Office for
National Statistics

Inclusive data taskforce

Findings from Roundtables and in-depth interviews with the devolved administrations, central and local government, and academic stakeholders

Table of Contents

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.

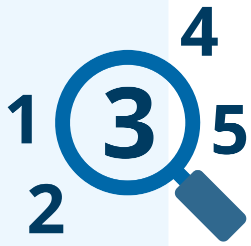
Research objectives and summary findings

Research objectives

The Office for National Statistics (ONS) held roundtable discussions and in-depth interviews with senior representatives from the devolved administrations, central government, local government, and academic and learned society organisations.

The roundtables focussed on:

- understanding gaps in the available equalities data
- whether and why certain groups are not included, recognised, or represented fully in the data and the potential impacts of this under-representation
- existing initiatives to address data gaps
- the required steps to improve inclusivity



Summary findings

Six lenses for viewing inclusivity were used for reporting findings which include engagement, concepts, methods, data, insights and learning from best practice. This section contains brief summaries of the findings under each of these lenses:

Participants advised that work is needed to improve trust in government and how personal data are collected and shared. The need for transparency in both data sharing and usage, as well as visible action resulting from data collected, were identified as areas for improvement. Utilising long-term engagement strategies with under-represented communities to maintain dialogue and allow time to build trust was seen as an important step to address these concerns.

Concepts used to collect and categorise data were discussed within a range of contexts. Issues with definitions and classifications were raised. These included perceived outdated and inconsistent use of definitions, which can result in a lack of comparability. Additionally, there may be inadequate representation within response options, so research participants must select options with which they do not identify. To address these concerns, participants suggested conducting appropriate reviews of concept definitions to keep up with the changing nature of society, as well as the development and consistent use of harmonised standards reflecting society today.

Insufficient sample sizes were said to lead poor granularity, hindering intersectional analyses around under-represented groups and preventing breakdowns by personal characteristics and local level geographies. A need for more inclusive data collection methods was highlighted, to reach under-represented groups. These issues were perceived as leading to specific groups being invisible within the data. Oversampling of under-represented groups was suggested, as well as developing strategies to reach people who are routinely excluded from data collection pathways.

Several issues with data were identified, including time lags, data gaps, access to existing data and a lack of personal characteristics data within administrative datasets, which were said to inhibit effective policy decision-making. Improving access to the data through user-friendly platforms was recommended, as well as improving the profile and usability of admin data.

A lack of insight into the needs and issues facing specific communities was said to hinder inequalities from being addressed. The way that data and statistics are presented was also said to undermine public understanding and access to the information. Qualitative research to gain insight into the lived experience of under-represented groups was proposed to better understand the priorities and needs of different population groups. It was also advised that exploration into how research findings are understood and how dissemination can best facilitate understanding is needed.

Use of best practice examples to inform data collection, analysis and dissemination practices in UK data was perceived as suboptimal. This was said to be a result of a lack of data sharing and poor communication across government and organisations. Establishing a cross-organisation statistics user group was recommended to improve statistical awareness and the use of statistics in the UK, particularly regarding equalities data.

Methods

Participants

Roundtable discussions were held with the devolved administrations, local government, university-based social research groups, economic research institutes, research funding organisations and learned societies. Alongside these, in-depth interviews were held with central government departments and learned society participants. These enabled deeper exploration of participants' individual views and facilitated better accommodating their availability.

Full details of the research participants are provided in Table 1. Participants reflected a range of research and policy areas and were selected based on the equalities work undertaken in their area. For the roundtable discussions, we aimed to achieve a sample of 5 to 6 participants per roundtable, to maximise the opportunity for participating stakeholders to contribute to discussions. However, this number was not met in some cases due to a lack of participant availability, or exceeded in others due to higher interest in participation in certain areas.

Table 1: Participating Organisations and Sample

Organisation type	Organisation name	Number of participants
Devolved administrations	Welsh Government	6 (Roundtable)
	Scottish Government	8 (Roundtable)
	Northern Ireland Executive Office Human Rights Commission	5 (Roundtable)
Central government	Government Equalities Office	1 (In depth interview)
	Equality and Human Rights Commission	1 (In depth interview)
	Race Disparity Unit	1 (In depth interview)
	Policy Lab	1 (In depth interview)
	Cross-Government Disability Unit	2 (In depth interview)
	Department for Work and Pensions	1 (In depth interview)
Local government	London Boroughs	9 (Roundtable)
	Combined Authorities 1	4 (Roundtable)
	Combined Authorities 2	5 (Roundtable)
	Non-metropolitan Local Authorities	7 (Roundtable)
Academic and learned society organisations	Social Research Group	5 (Roundtable)
	Economic Research Institutes	4 (Roundtable)
	Research Funding Organisations	4 (Roundtable)
	Learned Society 1	1 (In depth interview)
	Learned Society 2	2 (In depth interview)
	Learned Society 3	3 (Roundtable)

Research design

Roundtables lasted approximately 90 minutes and in-depth interviews lasted approximately 45 minutes, each following a semi-structured topic guide.

The topic guides remained fairly consistent throughout, however, they were adapted as data collection progressed to further explore particular areas and to better integrate a department's specific area of focus into the discussion. For example, local government topic guides were tailored to address specific issues in local area service provision, while central government topic guides were tailored towards policy decision-making. An example topic guide is shown in Annex A.

Key areas for discussion included:

Data sources

Which sources are most and least useful for addressing equality issues, an are there any inclusivity issues and concerns with these sources?

Data gaps

What are the key gaps in equalities data, what are the implications of these gaps for public policy decision-making and how can they be addressed?

Research and survey design

What are some of the key inclusivity issues relating to research and survey design, particularly regarding under-represented groups?

Harmonisation and coherence

What impact can the consistency of data and definitions have on the inclusiveness of data, particularly for under-represented groups; and how might harmonisation and statistical coherence be improved?

Engagement

What are the barriers to engaging with under-represented groups, and how can we address these to ensure everyone in society feels represented in data, analysis and outputs?

Outputs

How can data accessibility be improved for everyone in society, including the digitally excluded and under-represented groups?

Approach to analysis

Six lenses for viewing inclusivity were used as an analytical framework to code the verbatim transcriptions of interviews and focus groups. The coding framework was checked by a second analyst and summarised under each of the lenses for reporting.

The six lenses are:

- **engagement** with groups to ensure that everyone in society feels represented in data collection, analysis and outputs. Ways to improve the trustworthiness of, and confidence in, these processes and the people involved
- **concepts** and the extent to which the definitions that are used in data collection, analysis and outputs are harmonised, comparable, and aligned with current social ideas and identities
- **methods** of data collection and analysis, such as sample inclusivity and representativeness, and efforts to reach people who are routinely excluded from data collection
- **data** availability, including quality, gaps, timeliness, granularity, and the extent to which the available data facilitate intersectional analyses and meet user needs
- **insights** that may be generated through data and consultation, including how the findings are presented and shared, interpreted, and reflect lived experiences and needs
- **best practice** examples of inclusivity and factors which enable and encourage best practices to develop and be more widely adopted

Detailed findings from the roundtable discussions and in-depth interviews are presented in the following section under the six lenses for viewing inclusivity, including the key issues and potential solutions that were outlined by participants. Data were collected and analysed in adherence to Government Social Research Professional Guidance, following the principles for ethical best practice.

Findings on the engagement of society with data and evidence

Trust and trustworthiness

A recurrent theme relating to engagement, identified across all groups, was public mistrust of government, and therefore of government statistics, particularly described relating to under-represented groups. Past experiences were described as contributing to this.

> "People can be quite suspicious of local government and local authority… willingness to share data is just a sort of embodiment of that."
>
> **(Non-metropolitan local authority participant).**

> "Trust is a huge one. We, at one point, tried to do some ethnographic research with the Gypsy, Roma, Traveller community and I think that it really didn't land very well, because we represented government to those groups and we may be moved too fast… guess people sometimes might not have had very good experiences with government, or with services, and so then just to expect people to be happy to share information, or data, isn't always going to land very well."
>
> **(Central government participant).**

Participants suggested reasons for the UK's lack of success in gaining public trust, including referring to previous government policies that were perceived as discriminatory against certain groups, which may have resulted in hesitance to share personal data.

> "There are things happening in this country which perhaps are serving to reduce trust or create a lot of sensitivities in some of the research we're doing, particularly among migrant groups who are confronting what has been formerly described as a 'hostile policy environment', and some of them have no recourse to public funds and so on."
>
> **(Academic participant).**

> “There is a mistrust of government traditionally, which encourages people not to engage and not to give out personal information.”
> **(Northern Ireland Executive participant).**

Trust in government was linked to perceptions around how data is used and shared by the organisations responsible for collecting data, with views that transparency of these practices could be improved.

Participants discussed how certain communities feel their data have been used to discriminate against them, and that government is not interested in helping and supporting them. It was highlighted by a research funding organisation participant that certain under-represented groups have been hesitant to engage in data collection exercises historically due to **“negative consequences for themselves, or just a general sense that these things don’t make a difference”**. Additionally, it was mentioned that government is thought to not prioritise using existing data for certain under-represented groups, which was said to result in individuals questioning government’s data use.

> “If the government has the data, why aren’t they reporting on it and why are they leaving it to external bodies to basically point out that things aren’t as rosy as they might seem on a surface read of a government statistical release?”
> **(Central government participant).**

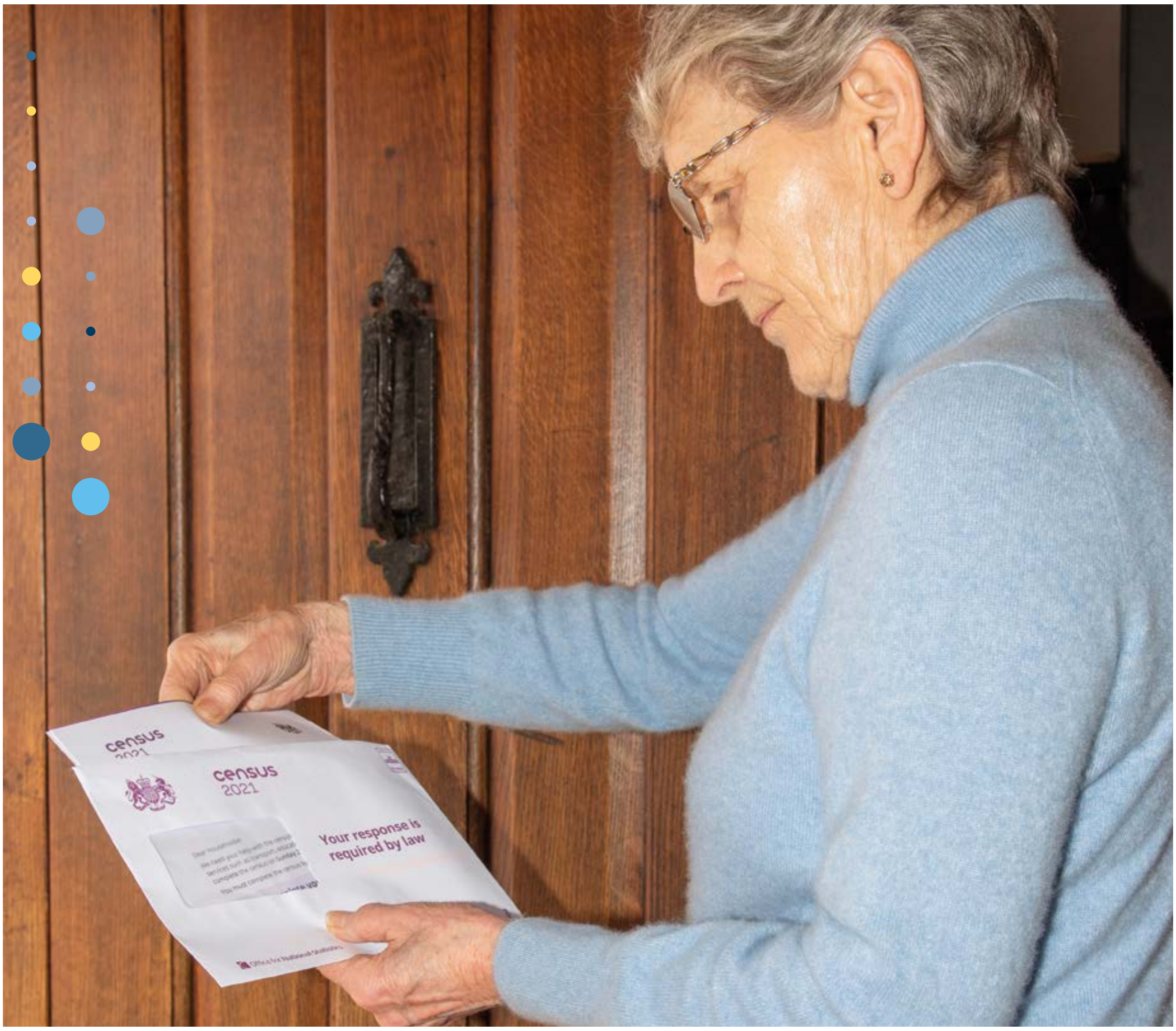
Academic participants also suggested that a lack of community ownership of data and large amounts of data being held by a few organisations or individuals has contributed to concerns among some groups about the influence these data may have had over their community.

Transparency

Public views about data sharing throughout government were highlighted by academic participants as a specific challenge to engagement. The census was provided as an example of how the purpose of data collection may be interpreted differently by different population groups.

> “[There is] scope for misinformation to fly around that actually if you’re filling in the census form, you’re telling the tax office where you are.”
>
> **(Academic participant).**

It was advised that further efforts on the part of data producers towards transparency, including publicising exactly how data would (and equally would not) be shared and used were necessary for these organisations to improve their trustworthiness. Academic participants suggested that Scandinavian countries have tackled similar issues by making data and linkage widely accessible, including the public availability of everyone’s salaries.



Effective engagement

Other issues raised around engagement related to under-representation, risk of invisibility of certain groups, and important issues affecting certain communities remaining unknown and therefore not being addressed in local authorities' policy decision-making.

> "We don't know what we don't know about these issues, and we then aren't able to ask the right questions... from the very outset."
> **(London borough participant).**

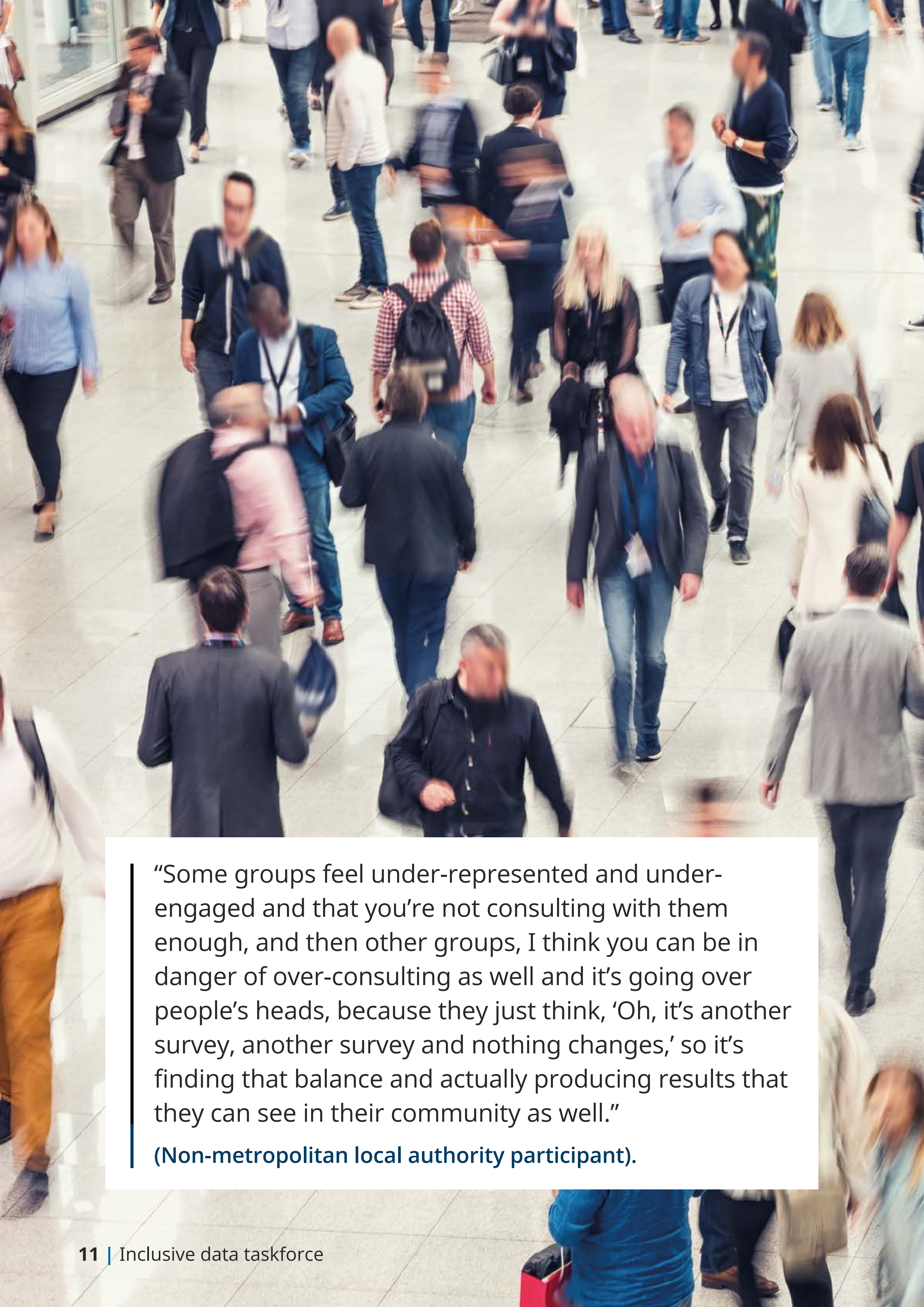
Effective engagement with under-represented communities was said to require substantial consideration and resourcing.

> "How do you find people in the first place when they're not appearing in a national sampling frame, and, or you don't have access to their personal details to contact them because of the legislation around it?"
> **(Academic participant).**

However, attempts to improve engagement with under-represented populations were said to be met with a lack of funding to undertake the necessary measures. An academic participant stated that **"we know what to do, it's just there's not enough money in it... to really scale things up."** Academic participants also called for government to think of new and innovative ideas for engaging with people to inform policy decision-making.

> "There are people out there who do not want to take part in surveys, they're not minded to and consequently, you need to be thinking about other ways of collecting information about them if it's necessary for policy or understanding."
> **(Academic participant).**

Training researchers from under-represented communities to undertake research was suggested by academic participants as a strategy to improve community ownership and representativeness. However, it was highlighted that past attempts to engage and consult with under-represented groups could result in over-consultation. This meant that researchers continue reaching out to the same groups that have already been accessed and continue to neglect the less well represented groups. This was said to be particularly problematic when community engagement in consultation does not result in tangible action.



> “Some groups feel under-represented and under-engaged and that you’re not consulting with them enough, and then other groups, I think you can be in danger of over-consulting as well and it’s going over people’s heads, because they just think, ‘Oh, it’s another survey, another survey and nothing changes,’ so it’s finding that balance and actually producing results that they can see in their community as well.”
>
> **(Non-metropolitan local authority participant).**

Suggestions for addressing barriers to engagement

Several participants suggested the importance of making efforts to improve perceptions around data sharing. For example, a participant from the Northern Ireland Executive mentioned that **"a wider, governmental campaign, linked into community and voluntary grass roots organisations"** could help to promote the importance of collecting data from under-represented communities, alongside best practice examples of evidence-led policy.

> "Having that initial conversation about why it's important and finding ways to build trust, I think is a precursor to then people feeling they're represented."
>
> **(Welsh Government participant).**

Using television adverts to demonstrate how personal data is used to benefit local communities was suggested by a Welsh Government participant, to improve knowledge and trust levels among under-represented groups, and to **"win people over and make them understand the value of providing that information and that they feel safe to do so."**

A need for visible, tangible benefits from sharing information, as an academic participant noted that research participants **"have to think that there's some point in them giving this information."**

> "If people can see the actual direct result of the information they're giving us on budgets relating to specific service changes, demonstrating that that engagement has actually done something for their community, people are far more likely to respond. So it's also communicating that work to then further that engagement in future."
>
> **(Non-metropolitan local authority participant).**

Being completely transparent about the use of public data and providing support to understand how information will be used was thought to be important. In particular, ensuring transparency around data collection practices and how personal data will be used to inform policy decisions. A central government participant proposed that, particularly when engaging with under-represented groups, doing it in **"a really transparent, really no punches pulled, honest way"** should be considered. Alongside the need for improvements in transparency, there were also calls for acknowledgement of the problematic misuse of data in the past.

> "[There is a need to] actually show how they are actively moving away from that, and what all of the different mechanisms are in place for that."
>
> **(Research funding organisation participant).**

Certain participants suggested utilising relationships with **trusted community groups who have greater understanding of the nuances of under-represented groups** to help gain access and build trust.

Using community representatives was seen as a better option than using **"smartly dressed people turning up with clipboards"** and could build trust and confidence and increase engagement. Conversely, participants mentioned that government logos might deter communities from participating. A successful example was provided by a Northern Ireland Executive participant of engaging with Irish Traveller communities, which involved **"feet on the ground, working with relatively local, trusted groups and individuals within certain extended families."** One of the combined authority participants also mentioned their previous collaboration with the charity Stonewall which enabled them to gain access and engage with a somewhat hidden population (business owners who identified as gay) and to build their trust, resulting in a successful research project. Taking a **"citizen science" approach was proposed by academic participants, which could involve funding community groups to undertake research in their own communities to address issues of interest to them."** approach was proposed by academic participants, which could involve funding community groups to undertake research in their own communities to address issues of interest to them.

Utilising long-term engagement strategies with under-represented communities to facilitate dialogue and allow time to build trust and ensure their voices are heard.

> "The collaborative efforts that are being done to work with communities who own their own data, and who are able to lead and influence how it's collected, used and analysed is really important."
> **(Research funding organisation participant).**

An example provided by academic participants was using approaches to co-produce research, engaging with communities on the subjects that matter most to them. This was described as essential for communities to be able to define themselves, rather than researchers doing this on their behalf without their input. However, a learned society participant stressed the importance of ensuring that co-production with communities is meaningful, undertaken in the correct language and timely, as well as **“getting groups and organisations associated in the design of the programme at the early stage.”**

Several participants highlighted the need for greater inclusion of under-represented groups within the research community. Participants felt this could help to gain trust when undertaking research and ensure that the research group is more diversified, which could help to break down some barriers. An academic participant suggested that this approach could result in a **“better understanding of, for example, a different culture, a different discipline or different ways of people thinking.”** Rather than using formal education to train researchers to conduct research in these communities, participants suggested upskilling individuals who are already part of an under-represented community into a research role.

> “[To create] communities of action as well, that then take through the policy and the ownership of actions that come off the data collection.”
>
> **(Research funding organisation participant).**

Another suggestion was to involve members of key relevant population groups on the boards of academic institutes, to improve direct communications with these groups, and encourage participation in engagement activities including public lectures.

Incorporating a welcoming tone to engagement and providing financial incentives, particularly for under-represented groups, could support participation in research. Although academic participants noted that this is often rejected and deemed inappropriate by ethics committees, **“thank you payments”** were proposed as a means to ensure that research participants are renumerated for their time, as, **“it’s well known that the more you incentivise it, the more you get participants.”** However, a note of caution was also added by an academic participant that **“there’s a trade off because if you incentivise it too much, you change the behaviour,”** and potentially risk the integrity of the research findings.

Findings on the appropriateness and inclusivity of concepts in data

Survey response options

Participants discussed the concepts and definitions used to collect and categorise data, within a range of contexts.

Concerns were raised around the problematic wording of survey questions and response options, which could lead to community members feeling excluded from data collection and potentially yield inaccurate results as respondents must select a response option which they do not identify with. An example was provided of surveys where gender questions only have male and female options.

> “So right from the outset, we’re not including people within our surveys. They’re made to feel like they don’t fall into one of those two categories. You’re immediately creating a not very inclusive environment and making people feel like they don’t belong as part of that survey.”
>
> **(Welsh Government participant).**

A Northern Ireland Executive participant mentioned that some individuals cannot accurately reflect their nationality in labour market statistics, for example, because they are unable to identify as dual nationals. **“You can be British or Irish, or both. You’re not allowed to answer both.”** It was noted that concepts, or the differences between concepts, may change over time, for example with sex and gender, or race and ethnicity. As people’s understanding changes this could lead to inaccurate data collection, or inaccurate reporting by respondents, it could also prevent people from responding in a way that reflects their identity or become a barrier to them responding at all.

A research funding organisation participant noted that recognition of the fluid nature of identities is not currently at the forefront of survey categorisations, and there was said to be a significant time lag between being able to **“identify yourself in the way that works for you, versus what is currently being rolled out.”** Local government participants also highlighted issues relating to the harmonisation of data definitions at a local level, and the importance of local data reflecting the evolution of definitions to represent changes in society. Academic participants reported the need to balance keeping up with society in terms of definitions and classifications with the potential loss of comparability over time if classifications and concepts are frequently revised.

> "You want to respect people's experience while on the other hand you need some level of consistency."
> **(Academic participant).**

Harmonisation and comparability

There were calls among academic participants for guidance to be put in place around the revision of harmonised definitions so this it is made explicit.

> "There is a challenge, do we stick with the standard definitions for some foreseeable future or do we kind of have some kind of properly thought through revisions protocol process, which says, we'll stick with these for a period and then here will be the way in which we may update them or not."
> **(Academic participant).**

Participants from London boroughs also stressed the importance of using tailored definitions in local surveys, so that the diversity and distinctiveness of each London borough can be better reflected in the data. It was mentioned that specific populations could not be identified for policy purposes.

> "We were unable to identify South-Eastern Asian communities within London and... in certain parts of North London they're very concerned that they're unable to identify some of the Cypriot communities."
> **(London borough participant).**

Academic participants discussed problems with the comparability of ethnicity across the Devolved Administration Censuses.

> "The census varies between England, Wales, Scotland, Northern Ireland, say, for ethnicity, with the smaller number of categories in Northern Ireland, meaning that if you want to do a UK analysis, you end up having to reduce the groups down to the categories you used in Northern Ireland across the UK [to be] compatible."
> **(Academic participant).**

Classifications and categorisations

Participants discussed the evolving classifications that are used to capture ethnicity in society, reflected that the 2001 Census categories for ethnicities are still widely used in data collection and that these do not adequately reflect the characteristics of under-represented groups today.

> "Ethnicity categorisation, we're still using the 2001 census. They [the health system] haven't updated it [ethnicity categorisation] in over a decade, rendering entire ethnic groups invisible... because they said they couldn't afford it. Well, that's a matter of prioritisation."
>
> **(Central government participant).**

The use of broad meta-categories for ethnic grouping, such as **"BAME"** (Black, Asian and Minority Ethnic), was also criticised as a term with which nobody identifies. The term was also seen to marginalise certain groups, making them **"invisible"** in statistics and aggregating groups which have very different characteristics, thus providing little meaningful insight. It was highlighted that such categorisation results in ethnic groups being **"lumped together"** irrespective of their differences.

> "The attitudes could be very different depending on whether you are Muslim, Jew, Sikh, Hindu or Christian.... the way ethnicities currently are grouped together... people do not see themselves in the data and that is a vital point of being inclusive."
>
> **(Learned society participant).**

Concerns were also raised about the perceived inconsistent use of the **"other"** category within data collection processes.

> "An APS [Annual Population Survey] 'other' is quite wide when you're looking at ethnicity, whereas in other cases, in other datasets, 'other' is just everything not listed above and usually it's a very long list."
>
> **(Combined authority participant).**

Within ethnicity data in particular, it was noted that there has been a substantial increase in individuals selecting **"other"** or **"mixed category"** within surveys.

> "That data is essentially useless at that point, if we're looking at particular answers to certain questions."
>
> **(Research funding organisation participant).**

Overall, **"other"** categories were described as problematic as they provide no understanding of an individual's identity, especially if they are from mixed ethnicity backgrounds.

Although issues were highlighted with ethnicity definitions and the need for classifications to keep pace with society, some participants promoted the potential for granularity with ethnicity classifications, and felt that this could be a good approach for disability to follow.

> "[For disability] we probably need to move closer towards our ethnicity classifications, where you have…multiple different levels of granularity."
>
> **(Central government participant).**

Additionally, some participants highlighted a lack of standardised recording of disability across the UK, with the vast number of different definitions and classifications causing significant issues for data analysis. A central government participant said that mixed usage of Government Statistical Service (GSS) harmonised standards could result in fragmented knowledge of what disability data is available **"which can cause difficulties when you're trying to compare data between datasets."** Welsh Government participants discussed how the social model of disability is used in Wales, which considers the personal barriers that people face. This therefore creates a need to **"capture information in a slightly different way, that would be inconsistent with what's happening across the rest of the UK."** Participants from the academic and government groups discussed how disability identification can depend on the context within which the data are collected.

> "[Disability can be] where someone identifies as disabled, where someone would fit the Equality Act definition of being a disabled person…anything around medical histories…and barriers facing day-to-day life as well."
>
> **(Research funding organisation participant).**

Social class and disadvantage were also identified as concepts causing issues for analysts, due a lack of consistency or harmonised definitions. **"Highest qualification of parents, self-professed socio-economic status"** were highlighted by a central government participant as examples of how class has been captured within data.

> "[It would be good to] come up with some easy way to ask about socio-economic class, that could be even asked as a census question and become more of a protected characteristics type question."
>
> **(Scottish Government participants).**

For children and young people, disadvantage was said to be defined differently depending on their age, with free school meals being a measure for younger pupils, and indices of deprivation being used for older young people and university students. However, a research funding organisation participant noted that **“area-based measures are actually quite a poor proxy for identifying poor children or children from groups that are disadvantaged.”** Local government participants reported that this was particularly problematic when trying to undertake analyses for specific areas.

Solutions for addressing conceptual issues

Participants recommended allowing individuals to self-identify and categorise themselves for personal characteristics, such as ethnicity and gender, and providing as many response options for these as is feasible. This could be done by providing free text boxes where **“other”** categories have been selected and undertaking analysis on these responses. This would inform researchers of how people are self-identifying and form an evidence base for revising classifications moving forward.

There was a suggestion that **“intersex”** should be regularly included as a response option under sex.

> “It is a legitimate category that people want to identify into, and it is almost always missed off as well.”
> **(Research funding organisation participant).**

Providing a clear steer and guidance from government regarding the correct categorisations to use in surveys, to maintain a consistent picture across UK data collection and ensure that it is not entirely the responsibility of research organisations to **“get this right.”**

Several participants suggested the need to improve the consistency and harmonisation of definitions throughout UK data sources and between government departments to ensure that all parts of the UK are sufficiently represented within data and to bridge the gaps caused by the lack of harmonisation. It was stressed that the Government Statistical Service harmonised standards should be used more frequently and more consistently, to better bring together definitions between different datasets. However, ensuring categorisations are appropriate and up-to-date with societal definitions was also said to be needed.

> “The more you can bring that data together the more you’ll get a more holistic picture.”
> **(Learned society participant).**

> “Pushing at the next boundary once you’ve got the most recent guidance out, rather than letting a long time-lag develop.”
> **(Research funding organisation participant).**

Irrespective of the described issues arising from lack of comparability over time with changing definitions, a research funding organisation participant called for the need to address out-of-date categorisations, as **“actually not acting could cause more harm than acting at this point**.” Suggestions were made for how to keep up with societal changes.

> “Horizon scanning, reading the news, trying to be aware of the emerging issues and thinking around whether this is going to be something which is impacting your local community.”
> **(Non-metropolitan local authority participant).**

Reviews of current classifications were also encouraged, particularly considering bigger societal events which may require changes, such as **“Brexit”.** Some participants suggested the need for a full review of ethnicity classifications in the UK, in conjunction with categories for religion.

> “[Need to create a] combination of almost a matrix of ethnicity and religion to identify the key groups or communities that need to be separately identified for statistical purposes to inform policy.”
> **(Learned society participant).**

Certain participants suggested that this would help to make visible the communities that have been hidden behind broad categories until now.

Findings on the inclusivity of methods for data collection and analysis

Accessibility

Issues resulting from data collection methods were discussed in depth. The approaches used for data collection and to involve individuals in research could reportedly impact the extent to which certain groups are reflected in the data. For example, several participants highlighted that that due to the coronavirus (COVID-19) pandemic, many surveys have been adapted to telephone or online data collection. However, many people do not have access to a landline or internet connection, and these modes could generate charges for the research participant. Additionally, surveys were described as inaccessible to some due to language, wording or format. People who are older, live in rural areas, have learning needs, are not well off, and without fluent English were identified at greater risk of being unable to participate in online research for these reasons. Digital exclusion was highlighted as an important consideration for data collection and the presentation of findings.

Participants highlighted the **consequences resulting from a limited understanding around the digitally excluded population**, in terms of the areas that are more likely to be excluded and the contextual factors surrounding this.

> “We don’t know whether or not we’re getting a sensible view on our policies, on what we’re doing, on whether we’re even targeting the right people.”
> **(Combined authority participant).**

> “I just don’t know whether we’re really on top of what this means for getting the digitally excluded in our surveys and in our data.”
> **(Academic participant).**

Participants noted the neglect of certain populations in official data collection pathways. A Northern Irish Executive participant highlighted that many groups who have **“dropped out of official channels”** are already facing marginalisation, such as young people not in education, employment, or training, or non-private household populations, including individuals in institutions or experiencing housing difficulties.

> “The bias is just reproduced, because if you don’t get people into the data on which then other decisions are made, it reinforces the exclusion.”
> **(Central government participant).**

Reaching under-represented groups

Additionally, lack of focus around the primary issues facing these groups, and consequently **“not asking the right questions,”** was identified by a research funding organisation participant as a major problem in survey design. This was said to result in an increase in non-response rates and data gaps, as well as potentially impacting communities in **“the way they respond to research, and the way that they’re involved in research going forward.”**

Retention and attrition rates among under-represented groups were highlighted by academic participants as particularly problematic.

> “Inclusion and retention of hard-to-reach populations in national, nationally representative datasets, and nationally representative longitudinal datasets is a major goal.”
> **(Academic participant).**

Attrition rates were reported to often be high among under-represented groups, but a lack of statistically sound methods to deal with the biases that this dropout causes could reportedly result in inaccurate assumptions being made. A research funding organisation participant highlighted that longitudinal studies, such as birth cohort studies, are particularly susceptible to issues around attrition, and as under-represented groups are also more likely to drop out.

> “We have those kinds of challenges of making sure that we’re maintaining continuity over time but while also wanting to be able to address some of the issues around [attrition].”
> **(Research funding organisation participant).**

Religion was mentioned as a characteristic which is rarely captured in data collection practices. It was highlighted that not collecting religion data could render some groups such as Jewish and Sikh communities invisible within data. Participants from the academic and government groups described the voluntary nature of the census religion question as problematic, as you cannot rely on the total number provided for various religious groups.

> “The last census said there were about 430,000 Sikhs for example, but that is how many answered the voluntary questions, so how many did not? We do not know.”
> **(Learned society participant).**

Data linkage

Data linkage was proposed as a strategy for capturing the relationship between different factors and characteristics, and overcoming challenges arising from a lack of intersectional data. This was described as particularly important due to the multidimensional nature of many personal characteristics, which policy makers seek to capture in data. However, academic participants described barriers to attempts at data linkage, including pushback from organisations on sharing and access agreements for administrative data, which could prevent particularly useful analyses from being undertaken. An example was provided of a previous research project which linked 15 years’ worth of data on homeless individuals from local authorities to significant health records over the same period. While this process was said to be extremely useful for their specific research objectives at the time, there were challenges with wider access and use.

> “The data could only be put together for that one project and then it would be destroyed or nobody else could access it to look at anything else.”
> **(Academic participant).**

Local government participants described the resource constraints that they faced when processing data, given their limited analytical capacities. Participants stated that reasonably affluent councils and those with strong research teams may be resourced to acquire and analyse the local level data needed themselves, while others cannot. This was said to be particularly challenging for multivariate analyses, where datasets are overlaid to capture intersectional characteristics.

> “It was easier when you had regional statisticians in place, because you could go to them, they understood the area, they’d built up that tacit knowledge and also, they knew who to talk to when things had to be checked...that function has gone almost completely now.”
> **(Combined authority participant).**

Solutions to methodological issues

To address methodological issues, participants suggested boosting sample sizes by oversampling to specifically target under-represented groups, enabling personal characteristics to be better captured. Aggregating years together to increase sample sizes was also considered as a potentially useful solution, although there were concerns that this might require a trade-off between timeliness and accuracy of data.

Participants recommended identifying what is important to communities and then putting specific mechanisms in place to measure these, so that the data that are produced focus on the questions these populations are interested in. Qualitative research was suggested to better define the parameters that are used within data collection tools.

> “[Introducing an] advisory group to actually pose the questions that the data would answer, so that we know what we know, and we know what we don’t know.”
> **(Welsh Government participant).**

Several participants suggested that efforts be made to ensure that questions are asked in a manner that allows all sectors of the community to understand and respond to them, to avoid excluding people or producing distorted results through the use of inaccessible language or data collection modes. Research funding organisation participants called for consistent guidance across the board for data collectors on what questions to ask and how to ask them, so that approaches are appropriate, consistent and sustained.

> “To generate good sustainable data, that is not just relevant for now, but could be relevant for questions we need to ask in the future.”
> **(Research funding organisation participant).**

Improving accessibility for people routinely excluded from data collection was said to be a necessity for inclusive practices.

Examples of potential strategies include:

- using printed materials and providing internet connection as part of a survey incentive for digitally excluded populations
- translating survey instruments into different languages
- distributing paper questionnaires in day centres to reach older populations
- systematic surveys of people living in institutions

However, academic participants noted that that different best practice approaches would be required across settings, due to the differing issues faced between institutional populations.

> “I think you need to develop strategies that are about how you engage with [under-represented groups] on their terms, in a way that they can engage with, which may be less efficient from our point of view but is necessary.”
>
> **(Welsh Government participant).**

Finally, reducing silos across government and organisations to improve data sharing and the capability for more data linkage to take place was suggested to allow for intersectionality of personal characteristics to be better captured. An academic participant outlined that if researchers were able to better access administrative datasets and were able to link data **“the gains from doing it are going to be major”**.

Findings on the appropriateness, availability and quality of data

Timeliness of data

Interview and roundtable participants discussed a range of problems with existing and future data, making it challenging to meet their needs for research, policy and decision-making. Timeliness of data availability was one of the issues identified.

> “Because quite often a lot of data is reported on a quarterly basis, so you were then looking for markers that you could look to try and understand, was there a change where there were more people likely to become redundant. […] The data was not frequent enough to be able to tell if there was a change in trend in the timeframe that we needed it.”
>
> **(Northern Ireland Executive participant).**

Census data were said to be widely used among government participants; however, they become out of date very quickly. Additionally, concerns were raised around the timing of the 2021 Census and how the data will be impacted by the current landscape.

> “We know people’s lives have changed in a huge way since the pandemic, but we can’t even tell if that census data will in any way represent the people that live in London at any other time, other than March this year.”
>
> **(Central government participant).**

> “The basis for so much of the work, and if it’s been done at a particularly unusual time in the labour market, and not just the labour market, socially and everything as well, then I do have some concerns about that.”
>
> **(Non-metropolitan local authority participant).**

Data access

Inadequate ease of access to, and availability of existing data were said to undermine inclusivity, as indicated in the methods theme, with reference to data linkage. Local government participants highlighted the issue of data being dispersed among various organisations, with differing access requirements. Hurdles put in place by some organisations to gain access to data were described as being more complicated than information governance guidelines require. National public databases, such as those provided by the Office for National Statistics (ONS) Secure Research Service, were described as difficult to access due to the limitations imposed, such as the need for a specific research proposal and the time required to determine whether data are suitable for use.

> "You have to have the funding in order to get the access, but then it takes ages once you've got the funding, in order to get the access, and it's not guaranteed that you will."
>
> **(Academic participant).**

This resource intensive process was reported to place severe restrictions on the amount of research that can be done. Participants from the combined authorities mentioned the potential usefulness of numerous HM Revenue and Customs (HMRC) administrative datasets, which could be used at local levels to consider issues surrounding wealth inequalities and long-term unemployed people with disabilities. However, this was described as counterproductive within a policy responsive atmosphere.

> "But the toing and froing over what you could do with it, how you could use it, and that process is quite lengthy, and every chart has to get signed-off."
>
> **(Combined authority participant).**

Additionally, data formatting was highlighted as a hinderance.

> "It's not a matter of data not existing, but a matter of the researcher not being able to access it in the format they need."
>
> **(Research funding organisation participant).**

Sample sizes and granularity

Small sample sizes within survey datasets were identified as a barrier to data being sufficiently granular to meet user needs. This was described as particularly challenging when trying to understand characteristics of a population, such as ethnicity, at a local level. For example, disclosure issues can occur due to small sample sizes when trying to breakdown data by personal characteristics. Participants found that they were unable to get the granular breakdowns they needed without grouping multiple minority groups together, which was said to hinder intersectional analyses and understanding, particularly when attempting to explore meaningful ethnic breakdowns in data.

> “When you want to look at issues around intersectionality, that’s almost impossible to do in surveys”
> **(Welsh Government participant).**

> “If you’re doing white, non-white, you might be okay” but that “if you wanted to look at within London versus outside of London, let alone England versus Wales or Scotland… for ethnic minority groups… there’s no chance of doing it in the surveys… it would have to be census.”
> **(Academic participant).**

Issues around granularity in household survey data were also highlighted.

> “How the cost of living is different for a household with a disabled person versus a severely disabled person or bringing in things around region and ethnicity and being able to look at combinations of those factors.”
> **(Learned society participant).**

Additionally, local government participants explained that local policy decisions were often made based on national level data, due to the quality and small sample sizes of local level data.

> “The data that we’re able to pull down from national sources…. isn’t representative of what’s happening in those areas.”
> **(Combined authority participant).**

For the devolved administrations, concerns were raised regarding the available data for each country and how these can be used. Scottish Government participants questioned the extent to which data published for England and Wales are relevant to Scotland; while Welsh Government participants noted difficulties finding data that solely represent Wales, often being combined with England. For example, data from the England and Wales Longitudinal Study was said to have sampling fractions too small to undertake effective analysis on issues relating to Wales. One academic stated that there is not enough information available to **“make informed policy decisions.”** Participants from Northern Ireland emphasised the need for their unique political and legal structures to be recognised within UK data and statistics, and that sensitivity was needed towards Northern Ireland’s history, described as **“a very contested space on human rights and equality.”**

Administrative data

Although administrative sources were generally seen as beneficial and useful for more timely data, certain participants noted that they often do not capture the information needed and advised that they be used with caution.

> "There is a fine line capacity to what you can get out of administrative records."
> **(Learned society participant).**

Administrative data were seen to not always enable an understanding of background characteristics or what is happening at the micro-data level. Some protected characteristics, such as ethnicity and disability status (as defined by the 2010 Equality Act), were identified as lacking within other important administrative datasets such as death registers and GP healthcare records. This was said to make it difficult to answer questions, draw any conclusions, or understand differing impacts or experiences from the data.

> "There is no data on the victim characteristics in Police Recorded Crime… you can get down to the level of offence, but we can't tell whether or not those are different for protected characteristics."
> **(Central government participant).**

For these reasons, academic participants suggested that there may be general over-reliance on census data. They further highlighted that while census and administrative data are valuable, they lack important subjective perspectives about how people feel or how they view the world. It was therefore recommended that administrative records be more effectively used when linked to other data sources, such as surveys.

Data gaps

A wide range of existing data gaps were identified, which were said to undermine inclusivity in data and evidence. Academic participants highlighted income as a key gap in census data, resulting in analysts having to use bank account data and other "creative measures"; which were described as less ideal. The lack of longitudinal data on income was said to prevent greater understanding of social mobility.

> "So you could have 10% of the population poor in two years. And [whether] it's a different 10% of the people or the same 10% of the people and the policy implications of that are hugely different. [...] Like what's keeping people in these groups and what's stopping them from getting out. I think that's a really important agenda to economists."
>
> **(Academic participant).**

Additionally, learned society participants noted that, with incomplete understanding of issues such as changes to cost of living in different areas, it is **"harder to make a policy in a reactive and appropriate way."** Concerns were raised around the lack of available data on income levels among caregivers and the scale and scope of children acting as carers. Data gaps were also identified in relation to informal care.

> "When you see the evidence that is available on the scale of informal care, it is absolutely massive, and if you compare that to what we actually know about that care in terms of who's doing it and what types of care they're providing and so on, it's just a huge mismatch there."
>
> **(Research funding organisation participant).**

Local government participants highlighted recurring gaps in local data, which were said to inhibit understanding of **"gateway communities"** (people moving between different local areas such as between home and work). These communities were seen as particularly vital to capture to effectively represent the dynamics of local areas.

Data bias

Bias in training data and the effect this has on machine learning algorithms was also highlighted as an issue by one learned society participant, who noted that machine learning systems have been trained on unrepresentative data.

> "There are challenges when machine learning systems are trained on data [which] perhaps may be not representative and might have biases in it, and therefore may have gaps in terms of not covering the needs of certain groups or may represent certain groups unfairly due to existing structural biases in society affecting the data that's collected through different processes."
>
> **(Learned society participant).**

Strategies to addressing data issues

To address data issues, participants proposed establishing a data source comparable to the census that is recorded more often, as census data are often not recent enough for policy analysis.

> "There is a fundamental need to have something that looks like census data more often."
> **(London borough participant).**

Non-metropolitan local authorities requested that timely small area datasets which cover social and labour market statistics be produced and made accessible.

> "So more small area data, which I know is hugely expensive, but that's the wish list… so that we're not relying on ten-year census data."
> **(Non-metropolitan local authority participant).**

Participants recommended adding and maintaining accessible and user-friendly interfaces for extracting local data, such as NOMIS (an ONS service providing access to official UK labour market statistics) and removing or easing the requirements for local government and civil society organisations to access data from central government.

It was said that the **ONS could provide a role of bridging the gap between central government and local government** to reduce reliance on individual central government departments for specific datasets.

> "Creating a place where people could at least learn to see how their datasets relate to a set of other datasets somewhere else, and how they've built up and developed, that would be a step forward."
> **(Central government participant).**

It was also suggested that investing to increase sample sizes in surveys would help to improve the geographical granularity and quality of the data at local levels to accommodate local decision-making.

> “The small sample sizes will require a big investment to be able to collect more data.”
> **(Learned society participant).**

A big step-change was said to be necessary to better coordinate and combine locally held datasets to build a national picture. An academic participant suggested that while research on vulnerable populations, such as people with mental health challenges has been possible at a local level due to collaboration between academics and individual trusts, this did not provide a national picture.

> “Being able to do that [research] in terms of understanding national trends and what’s really happening rather than sort of pockets of deprivation and deep kind of relations, we really need a step change in a lot of those areas.”
> **(Academic participant).**

Improving the quality, accessibility and use of administrative data sources was advised.

> “There is a need to really optimise the administrative data sources. We need to streamline them. We need to make them transparent in the way they’re constructed and the way they are used.”
> **(Central government participant).**

It was noted that despite access issues for these datasets, they do have great potential. Creating new sources of administrative data was also suggested by academic participants, such as using financial data and mobile phone data. However, it was warned that holding these data may be accompanied by ethical and privacy issues and may not capture certain groups.

> “Administrative data may capture more of some of the groups that we’re interested in than the surveys do.”
> **(Academic participant).**

> “Particularly relevant in something like a pandemic and when you’re trying to track the impacts of policy interventions and trying to look at things like how people’s behaviours have changed, how people move around, [what the data] tells you about things like transmission patterns, etcetera.”
> **(Academic participant).**

It was suggested that **developing a greater understanding of data gaps and prioritising how to fill these** data gaps is increasingly important. A complete **"data gaps exercise"** was advised.

> "Undertaking a comprehensive gaps exercise to assess the statistical requirements…of the government's equality agenda, the Equality Act for example, and all the protected characteristics."
> **(Learned society participant).**

A learned society participant recommended reviewing biases based on historical data fed into machine learning algorithms for decision making; and improving privacy enhancing technologies to lower the exposure risk for certain minority groups and better enable the use of data. An example of this would be using differential privacy techniques such as creating synthetic data, which would help create an accurate representation of society without using personal information.

Findings from insights generated from data and their presentation

Understanding personal experiences

Ideas were discussed for gaining better insights from the data and engagement activities. These included approaches to better understand people’s life experiences, such as ethnography and other qualitative methods. Surveys were seen to predominantly focus on recording descriptive characteristics, providing little insight into the needs of individual.

> “A lot of the data sources we’re working with are not based on the lived experience of disabled people but rather just simply outcomes, or management information, which usually are related to benefits or others and how people are given money, rather than what they need to survive.”
>
> **(Central government participant).**

> “Less weight is given to perceptions and emotions, and basically, people’s perceptions of experiences, [...] perceptions and feelings have great power, and we need to understand that to make better policy.”
>
> **(Central government participant).**

> “What would a human-centred approach to data collection look like if it started with people and their life experiences, and how would that re-categorise the labels that we use to describe people and the topics and areas of interest that we even collect information on? What are the interests that are important to marginalised groups, that don’t sit on the other side of the room? What would they want to see studies on? I just really actually think it’s vital to policy development.”
>
> **(Central government participant).**

Understanding data and community needs

Learned society participants identified a need for government offices and departments to undertake more frequent consultations on key changes to the publication of official statistics, to gain insights into data user needs and ensure they are provided for.

> "Some publications provide a lot of information on a certain vulnerable group or protected characteristic, and there needs to be a higher level of consultation or engagement before that is stopped or changed in some way."
>
> **(Learned society participant).**

Local government participants highlighted issues that arise from making decisions based on an incomplete understanding of what is happening in their communities. This lack of insight was said to result in community needs not being effectively captured when developing policies, which can thereby exacerbate inequalities.

Devolved administration participants highlighted that without better data enabling experiences between diverse groups to be distinguished, policymakers are not fully able to clearly define who is being left behind. Welsh Government participants also highlighted how important insights from available statistics on inequalities are not sufficiently driving impact, change and accountability conversations.

> "That connection with the information we do have doesn't always seem to happen. Because there's some pretty stark statistics in there, even with the stuff we've got, so why is that not driving impact and change."
>
> **(Welsh Government participant).**

Data presentation

Linked to the accessibility of data outputs, interview and roundtable participants discussed how the ways in which data are presented and explained can impact their interpretation and therefore potential to provide useful insight.

> "Explaining that difference is a real difference, or at least it's a quite strongly inferable one, that's the point where we hit the challenge, because that's the point where we are convincing audiences that what we are saying is credible, should be considered and isn't just luck of the sampling bias, or luck of measurement."
>
> **(Central government participant).**

It was agreed that social media had the potential to reach a wide and diverse audience. However, issues with using social media to present findings were discussed, including the risk of reports being misread, key data being distorted and used to communicate messages suited to the reader's interest, or the full report not getting its due regard.

> "Twitter in itself and that short byte size nature of it means that there is real distortion and potential for distortion of data."
> **(Northern Ireland Executive participant).**

Academic participants presented one of the main reasons for data inaccessibility as a lack of mandatory provisions for data outputs within the research community and the means to ensure inclusivity of access to research findings.

> "There is a real balance about trying to make the data accessible so it can be accessed by people who have visual impairment or disabilities but also that the information is easy to access for people who are not statisticians or researchers."
> **(Learned society participant).**

Solutions to improve insight

To address the insight issues described, participants suggested using qualitative approaches to capture lived experiences and develop insights into what matters to communities to improve inclusivity. Participants suggested that the collection of qualitative data could inform policy work in a different way to statistics, help communities feel more included, and help address data disaggregation issues. However, it was stressed that qualitative data should be captured in a structured manner, and findings disseminated efficiently across different organisations and government departments to ensure groups are not repeatedly asked the same questions.

The creation of a **centralised database for local government policy analysis** was also recommended.

> "[This could be] interrogated in a way that would give you multi-faceted ways of looking at it through different lenses."
> **(Central government participant).**

Participants recommended improving the accessibility of data that are presented to the public through a variety of methods, including:

- clear bullet points
- lay language
- clear and simple tables
- maps and charts
- downloadable data in CSV format
- producing accessible websites, using social media platforms
- making the information as easily digestible as possible

It was also suggested that researchers need to consider people who speak English as a second language and others who may be less proficient in English when disseminating research findings. A learned society participant highlighted the importance of using data visualisation techniques to improve the clarity of key messages, **"but also picking up the key points and limitations of the data."** In addition, ensuring access to easy-read versions or screen-reader compatible versions of statistical publications where required, and producing offline versions of certain publications where possible.

Conducting cognitive interviewing to understand how users comprehend survey data was suggested by a research funding organisation participant, as well as working with certain under-represented groups. Participants felt this would improve the way data and evidence are presented in the future, making insights more meaningful for diverse groups.

> "[To test] different ways of showing the same evidence and seeing whether that resonates with how they best connect with it."
> **(Learned society participant).**

> "Funders could probably help each other to come to some standards around this kind of activity."
> **(Research funding organisation participant).**

Certain participants highlighted the importance of being transparent about data gaps to help contextualise the disparities that are observed and make the findings less open to misinterpretation.

> "Make visible the gaps, rather than just make visible the data?"
> **(Welsh Government participant).**

Findings from examples of best practice

Learning from others

Participants shared examples of best practice for inclusivity, and ways that learning from others could be improved. Academic participants felt that often there are so many questions surrounding which intersectional lenses should be applied to datasets that researchers may avoid these types of analyses altogether. Exemplary practice was identified within the United States.

> "Some great research that's come out from researchers looking at engineering STEM cohorts in America, that have provided some intersectional lenses to student identity."
>
> **(Academic participant).**

However, it was suggested that despite being aware of these insightful research projects **"[The UK] just haven't pulled that knowledge across into the work that we've done." (Academic participant).**

Additionally, the United States Census was praised by academic participants for their inclusion of an income variable, which participants felt would enable better analysis of income dynamics and monitoring of socioeconomic mobility.

> "[That] would be extremely helpful and be super interesting to look at how different groups experience income volatility."
>
> **(Academic participant).**

Knowledge sharing across government

Overall, participants noticed that coordination, and information knowledge and skills sharing across government and organisations could be much more efficient and effective. For example, through sharing resources and learning from one another, and sharing examples of best practice. Local government participants shared that if one local authority acquires, produces, and analyses a set of data to meet their local policy needs, these may not be communicated well across other local authorities, who may also have an interest in these data for their respective areas, due to limited coordination between them. Better communication and coordination were said to lead to better use of data and reduce risk of duplication of efforts.

Local government participants also noted that **not enough data sharing takes place between central government and local government,** which leads to the exclusion of the latter from valuable information and data sources.

> "As a combined authority, we don't have access to local authority data, only if the local authorities share it with us."
> **(Combined authority participant).**

It was felt important for local authorities to work collaboratively with their combined authority to share the data that they have, so that full programmes of support can be developed for certain minority groups.

Another barrier to shared learning and making the most of resources, reported by a central government participant, was a **"lack of transparency"** across government around existing data. This included **"what is where, what is covered, how it's being collected and how it's being published."** Some participants felt that this can make it difficult to ascertain whether harmonised definitions were being used across government areas, and to understand the underlying data that analysts and policy makers need to work with.

> "Knowledge within departments of what data they had on disability was very fragmented and there wasn't a single point of contact who could signpost people to what they might have, but then, secondarily to that, in many cases the people in the department didn't know whether a particular data set used GSS [Government Statistical Service] harmonised standards or not."
> **(Central government participant).**

Suggestions for learning from best practice examples

To improve and better learn from best practice, participants suggested using the precedents set by international good practice examples to improve the regularity and simplicity of linking datasets.

> “Which is bringing together social care data, health data and Department for Education data, and has managed to create a mechanism to allow that data to flow.”
>
> **(Research funding organisation participant).**

Another suggestion was to set up broad partnerships with a wide range of stakeholders who have different needs and create an organised user group, particularly for equalities data. Participants thought that this could increase statistical awareness and make better use of statistics in the UK. Non-metropolitan local authority participants in particular noted that the partnerships developed since the start of the Coronavirus pandemic had been particularly valuable.

> “Out of a crisis may well come an element of improved partnership working, so that we can actually share knowledge and identify issues.”
>
> **(Non-metropolitan local authority participant).**

Scottish Government proposed planning **“analytical gatherings”** to enable analysts to share their successes in producing impactful data on protected characteristic groups, as **“that would allow others to see if that would be applicable in their own areas.”**

It was also suggested that knowledge and learning could be taken from the development of the 2021 Census to improve harmonisation of measures for protected characteristics groups across the Government Statistical Service (GSS).

> “Making sure that there’s a real commitment across the GSS to actually using that knowledge and learning.”
>
> **(Welsh Government participant).**

It was advised that learning can also be taken from the UK Government’s citizen assembly on climate change, which a Northern Ireland Executive participant cited as a good practice example of **“genuine random bringing together of people in a quite scientific way,”** to address complicated societal issues. It was also reported by a Northern Ireland Executive participant that lessons could be drawn from the Republic of Ireland’s successful approach of “testing and refining of key questions” for use in their referendum, which ensured that the questions were appropriate before being posed to the general public.

Central government participants suggested that analysts across government who are working with equalities data should be able to receive training from those who are generating survey data, such as the Annual Population Survey. In addition to knowledge sharing and improved communication, this could then allow microdata to be more widely shared across government.

Producing best practice guidance for local authorities conducting their own research was recommended, for local areas to be more easily compared. Certain participants thought that this would ensure equalities information and best practice guidelines are harmonised across local areas.

> “Having some harmonisation across the departments would be helpful, and just best practice guides, or, ‘Here’s the questions that we ask,’ and maybe you want to be using[…] The ‘what works centre’ type approach to asking questions, or gathering the information that you need.”
>
> **(Combined authority participant).**

Annex A. Topic guide structure for roundtable discussions and in-depth interviews

Introduction

- What is one success you've had improving equality and inclusivity of data, statistics, analysis and their presentation (if any)?
- What is one key question around inclusivity and representation you would like to be able to answer?
- What successes have you experienced in trying to answer this question?
- What struggles have you experienced in trying to answer this question?

1. Data Sources

a. What sources of data do you find most useful for addressing equality issues in your role?
b. Any challenges with these sources?
c. What sources of data do you often have issues with, in terms of inclusivity?
    i. What are these issues?
    ii. If applicable: What are your views on the potential for administrative data sources to address your data needs?
    iii. Any concerns about administrative data?

2. Identified gaps and implications

a. Are there any specific gaps in the data that you use that could be addressed?
b. What barriers are there that prevent filling these data gaps?
c. What are the impacts of these gaps on:
    i. Your organisation (such as in developing policy, communication and your organisational aims?)
    ii. Your key stakeholders?
    iii. The equality groups that the data affects?
d. What can be done to fill these data gaps or create more inclusive data?

3. Research and survey design

- **a.** How does research and survey design create barriers to equality and inclusivity of data?
- **b.** How can we ensure everyone is included in data collection?
- **c.** Can we resolve these barriers? How?
- **d.** How do we ensure we're asking the right questions?

4. Engagement

- **a.** What steps do you take to encourage wide engagement and participation in research?
- **b.** What are the barriers to engaging with under-represented groups to encourage participation in research?
- **c.** How [else] can we address these?
- **d.** How can engagement be used to improve inclusivity?
- **e.** How can we ensure that the learnings from engagement are used widely?
- **f.** How do you engage with issues of trust about participation in research and data sharing among the general public or under-represented groups?
- **g.** How do we ensure all members of society feel represented in data, analysis and outputs?
  - **i.** Any specific examples?

5. Harmonisation and coherence

- **a.** What impact does harmonisation and statistical coherence have on the inclusiveness of data?
- **b.** What issues (if any) have you experienced with harmonisation and statistical coherence in your role?
- **c.** Are there changes that could be made to address harmonisation and coherence and improve inclusivity of data?

6. Outputs

a. How important is it that analytical outputs are accessible to all? (Produced by your and other organisations.)
b. How represented do different sections of society feel in your analytical outputs?
    i. What more could be done? Current challenges?
c. How can we make sure that even the digitally excluded are included in data collection and able to access the outputs?
d. Do you have any strategies in place to reduce the barriers faced by the digitally excluded?
e. Are there any further thoughts you had on how to make data, analysis and outputs that you use or produce more inclusive?

Finally, would everyone be able to tell me the top issue that they would like considered by the Inclusive Data Taskforce?

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