Office for
National Statistics

Inclusive data taskforce

Findings from the Online Inclusive Data Consultation

Contents

Research objectives and summary findings

Disclaimer: Within this report, we aim to portray the views of participants and to reflect their words as closely as possible. The findings that are presented therefore reflect the opinions and experiences of a range of individuals and may not be shared by others within the same or other institutions, including the Office for National Statistics. Some quotes have been edited for language and grammar to improve accessibility, without changing the content or meaning.

Research objectives

The UK Statistics Authority's (UKSA) Online Inclusive Data Consultation was open to the public for 12 weeks from 5 January to 26 March 2021. Its purpose was to support the work of the **Inclusive Data Taskforce** in considering how best to ensure that:

> "...our statistics, [analysis and publications] reflect the experiences of everyone in our society so that everyone counts, and is counted, and no one is forgotten." (UKSA strategy - Statistics for the Public Good, 2020).

We consulted to gain views on:

- what was needed to improve the inclusivity of UK data and evidence, such as:
  - where there are data and evidence gaps
  - where data and evidence are currently lacking or partial (regarding topics and quality)
  - where further work is needed
- where to make improvements and what is currently working well

Summary findings

Theme-based analysis of responses generated four main themes, each with two sub-themes, in relation to the inclusivity of data and evidence across the UK. These consisted of accessibility of data, inclusivity of methodological practices, inclusivity of existing data and evidence, and trust, transparency and engagement. Within these themes, there were several issues which were common across both individuals and those responding on behalf of an organisation.

Participants identified some problems with the accessibility of data and evidence needed for research purposes because:

- it was not freely available
- it was not available quickly enough to keep up with current topics
- concise data was difficult to access due to being spread across a variety of statistical organisations and their sources, for example, websites and datasets; users must access each source separately to obtain the required data
- data were identified as not being user friendly, such as being presented unanalysed in Excel, or not designed to support people with visual impairments

Inclusivity of methodological practices were also seen as important to address. The classifications used in survey questionnaires was seen as one of the main challenges for data collection. Relying on quantitative data alone to explore complex inclusivity topics was also raised as an issue. Using qualitative or mixed methods, which would enable deeper understanding of people's experiences and the circumstances that influence representation and inclusion, was suggested.

In terms of the inclusivity of existing data and evidence, specific data gaps were identified. These mainly related to digital poverty, socioeconomic inequality, education and housing inequalities, and disability.
These gaps were said to result in:

- under-representation of groups
- a lack of granularity within the data (particularly for dimensions which overlap)
- an inability to address relevant issues and inequalities

Issues around geographical coverage at the local, national, and international level were outlined, including insufficient coordination and consistency to enable effective comparability between areas.

Finally, trust, transparency and engagement were also considered important by participants. Some participants believe that research agencies lack trustworthiness, and distrust was cited as a major barrier, preventing the participation of certain groups in data collection. **One reason provided for this was a lack of transparency within research processes, for example, not explaining:**

- why data are collected
- how data may be shared
- the process of data collection itself

Previous engagement and consultation activities were appreciated, but several participants felt that further efforts were needed, with a focus on collaboration and follow-up action. Inclusion and representation are considered essential, especially for those often excluded from data collection.

Methods

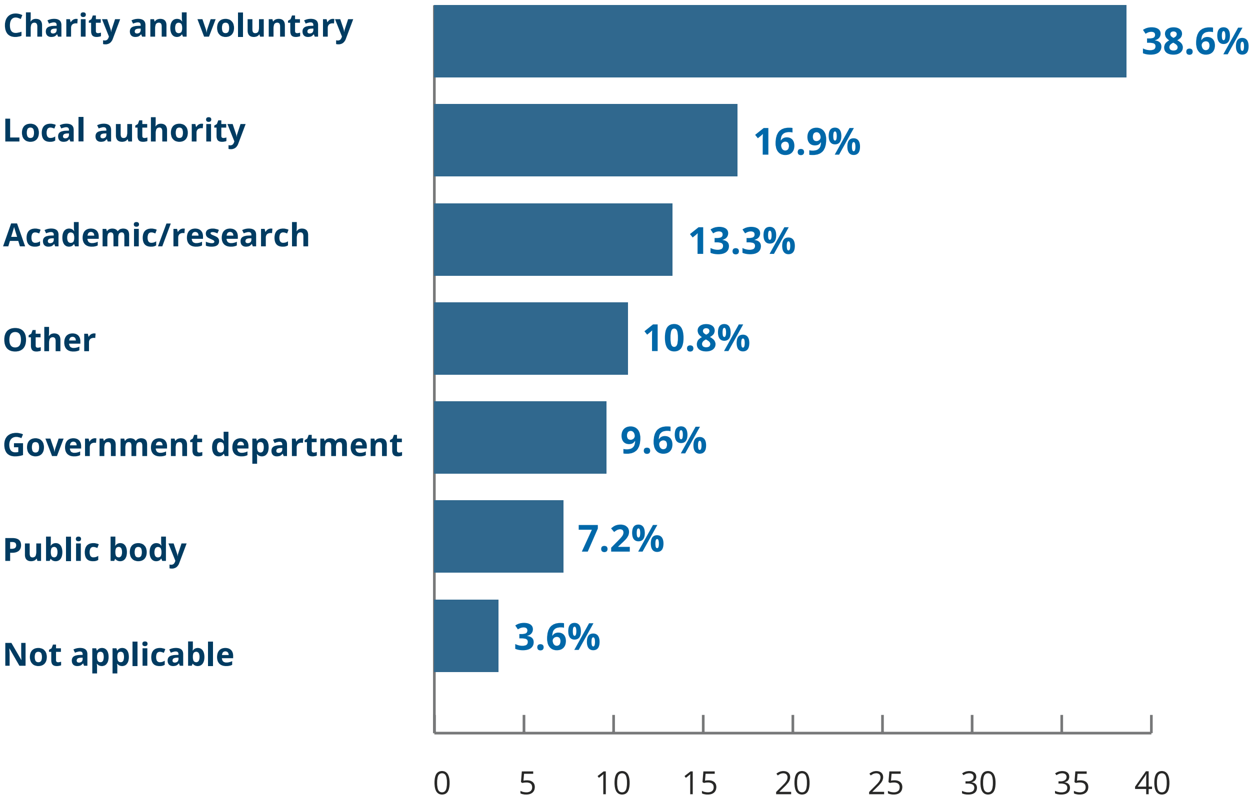
Participants

The consultation received a total of 185 responses:

100 responses (54%) were from individuals
83 responses (45%) were from organisations **(outlined in Figure 1)**
2 responses (1%) did not answer this question.

Figure 1: The charity and voluntary sector made up 38.6% of organisational responses

Percentage of responses to the online consultation by organisation type



Source: UK Statistics Authority Inclusive Data Online Consultation

Research design

The consultation survey was made up of four sections (for a full list of consultation questions see **Annex A**):

1. **“About You” asked about participants and their interest in inclusive data and evidence**
2. **“Current Data or Evidence” collected views on what is currently measured, how it is measured, and in what detail**
3. **“Data and Evidence Accessibility” explored views on how accessible data and evidence currently are and how clearly they are presented**
4. **“Making Improvements” collected suggestions for how improvements could be made and learning from what is already working**

Online responses were gathered, with:

178 (96.2%) responses submitted via the online portal, **Citizen Space**
7 (3.8%) responses submitted via email

Participants were given a range of possible improvements and asked to select which they would like to see implemented to address the issues they had experienced with data and evidence. They could select multiple options. The most popular responses to this question from across both groups (see Figure 1) were:

to “Improve the level of detail available”, with:

62 responses (75%) from organisations
66 responses (66%) from individuals

“Filling data gaps”:

58 responses (70%) from organisations
60 responses (60%) from individual

Responses in the accompanying textbox were included in the theme-based analysis.

Figure 2: Both organisations and individuals prioritised improving the level of data available

Percentage of participants by type of improvement they would like to see to address issues with the data and evidence

---

Improve the level of detail available:

Organisations: 75%
Individuals: 66%

Fill gaps in current data:

Organisations: 70%
Individuals: 60%

Enable greater access to data:

Organisations: 53%
Individuals: 44%

Address problems with measurement:

Organisations: 51%
Individuals: 40%

Improve presentation of the evidence:

Organisations: 40%
Individuals: 33%

Anything else:

Organisations: 13%
Individuals: 18%

Source: UK Statistics Authority Inclusive Data Online Consultation

Approach to analysis

Two approaches were used to analyse the findings:

- qualitative theme-based analysis for textbox responses
- quantitative analysis for closed or tick box responses

The theme-based analysis involved developing inductive codes (or “labels”) for the responses to each question, which were then grouped into common themes and refined as the coding process continued. To ensure reliability and consistency of coding, ongoing discussions took place between coders.

Quotes from participants have been included in this report to **illustrate the findings** in their own words and to **give additional context and clarity.**

Participants were asked if they consented to their name being published. For those who consented, the organisation name is shown next to the quote. Where organisations did not provide consent, either by declining or not answering the question, quotes have been attributed to the organisation’s sector where possible. Names of individuals have not been used, though quotes derived from individual participant responses are labelled as such.

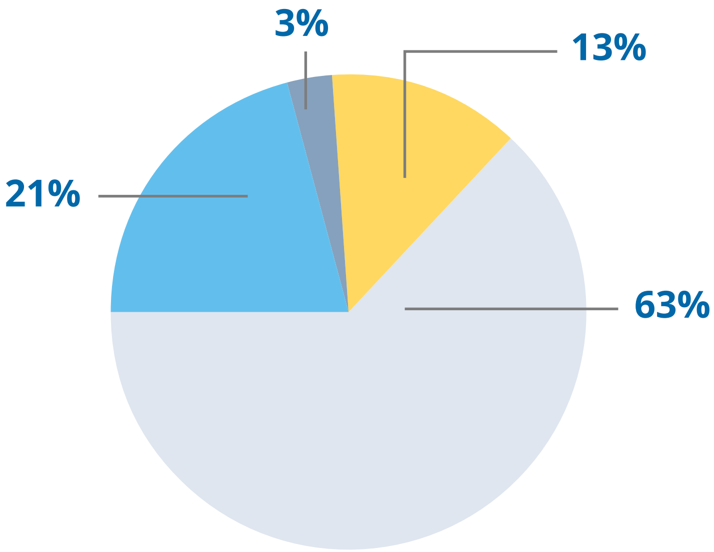
Findings on the accessibility of data and evidence

Data sharing, timeliness, and ease of access

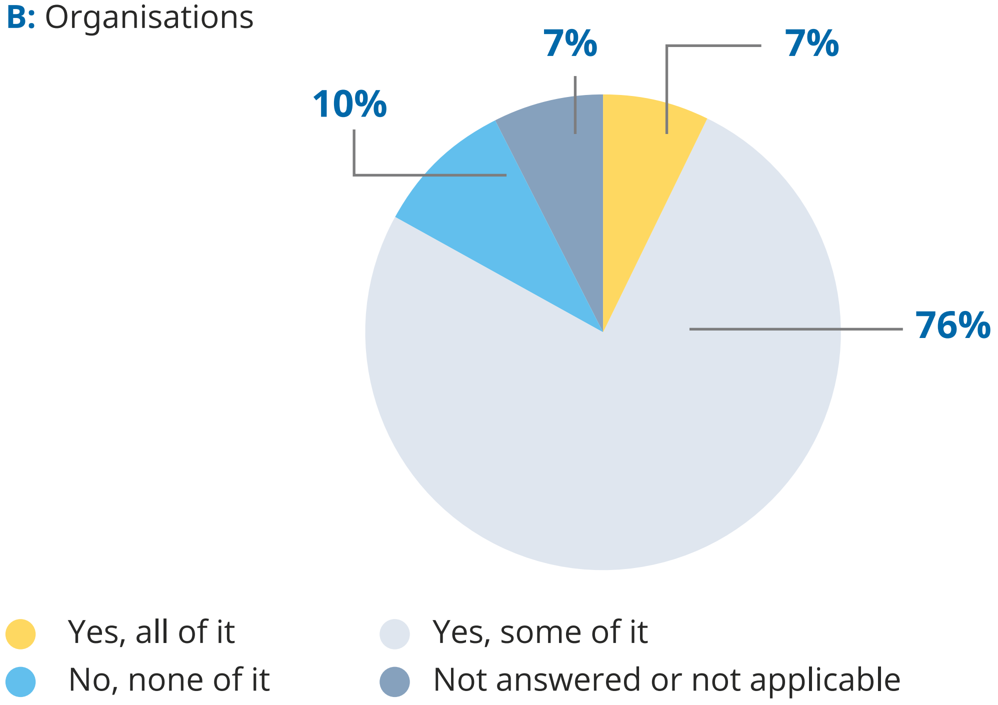
Figure 3: Only 13% of individuals and 7% of organisations reported being able to access all of the data they needed

Percentages of individual and organisational experiences of trying to access the data they needed for their purposes

A: Individuals



B: Organisations



Source: UK Statistics Authority Inclusive Data Online Consultation

Difficulty accessing data was an issue raised particularly by those with research interests in education, young people, and socioeconomic inequalities. Several barriers to data access were identified. Firstly, a lack of co-ordination between different major data providers was raised by both individuals and organisations.

> “Improved access and integration of data, for example the ability to combine and compare data from different government agencies.”
> **(Disabilities Trust).**

> “It would be great if there were a mechanism to allow academics and public bodies to share data and construct complementary analyses.”
> **(Individual).**

Secondly, long procedures and waiting times to gain access to data were described, which may be linked to the challenges noted around lack of collaboration between data producers. Again, this was mentioned by both individuals and organisations, with an individual reporting **“waiting for many months, if at all, for the result”** when trying to gain access to the data they needed.

The application process was described by organisations as a burden, being too long and resource intensive. One participant gave the example of Freedom of Information Requests (FOIs), which they said were

> “very time-consuming and do not always yield clear results”
> **(Refugee and Migrant Children’s Consortium).**

Organisations noted that they needed access to sensitive data more quickly for time-sensitive projects. The National Pupil Database and the Individual Learner Records were cited by an individual as examples of data sources which require an **“extremely lengthy application procedure”,** alongside general data on income for the purposes of socioeconomic inequalities research.

The length of time taken to access data was said to have a negative effect on researchers and interested parties, limiting their ability to analyse and report on data for current and newly emerging topical issues.

This was said to be caused by

> “undoubted imbalance in power between data owners and data users in practice, despite the good intentions of legislation.”
> **(Individual).**

Participants mentioned that to access datasets from the UK Data Service they had to agree to limiting conditions, such as destruction of data after a set time-period. This was seen as particularly difficult for government bodies carrying out statutory functions on a longer-term basis. Access to the Secure Data Service (SDS) requires attending office sites, which can be an issue for organisations and individuals. The inability to access these sites, and so the SDS, was perceived as particularly problematic during COVID-19. Several participants requested more flexible arrangements for accessing secure data.

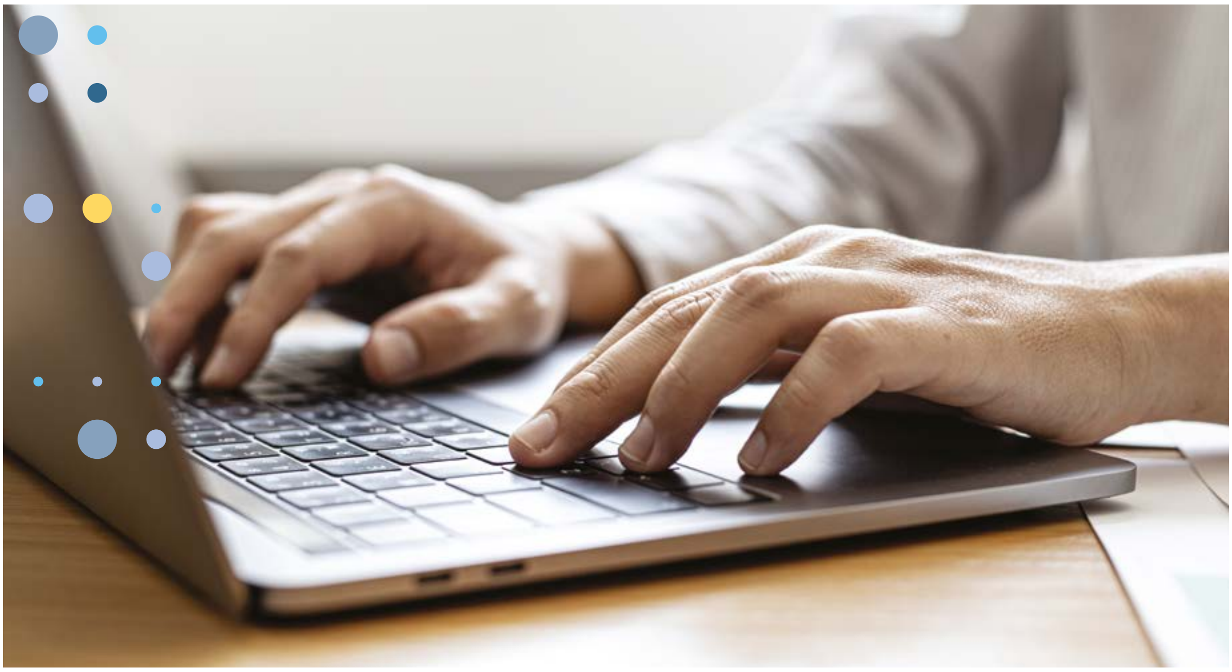
Access to some historical data was described as problematic by several individual participants, as **"huge amounts of historical data have been collected but there is not a clear or comprehensive way to access this data" (Individual).** It was highlighted that data may not have been digitised and therefore are only available **"in hard copy from the individual local authority offices, if at all" (Individual).**

> "Comprehensive air quality data over long time frames. This data almost certainly has been recorded, but it does not appear to be accessible or digitised."
> **(Individual).**

Organisations highlighted the challenges around exploring data which have not been analysed, meaning they had to download and analyse the raw data themselves. Additionally, data presented in traditional data tables as opposed to open data were said to be time intensive to convert into a usable format for analysis.

> "A lot of time is needed to change the format of the data we use before any analysis is performed."
> **(Government Department).**

Linked to the time it takes to access data that have already been collected in the necessary format, participants raised problems caused by lags in data series and the wait for datasets to be updated with available, already collected data. Timeliness of data availability was an area identified by organisations as needing improvement.



> “Where there is a significant lag in reporting, and data are only published on an irregular basis (particularly data for which multiple years are aggregated to boost sample sizes), it would be very beneficial if the lag could be reduced, timeliness improved, and single year data provided.”
> **(Greater Manchester Combined Authority Research Team).**

> “ONS publishes data on energy expenditure for different household types including disabled people. Unfortunately, the most recent period available is from 2018, a two-year lag in representing the current situation.”
>
> **(Scope).**

Requests were made for ONS to prioritise the release of key local data from Census 2021.

> “It takes far too long for local level results to emerge (based on Census 2011) and greater focus on getting these results to users could address some of the issues we have articulated in this response.”
> **(Local Authority).**

“User-friendliness” and inclusivity of access

Lack of “user-friendliness” or usability of the sites or platforms which hold public-facing data were reported as a barrier to data inclusivity. Difficulties with locating datasets and navigating websites were also repeatedly mentioned as barriers to accessibility.

The ONS website was specifically mentioned in relation to these issues, with some participants noting that data release pages are unclear as to which populations they cover.

> “Navigating to data can be very tricky on the [ONS] website. I often have to ‘Google’ the type of dataset I am looking for in order to find it on your website.”
> **(Local authority).**

> “It’s knowing where to find the data sets as a first port of call. The ability to go to one place, that’s reliable and easy to navigate would be a huge step forward.”
> **(Private trading organisation).**

> “We regularly help people navigate the ONS website and other data repositories to help them find the information they need.”
> **(The What Works Centre for Well-being).**

> “Better ONS search facility with intuitive search. Search facility that provides options. [Being] visually impaired, I am not always aware I have a letter wrong, for example.”
> **(Individual).**

Whilst there was only one report of usability issues within the individual responses, this may represent a wider issue to be addressed. The lack of available data to people at risk of digital exclusion was also mentioned. As an individual participant suggested, “there should be research in ALL local archives and libraries.”

Different data formats, sources and multiple platforms were said to make access to data and evidence more problematic. A centralised database or more cohesion and navigation between databases was recommended. It was suggested that an improved, cohesive database such as this could be held by ONS or The UK Data Service.

> “One option for increasing access to official survey data is to make more of it available through the UK Data Service’s Nesstar interactive website, although this can be slow and cumbersome to use. A better option could be to create an API [Application Programming Interface] to restricted survey data that prevents disclosure of sensitive data but allows non-disclosive calculations to be straightforwardly carried out.”
> **(Local authority).**

Mindful of the complexity of bringing all data sources across different providers together, an alternative suggestion was that it would be beneficial to have an “Inclusion” section on the ONS website. This would bring together all ONS data on inclusion and personal characteristics in one place.

Participants were keen for data to be presented in a way that ensures key messages and stories from the data are clearly and easily visible, with 66 participants (35.7%) indicating an interest in improving the presentation of available evidence.

> “The Ethnicity Facts and Figures website presents data in tables and bar charts that make it easy to digest the key messages and examine any disparities. We would like to see more ONS data presented in this way, as some of the issue briefing sheets have begun to do.”
> **(Business in the Community).**

The way data are often presented, as discrete categories, was said to mask any understanding of intersectionality (how various social categories might inter-connect or overlap).

> “Excel data tables can be reductive for their separation of different characteristics, for example presenting “gender” and “ethnicity” side by side, rather than allowing for an overlap of these categories and intersectional analysis.”
> **(Charity organisation).**

There was a call for more interactive data, so that users would be able to examine it from different angles. Several participants felt that a tool that enables users to navigate, aggregate and compare all personal characteristics data would be ideal.

Findings on the inclusivity of methodological practices

Data collection, research design and analysis procedures

Participants reported a general awareness of the impact of the data collection methodology on the inclusivity of data. When data collection is carried out with an effort to include everyone, this was said to help people feel seen, respected, and valued for who they are.

It was seen as i**mportant to take a collaborative and thorough approach when collecting and analysing data, with sensitivity to diversity.** Although the need for standardised data collection was acknowledged, there was concern around **a "one-size-fits-all inclusion model" which may unintentionally reinforce practices that exclude the most under-represented minority communities from "being counted".**

Linked to this, a lack of diversity within the research profession was raised as an issue. For example, the personal identity and characteristics of those designing and carrying out the data collection and analysis was raised as a concern by some groups as was a lack of awareness of issues affecting particular groups.

> "There exists a current lack of queer data competence among those responsible for the design and execution of data practices."
> **(Individual).**

On the other hand, concerns were also raised regarding research conducted by advocacy groups, with some participants asserting that these groups may have political agendas which do not necessarily reflect the views of all the individuals they represent. It was instead advised that research about under-represented groups should directly involve people from those groups, including research carried out by public bodies.

> "Employ a deaf-blind individual, don't presume your data from Sense or Deaf-blind UK defines a community."
> **(Individual).**

For more detail on how this links to trust in those collecting the data and the need for better representation, see the **findings on trust.**

Involving people from affected communities or population groups in all aspects of the research process, referred to as data co-production, was also suggested as a means to improve the appropriateness and relevance of research. For example, this could make data collection methods more accessible and avoid possible confusion and inaccurate recording.

> "It is concerning that we have heard a number of reports that people are confused about how to complete questions relating to disability in the new census, as these are framed using the term ill-health. Many people have stated that they are disabled, and not suffering from ill-health. Developing measures and data collection methods in coproduction with disabled people can help to mitigate such issues and support appropriate and accessible data and research."
>
> **(Scope).**

Participants highlighted the benefits of using qualitative and mixed research methods to provide more context and insight about inclusion and personal characteristics. Additionally, where quantitative analysis is limited due to small sample sizes, several participants suggested that more weight be given to qualitative approaches. This is also described in the section on **extending outreach and representation.**

> "Limitations of survey evidence can sometimes be countered with qualitative research that can explore the rich detail missed in quantitative data collection. For example, when collecting data on other related features of social identity and participation, such as religion, more nuanced information on religiosity and religious practice is not collected, meaning that the entire picture is not captured in survey data."
>
> **(Individual).**

However, it was advised that qualitative approaches should be fully examined for forms of structural exclusion or power dynamics such as unconscious bias in the assumptions and framing of research objectives. This links back to the issues around lack of diversity within the research profession, and the need for better representation and cultural awareness within the social research workforce.

Longitudinal data and data linkage were said to provide a fuller picture of the dynamic nature of inequalities, although wider topic coverage is required in existing data sets to facilitate this.

> “To study inequalities among subgroups of the population (such as women and men, people with disability, ethnic groups, etc.) in a comprehensive and meaningful way, we need data which contain longitudinal information on individuals’ characteristics (such as sex, ethnicity, disability), family background characteristics (parental education, occupation, income), education (qualifications and training), labour market outcomes (employment status, occupations, earnings) and other life outcomes (such as health).”
> **(Individual).**

Participants raised concerns around survey sampling, with solutions to under-representation within sampling frames offered, such as boosting and oversampling.

> “We acknowledge the inherent limitations around population size which lead to the limitations of sample size in national surveys. However, we think that there is a strong case for much more frequent uses of over-sampling and boosts for different ethnic groups, which will give us a richer and more accurate image of the situation across the country.”
> **(Centre for Aging Better).**

It was noted that guidance was needed on how to deal with both over- and under-representation.

> “All survey methods are likely to over represent certain groups and underrepresent others. It would be really useful to have clear guidance from the UK Statistics Authority of how limitations of representativeness can be managed if using online surveys. For example, should a new standard be being set in over representing traditionally under-represented groups?”
> **(Natural England).**

Participants also shared perceptions of exclusionary data collection processes and practices. Some organisations noted that common modes of data collection and survey sampling tend to exclude specific groups. The shift to mixed-method and online approaches to research due to the COVID-19 pandemic were also believed to miss out those who are digitally excluded.

> “If data are collected online only this will exclude older people who are not online, who are disproportionately older and more socioeconomically disadvantaged.”
> **(Age UK).**

> “[Online] data collection modes have lower response rates and may result in increased biases due to non-response, thus the views of specific population groups may be even harder to ascertain. Online surveys, by design, exclude the offline population, so should always be combined with another survey mode, to ensure that the data collected is fully inclusive.”
>
> **(Local authority).**

Participants raised further issues relating to survey sampling and data collection which may result in the exclusion of disadvantaged groups. For example, survey samples often focus on private households, excluding those with different living circumstances, such as people in temporary accommodation, residential care facilities, or without a fixed abode. The Refugee and Migrant Children’s Consortium called for further efforts to address this, as **“counting all children, including undocumented children, is vital.”** Household surveys were also viewed as potentially generating inaccurate data as some people might not wish to disclose aspects of their identities when other membersof their household are present, such as their sexual orientation. Additional sampling concerns were raised around people deemed unable to provide informed consent, such as those lacking mental capacity who are often excluded from surveys. It was advised that further efforts be made to obtain proxy consent wherever possible.

> “All these restrictions can lead to underestimates of the levels of disadvantage, poor health, disability and need for care among older people.”
>
> **(Age UK).**

Participants also mentioned inclusivity issues around data collection for administrative purposes. Administrative data from service use was said to exclude those who do not, or cannot, access these services, rendering them invisible in the data. The voluntary nature of some administrative data response categories was also flagged as problematic by some participants. They noted that when people are given the opportunity to choose whether they disclose demographic information, response levels decline and sample sizes become too small to be used in analysis.

> “This undermines the potential benefits that the information could provide. For example, response rates for data on ethnicity collected for claims for Universal Credit sometimes fall to under 50%.”
>
> **(Charity organisation).**

Question design, categorisation, concepts, and definitions

Individuals and organisations outlined specific issues around survey questions and response options which restrict the collection of accurate, inclusive data.

> "The current Personal Wellbeing measures (ONS4) do not work well for people with learning disabilities, and more validity and cognitive testing is needed to address this."
>
> **(What Works Centre for Wellbeing).**

Problems with question design and response categorisation were said to cause exclusion of certain groups and render others invisible. Specific concerns were raised around the ethnic group categories being too broad, and therefore excluding groups, or limiting their response options. Some organisations were reported to still rely on the 2001 Census classifications, despite this classification framework excluding key ethnic groups.

> "Ethnicity classifications using older Census classifications miss out some groups. COVID-19 analyses of NHS data and Hospital Episode Statistics are based on 2001 Census classifications that do not include Gypsy, Roma and Traveller, and Arab ethnic groups."
>
> **(Government department).**

The use of **"mixed"** and **"other"** ethnic group categories was also highlighted as problematic, particularly with regard to ethnicity and health outcomes. As each category covers a broad range of individuals, there is a risk of masking any similarities and differences, both within and across categories.

> "The tension rests in being unable to fully understand 'who' this category covers and which sections within it are disproportionately affected by COVID-19."
>
> **(Individual).**

> "Review of the continued suitability of the standard ethnic group classifications [is needed], from the starting point of how ethnicity is conceptualised."
>
> **(Centre on the Dynamics of Ethnicity).**

There were calls among some participants to include additional categories within the ethnic group responses, particularly where surveys do not include a religion question.

> “Over the last several years and increasingly over the last year, different services are using ONS ethnic group lists as part of their criteria for funding or planning. Without having an opportunity for Jews to tick a box for Jewish, we become invisible.”
> **(Individual).**

Related to this, a general lack of data on religion was highlighted, with expanded collection of religion data viewed as crucial for understanding specific groups.

> “UKSA and ONS should strongly encourage the collection of data on religion wherever possible, as this is likely to be highly relevant to understanding the needs, preferences and identity of the population being surveyed.”
> **(The Board of Deputies of British Jews).**

With regards to the collection of data on disability, concern was raised about the conflation of disability with ill health within surveys.

> “The basic problem is that disability is currently wrapped up within the definition of the Disability Discrimination Act (DDA). Typically, questions are along the lines of ‘Do you have a physical or mental impairment that impact your capacity to have a normal life for the next 12 months’. What this does is conflate disability with health issues. […] As a disabled person myself, there is a world of difference between someone who is disabled and someone with health issues.”
> **(Individual).**

It was suggested that that use of the word **“impairment”** in tools which collect data on disability should be clarified and used only for describing either disability or ill-health. This also links back to the importance of co-production in question development to ensure relevance and appropriateness, as discussed in **Data collection, research design and analysis procedures.** Organisations called for disability data collection to shift from the medical model, which focuses on an individual’s impairments or differences, to the social model of disability focusing on barriers which may limit their participation in society.

> “Data on disability should also go beyond looking at just functioning and exploring it further to focus on participation or barriers to participation.”
> **(Charity and voluntary organisation).**

Additionally, some participants felt that the definition used in the **Measuring Disability for the Equality Act 2010 Harmonisation Guidance,** often adopted across government surveys, may have led to an increase in disability prevalence.

> "The socio-legal nature of the GSS harmonised definition of disability means that it has been subject to expansion. In order to unpick the effects of changing disability disadvantage from the effects of increased prevalence, a supplementary measure which is based more on a functional definition of disability, and less subject to the effects of changing social norms and values, is needed."
>
> **(Disability at Work).**

The collection of data which separated out whether a person was deaf from whether a person was blind, rather than collecting information on whether they were deaf and blind, was noted by an individual participant.

In addition, they suggested that disability should be further sub-divided into four categories:

> 1. **Physical -** those with physical impairments, including many wheelchair users.
> 2. **Sensory -** those with sensory impairments, for example, deaf, blind.
> 3. **Mental -** those with mental or learning impairments
> 4. **Emotional -** those with emotional or social impairments, for example, **"autism"**
>
> **(Individual).**

Looking at concepts used to measure other protected characteristics groups, some participants noted that definitions of sex, gender and gender identity should also be clearly distinguished. There was concern that these different concepts may be conflated in data collection, impacting the utility and meaning of the data. Some noted that making distinctions between these concepts clearer would better enable monitoring of any inequalities facing those who are cisgender and those who are transgender

> "We are very worried about the trend to conflating sex with gender identity. It is important for us to have data separately about, for example, the employment experience of females or trans women."
>
> **(Fair Play South West).**

There were differing views on the issue of self-identification of individual characteristics. For example, some individuals and organisations raised concerns about self-identification of sex. This was linked to the view that the focus should be on collecting biological sex information (sex at birth). They felt that conflating biological sex with other concepts such as gender or gender identity could lead to potentially inaccurate data, (for example, crime rates by sex), the ability to carry out statistical analyses to answer research questions on sex discrimination (for example, within sport), and an overall inability to **"accurately reflect outcomes for males and females, and for transgender women and transgender men"** as any differences between these groups would be masked.

In contrast to this position, others felt that self-identification is fundamental to collection of data on individual characteristics, and that this should be acknowledged.

> "All data about an individual's identity characteristics is self-identified. The UK Statistics Authority has a role to play in addressing the spread of misinformation about the concept of 'self-identification' and the incorrect view that this only relates to data about trans women and the collection of data on gender and sex."
>
> **(Individual).**

One participant commented on the fluidity of gender and how self-identification was important for inclusivity.

> “Gender is fluid in everyone. We should be breaking down societal gender bias, not reinforcing it.”
> **(Individual).**

In relation to the collection of sex data, there was also concern about the lack of harmonised measures for certain characteristics, especially variations of sex characteristics (sometimes referred to as intersex). The lack of standard data collection approaches contributed to the view that these populations may be overlooked. This is further discussed in **Findings on the inclusivity of existing data and evidence.** Linked to this was the idea that trust and engagement in data collection may also be impacted when participants are not provided the opportunity to select response options which align to their identity.

Another issue raised with measurement of personal characteristics relates to age definitions and categorisations. Some described the inconsistent grouping of age bands across organisations as negatively impacting the utility of the data as it obscures differences between age groups and results in difficulties producing relevant and effective analyses. This was noted as particularly affecting data for children and young people, due to inconsistent definitions and large age groupings.

> “We are often unable to answer questions in a satisfactory way due to the ways in which current data are presented. Notably in inappropriate age bandings. This relates to a general inconsistency in definitions of a ‘child’ or ‘young person’ within agencies and organisations. [...] young people are grouped into large bands (such as 0-18 or 16-24), which obscures nuances between different age groups. We recommend all data are reported in quinary age bands (0-4, 5-9, 10-14, etc.).”
> **(Association for Young People’s Health).**

Finally, the limited scope of **“Protected Characteristics”** was another issue identified raised in relation to classifications. Many groups who may also experience discrimination, exclusion, or greater risk of disadvantage are not included in **the Equality Act (2010)** and therefore may neither be protected by this equality legislation in Great Britain, nor be a significant focus of research on inequalities. Examples of this included socioeconomic status, immigration, employment, and housing status.

> “This may mean that data about them is not collected when ‘protected characteristics’ set the boundaries of societally acceptable and unacceptable discrimination.”
> **(Charity organisation).**

Findings on the inclusivity of existing data and evidence

Gaps in data and evidence

Figure 4: The most common data concern was about data gaps

Percentage of participants giving each reason for being unable to answer their questions, by type of participant

Gaps in the current data:

Organisations: **70%**
Individuals: **61%**

The level of detail available:

Organisations: **64%**
Individuals: **55%**

Problems with measurement:

Organisations: **45%**
Individuals: **40%**

Something else:

Organisations: **30%**
Individuals: **25%**

Source: UK Statistics Authority Inclusive Data Online Consultation

Further analysis of the textbox responses identified specific gaps in data and evidence relating to broad topic areas. The implications of this were reported to include hindering ability to create effective policy, to use statistics and data to improve society, and to understand the people of the UK in greater detail.

Specific gaps around equalities data were highlighted by participants including:

- digital poverty and health
- child asylum seekers and immigrants
- socioeconomic status and food poverty
- health, education, and housing inequalities
- sexual orientation and gender diversity
- carers, and those with learning disabilities.

> “The inability to address important questions from policy and practice on the state of wellbeing of some populations, especially people with learning disabilities, LGBT+, and unpaid carers”, meant that “as a result… the full extent of wellbeing inequalities in the UK is not understood and may be underestimated”. This could consequently “lead to a widening of inequalities, and the design of policies that leaves some populations behind.”
> **(The What Works Centre for Wellbeing).**

There was a perception among some participants that gaps in data relating to certain personal characteristics exist because they are **“not prioritised or mandated” (Individual).**

Participants also suggested that further data be collected around the specific circumstances of populations at greater risk of disadvantage to gain greater insight into their lives and experiences. Examples of this focused on the living and accommodation conditions of disabled people; recipients of social care; and progression and attainment post-16 years of age, particularly for those in vocational courses or apprenticeships, including course opportunities and entry requirements.

Other topics participants noted where it would be helpful to collect further detail of people’s life experiences included: redundancies, pupil exclusion (especially relating to Gypsy, Roma, and Traveller populations), prejudice, bullying, poverty, hospitalisation, mortality, prevalence of Covid-19, and prosecutions and convictions. It was suggested that affirmative steps needed to be taken to ensure better representation of under-represented communities.

> “Making sure people or groups are not excluded would demonstrate to those groups and others that many other structural conditions are ethical and fair.”
> **(Equalities Hub).**

> “I am primarily interested in promoting accurate representation of those who have previously had their relationship to sex, gender and sexuality overlooked by UK data collection exercises. These people tend to identify as part of the LGBTI+ community. It is hard to answer this question when the data is not available to think of what COULD be answered (i.e. how do you predict unknown unknowns?).”
> **(Individual).**

The impact outlined of not having sufficient data and evidence around certain groups, as noted by individuals, was twofold.

Firstly, a lack of data and evidence was said to lead to invisibility and under-representation of specific groups and their experiences. It was highlighted that this lack of awareness and understanding of certain populations can lead to many questions going unanswered.

This was said to make it difficult to understand discrimination based on personal characteristics and whether some groups have been systematically disadvantaged. In addition, the inability to draw robust conclusions, as well as to identify where bias exists, can yield misleading results.

Secondly, where discrimination and systematic disadvantage cannot be identified due to data gaps, this hinders ability to formulate effective responses.

> “Young people who do not achieve expected or benchmark levels at GCSE are least likely to make good progress in education and employment post-16. Understanding their situation is key to reducing educational inequalities but at the moment they are quite invisible while policy and data have focused on higher achievement and widening access to university.”
>
> **(Individual).**

> “Getting to the bottom of these evidence gaps is critical for understanding how to make more progress in addressing discrimination and of profound importance for the framing in law and policy of protected characteristics.”
>
> **(Individual).**

> “Due to the lack of data collection on sexual orientation and gender identity in adults, it is impossible to assess inequalities in this population and to assess whether the inequalities in individual research papers are playing out at a population level. This is leading to these issues being ignored in policy and in-service implementation.”
>
> **(Individual).**

Different organisations shared a common view that the populations they work with are not **“nationally represented”** or **“reliably identified”** due to data gaps. They saw this as hindering their ability to understand and meet the needs of these groups. Examples included those working with young offenders, those dealing with young homeless people who are **“not recorded anywhere in official statistics”**, migrant children, asylum seekers, and people working with children excluded from school.

> In the case of migrant children, lack of data availability "makes it difficult to answer questions about children and young people's well-being and outcomes"
> **(The Refugee and Migrant Children's Consortium).**

Data gaps in this context mean that comparisons between migrant children and non-migrant children are not possible. Organisations described how **gaps in data affected their ability to inform policy development and evaluate programmes and interventions.**

> "Without robust data, it is not possible to adequately plan for and allocate the necessary resources to ensure that government programs achieve their objectives."
> **(Charity organisation).**

> "Additional measurement and assessment are needed to capture hidden forms of homelessness and provide greater detail and disaggregated data [...] Due to the difference in methodologies and reporting, there can no comparison in data or evaluation of the impact of strategies and programs."
> (**The Institute of Global Homelessness).**

More, and better, intersectional data were described as crucial for understanding inequalities so that any inter-connections between social categories can be identified.

> "It should be possible to interrogate data by multiple characteristics more readily, for example, to understand the experience of Muslim women or African-Caribbean heritage men."
> **(Individual).**

Data gaps relating to **intersectionality impact a large range of communities and individuals,** including those of older ages.

> "We are often not able to answer research questions regarding intersectional issues, for example how older people of particular minority ethnicities are impacted by an issue. This can mean that we are not fully able to describe inequalities between different groups of older people."
> **(Age UK).**

Further examples of the need for more intersectional data in certain areas include science, technology, engineering, and mathematics sectors, to understand how diverse groups of women's career trajectories differ from those of men. Also, better health data is required to understand susceptibility to certain conditions across both age and ethnicity.

Geographical coverage and data comparability

Issues relating to geographical coverage and granularity of data were commonly reported **with over half of participants (54.1%)** indicating it had impacted their ability to answer their questions. The geographical level of available data was the most frequently raised area requiring improvement, with **129 participants (69.7%)** noting this.

Issues raised in the textbox responses also included a lack of comparable data at the local, national, and international level. This reportedly made sub-national analysis difficult as and statistics broken down to a more detailed level challenging.

The importance of timely data at the local level, with sufficient demographic information to understand the experiences of different population groups, was another recurring theme across participants. Participants noted these data would enable organisations to identify and address equality issues, and to inform policy decision-making and service provision. However, they reported that analysis of existing datasets may not provide sufficient demographic information at lower level geographies due to statistical disclosure control and small sample sizes

> "There is a trade-off between granularity of data in terms of demographic characteristics and granularity of geography."
>
> **(Local authority).**

Organisations highlighted a range of different areas where data gaps exist owing to a lack of data at lower level geographies.

Examples given included: personal characteristics, digital inclusion, household income, tenure, homelessness, transport, the labour market, the environment, migration, finance, wellbeing, and crime. More local data was also deemed as essential to allow better service planning in areas with higher-than-average national rates of mental ill health and substance abuse.

> "We would argue that there is a strong public interest and duty argument in being able to understand such areas at the local level, so we can track change over time, understand how policy choices and investment might be playing out, and compare to other places."
>
> **(Local authority).**

> "Reliance on national level figures may underestimate the extent of mental ill health in urban contexts such as Lambeth and Southwark in London, home to hugely diverse communities."
>
> **(Local authority).**

A range of examples were also given to demonstrate the need for, and value of, geographically comparable data.

Comparability was seen as vital to measure service performance across areas, identifying what is working best, and where improvements can be made.

> “To look at how issues vary within constituencies and how figures for their constituencies compare to those for other constituencies to learn from areas with figures that show they are doing well” or to “compare approaches to police interaction with children across the UK in order to identify prevalence (or not) of practices that breach children’s rights, particularly in relation to certain groups of children for example Black or Gypsy, Roma and Traveller children.”
> **(Just for Kids Law).**

Some local authority participants highlighted that being able to compare city regions is **fundamental to understanding relative performance levels in certain areas.**

> “It is also important to be able to compare the city region to national and regional averages for benchmarking purposes.”
> **(Greater Manchester Combined Authority).**

Additionally, some noted the importance of being able to compare data between urban and non-urban areas in order to understand the social and economic differences faced by residents in these areas.

Organisations discussed how, **without comparable small-area data, they are unable to understand trends at a “neighbourhood level”,** making it more difficult to consider the needs of different groups in that area.

> “This may mean that areas which are performing less well are masked by the overall local authority average, and the experiences of different population groups are masked by the overall experience.”
> **(Combined authority).**

The lack of data at a local level was viewed as impacting both specific groups and communities in general, particularly where data gaps have affected planning and resource allocation decisions The Strategic Housing and Regeneration Service (SHARP) commented on the number of negative impacts they had experienced owing **"to a lack of comparable data or evidence".** For example, they had to use datasets from **"less robust"** sources, such as data purchased from Rightmove, to answer questions because Land Registry data could not provide what they needed. They felt this **"can lead to less informed decision making".**

Another organisation emphasised that

> "comprehensive evidence at as local a level as possible is paramount"
> **(Local authority).**

Christian Aid also detailed the importance of comparable local data across all geographic levels for the work they do to support communities and local and national governments.

They described using a **"bottom-up approach by understanding the communities first before aggregating to local and national levels."**

Further challenges were noted in comparing data across the UK. A range of causes and impacts of issues around geographical comparability were offered.

Firstly, participants noted reduced comparability due to variations in how data are collected and reported as well as differences in the wider legislative contexts for equalities.

> “There are difficulties comparing data across the UK’s four nations. Data is often collected and reported differently, with slightly different definitions used and questions asked that reflect different research and policy contexts in each of the settings.”
> **(Association for Young People’s Health).**

> “The equalities guidance used to collect the data, differs between England, Scotland and Wales.”
> **(Anonymous organisation).**

Secondly, geographical boundaries are defined differently by different institutions and agencies.

> “Differences in the reporting of the exact ‘locality’ of each place can make comparing data from different agencies difficult (for example, social care statistics against healthcare statistics).”
> **(Association for Young People’s Health).**

Thirdly, combining England and Wales within certain statistics (for example, healthcare) means that differences between the two cannot be easily distinguished.

> “Data being available at a Wales level is extremely important as there can be major differences in the experiences of minority and marginal groups between England and Wales thereby making England and Wales combined data far less useful.”
> **(Government Department).**

There were calls from both individuals and organisations for better coordination and collaboration to improve comparability of data evidence across the countries of the UK.

> “While we need to see the data broken down by devolved administrations, we also need to talk the same language as each other so that we can share best practices; we are not even at that stage yet in digital poverty, but ONS can help by providing comparable broken-down data.”
> **(Individual).**

Included in this were requests for more coordinated efforts across the countries of the UK and their statistical offices to collect and link comparable longitudinal data. This would better enable the study of inequalities across different groups and different geographical areas of the country. Some noted this would aid evaluation of whether certain policies are more effective than others.

> “It is essential to collect comparable data across the UK, since policies are different in the devolved administrations, in particular in education, health and justice, and we need to know how these different policies can impact on people’s life chances and outcomes, especially in the case of groups with protected characteristics”
> **(Individual).**

> “We need to be able to make comparisons across the whole country, so need all 4 nations to be applying the same rigorous data definitions.”
> **(The LGB Alliance).**

Concerns were also raised about how the delay to the 2021 Census in Scotland would affect comparability with data from the censuses in the other countries of the UK. As the censuses in England, Wales and Northern Ireland happened amidst the pandemic, the context of people’s lives may be quite different to the situation in 2022 when the census will be held in Scotland. Additionally, it may also affect data quality due to difficulties linked to migration within the UK:

> “this will create future problems where individuals could be missed (or double counted) if moving to different areas of the UK.”
> **(Individual).**

It was suggested that in the absence of comparable data across regions, people need to find ways of assessing comparability across the UK on a case by case basis.

> “It would be lovely to have totally comparable data across regions and nation states, but that is not realistic. Instead, what is needed is excellent documentation so that individual researchers can better assess how comparable data sets are. At present we need to do this as individuals, which is inefficient.”
> **(Individual)**

Findings on trust, transparency, and engagement

Extending Outreach and representation

Inclusion and representation were described as essential by participants, especially for those regularly excluded from data collection. Individuals reported that effective engagement which incorporates a wide range of views is critical.

> "Ensure that you are consulting widely with ALL parties, listening carefully to the arguments, references, evidence and concerns."
> **(Individual).**

Some participants noted a perceived lack of engagement with certain groups, particularly those at higher risk of disadvantage, which they felt resulted in their **under-representation in data.** They felt this could also produce ill-informed conclusions.

For example, a participant noted that predictions around the result of the 2016 United Kingdom European Union membership referendum may have been incorrect because certain groups were under-represented in the data due to living in circumstances not covered by traditional data collection methods, such as those with irregular migration status.

> "This renders participation in research and data exercises almost completely inaccessible."
> **(Individual).**

Participants noted that **further efforts should be made to reach under-represented groups in data and evidence,** with the onus on those collecting the data.

> "We need to stop thinking of populations as 'hard to reach' but instead as 'seldom reached' or 'routinely excluded from data collection'."
> **(Individual).**

Organisations gave several examples of groups who may be left out of traditional data collection pathways including:

- the “hidden homeless”, such as sofa surfers
- those living in precarious circumstances who make efforts to hide away or keep moving due to risk of violence
- people who find alternative, potentially unsafe, arrangements to avoid homelessness, for example “engaging in so-called ‘sex for rent’”
- people who do not access support or services needed due to fears of their personal situations worsening, such as being left homeless or having a child removed due to accessing support for domestic abuse

This links to the **inclusivity of data collection procedures** discussed under the methodological practices theme.

Current efforts to improve representation through wider engagement and consultation were widely appreciated by participants though organisations particularly suggested a greater focus is needed on collaboration and subsequent action. This included calls for more frequent consultation with organisations (such as those advocating for equalities) and the public. More effective inclusion of under-represented groups, whose voices may not be heard and whose experiences may differ from others, was also suggested as was collaboration with organisations with relevant expertise. However, engagement and consultation were viewed as only the first steps, with subsequent action required to make real progress.

> “This consultation, and the broader work of the Inclusive Data Taskforce [should] lead to substantive action to address the lack of inclusive data in Scotland and across the UK.”
> **(Close the Gap).**

> “We acknowledge the efforts the ONS has made since 2019 to improve how we look at disability, including outcomes for education, employment, housing, crime, well-being, loneliness and social participation. The [Disability] Trust, however, would appreciate extending this data to capture and present data on hidden disabilities, including the behavioural, cognitive, and emotional consequences of brain injury. This may be facilitated by linking up with key organisations to facilitate collection of data from individuals with disabilities living in communal settings, and to ensure their views and experiences are included.”
> **(Disability Trust).**

Some participants also felt that contact with ONS and other government representatives could be improved, as attempts to reach out were not always met with a response. They perceived this as a barrier to sharing knowledge and expertise, as well as limiting opportunities for external quality assurance.

> “RMCC welcomes working collaboratively with researchers and other organisations to support and inform participatory research and co-production work.”
>
> **(Refugee and Migrant Children’s Consortium).**

Some individual participants also wanted better communication, promotion, and visibility of data and evidence. They felt it would be helpful to raise awareness of what is available, as well as clearly outlining existing evidence gaps. Some participants suggested how communications could be improved, such as by ensuring inquiries are addressed in a timely fashion and that more outreach work is done

> “Newsletters could be created by the ONS and sent out to Local Authorities to keep us aware of changes, new datasets or reports and other changes.”
>
> **(Local Authority).**

Not all participants felt that outreach would be helpful, with one individual noting a general

> “level of apathy to statistics, even if they convey a damning message”
>
> **(Individual).**

Another participant suggested that more qualitative data, including ethnographic research into different demographic groups and cultures, be used and published alongside figures for **“humanising these stats”.** This was offered as a way to increase public engagement and stakeholder buy-in.

Trust in data and data collection

Organisation and individual participants felt that some people and communities were less likely to engage in data collection activities due to a lack of trust around this interaction and concerns about potential consequences. These factors were thought to have a negative impact on the integrity and useability of the data. As noted earlier, some participants perceived a

> “widespread issue of trust when trying to obtain reliable data”
> **(Individual)**

from under-represented minority ethnic groups, particularly those from the Gypsy, Roma and Traveller, and migrant communities.

Both individuals and organisations reported that a lack of trust can sometimes stem from fears that engagement in data collection processes will worsen participants’ current situations, such as the risk of detention to those without documentation.

A charity organisation stated that fear of the use of immigration status data by agencies may dissuade victims from reporting violent crimes. The result of this mistrust may mean that some communities, who have specific needs and circumstances, are regularly omitted from survey data.

Engaging in data collection was also viewed by some participants as potentially harmful, especially when people are presented with response options which they feel do not represent them. To avoid this, participants highlighted the importance of **ensuring that questions and responses are inclusive and enable individuals to answer in a way that best reflects their identity.**

> “Data collection can trigger isolation and significant alienation, such as dysphoria in intersex and non-binary people who are forced to report themselves as having a binary sex (as happened to intersex and non-binary people aged under 16 in the 2021 Census). Therefore, what is important is to enable and support those who are collecting, using and sharing data to properly examine such considerations as part of their data collection and service design processes.”
>
> **(Charity organisation).**

This also links to diversity within the research community, with some participants noting that trust in research may be improved if researchers themselves were more like those involved in the research.

> “The act of data collection about people with protected characteristics by those without them can itself be seen as a form of disenfranchisement, disempowerment and exploitation, which leads to some communities becoming unwilling to take part in data collection, or distrustful of data-driven applications.”
>
> **(Individual).**

Collaboration with communities throughout the research process was seen as crucial to develop trusting relationships. Perceptions of independence and impartiality of the data collector were –also viewed as important to public trust. People may perceive data collection to be influenced by political factors, particularly where the topics are the subject of ongoing political debate.

Some participants also noted that that the ways statistics are presented can be misleading and called for more information to be shared about the data, and the level of uncertainty associated with it.

> “Both the ONS and by Government departments publish statistics that may well be misleading, but with limited warnings that this may be the case. For example, the area profiles on Nomis (an ONS service providing access to official UK labour market statistics) often provide figures with a lot of uncertainty around them, but they don’t flag this is the case. Another example is the data published in the Ethnicity Facts and Figures on employment by ethnicity, which again does not acknowledge the level of uncertainty.”
> **(House of Commons Library).**

An academic and research sector organisation noted that the media and politicians, may fuel this public distrust towards government and the data it collects in seeking to discredit the policies of their opponents. Both individuals and organisations also suggested that a lack of transparency from government around why data are needed and collected, and how data are used, may prompt further distrust in data collection and outputs. It was suggested that transparency could be improved in areas such as the aims of data collection, the data collection design process, and the supporting information to explain the published statistics.

A lack of understanding around reasons for collecting data and how it is used may undermine the trust of research participants, particularly when positive action does not seem to result from the data collection.

In relation to holding and sharing data, some participants advised that all organisations should be encouraged to carry out Equalities Impact Assessments (EIAs) and make them publicly available to increase public trust in data collection and statistics. To build trust in data collectors and public willingness to share their data, participants also noted that people need to have a strong sense of confidence in the way their data will be used, that positive outcomes will result, and that action is taken to reduce any risk of harm.

> “Address the harmful data sharing practices between immigration enforcement and other public services which deter individuals from using services as well as participating in data gathering exercises.”
> **(Refugee and Migrant Children’s Consortium).**

Annex A: Questions asked on the Inclusive Data Online Consultation survey

Section 1: About you

This section asks about you and your interest in inclusive data and evidence

1. What is your name?

- Feedback type: textbox

2. What is your email address?

- Feedback type: textbox

3. Are you answering on behalf of an organisation or as an individual?

- Feedback type: radio button, single selection
- Possible answers:
  - Yes
  - No

4. If you are answering on behalf of an organisation, what sector do you work in? This will assist us in monitoring the range of people that have responded to this survey

- Feedback type: radio button, single selection and textbox for “other”
- Possible answers:
  - Government department
  - Public body (for example health, transport, emergency services)
  - Local authority
  - Housing
  - Academic or research
  - Charity and voluntary
  - Commercial
  - Utility
  - Journalist or media
  - Other (please specify below)
  - Not applicable

5. **What is the name of the organisation that you represent?**

- Feedback type: textbox

6. **We may wish to contact you in relation to your response or to invite you to attend follow-up events. Would you be happy for us to do so?**

- Feedback type: radio button, single selection
- Possible answers:
  - Yes
  - No

7. **What is your main area of interest or your broad area of research, in terms of topic and groups that you're interested in?**

- Feedback type: textbox

8. **To support transparency in our decision-making process, responses to this consultation will be made public. This will include the name of the responding organisation or individual. Please confirm that you are content for your name to be published. We won't publish personal contact details.**

- Feedback type: radio buttons, single selection
- Possible answers:
  - Yes, I consent to my name being published with my response
  - No, please remove my name before publishing my response

Section 2: Current Data or Evidence

This section focuses on current data and evidence, what we measure, how we measure it and in what detail

9. **Are there any questions you are currently unable to answer because of a lack of data or evidence? If there are any, please tell us what they are.**

- Feedback type: textbox

10. **Please tell us the reasons why you are unable to answer these questions. (Please provide further details below)**

- Feedback type: checkbox, multiple selection, and textbox for "something else"
- Possible answers:
  - Gaps in the current data?
  - Problems with measurement
  - The level of detail available?
  - Something else?

Section 3: Data and Evidence Accessibility

This section focuses on how accessible data and evidence are and how clearly they are presented.

11. **Are you currently able to access the data you need for your purposes? If you are not able to access all the data or evidence you need for your purposes, what data are you unable to access and what are the barriers to you accessing this data?**

- Feedback type: radio button, single selection and textbox for “Yes, some of it” and “No, none of it”
- Possible answers:
  - Yes, all of it
  - Yes, some of it
  - No, none of it

12. **Are there any issues with how the data or evidence that you currently rely on are presented? If so, please provide details**

- Feedback type: textbox

Section 4: Making Improvements

This section is about how we can make improvements and learn from what is already working

13. **Thinking of all the issues you may have experienced with the data and evidence, which of the following improvements would you like to see? Please provide details.**

- Feedback type: checkbox, multiple selection, and textbox for “Anything else”
- Possible answers:
  - Fill gaps in the current data?
  - Address problems with measurements?
  - Improve the level of detail available
  - Enable greater access to data?
  - Improve presentation of the evidence?
  - Anything else?

14. Please tell us about how important it is for your purposes that data or evidence are comparable across different geographies, for example, across the 4 countries of the UK, internationally or at a more local level? Please give details of what geographies you would like to be able to compare across.

- Feedback type: textbox

15. Please tell us about any impacts you've experienced due to a lack of comparable data or evidence.

- Feedback type: textbox

16. What change to the current data or evidence would you most like to see to be able to answer the questions that are most relevant to you?

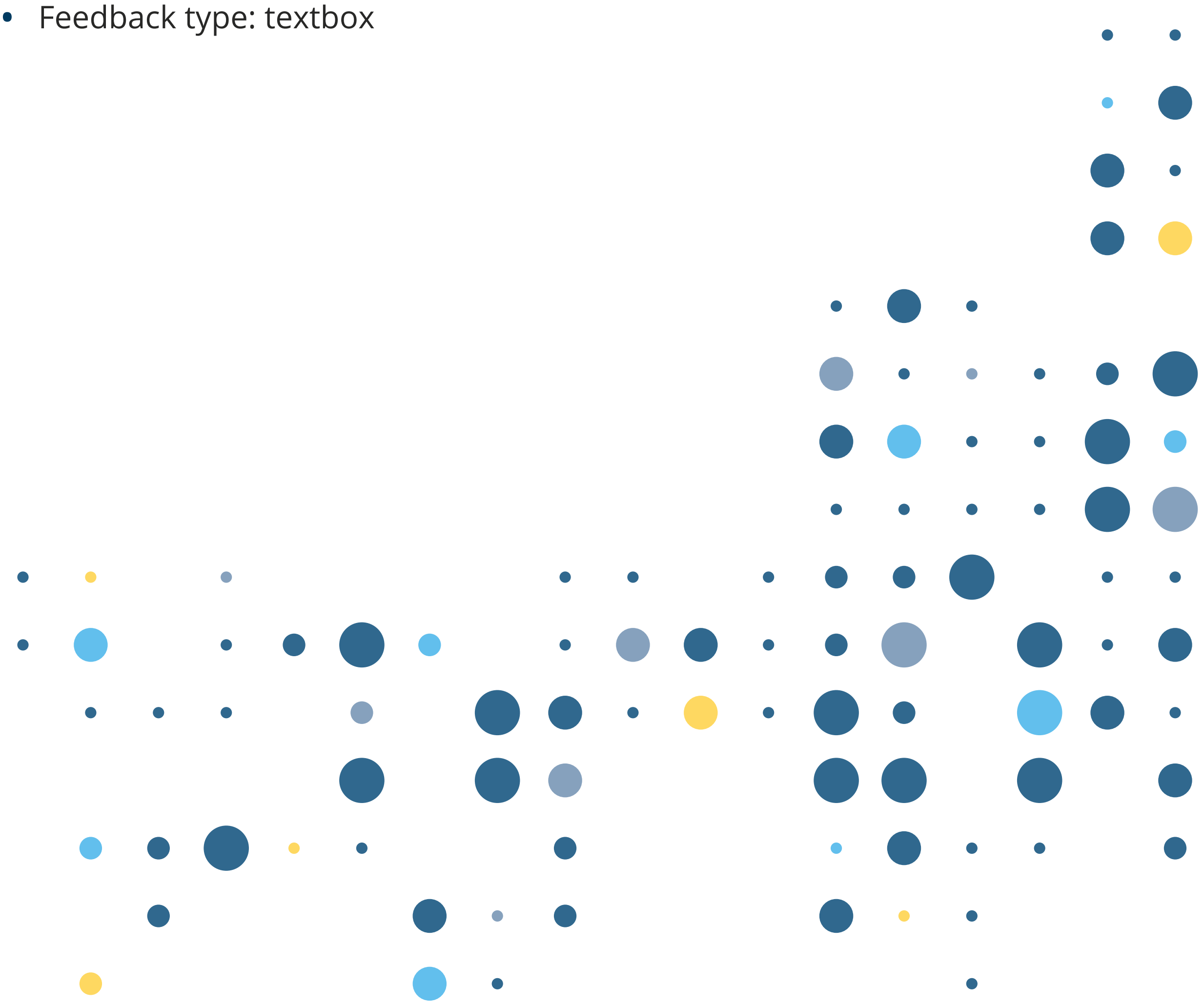
- Feedback type: textbox

17. Please tell us about any examples of inclusive data and evidence that you think work well. If relevant, please include a link.

- Feedback type: textbox

18. If there is anything else you would like to add that hasn't already been covered, please share your views here.

- Feedback type: textbox



Contact us

If you would like to get in touch, please email us at **equalities@ons.gov.uk**
Alternatively you can write to us at:

Government Buildings
Cardiff Road
Newport
South Wales
NP10 8XG