groups'**, or people considered to share characteristics that are assumed to be on the basis of shared 'race'. The term 'racialised' highlights the social nature, rather than biological foundation, of this assumption.



# What are the critical data gaps?

Across the UK data infrastructure, a considerable amount of data exists to explore the experiences and outcomes of a range of groups of people with different characteristics.

As a core demographic variable, **sex** is collected in most administrative and survey data. The UK collects, reports and produces considerable analyses of data on **disability**. Furthermore, the UK has, compared with most other European countries, rich data on ethnicity based on detailed consultation and piloting. **Ethnic group** data are available across many official data sources, including administrative, survey and the UK censuses, though data quality and granularity varies across sources. The UK also has good data on **religion** from the censuses and some government surveys, although the religion question is a voluntary one, and therefore not answered by all respondents. The harmonised question response options for both religion and ethnicity at the most granular level are also not the same for all the countries of the UK.

Nevertheless, our consultation activities identified a number of areas where participants feel there are critical data gaps, with 70% of organisations and 61% of respondents to the online consultation stating that data gaps had impacted their ability to answer the questions of most importance to them. These critical data gaps can be broadly separated into groups or characteristics that are **missing completely from the data** and those where **insufficient data** are available or are of **insufficient quality or granularity** to meet user needs.



## Groups or characteristics missing from the data

Across the consultation activities, a number of groups were repeatedly identified for whom even basic demographic information is missing. These included **transgender, non-binary and gender-diverse people, non-household populations** (for example, members of residential establishments such as care homes or prisons, and homeless people, particularly those who do not access any services for rough sleepers) and **groups often deemed 'harder to reach'** (for example, Gypsy, Roma and Traveller groups, ex-prisoners, asylum seekers, victims of domestic violence and undocumented migrants or victims of human trafficking). Some of these groups include the most vulnerable and disadvantaged people in the UK, rendering the absence of data reflecting their lives and experiences as especially critical.

**Children** are another group that many identified as missing from the data. Where we do have data for them, this is often collected from people other than children themselves and therefore children's own voices may not be heard. The Nuffield Foundation has identified a number of critical gaps in the data on children. This includes a lack of information on all areas of life for looked-after children as well as under-representation of children who have experienced abuse or neglect in early childhood and a lack of information on their outcomes. A report from the London School of Economics also noted the lack of data to understand child poverty and multidimensional disadvantage among children, specifically identifying young carers, migrant children, Gypsy, Roma and Traveller children, and children at risk of abuse or neglect as groups that are 'missing' from or 'invisible' in existing data. These groups were also highlighted by participants in our consultation activities.

Various governmental participants described gaps in understanding the **digitally excluded population**, the reasons why and in what circumstances people may be at risk of digital exclusion, and the extent to which they are not represented in routine data collection. Participants felt that this has likely been exacerbated by measures put in place in response to coronavirus (COVID-19), for example with many surveys now being adapted to online platforms, potentially generating charges for participants, and restricting access.

While there will be many variables and types of data that could potentially be useful in understanding inequalities across the population, participants in our consultation activities identified two key variables as having policy and explanatory relevance. The first was **income data** which is rarely collected alongside personal characteristic information and was identified as a critical gap in census data, seen as essential to understanding disadvantage. The second was **socio-economic background**, an important variable for understanding topics such as educational inequalities, though rarely available in government

data and not included in the censuses; it is often only available in the form of rough approximations, such as the binary Free School Meals' eligibility measure in the Department for Education administrative data. Participants suggested that these should be included more regularly within data that are collected, alongside personal characteristics.

## Groups for whom there are insufficient data

Even for those who are included to some extent in the data infrastructure, there are gaps in the information that is collected. Although data on **sexual orientation** are collected in several UK data sources, there is a scarcity of information on the differing experiences and outcomes of people in terms of their sexual orientation. In addition, despite **pregnancy and maternity** being protected characteristics under the equalities' legislation in Great Britain, information on inequalities in pregnancy and pregnancy outcomes is partial. For example, recognition of racialised differences in maternity outcomes was only made possible by the collection of data by a charitable organisation over the last few years. Gaps in the data on **religion** were noted by both academics and learned society participants and respondents to the online consultation. Specifically, a number of participants said that **religion** is often not collected in surveys, and when collected, is not routinely reported or is often conflated with beliefs and practices, which can obscure inequalities.

Across several consultation activities, participants also described a lack of data on relevant personal characteristics in administrative data sources to help with understanding equalities issues. To our knowledge, no government data at all is collected on the operation of caste in the UK despite qualitative evidence on caste discrimination.

## Groups for whom data are of insufficient quality

Even where relevant groups are included in survey or administrative data, there are risks that **the quality of these data is poor**. First, as noted in the section on groups or characteristics missing from the data, information on children's characteristics may not be collected directly from the children themselves but provided by their parents, carers, teachers or others responsible for children. The same may apply to household residents who are temporarily absent, for example in a communal establishment. Information provided by proxies may be inaccurate. In the case of administrative data, it is often unclear whether the data were reported by the individual concerned or not.

Second, there can be potential **problems of missingness in data on items relevant to inclusion**. For example, census questions on religion are voluntary and therefore have lower response levels than compulsory questions. There also appear to be relatively high levels of non-response on ethnic groups in some administrative data sources. The government's Ethnicity Facts and Figures website

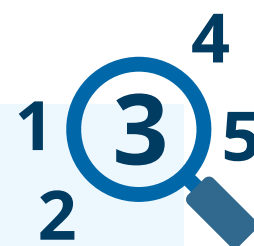
shows that for some administrative data, ethnic group was not available for over twenty percent of cases. This may lead to bias in the data but, without clarity on who this information is missing for, it is not possible to ascertain the direction of any bias.

Third, information from previously collected data sources (for example in a panel study or when data from different sources are linked) can become inaccurate over time as few characteristics are permanently fixed. Characteristics can change as groups express preferences for particular categorical groupings or because different groupings are suitable for different purposes. There are, for example, occasions when the use of 'Black' or 'Asian' categories may be appropriate, but more granular terms are often necessary. In addition, many important characteristics, such as disability status (for example as a result of a stroke or an accident) or socio-economic position (as a result of redundancy) can change unpredictably.

These three issues all potentially lead to greater 'noise' in the data, leading to a more blurred picture, or to bias, which can produce a misleading picture. Sub-optimal data quality and the under-representation of certain populations or groups within data could result in a range of impacts, including discrimination, misrepresentation, reduced life chances, hidden harm and potentially even loss of life for those in highly vulnerable circumstances. As a result, addressing these gaps is viewed as a priority.



# How inclusive are the UK's current approaches to data collection?



Data are crucial for understanding the needs and circumstances of different groups of people, enabling the translation of information into insights, from which action can be taken. This is only achievable if we collect complete and appropriate data, to ensure that everyone counts and is counted, and no one is left behind. Our consultation activities identified a range of issues that need to be addressed for us to be able to make this happen.

## Addressing participation in data collection activities

Research with Civil Society Organisations (CSOs) and individuals from relevant groups and populations indicated that there are a wide range of practical, cultural and emotional factors that impact on people's willingness, ability and opportunity to provide their personal information and participate in formal research exercises. These can broadly be separated into issues around **trust and trustworthiness**, a **willingness to participate** and **accessibility of data collection**. However, these concerns and issues are not mutually exclusive, and a combination of these factors may affect people's ability to engage with data collection activities.

**Trust** emerged across all of our consultation activities **as a barrier to participation in data collection**. This included a perception among several participants that there is a general sense of distrust in the government, as well as in government statistics, particularly, though not exclusively, among under-represented groups (specifically described as affecting those from Gypsy, Roma and Traveller communities, other minority ethnic groups and documented and undocumented migrants). Participants described how this could result in some groups being under-represented or effectively invisible, ultimately leading to policy decisions which may not adequately reflect these populations and increasing their distrust.

Participants noted a degree of uncertainty and apprehension among relevant groups and populations about how their data may be used by the government. Many identified fears that their participation in data collection or in accessing certain services could lead to unequal treatment, discrimination, or worsen their situation. This was described by groups involved in collecting data, by individuals from relevant groups and populations and CSOs. Individuals described not feeling able to report violent crimes due to a perceived risk of being detained because of disclosure of their immigration status. Respondents to the online consultation felt that the onus for reaching and reassuring these members of society was on those collecting data, to ensure their safety and that the needs of these groups are understood and reflected.

Our consultation activities also identified additional barriers to individuals' participation in data collection which may affect their willingness to participate. These included:

- being unable to identify themselves in the options included within data collection tools, and feeling excluded by the use of inappropriate wording (for example, individuals not having the opportunity to express transgender, non-binary or gender-diverse identities, dual nationality or multiple ethnic categories).

“I like to say, ‘I’m from an African Caribbean [background], my mum’s Ghanaian and my dad’s from Barbados,’ but there’s no form that can get me to say that. Then I just feel outcasted.”

#### Individual

- exhaustion from over-research (for example, those with mental health issues and racialised and disadvantaged ethnic groups)
- experiencing competing pressures in their day-to-day lives (for example, managing with a physical disability, managing paid and unpaid work)
- there being little or no perceived personal or community benefit from participation, especially where previous consultation activities had not led to action or where tangible results from participating were not seen.

Academics and learned society participants similarly noted that some people may not want to participate in data collection activities. This results in a ‘survivorship bias’ whereby researchers focus their efforts towards groups who have already participated and neglect those who are under-represented, though this problem is not specific to vulnerable or marginalised groups.

Even where there are no issues with trust or trustworthiness and the individual is willing to participate, we found that people may still be prevented from doing so due to **lack of accessibility of data collection exercises. First, online data collection instruments can exclude those who have no or limited digital access or lack the necessary digital skills.** Digital exclusion was described by government participants as reflecting technology and skills gaps, and costs



associated with participation. Lack of Wi-Fi availability, and costs incurred through telephone engagement were reported barriers to participation for digitally excluded groups and those experiencing financial hardship. Organisations who participated in the online consultation raised concerns around the move from face-to-face approaches to online surveys during the pandemic, and that specific sub-groups may be excluded as a result. Concerns were expressed by respondents to the paper-based consultation around the growing number of surveys being hosted online and an awareness of the implications of their own lack of digital skills. Respondents highlighted that decision-makers must recognise and respond to the difficulties relating to digital exclusion to avoid further exclusionary practices.

Second, **methods may not consider the language, literacy or comprehension needs of different population groups**. For example, participants from government stressed that older people, disabled people and those unable to participate due to language or interpretation barriers are at greater risk of exclusion from research. Those who participated in the online consultation also raised concerns around individuals deemed unable to provide informed consent often being excluded from surveys, such as older people or those lacking mental capacity.

Third, **personal identity and characteristics of those responsible for designing and carrying out the data collection and analysis** was a concern for some groups. There was a perceived lack of diversity within data collection organisations they felt that greater representation across relevant groups and populations within the research community would ensure better understanding of different cultures, address barriers to participation and reduce the risk of burdening participants with duplication of research.

## Ensuring that the data collected meet respondent and user needs

The need for data collectors to provide **meaningful categorisations that respondents can recognise and use to describe themselves and their circumstances** emerged consistently from our consultation activities. The labels used to capture individual characteristics within data collection were perceived as critically important to enable people to **select categories in surveys and on forms that reflect their personal characteristics and circumstances** and to ensure that the data allows for **an accurate understanding, and actions can be taken in response**. Restricting the presentation of data under labels which could homogenise diverse and distinct groups was viewed as highly problematic, misleading and potentially offensive. If the value of data is to enable people to be better understood, and represented in services and policies, data must accurately

represent people's circumstances and identities. Within survey data in particular, the **number of respondents selecting an 'other' or 'mixed'** ethnic group category has **increased**, restricting effective understanding of an individual's identity.

**Questions were often considered outdated** by CSOs and individuals from relevant groups and populations. This was particularly noted for questions and definitions around ethnicity, which may conflate race, ethnicity and nationality and provide broad categories that cover a range of ethnic groups and national origins. This was also noted within questions on **disability** based on outdated, deficit concepts that do not sufficiently capture the experiences of individuals (including needs, structural barriers encountered and experiences of overcoming barriers, or the diversity of disabilities, for example). The organisations and individuals who took part in the online consultation specifically called for **disability data collection to shift from a focus on the medical model**, which looks at an individual's impairments or differences, towards the social model of disability, exploring the individual's needs and perspectives and viewing society as a major contributor to incapacity. This could better address the organisational and structural barriers which limit people's participation in society.

## **Ensuring data collected are of sufficient quality to accurately count everyone in society and monitor their outcomes**

Those who participated in our consultation activities identified various quality issues in relation to data collection, particularly in terms of **conceptual challenges and lack of harmonisation and coherence**.

The need to harmonise the data that are collected on personal characteristics was stressed so that the characteristics and circumstances of minority groups are reflected in all UK administrative and national survey data. Participants also highlighted the need for the definitions, categories and types of questions used to collect data on personal characteristics to be more inclusive. A lack of harmonisation was seen to hinder the ability to disaggregate as well as to compare data across different countries of the UK. Multiple definitions, classifications and response options for ethnicity, disability, sex and gender were highlighted by those who participated in the consultation activities as presenting particularly challenging conceptual issues. It was stressed by CSOs that the lack of harmonisation in the administrative data collected from public services (for example schools, police forces, health services) has resulted in an **inconsistent picture of particular subgroups** (notably relating to faith and ethnicity) and misalignment between "official" data and those collected by CSOs on the ground. Government participants and online respondents mentioned that the 2001 census categories for **ethnicity** remain widely used in data collection, even though it has been recognised that this does not always adequately reflect the ethnic diversity of the population at the present time.

The lack of consistency in the use of **disability** definitions across the UK was highlighted, resulting in disability information being captured in different ways. Inconsistencies in **definitions for the term 'disadvantaged'** were also mentioned as a significant issue for analysts, which was particularly problematic when trying to undertake analyses for specific local areas.

The issues arising due to **small sample sizes** in household surveys were raised across almost all of our consultation activities. These create a **lack of granularity** within the data, which undermines understanding of specific sub-groups of the population, whether that be by sector, geography or characteristic and can render entire groups invisible in data. Achieving local level information on specific populations or group characteristics often involves aggregating smaller groups into larger categories. Such larger categories may not adequately reflect populations of interest and individuals may not identify with larger aggregations; the continued use of the broad 'BAME' (Black, Asian and Minority Ethnic) category was said to have the potential to marginalise and alienate relevant groups and populations.

Participants noted that small sample sizes also hinder our ability to undertake **intersectional analyses**. For example, even though information on age and sex distributions may be available at small geographies, analysing the experience of older, migrant women in specific areas will be inhibited by small sample sizes. Respondents to the online consultation suggested that the issues with sample sizes within household surveys could be overcome through the use of qualitative approaches for specific populations, noting that **quantitative approaches** are limited for understanding the lived experiences of different groups of people.

**Frequency of data collection** was also raised as an important issue. It was acknowledged that the UK Censuses are a valuable source of inclusive data, providing insights not achievable with other data sources, but the 10-year gap between censuses means that the resulting data are often several years out of date. The contextual data provided by the census on local area characteristics can help in understanding the experience of disadvantage but is more valuable when supplemented by the small-area Indices of Multiple Deprivation (IMD). However, the IMDs are country-specific and updated at different intervals so cannot provide a harmonised understanding of local area deprivation across the UK. Additionally, there is frustration that publicly funded deep-dives on key equality areas (such as sexual orientation) are undertaken infrequently and seemingly with no long-term strategic intent.

Finally, a concern raised by academics and learned society participants related to the **use of existing data as learning data in machine learning algorithms**. Any biases in that existing data, for example where they mis-represent certain parts of the population, will result in unrepresentative or biased predictions. This was said to have the potential for biases to be perpetuated in future decision-making.

# How can we ensure everyone in society is represented in data analysis and outputs?

A clear message that emerged from our consultation activities **data robustness**, in terms of representativeness and statistical validity. Additionally, the **relevance and depth of the data** are essential aspects that should be adequately considered. For providers of local services, this might mean having data about the local population which their specific services cater to. For users involved in regional or national advocacy, it could be having data to enable understanding of issues that impact a particular equality area or group in a manner that allows for disaggregation by characteristics of interest. Being able to explore how different characteristics intersect to create different experiences for individuals within society is crucial to ensure that we 'leave no one behind'. Similarly, the availability of geographically granular data is critical to provide policy-makers at a local level with the necessary evidence to inform their decisions. This is particularly relevant when considering data to inform the UK government 'levelling up' agenda but is also essential for specific locally located institutions and services.

Individuals and groups who participated in our consultation activities **identified a wide variety of issues as impacting on their use of data and inhibiting the extent to which statistics reflect the experiences of everyone in society.**

**There was a general recognition that organisations and individuals could access a high volume of data on some but by no means all topics. However, deriving actionable insights from the available data was perceived as problematic, due to data gaps or data aggregation inhibiting value and usability.** The issues that were consistently reported across the different consultation activities follow.

## Lack of granularity in data to enable an understanding of specific subgroups and the intersectionality of personal characteristics

“While it might be possible for me to say the Asian communities within our local area are X or Y or Z, it becomes more difficult then to say ‘Okay, what about just the Asian women?’ or ‘What about Asian women with a disability?’”

Civil Society Organisation (CSO) working in support of race and ethnic equality

This challenge has been explored by participants in terms of the extent to which the data are **fit for purpose**, recorded **in standardised formats** and held in a condition that is **findable, accessible and usable for analysis**.

Throughout consultation exercises, participants emphasised the issue of **data not allowing for disaggregation around key characteristics** and the importance of understanding the intersection of different characteristics and the influence this has on various outcomes. For example, organisations who took part in our engagement activities reported that it is not enough to see data of interest disaggregated by ethnic group, but other characteristics are also required, such as ethnic group by age, or ethnic group by sex.

The current practice of aggregating personal characteristics into broader age, ethnic, faith, or sexual orientation categories was seen both to severely inhibit the inclusivity of statistics and to potentially misrepresent the issues and needs of smaller and more marginalised communities. The CSOs participating in this research highlighted that even within the data that are considered of the highest standard, such as the Census, the lack of granularity meant they were unable to develop an inclusive understanding of groups across the range of personal characteristics. This was seen to impact not just their services but the efficacy of public services and public policy, and ultimately people’s life outcomes.

CSOs and those organisations who took part in our online consultation mentioned their efforts to address the lack of granularity, by developing other sources of data to fill the gaps. These included:

- relying on qualitative research
- using administrative data collected by local services
- conducting their own surveys with service users
- trying to obtain data through Freedom of Information requests
- utilising published data to extrapolate national level data to the local context

This process was seen as potentially introducing error and bias, which were acknowledged as impacting on the quality and accuracy of insights available. Such approaches also were recognised as placing a high burden on organisations and individuals who typically have limited skills and resources to collect, collate, and/or analyse data.

Other reported areas of particular concern were **missing data on marginalised and minority groups or those where the relevant characteristic may be temporary but there may be a long-term impact**. CSOs highlighted that data were particularly lacking for:

- more recently defined or less understood groups (for example, those categorised as neurodiverse, and particularly neurodivergent women)
- those described as “harder-to-reach” (for example, Gypsy, Roma and Traveller groups, ex-prisoners, asylum seekers, young African Caribbean men)
- members of some types of communal establishment, for example prisons
- groups where data are collected in a non-standardised manner (and therefore expensive to collate and publish)
- those whose characteristics or circumstances are temporary, making it harder to capture intersectionality with outcomes of interest (for example, the number of pregnant people in employment)
- groups where it is perceived that there is limited political interest in publishing data (for example, numbers of asylum seekers, short-term or female prisoners or people who are homeless)



It was also felt that **statistical data on its own cannot provide a comprehensive understanding of lived experience**. Given the complexity, richness and intersectionality of issues affecting people's lives, those who participated in the consultation activities consistently viewed it as crucial that qualitative analysis is carried out in addition to the production of statistics. This was highlighted as a means for providing fuller and more in-depth understanding of individuals' lived experiences, and participants felt that greater inclusivity could be fostered through developing a more holistic picture of communities through the use of a combination of qualitative and quantitative analyses. It was also reported that, due to the **lack of timely outputs**, many services are reliant on qualitative research and more anecdotal data to understand the needs and circumstances of relevant groups and populations. Such insights are valued by CSOs, though recognised as less robust data for service planning in terms of estimating absolute levels of service need.

Some organisations and individuals who participated in the consultation activities also reported that the **lack of linkages between current administrative datasets** could be a missed opportunity to better understand under-represented groups. This was seen to result from the lack of a common, unique identifier (such as an NHS number) to allow for individual and household comparisons to be made across different datasets, and a "siloes mentality" in government policy and services. In some cases, this was felt to be exacerbated by weak or missing legislation and some organisations expressed that the framework for sharing personal data introduced under the Digital Economy Act was not explicit enough to enable the dataflow into the Office for National Statistics (ONS), for example. Given the resource and skills that exist within ONS there was a call, for ONS to play a stronger role in collating and connecting data in a meaningful, General Data Protection Regulation (GDPR)-compliant manner. It must be recognised, however, that there is resistance to data linkage from those who fear the misuse of data could be used to harm particular groups.

## Lack of resources and/or skills to find and analyse data

"What we find is that, overall, the data exists, it's having the resource capabilities within the organisation to be able to spend the time, and data-science expertise, aggregating that into the specific data models that we'd need, to be able to then use."

CSO working with children and young people

**Dedicated research capacity and the skills needed to source and analyse data** was a common issue raised across the different consultation activities, in particular by CSOs, individuals from relevant groups and populations, and local government participants. Organisations and individuals with more experience

working with data recognised the tension between the desire to have access to more granular data and the implications this has for data users. For example, most CSOs and many smaller local authorities do not have sufficient budgets to be able to employ people with the data analysis skills they need and, therefore, they do not have access to research skills in-house. For this reason, they would prefer summative data split by key socio-demographics and personal characteristics at a sufficiently granular level. This is particularly the case for those working more directly with service users. Participants requested data be provided to enable users to examine the intersection of personal characteristics, allowing the data to be broken into smaller sub-groups (for example, showing religious differences within ethnic groups, age and sex breakdowns across ethnic groups and sexual orientation by disability status).

## Lack of suitable and accessible outputs

“It is just no use saying, ‘BME people are this,’ or whatever. It is just absolutely useless. [You need] to have it broken down in a way that is meaningful for people to identify with... down to as granular a level as possible, so you can actually say, ‘If we do this action, it will have a positive impact for this community.’ Rather than saying, ‘If we do this, it might impact this massive group of people with such different lives.’”

CSO working for race and ethnic equality

“I think sometimes it can be a challenge to know what information the government holds. Particularly when we do Freedom of Information requests or parliamentary questions, for example, data might be released and it doesn’t really tell you exactly what is held and what the scope of the evidence is that’s there, that we could draw on, if that makes sense. I think we just don’t always know what is being collected.”

CSO working with individuals who are pregnant or on maternity leave

In relation to the reporting of equalities data, one of the main points stressed by individuals, groups and organisations who took part in our consultation activities is the **difficulty for relevant groups and populations to relate to data outputs, due to broad categorisations and lack of granularity**, as mentioned in the previous sections. The **lack of readily available and easily accessible data for analysis** was also reported. For example, data presented in traditional data tables were said to be time intensive to convert into a usable format for analysis. While some organisations were seeking raw data available for their own analysis, others required data that was already analysed in particular ways to suit





their needs. For example, organisations who took part in the online consultation described difficulties exploring data which are not already analysed and having to download and analyse raw data. Larger national CSOs, particularly those with an advocacy remit, tended to have a good understanding of the key data available on the groups which they worked with, while smaller and more local CSOs reported greater uncertainty as to exactly what data were available and where and from which organisation.

“We look at it and it can be quite hard to interpret sometimes. Issues that have already been mentioned about you’ve got to read reams of explanatory notes and even then, that might not fully explain things. It’s very technical stuff.”

#### CSO working with older people

The use of **simple language** to help the interpretation of statistics and reassurance that the figures are produced by “trusted sources” were noted as key elements in accessing and feeling confident about information. To ensure inclusivity, participants stressed that **facts and figures should be available in a range of formats** to enable access for diverse audiences. The dissemination of findings in an exclusively online format, whether as data tables, online articles or statistical bulletins, was seen as preventing access for those who are digitally excluded, to whom the data may relate, and who instead have to rely on third parties, such as the media, to disseminate findings by more traditional means.



# What does good practice look like and what are the conditions that enable inclusivity to thrive?

One of the questions the Taskforce considered was about how we can learn from experiences here in the UK and more widely in improving our approach to equalities and inclusion going forward. In this section, we highlight examples of promising practices that have been shared with us by participants in the consultations, or that we are aware of through our own experience. We hope they may help to demonstrate some ways in which our recommendations can be put into practice.



## Building trust through engagement

An important message we heard from participants in our consultations was the need for a trustworthy system, supporting all groups across society to understand both the benefits and the risks of data sharing.

In the UK, the production of government statistics is underpinned by the Statistics Code of Practice requiring statistics producers to think about the 'rights' of those people whose data are being collected, shifting the emphasis from the data to the individuals providing their information. Alongside the Code of Practice, the UK statistics regulator, the Office for Statistics Regulation, has issued guidance on building confidence in the handling and use of data. This urges producers to proactively consider the rights of people providing their data to support the public good. Respecting these rights is crucial to protect and empower citizens within this exchange.

There are numerous good practice tools that statistical producers can draw upon as well. For example, the Open Data Institute: Data Ethics Canvas visually distils the key ethical questions that researchers should consider.

## Answering the right questions

To build an inclusive statistical system, roundtable discussion participants highlighted the importance of having strong foundations. Good practice was identified in how analysts in government see themselves, understanding that their primary role is not to produce data and evidence, but to work with others to identify and answer questions that are important to society. This is a subtle but important distinction. From this foundation it is easier to ask ourselves whether people or experiences are missing in the way in which those questions are answered.

Looking beyond the UK, the notion that statistics add most value when they answer society's questions was echoed in a review of measuring social exclusion by Statistics Canada and the National Institute of Statistics and Geography of Mexico. The researchers highlighted that in developing indicators of social exclusion, it is important to understand and address the main social and policy questions from the start.



## Engaging all communities

The extensive community engagement process that goes into delivering the censuses across the UK, is another example of good practice. Participation is encouraged by creating networks of organisations, community leaders and charities that facilitate direct links to relevant groups and populations. Engagement is also undertaken with members of the public directly to ensure a balanced view of concerns and needs. Developing knowledge and understanding about what matters to people has helped researchers to communicate more effectively about the importance of the data they provide. Working with community experts who can advise on the best channels to communicate with local groups and the language that would most resonate is also very important.

“Knowledge and understanding don’t come from us shouting. It comes from the acknowledgement of the community”

**CSO supporting the Chinese community**

The Government Statistical Service has also published a new User Engagement Strategy, to provide practical guidance on user engagement. The strategy emphasises that: “we don’t need a few people thinking about inclusivity (or user engagement) perfectly, we just need lots of people trying their best to do it well and working together to make it happen as part of business as usual”.

## Partnerships

Partnerships work well when they bring diverse people together. This can challenge us to think differently and set higher expectations. In Colombia, the national statistical office has created a multidisciplinary group to help mainstream an intersectional approach. Members include statisticians, economists, a psychologist, an anthropologist, and other advisers. This group has started developing data disaggregation guidelines and piloted questions to measure sexual orientation and gender diversity in surveys.

## Inclusivity in routine data collection

When undertaking surveys, the standard sampling frame is households. However, it is equally important to consider non-household populations such as residents of communal establishments (for example, care homes, prisons), Gypsy, Roma and Traveller communities, and homeless people. An example of comparative approaches to these issues comes from Australia. Here effort is made to include non-household residents using a list sample of non-private dwellings such as hotels and motels. Similarly the US equivalent of the Labour Force Survey (the ‘Current Population Survey’) also aims to include non-household residents; the stratified sampling frame includes a ‘group quarter’ stratum containing those housing units where residents share common facilities or receive formal care.

## Developing new approaches to inclusive data collection

Statistics Canada’s efforts to be inclusive and timely in providing insights into experiences during the pandemic have been highlighted in a public policy forum looking at innovation and leadership during the pandemic. This showcased how Statistics Canada and the Canadian Chamber of Commerce worked together on the design of the Canadian Survey on Business Conditions. Their approach was to crowdsource data, using email to get people to answer an online survey. This was launched within days of the economic shutdown to collect “real time” data on the pandemic’s impact. Statistics Canada put a special focus on how minority-led businesses were coping, asking business owners to identify the percentage of the business owned by race, ethnicity, sex, whether they are Indigenous and whether they identify as lesbian, gay, bisexual, transgender, queer and/or two-spirited. Patrick Gill of the Canadian Chamber of Commerce said: “The impact of this recession is being felt differently by different demographic groups than the

last recession and the traditional forms of asking questions weren't capturing that story, instead of taking a long time to create its own survey, it moved faster by using crowdsource methodology to get it out and actually worked with the business community on designing the questions that matter most."

Statistics Canada also engaged with disabled people for a crowdsourced survey on how those with disabilities were faring during the pandemic. It worked with organisations including Children First and Vanier Institute of the Family to understand experiences of parenting during the pandemic. Statistics Canada will be developing principles on the appropriate use of crowdsourced data to share promising practice with others.

In the UK, the Government Equalities Office ran the National LGBT survey in 2017 to gather information about the experiences of lesbian, gay, bisexual, and transgender people in the UK. The survey response was unprecedented and over 108,000 people participated. This provided rich insights into experiences of these groups in the areas of safety, health, education, and employment.



## Making the best use of existing data to improve inclusiveness

A good example of the use of administrative data and data linkage for improving inclusiveness is the Scottish Government's health and homelessness in Scotland project. This linked local authority data on homelessness with NHS data on hospital admissions, outpatient visits, prescriptions, drugs misuse and National Records of Scotland information about deaths to explore the relationship between homelessness and health in Scotland. It included transparency around the risk assessment process which enhanced Scottish Government's trustworthiness to those involved in sharing and using the data. They published their data privacy impact assessment alongside the main analysis report, including the original application for the data, how it would serve the public good, details of the application's approval and how others could access the data. This approach is now standard practice for all Scottish Government publications based on linked data.

Another example of using existing data to provide more inclusive insights is ONS's Coronavirus (COVID-19) related deaths by ethnic group, England and Wales. Ethnicity is not recorded on death certificates and therefore to undertake this analysis, deaths involving coronavirus (COVID-19) were linked to the 2011 Census for England and Wales. This allowed ONS to produce statistics on mortality due to COVID-19 by ethnic group, revealing which ethnic groups were at greater risk of dying from COVID-19, and helping to identify groups that have been disproportionately affected by this disease. Linked data have also been used to explore Deaths involving COVID-19 by religious group in England.

## Inclusive dissemination and communication

"I'm sure all of us have been to endless meetings where we give our opinions and the service we help says, 'Thank you very much.' And that's the last you hear of it, and they mainly ignore what we recommended."

**CSO that works with transgender, non-binary and gender-diverse individuals**

Trust and transparency can be improved by more effective feedback. Good practice in communication of how data will be used or why suggestions were not adopted, was highlighted as a way to improve participation in, and inclusivity of, research.

"They're really good at how they conduct this at Essex [Understanding Society, University of Essex] because they send you this newsletter a couple of times a year saying what they've done with the data. And you think you're contributing, and they do some amazing things with the data in terms of looking at the population and that feeds into, I mean it's very academic as you imagine if it's from the University of Essex and that feeds into all sorts of government policies. And then you feel you are contributing to these policy decisions by just filling in this survey once a year online."

**Individual**

This shows the importance of communicating effectively with prospective research participants, by demonstrating how the confidentiality of their data will be maintained and in providing feedback on how their data have been used.

The ZOE COVID-19 symptom tracker app, created by a collaboration of academics, including King's College, London, is another example where engagement is enhanced through feedback to participants. Users of the ZOE app regularly report on their health and symptoms and whether or not they have tested positive for the virus. Participants are provided with updates and alerts which show how their data is shaping the latest coronavirus evidence. The tracker helps provide real time intelligence on the scale of COVID-19 outbreaks and how it affects different demographics and the information is also shared with the National Health Service.



# Our recommendations

**How can we be more inclusive  
in our data?**

## **1. Create an environment of trust and trustworthiness which allows and encourages everyone to count and be counted in UK data and evidence.**

1.1 Trust is crucially important for the collection and use of data and for inclusion in statistics. People are happy to provide their personal information when they believe that (1) their data matters and will be used to improve people's lives and are convinced of the (2) reliability, (3) responsiveness, (4) openness and inclusiveness, (5) integrity and (6) fairness of the data producer. To enhance trust and trustworthiness in the provision and use of data, data producers should develop a social contract with those who provide their data (the respondents). This should include:

- 1.1.1 a clear explanation of why the data are being collected and how they will be used
- 1.1.2 the confidentiality and anonymity of respondents, if and why their information will be shared with third parties and under what circumstances, if any, de-anonymisation might occur
- 1.1.3 the provision of timely, free and accessible feedback to respondents
- 1.1.4 engaging with relevant groups and populations across the whole data process, seeking their advice and support with conceptualisation and planning, data collection, analysis and distribution
- 1.1.5 the public interest should prevail over organisational, political or personal interests at all stages in the production, management and dissemination of official statistics

This will help to address the most important issues for participants of data collection, to ensure that there are demonstrable benefits, and that the risks and costs to participants have been minimised.

- 1.2 Data producers should work together to undertake long-term engagement activities with relevant groups and populations in order to maintain open dialogue and build trustworthiness. This could be achieved through outreach, local-level knowledge building and recognition, reporting costs and benefits of engaging with data collection activities, and learning from previous data collection activities to address the costs and barriers to participation, such as the 2021/22 Censuses. [see also recommendation 1.4].
- 1.3 Data producers should facilitate trust among potential participants and demonstrate their own trustworthiness by increasing diversity among their staff, including those directly collecting data from the public, and by ensuring that participants are all treated with equal respect.
- 1.4 Data producers should undertake appropriate research to identify the practical barriers to participation and implement best practice in data



collection, including ethical considerations, to enhance the inclusiveness of the approaches taken. This might entail providing internet access to address the barriers for digitally excluded groups, and translators for those not fluent in English.

- 1.5 Data producers should ensure that data collection instruments are accessible to all, recognising differences in language, literacy, and the relative accessibility of different modes and formats. For example, using multi-mode surveys as standard practice and implementing additional adjustments to enable the participation of adults and children with a range of disabilities, and those who experience other forms of exclusion, including digital exclusion.
- 1.6 Data producers should avoid the use of proxy responses and ensure that the default approach is for self-reporting of personal characteristics, including, where appropriate, collecting information directly from children.
- 1.7 Practical barriers to the access and use of ensuing data should be investigated, as well as ways of promoting confidence in these data.

## **2. Take a whole system approach, working in partnership with others to improve the inclusiveness of UK data and evidence.**

- 2.1 ONS should establish a clear mechanism and timetable for monitoring and reviewing the recommendations of the Taskforce, reporting on how far they have been implemented and outlining strategies to ensure their implementation going forward.
- 2.2 Data producers should institute a continuing user forum to embed the work of the Taskforce into regular workstreams.
- 2.3 Data producers should engage with academics, user groups and others outside of government with experience of key equalities issues relevant to the UK on an ongoing basis, to share knowledge and approaches to measurement.
- 2.4 ONS should undertake a systematic review of how other National Statistics Offices undertake the collection, analysis and reporting of equalities data. They should work with other countries with promising practices to share knowledge and approaches to measurement and reporting.
- 2.5 Data producers should consider the joint financing of data collection across the UK data infrastructure, to ensure that the costs of addressing data gaps and under-representation are shared and sustainable, and that cost-effective solutions are developed.
- 2.6 ONS should play an active role in international initiatives to improve the inclusivity of statistics, including, but not limited to, the UN Committee on the Rights of Persons with Disabilities and the revision of the UNSD

Guidelines and Principles for the Development of Disability Statistics. To take this forward, ONS should seek the establishment of a UN City Group on Inclusive Statistics.

2.7 ONS and other data producers should share, evaluate and publish effective innovative practices to enable wider learning.

### **3. Ensure that all groups are robustly captured across key areas of life in UK data and review practices regularly.**

3.1 The Inter-Administration Committee and the UK Census and Population Statistics Strategic Group should set up a mechanism to regularly review who is under-represented in UK statistics or data collection exercises, and lead work to address this. This would enable them to respond to changes in coverage, alongside changing social composition, social categories and social understanding.

3.1.1 Particular current priorities are those who are digitally or linguistically excluded, disabled children and parents, victims of intimate partner and domestic violence, particularly older and migrant victims and minors, women experiencing pregnancy and maternity, children in food poverty, residents of communal establishments such as prisons, Immigration Removal Centres, hostels and care homes, the 'hidden homeless', and small groups such as Gypsy, Roma and Traveller communities, scheduled caste and tribe Asian and African groups who, at present, are largely invisible in published statistics.

3.2 Data producers should review the representativeness of key surveys and administrative datasets (initially by benchmarking against the UK 2021/22 Censuses) and take swift action to address identified issues, particularly as they relate to historically under-represented populations (for example young black men) or more marginalised groups (for example children), including but not restricted to those noted in recommendation 3.1.1.

3.2.1 Such a review should ensure that key surveys and datasets are using consistent measures [see also principle 5 on appropriateness and clarity over the concepts and principle 7 on harmonised standards], and that inconsistencies are not distorting comparability.

3.3 Data producers should explore how to improve the collection of administrative data on characteristics that are legally protected in equalities legislation in England, Wales and Scotland with users and relevant government departments. Such as, religion and belief, gender reassignment (gender identity), marriage and civil partnership, maternity and pregnancy, other relevant characteristics such as socio-economic background and migrant status. Additionally, regularly collected (and also legally protected in England, Wales and Scotland) characteristics such as sex, ethnic group and disability status should continue to be

comprehensively and appropriately recorded.

- 3.4 Sex, age and ethnic group should be routinely collected and reported in all administrative data and in-service process data, including statistics collected within health and care settings and by police, courts and prisons. The quality of these data should be regularly reviewed to provide information that better reflects those in contact with these settings.
- 3.5 Where it is impractical or inappropriate to collect characteristics that are legally protected in England, Wales and Scotland in administrative data, or where such data provide insufficient information on groups and populations' experiences, data producers should consider large-scale survey exercises to supplement understanding of these issues.
  - 3.5.1 For example, supplementary data are likely needed to better capture information on sexual orientation, especially among those not of working age and outside of cohabiting couples. Additionally, data on psychological well-being across the age and sex spectrum are required, particularly on the mental health of older men, girls and young women.
- 3.6 Data producers should work in partnership to ensure that UK administrative data sources appropriately reflect relevant characteristics as much as possible [see also principle 5 on appropriateness and clarity over the concepts and principle 7 on harmonised standards]. This includes working to link various administrative datasets, and to survey data where appropriate [see also principle 6 on methods used] to ensure coverage. This will enhance the potential of (linked) administrative data to fill the gaps for relevant characteristics, and their intersections across and within the different countries of the UK. This would also improve the understanding of relevant groups and their experiences over time and across settings.
- 3.7 Data producers should evaluate the coverage of non-private household population groups in UK data and take the necessary action to address those missing from the current data. In particular, ensuring longer-stay residents in care homes, hospitals, and prisons, and the turnover of people between private households and other (or no) residences is reflected.
- 3.8 Data producers should recognise the diverse data needs of different users in the collection of data about specific populations and groups and the intersection between population and group characteristics, and put mechanisms in place to ensure that data collection and reporting serves a variety of user and respondent needs.

#### **4. Improve the UK data infrastructure to enable robust and reliable disaggregation and intersectional analysis across the full range of relevant groups and populations, and at differing levels of geography.**

- 4.1 Data producers should ensure sufficient granularity of data to enable meaningful disaggregation. They should avoid the use of meta-categories which can disguise heterogeneity between groups, within them, and with which people may not identify (for example, White, BAME, LGBTQ+).
- 4.2 ONS and the Cabinet Office should actively promote an intersectional approach to exploring and presenting equalities data across the UK. Potentially misleading single characteristic analyses should be avoided, and 'like for like' analyses controlling, for example, for age, sex, sexual orientation, racialisation, socio-economic background and position<sup>1</sup>, and place be readily accessible.
- 4.3 ONS, National Records of Scotland (NRS) and Northern Ireland Statistics and Research Agency (NISRA) should carry out detailed intersectional analyses from the 2021/22 censuses, to provide granular insights into the nature of disadvantage. To include socio-economic background in the analysis of intersectionality, the analysis should also draw on other authoritative sources such as the Labour Force Survey/Annual Population Survey, since the important measure of socio-economic background was not included in the 2021/22 Censuses.
- 4.4 Producers of existing online tools should adopt intersectional approaches, enabling analysis of different characteristics together (such as ethnicity and religion) to improve understanding of inequalities.
- 4.5 Data producers should use targeted oversampling of under-represented groups as an approach to address specific gaps in knowledge that result from small sample sizes and to facilitate intersectional analyses.
- 4.6 ONS must ensure that the 2023 recommendations on the future social statistics system provides an enduring solution, to meet the full range of inclusivity data needs that are included in the recommendations, including for those groups identified as priorities in 3.1.1.

---

<sup>1</sup> Throughout the document, socio-economic background refers to the socio-economic position of the co-resident parent(s) or carer(s) at the time when the target respondent was growing up, age 14 usually being the illustrative age when collecting such data. It is thus a measure of the socio-economic position of the respondent's family at the time the respondent was growing up, not their current socio-economic position. This is analogous to the distinction between one's current religion and the religion one was brought up in.

## 5. Ensure appropriateness and clarity over the concepts being measured across all data collected.

5.1 Data producers should review the conceptual foundations of their measures for relevant populations and groups, ensuring the measures that are used accurately reflect the current standards and legislation. Data providers should ensure that measures are conceptually robust and do not incorporate formulations that might be deemed to be derogatory, inappropriate or misleading.

5.1.1 As a priority, ONS should transition its measures of disability to approaches more firmly based upon the WHO ICF and ICF-CY biopsychosocial model conceptual frameworks.

5.1.2 ONS should transition its measures of ethnicity and religion so that they better correspond to the current conceptual understandings, reflect the diversity of the population and are recognisable and meaningful to those from specific ethnic and religious groups.

5.1.3 The robustness of measures to capture the experience of populations and groups should be considered. For example, to measure poverty more effectively, ONS should review income equivalisation methods, improve estimates of income poverty and fuel poverty amongst people with disabilities and other affected groups.

5.2 In cooperation with the Devolved Administrations, ONS should develop (and evaluate) a set of measures of socio-economic background that are suitable for collection in administrative datasets and surveys. At a minimum, this should include measures of parental occupation and parental education and be sufficiently granular to capture a range of occupational classes and educational levels, while not being burdensome for respondents.

5.3 Data producers should ensure that survey and question design is based on a clear conceptual understanding of the information that is required, drawing upon best practice to translate this conceptual understanding into accessible and appropriate data collection [see also recommendation 1.5 under principle 1 on trust and trustworthiness].

5.4 Data producers and analysts should ensure that the language used in the collection and reporting of all characteristics is clear. For example, clearly distinguishing between concepts such as sex, gender and gender identity; or ethnic identity and ethnic background. This would help to avoid ambiguity and confusion among respondents and data users, which can undermine data and analytical quality, as well as belief in the validity and reliability of data.

5.5 When sharing or reporting data, data producers should be transparent about how the data have been collected (for example, the questions, modes and mechanisms for providing responses, including clarity around the use of proxy responses). Comprehensive metadata should be published alongside their data, which are accessible to respondents and data users to enable them to assess the quality and suitability of data.

## **6. Broaden the range of methods that are routinely used and create new approaches to understanding experiences across the population of the UK.**

6.1 Data producers should explore opportunities to utilise more varied, innovative and flexible approaches to data collection and combination, where this will be of particular value for enhancing our understanding of the experiences of relevant groups and populations and/or for enabling the inclusion or voices of groups currently under-represented or missing from existing data sources – such as undocumented migrants, those with disabilities, the ‘hidden homeless’, and children. Such approaches will also be relevant for providing more comprehensive information on the characteristics and experiences of those priority groups identified in points 3.1.1 and 3.2.

6.2 A wider range of methods should be considered for capturing those temporary experiences that are not often well recorded – but which may be important for inclusion. These include pregnancy, hospital stays, school exclusions, periods children spend looked after by the local authority, ‘sofa surfing’, and periods in prison or on remand. It also includes experiences that are sensitive and poorly covered for some groups, for example intimate partner violence and other forms of domestic violence, especially as experienced by older women.

6.3 Some of the currently underutilised methods that would provide valuable additional insight include: ethnographic methods to understand lived experiences, field experiments to understand more about discrimination, comparative studies across the UK to examine ‘what works’ in promoting inclusive data collection, linking administrative data to survey data or other administrative sources, and better using and enhancing longitudinal and panel data collection.

6.3.1 For example, for those surveys currently collecting information about children, data producers should consider what information can be collected directly from children themselves, using appropriate instruments and diverse forms of data collection (for example, pictures and diaries), drawing upon best practice in data collection

and ethical approaches, while recognising the potential additional time / burden involved and the privacy needed for children to be able to take part. Data should also be collected to reflect more marginalised children (for example, disabled children, children of prisoners, Gypsy, Roma and Traveller children, looked after children, refugees and unaccompanied migrants) as a priority.

6.4 In all innovations, such as those noted, adequate attention must be paid to issues of consent, trust and trustworthiness [see also principle 1 on trust] and risks of disclosure.

6.4.2 For example, in relation to data linkage, ensuring that the data to be linked, and the linking process, have been demonstrated to be of appropriate quality and accuracy and that safeguards have been put in place to protect respondent privacy, confidentiality and anonymity, in line with existing guidelines such as the UK Statistics Authority Office for Statistics Regulation's Systemic Review Programme on Joining Up Data for Better Statistics. Open communication about the use of data and safeguards is also essential. Linkage and data security should follow best practice and be justified in relation to public interest concerns, while not using safeguards as a means to restrict information about or analysis of populations of interest [see also principle 3 on groups being robustly captured].

## **7. Harmonised standards for relevant groups and populations should be reviewed at least every five years and updated and expanded where necessary, in line with changing social norms and respondent and user needs.**

7.1 Data producers should research user and respondent data requirements and draw on best practice standards and guidelines from other countries and relevant international bodies, to ensure that harmonised standards remain appropriate and relevant.

7.2 Data producers should undertake research into the user and respondent needs for data on groups of interest and provide guidance on how to collect this across different modes, continually reviewing approaches to maintain relevance. This information should then be used to update the existing harmonised standards.

7.3 ONS and others across government and the devolved nations should work together to improve the harmonisation and comparability of data sets across the UK, between regions and over time. They should ensure that the basic data are sufficiently granular in each part of the UK to avoid situations where the only harmonisation that is possible involves unsatisfactory 'lowest common denominator' meta-categories such as 'white' and 'non-white'. This is particularly important since such binary

categorisations are often experienced as pejorative in taking whiteness as the norm, rather than recognising diversity [see also recommendation 4.1 under principle 4 on UK data infrastructure].

7.4 Data producers should use harmonised standards when collecting data, or more granular systems which are compatible with the harmonised standards, to improve comparability and better use existing data.

## **8. Ensure UK data and evidence are equally accessible to all, while protecting the identity and confidentiality of those sharing their data.**

8.1 ONS should work with others to create a centralised, explorable and accessible UK-wide 'one-stop-shop' database of equalities data and analysis.

8.2 Data producers should make administrative data accessible to a wide range of users, including to non-experts. This should include both outputs and non-disclosive 'raw' data to enable alternative analyses. In particular, data collected on residents of communal establishments, such as prisons and care homes, need to be made available in order to improve services and transparency.

8.3 Data producers should develop additional, user-friendly online tools for non-experts to explore existing datasets. Users should not be limited, as in some existing online tools, to pre-set tables provided in advance by the data provider. Rather, they should be able to explore the data so that it can meet their needs, subject to automatic disclosure safeguards, as for example, with Statistics Finland. On the principle of 'generalised reciprocity', where members of the public have provided their data voluntarily as respondents, data providers should not charge other members of the public needing access to the data.

8.4 Data producers should consider language, literacy, format and comprehension when presenting analysis and evidence, in line with the 2018 Accessibility Regulations, and produce accessible websites and outputs for diverse audiences, including the digitally excluded.

8.5 Where relevant administrative data exist that enhance the understanding of inclusion/exclusion, the responsible departments should be required to publish these. Data producers should, as far as possible, adopt an open data model, to help ensure that data are freely available and usable by everyone.



# Supporting evidence

The Office for National Statistics (ONS) engaged with a variety of audiences on behalf of The Inclusive Data Taskforce (IDTF) to understand how data and evidence can be more inclusive, and the specific requirements for this. This included data collection practices, how data are used, how evidence is presented, and where there are gaps in data or evidence that could be filled to improve inclusivity. The Taskforce considered all the evidence gathered in making their final recommendations.

The Centre for Equalities and Inclusion within ONS led, designed and implemented several packages of research for the Taskforce. This aimed to engage members of the general public, equality groups, academics, civil society organisations (CSOs), think tanks, the devolved administrations and central government departments on the topic of inclusivity in data.

In early January 2021, the ONS launched an open online consultation seeking views about the inclusivity of UK data and evidence, including areas for improvement as well as examples of good practice. The online consultation was widely promoted among internal and external stakeholders, as well as members of the public. This consultation closed on the 26th of March 2021.

ONS also held roundtable discussions and in-depth interviews with representatives of devolved administrations, local authorities, central government, academics and learned societies. These groups and interviews ran from January 2021 through to April 2021.

The Taskforce also commissioned Basis Social to conduct focus groups and in-depth interviews with civil society organisations (CSOs) and members of the public with lived experiences of a range of equality issues. This research took place between February 2021 and April 2021.

A paper-based consultation was also conducted as another opportunity to gather the views of the public, particularly those groups at greatest risk of digital exclusion. Capturing the views of these groups was important to the inclusivity of the consultation itself, particularly in light of the widespread use of online data collection methods since the onset of the pandemic. This research took place in April 2021.

An official mailbox was also set up specifically for interested parties to correspond with the Taskforce. All e-mails and attachments were logged, which included responses to the Online Inclusive Data Consultation, and reports and ongoing initiatives relating to the topic of inclusiveness of data and evidence. These were forwarded to members of the Taskforce and considered when drafting their recommendations report. E-mail responses to the Online Inclusive Data Consultation received by the closing date (26 March 2021) were included in analysis and reporting alongside responses submitted online via Citizen Space.

# Contact us

If you would like to get in touch, please email us at [equalities@ons.gov.uk](mailto:equalities@ons.gov.uk)  
Alternatively you can write to us at:

**Government Buildings**  
**Cardiff Road**  
**Newport**  
**South Wales**  
**NP10 8XG**

