Consultation on Inclusive Data

Research by Basis Social with Civil Society Organisations and members of the public on behalf of the Office for National Statistics

May 2021

BASIS
SOCIAL

Contents

Executive Summary ______________________________ Error! Bookmark not defined.

1. Research background ______________________________ 122

2.Data use by CSOs working with individuals with protected characteristics __________ 188

3. Improving the inclusivity of data on individuals with protected characteristics – the CSO perspective ______________________________ 222

4. Data use, challenges and opportunities disaggregated by equality area – the CSO perspective ______________________________ 300

5. Preferences and principles for the collection of more inclusive data __________ 655

6. Building a more inclusive picture of individuals with protected characteristics or from underrepresented groups, disaggregated by equality areas ______________________________ 777

7. Implications for improving inclusivity in data amongst individuals with protected characteristics ______________________________ 96

List of Annexes ______________________________ 99

Acknowledgements

We would like to thank all those civil society organisations and members of the public that gave their time to participate in this research. Without their honest and open-reflections on the barriers and opportunities for greater inclusivity in not just UK statistics but UK society, it would not have been possible to produce this report for the Office for National Statistics (ONS).

We would also like to thank Rosie McLeod, Head of Data and Learning at New Philanthropy Capital for her advice and support throughout the project, and to the team at ONS leading the consultation activities. This included Dawn Snape, Silvia Manclossi, Emma Jones, Sarah Coates and Alex Buckley. Finally, we would also like to acknowledge the valuable input received from members of the Inclusive Data Taskforce in helping shape the methodology and feed into the final report.

The views, opinions and recommendations expressed in this report are those of the authors and do not necessarily reflect those of ONS, the Inclusive Data Taskforce or the UK Statistics Authority.

List of abbreviations

BAME	Black and Minority Ethnic
BEIS	Department for Business, Energy and Industrial Strategy
CCG	Clinical Commissioning Group
CHAIN	Combined Homelessness and Information Network
CSO	Civil Society Organisation
DfE	Department for Education
DHSC	Department of Health and Social Care
DWP	Department for Welfare and Pensions
DVLA	Driving and Vehicle Licensing Authority
EHRC	Equality and Human Rights Commission
FOI	Freedom of Information
GDPR	General Data Protection Regulation
HMRC	Her Majesty's Revenue and Customs
IJPR	Institute for Jewish Policy Research
LGBTQ+	Lesbian, gay, bisexual, transgender and queer (or questioning) and others
MCB	Muslim Council of Britain
MHCLG	Ministry of Housing, Communities and Local Government
MoJ	Ministry of Justice
ONS	Office for National Statistics
PACT	Prison Advice and Care Trust
PHE	Public Health England
PTSD	Post-traumatic stress disorder
SEG	Socio-Economic Grouping
UKSA	UK Statistics Authority

“We need to see changes. That is most important. If they collect the data to have a certain statistic, then we need to see the changes. If there are no changes, then what is the point?”

Photo by Maatla Seeteloon

EXECUTIVE SUMMARY

Background and approach

The Office for National Statistics (ONS) is an independent government department and the UK's largest provider of official statistics. As part of a drive to ensure that statistics reflect the experiences of everyone in society, ONS established an Inclusive Data Taskforce. The Taskforce is to consider evidence on improving data inclusivity gained from a range of engagement activities before making recommendations to the UK Statistics Authority, who hold responsibility for overseeing the production of national and official statistics. On behalf of the Taskforce, ONS commissioned Basis Social to undertake research with a range of civil society organisations (CSOs), as well as members of the public that have different protected characteristics or are under-represented in the UK statistics.

This research involved two phases of qualitative research:

1. Research with over 80 CSOs working across 15 different equalities areas to understand: their data and evidence needs; the relevance and utility of current data and evidence available to them; evidence gaps and the impacts of these on their work and wider society; and opportunities for improving the inclusivity of data
2. Research with over 90 members of the public recruited as having different protected characteristics or who are under-represented in UK statistics, to understand: if and how they feel that they are (and should be) represented in UK statistics; if and how they want to be engaged with in providing their 'data', including through what mechanisms and for what purposes; and their views about the potential benefits and risks in providing their data

'Lenses' on inclusivity

In undertaking this research, it was evident that there are a number of ways in which CSOs and members of the public view data inclusivity. These different lenses on the inclusivity of data included:

- the way in which individuals, organisations and communities are engaged in the process of providing their information or 'data';
- the concepts, constructs and data-points gathered, and the extent to which these are inclusive of different personal and social identities;
- the methods through which data is gathered;
- the accessibility of data and appropriate dissemination of findings; and
- the insights that are generated as result of that data, and the extent to which these align with the lived experiences and priorities of people that the data is seen to represent.

This report covers each of these in detail, from the perspective of organisations using data to provide services and advocate on behalf of different communities, and from the perspective of individual members of these communities themselves. There is however, one additional lens through which data inclusivity should be viewed, and this relates to the influence of under-represented groups in the purpose of data collection and on the actions taken in response to this evidence.

This research purposefully involved more under-represented (and disenfranchised) communities. In consulting with these groups, it was evident that - beyond basic administrative data given to receive a public service - there was a mixed awareness, understanding or appreciation as to the purpose, use or benefit of providing their 'data'. Instead, there was a sense of caution around the trustworthiness of 'authority' (both Government and public services) amongst many groups, and in participating in what could be seen to be an extractive process to provide information that could be used against them, either individually or collectively. The value of data is in the actions that it then informs. Without inclusivity stretching to the role of minority communities in shaping the purpose of data collection and influencing actions taken in response to this data, the whole premise for collecting data is undermined. Addressing this power imbalance will be a foundational step for more inclusive data.

Facilitating participation

A wide range of factors impact on the ability and opportunity for people to provide their personal information and participate in formal research exercises. These range from concerns and fears around the collection and use of data, through to practical challenges of access, language and literacy. The specific issues vary both within and between groups and are more or less evident in different data collection exercises and who is undertaking this data collection. Rather than see minority groups as 'hard to reach', to aid inclusivity it will be important to understand how to best support people in providing their data, across different circumstances (e.g., different forms of administrative data collection or population surveys), and to work with communities to ensure any potential barriers are minimised or removed entirely.

CSOs working with minority groups at both national and more grassroots levels, often have a strong understanding of the different ways through which to best engage and support people to participate in research. There was also a strong desire expressed amongst CSOs involved in this research to contribute toward a process of increased dialogue and understanding. Tailoring approaches for engaging with communities and adapting methods to better enable and empower people to provide their data will promote greater inclusivity. This will require greater effort on the part of researchers (or public services), working in collaboration with CSOs, to account for the different needs and preferences of those individuals from whom data is to be collected, and to adapt data collection approaches accordingly.

From the perspectives of members of the public recruited as having different protected characteristics or being under-represented, there were a common set of principles that were identified as important for organisations collecting data to adhere to. These included:

- Transparency on what data is being collected, the purpose for collecting it, how it will be used and/or shared, and around confidentiality
- Inclusive approaches to data collection
- Informed consent through the tailoring of information to the needs of individuals
- Ensuring individual anonymity in data which is published in aggregate form
- Encrypting data and keeping data securely to protect against data breaches
- Ensuring that people who provide their data, particularly within research exercises, receive feedback as to the outcomes of sharing their data

"I haven't been presented with any evidence that tells me that them having that kind of information about me as an individual benefits me in any way, shape or form."

Utility and use of data from under-served groups

'Data' used by CSOs range from anecdotal professional opinion and qualitative case studies through to surveys of service users, national population surveys and administrative datasets. Data provided the foundations on which CSOs identified the need to then develop strategic plans and services, to advocate on behalf of vulnerable groups, and to evaluate provisions.

The robustness of data was integral to this in terms of its representativeness and statistical validity, but as important was the relevance, depth and nuance of the data. For local support services, this meant data about the local population which their specific services catered to. For CSOs involved in more regional or national advocacy, this meant data that enabled them to look at issues impacting a particular equality area or group in a manner that allowed for disaggregation by characteristics of interest. Both quantitative and qualitative research held value. Statistics for enabling an overall assessment of need, and stories for understanding lived experiences and how different factors interacted to impact individual's lives.

A point repeated by CSOs throughout this research was the importance of understanding the intersectionality of different characteristics and the influence this had on outcomes for people. For example, it was not enough to see data of interest broken out by ethnicity, but by other characteristics such as ethnicity by age, or ethnicity by gender.

"[You need] to have it broken down in a way that is meaningful for people to identify with… down to as granular a level as possible, so you can actually say, 'If we do this action, it will have a positive impact for this community.'"

The feasibility of disaggregation was reliant on a number of factors, including: data on protected characteristics being collected in a consistent, harmonised format (i.e. uniform questions and response options across surveys and administrative forms); data being recorded digitally, and recorded on interoperable systems that allowed for collation at a national level; and data being made publicly available in a format that enables data to be analysed according to user-defined levels of interest. There is a challenge, in particular with administrative data, of: inconsistencies in the collection and reporting of data; of exclusive definitions, which results in gaps in data around key characteristics (for example faith data); or specific issues where understanding intersectionality and nuances in the data is key (for example counts of homeless young people or reconviction rates of female prisoners imprisoned under 12 months).

The lack of granularity in the statistics currently available - even in the data considered to be of the highest standard, such as the Census - meant that the majority of CSOs participating in this research were unable to develop an inclusive understanding of groups, across the range of protected characteristics. This was seen to impact not just their services but the efficacy of public services and public policy, and ultimately on people's life outcomes. Organisations looked to compensate for this by developing other sources of data to fill the gaps, often relying on qualitative research, using service-level administrative data, conducting their own surveys of service users, trying to obtain data through FOI requests, and utilising published data to help extrapolate more localised findings to a national level. This was seen as having the potential to introduce a process where error and bias were acknowledged as impacting on the quality of data and insights available, as well as placing a high degree of burden on CSOs who typically had limited skills and resources to collect and/or analyse data.

Looking forward there was a desire amongst CSOs for data which were more 'findable, accessible and interoperable and reusable' – facilitated both by better curation of data and through data linkages made possible through a common agreed unique identifier. This is fully aligned with the UK's National Data Strategy (2020).

Classification, labelling and categorisation

Members of the public were typically happy to provide their personal information, including more sensitive data such as their faith or sexual orientation, where there was a clear rationale (assumed or explicit) for providing that information. Typically, this was in respect of their gaining access to a service, and where those data helped to ensure a more appropriate, tailored service was received. In this sense, the majority of people felt that sharing their data was important to receive a personal benefit, and a smaller minority could make the link between their data and wider public policy or public benefit. The personal information which people were least comfortable sharing was security or financial data (except with banks), "private" personal data (such as biometric data, photos/videos, information about friends and family, and details of conversations), and political views. This data was seen to be more uniquely tied to them as an individual and had the potential to result in negative repercussions if used in unintended ways. Additionally, gender diverse individuals, and those from Gypsy, Roma and Traveller communities, were generally less comfortable with sharing their personal information than others.

In all cases, willingness to share was linked to the trustworthiness of the organisation (or individual) requesting the data. Trust could be influenced by the nature of the data being requested or mode in which it was collected, or through the existing relationships and power dynamics between the data collector and the individual (or group). Most participants were happy to share data with public services, though some groups were more wary than others due to historic and/or current tensions with either government or public services.

The types of data gathered by public services were felt to be broadly accurate by participants, though incomplete in the picture this presented of individuals and their circumstances. While they may capture some basic characteristics (e.g. around age, ethnicity, gender etc.), these were felt to be at such an aggregated level that they lacked nuance or power in helping to explain individual circumstances and outcomes. For most people – at an individual level - this was not a concern, and there was a general reluctance to share more information than was necessary. This was particularly the case for members of those groups who were regularly asked to provide their data (e.g. disabled people or those with experience of homelessness) or who had cause for concerns around how their data might be used to their disadvantage (e.g. Gypsy, Roma and Traveller and trans, non-binary and gender diverse communities).

To provide a more inclusive picture of individuals there was an expectation that this would involve asking people to provide answers to more open questions in respect of their lives (including their attitudes, priorities, needs and desires). It would also involve ensuring that where closed questions were asked - particularly in relation to socially constructed categories such as ethnicity, gender, sexual orientation and disabilities - that these are reflective of the fluidity and nuances of how people see themselves. People recognise these are important data, and there are sensitivities around providing these data. It will be important to acknowledge that - with more detailed data collection - there is a greater need to address concerns around how the data are to be used, with implications for engagement and communications.

> *"It's like there's certain boxes they put you in. You can't be a bit of everything, you're either that or you're that. It's like well what am I? Where do I fit into all of this then?"*

The labels used to represent different individual characteristics were seen to be critically important both to enable people to select categories in surveys and forms that reflect their personal identities, and to ensure that the data allows for an accurate understanding and to enable actions to be taken in response to the data. Whether this is 'BAME', 'Asian', 'physical disabilities' or 'LGBTQ+', restricting the presentation of data under labels which are seen to homogenise diverse and distinct groups was seen to be highly problematic, misleading and potentially offensive. If the value of data is to enable people to be better understood, and represented in services and policies, it is imperative that these data are an accurate representation of people's circumstances and identities. Both from the perspective of CSOs and of members of the public, the current categorisations used in administrative data and population studies would benefit from refreshing, in consultation with minority groups to also support data harmonisation.

Implications

This report details the ways in which ONS, and other organisations collecting data, can best work with CSOs and individuals to ensure that UK data better reflects the UK population. Key opportunities, detailed in Section 7, include:

- Harmonising the types of data collected on protected characteristics to enable the needs and circumstances of minority groups to be identifiable in both administrative and national survey data. This requires definitions, categories and forms of questions asked around protected characteristics to be more inclusive and harmonised across data collection methods. It also requires public services and agencies adopting these standards in (digitised) data collection and making these data publicly available in a standardised format.
- Recognising that many protected characteristics are not static and can change over time. It will be important to work with relevant CSOs in determining the best ways of capturing data which are timely and accurate.
- Researchers providing clear communications around the purposes of data collection, with suitable reassurances around how data will be stored, managed, shared and used to improve public trust in data, and the trustworthiness of those organisations collecting and using data.
- Researchers to better match the channel and form through which data are gathered to the audience (including face-to-face, telephone, online, peer-to-peer). It is the author's view that this should involve focusing less on standardisation of mode and more on the purposes for which data are being gathered. This will involve greater collaboration with CSOs to understand needs and preferences of different audiences, and how to tailor approaches to maximise participation. This would also involve greater awareness amongst both CSOs and policymakers as to the opportunities for qualitative research, and ways to increase the robustness and representation of qualitative research.
- Researchers and research commissioners publicising the way in which data are being used, the decisions taken in response to data gathered, and what this means for changes to policies and services and ensuring that the way this is communicated is accessible for and is reaching all communities which are involved in the research.
- Researchers and research commissioners providing administrative and population survey data in such a way as to enable (GDPR compliant) manipulation of disaggregated data by different protected characteristic to allow for more contextualised, intersectional and actionable understandings of groups, at a localised level.
- Making training, resources and funding available for CSOs (and other organisations) who want to undertake their own collection and/or analyses of data relating to specific equality areas and/or protected characteristics more accessible. ONS (or another relevant organisation such as the Information Commissioners Office) should also provide information, advice and guidance for CSOs on GDPR and best practices in collecting and collating information on service users or members.
- Researchers and those using research minimising, wherever possible, the use of homogenous terminology for grouping individuals into meta-categories which may misrepresent or conceal the needs of specific groups within these categories (e.g. 'BAME', 'Asian' or 'LGBTQ+').

- ONS to help facilitate discussions around the use of a unique identifier and data linkages in administrative data sets, both for the purposes of research but also more effective public services. Related to this would be putting in place legislation that strengthens ONS powers to request and receive equalities-related data from public authorities, for data linking, analysis and reporting, helping to mitigate concerns around data being withheld due to political sensitivities.
- Allied to data linkage, ONS to help facilitate discussions with CSOs and members of communities where a balance needs to be struck between capturing data on protected characteristics and respecting the wishes of people not to be identifiable by these characteristics (due to concerns around the potential for discrimination).
- Engaging members of the public, CSOs, public sector authorities and private organisations, where possible, in the design of research, to ensure the research is relevant and contextually correct. This includes discussions around the principles of ethical data collection and collation.
- For ONS, and other public bodies to work toward the principles set out in the National Data Strategy, where data are ‘findable, accessible, interoperable and reusable’.

Accounting for the implications of this research will go a significant way in helping improve the inclusivity of UK data and evidence, but more is needed to realise the goal of making a step-change. To ensure everyone counts in UK statistics and evidence it will be important to increase people’s participation in shaping the agenda that determines what data are collected and how they are used. Greater inclusivity is about providing people with greater understanding, ownership and control over not just their data, but what their data represents: their lives and circumstances.

1. Research background

1.1. Research objectives

The UK Statistics Authority (UKSA) holds responsibility for supporting the production of national and official statistics, which are of fundamental importance to the development, implementation and evaluation of public policy at national, regional and local levels. The UKSA remit, therefore, includes addressing challenging questions for standard population surveys (fully detailed in the Equalities Data Audit[1]), such as how to include those who are under-represented in UK statistics. There is rightfully an ambition – in part accelerated by COVID-19 – to produce data at the level of granularity and interoperability required to fully inform policy decisions and service provision. This includes both data at a smaller geographic scale and inclusive of the variety of people and characteristics within the UK.

The Office for National Statistics (ONS) is an independent government department and the UK's largest provider of official statistics. As part of a drive to ensure that statistics reflect the experiences of everyone in society, ONS established a new Inclusive Data Taskforce. The Taskforce is undertaking a range of engagement activities to hear from people with different perspectives on data inclusivity including CSOs, local government, academics think tanks and the public. On behalf of the Taskforce, ONS commissioned Basis Social to undertake research with a range of civil society organisations, as well as members of the public that have different protected characteristics or are under-represented in UK statistics.

In summary, this research sought to:

1. understand the data and evidence needs of civil society organisations (CSOs) who use information about equality for public good; how relevant and impactful current data and evidence are for them and the communities they represent; gaps they have identified in the collection and presentation of evidence, and the impacts of these on their work and wider society; what needs to happen to fill these gaps; and the implications of this for the groups they serve
2. understand, from the perspectives of members of the public - across a range of key equalities areas – if and how they feel that they are (and should be) represented in the facts and figures of the UK; if and how they want to be engaged in providing information on their situations, attitudes and behaviours, including to whom, through what mechanisms and for what purposes; and their views about the potential benefits and risks in doing so

The findings from this research will form part of the evidence considered by the Inclusive Data Taskforce in making recommendations to the UKSA.

1.2. Methodology

Stage One involved 20 two-hour online workshops with over 80 CSOs, each involving between two and seven organisations representing a different equality area. Table One (below) provides details of those organisations that participated in this research.

For each equality area, a minimum of 12 CSOs were identified as providing a representative spread of CSO size and location (including coverage of England, Wales, Scotland and Northern Ireland). Invitations

---

[1] ONS (2018) Equalities data audit. Accessed at: https://www.ons.gov.uk/methodology/methodologicalpublications/generalmethodology/onsworkingpaperseries/equalitiesdataauditfinalreport

were sent with a minimum of one weeks' notice to invite expressions of interest in participating. For the majority of equality areas,[2] a minimum of five CSOs accepted the invitation to attend. We aimed to place a limit on five CSOs per equality area to maximise the opportunity for participating stakeholders to contribute to discussions.

Table One: Participating CSOs

Equality area	Participating organisations
Physical disability or impairment	• Disability Rights UK • Scope • Leonard Cheshire • Limbless Association
Learning disability, neurodiversity or dementia	• The Brain Charity • AchieveAbility • Learning Disability Wales
Race or Ethnic equality advocacy and support groups	• CRER Scotland • Race Equality Foundation • Dalit Solidarity Network • SCORE Scotland • Leeds Chinese Christian Church • Hackney Chinese Community Services • Chinese in Wales Association • Chinese Wellbeing • North East Chinese Association
Gypsy, Roma and Traveller advocacy groups	• Gypsy, Romani Arts • Travelling Ahead • CGTP • Gypsy Traveller • Traveller Movement • TGP Wales • Derbyshire Gypsy Liaison
Homelessness	• Emmaus UK • Simon Community NI • Centrepoint • St Mungo's
Faith	
***Buddhist faith organisations*[3]**	• Adisthana • East Midland Buddhist Association
Christian faith organisations	• Church of England (Stats)

---

[2] Excluding the Buddhist faith equality area, for which it was more challenging to engage organisations in the research. The feedback received as part of the recruitment process suggests that Buddhist organisations tend to work and collect information predominantly on those individuals that are part of their specific Temple or Centre for the purposes of teaching rather than addressing a community need

[3] Note that these were interviews in place of a workshop

	• Church of Scotland (Stats) • Methodist Homes
Jewish faith organisations	• London Jewish Forum • Board of Deputies of British Jews • Partnerships for Jewish Schools • Leeds Jewish Welfare Board • Institute for Jewish Policy Research[4]
Hindu faith organisations	• Hindu Forum of Great Britain • Hindu Council of the North • South London Hindu Council
Muslim faith organisations	• Amina MWRC • QED Foundation • MEND • Muslim Council of Britain • Muslim Council of Scotland
Sikh faith organisations	• Sikh Council of Wales • Sikh Sanjog • Nishkam SWAT
Migrants, asylum-seeking and modern slavery	• Migrant Help • Anti-Slavery International • Migrants Rights Network • The Refugee Council[5]
Children and young people	• Children's Rights Alliance for England • Princes Trust • Children's Commissioner for England • Young People's Health
Vulnerable children and young people	• Children's Society • Family Action
Pregnancy and maternity	• Gingerbread • Maternity Action • Mums Aid • Henry • Approachable Parenting
Prisoners and ex-prisoners	• Working Chance • Switchback • Unlock • Prison Reform Trust

---

[4] Included as an additional interviewee

[5] Included as an additional interviewee

Older people (aged 70+)	• Alzheimer’s Society • Independent Age • Friends of the Elderly • MS Society
Sexual orientation	• MindOut • Diversity Trust • BlackOut • Albert Kennedy Trust
Transgender, non-binary and gender diverse	• SAIL • Transgender NI • Gender Identity, Research and Education Society • Unique • The Beaumont Society
Women’s equality	• The Young Women’s Trust • Welsh Women’s Aid • Rosa • Women’s Resource Centre • Fair Play for Women • The Women’s Budget Group

While there was a relatively short fieldwork window within which to engage with CSOs we looked to accommodate stakeholder availability in the scheduling of workshops where this was sought.

A standardised topic guide was developed for use with each equality area. This is included in **Annex A** of this report, but in summary, covered:

- Current use of data and evidence within each organisation
- Key sources of information, including how this is accessed and what data are collected internally
- Detailed exploration of the findability, accessibility, interoperability and replicability of data, including geographic and temporal comparability
- Perceived accuracy and fairness of data analysis and presentation
- Examples of inclusive data and evidence seen to work well (and less well) in shaping policy and services
- Gaps in data or evidence, how these are currently managed, and the impact of these on policies, support services and outcomes for specific groups
- Views toward both the risks and opportunities relating to more inclusive data
- Key improvements desired for the data and evidence available, and the conditions needed for these

Stage Two involved a mixture of interviews and group discussions with over 90 members of the public recruited as having different protected characteristics or who are under-represented in the UK statistics. The approach used and achieved sample is detailed below and divided by equality area (see Table Two). Different approaches were taken to ensure that participants felt most comfortable and confident in participating in the discussions. These included Zoom (video-conferencing) group discussions and in-depth interviews, and in-depth interviews undertaken by telephone.

Table Two: General public sample

Equality area	Method and Sample
Physical disability or impairment	Group discussion (n =1), including at least one participant with: • Visual impairment x 2 • Hearing impairment x 1 • Physical disability x 2
Learning disability, neurodiversity or dementia	In-depth interviews (n = 6), including participants self-defined as having: • Mild disability x 2 • Moderate disability x 2 • Severe or profound disability x 2 carers[6]
Race or Ethnic equality	Group discussions x 7 • Black British (n = 5) • Black Caribbean (n = 5) • Chinese (n = 5) • Indian (n = 5) • Pakistani (n = 5) • Gypsy, Roma and Traveller (n = 5) • Somali (n = 5)
Homelessness	In-depth interviews with participants currently experiencing homelessness (n = 6): • Those using hostels x 2 • Those living in insecure housing like bed and breakfast x 2 • People who move around friends and family – 'sofa surfing' x 2
Faith	In-depth interviews (n = 12) • Buddhist faith x 2 (male/female) • Christian faith x 2 (male/female) • Hindu faith x 2 (male/female) • Jewish faith x 2 (male/female) • Muslim faith x 2 (male/female) • Sikh faith x 2 (male/female)
Migrants, asylum-seeking and modern slavery	In-depth interviews (n = 6), including participants identified as: • Migrants (including EU) x 2 • Refugees and those seeking asylum x 2 • With experience of modern slavery x 2
Children and young people	In-depth interviews (n = 6), to include children and young people aged: • 8-10 years old and their parents x 2 • 11-16 years old x 2

---

[6] For individuals with severe or profound disabilities we undertook interviews with individuals' carers

	• 17-18 years old x 2 Within this sample were two children and young people with care-experience
Older people (aged over 70 years old)	Interviews (n = 6) with individuals: • Living independently alone x 2 • Living independently with a partner x 2 • Dependent on outside care x 2
Sexual orientation	Group discussion (n = 4), including representation from individuals identifying as lesbian, gay and bisexual
Transgender, non-binary and gender diverse	Group discussion (n = 5) with transgender, non-binary and gender diverse individuals
Women's equality	Group discussion (n = 5) with women that have reported experience of inequality.

All members of the public were recruited via a professional, independent recruiter (Acumen) to enable this research to cover off all the quotas within the project timescales. Participants received incentives of between £25-£60 depending on the nature of their engagement with the research team. Across the sample we aimed to achieve a balance of age, gender and locations from around the UK.

Again, a standardised topic guide was developed for use with each equality area. The master copy topic guide for the focus group format is included in **Annex B** of this report, but in summary covered:

- Understanding and experience of providing personal data
- Perceived benefits and harms of sharing personal data
- Judging when and with whom to share personal data
- Using personal data to build a more inclusive picture of [NAMED] equality area
- Acceptability of different uses of personal data
- Rules and principles for the collection and use of data
- Top issue considered by individuals in respect to collecting, analysing and presenting data on [NAMED] equality area

All interviews and group discussions were transcribed and coded in NVivo. A coding frame was developed at the outset of the research, utilising the key themes covered in the workshop, group discussion and interview guides. This was refined on a group-by-group basis, to enable emerging themes to be included under new nodes. Interviewers individually reviewed transcripts and NVivo project file summaries before completing pro-formas for each workshop, group discussion or interview undertaken detailing key findings and supporting quotes aligned to the primary areas for discussion (and the objectives for this research). These were then mapped out on a Miro board and reviewed as part of group analysis sessions following each stage to agree the key findings and narrative for the report.

2. Data use by CSOs working with individuals with protected characteristics

2.1. How data are being used by CSOs, across equality areas

How CSOs engaged as part of this research varied in their remit both in terms of the equality areas served and in the way in which the organisations worked to meet their objectives. Many CSOs were working directly with individuals, as members or service users, to provide individual support as well as providing more group or community-level advocacy support. Some had a mixed remit, incorporating both service delivery and advocacy, while others focused more heavily on advocacy.

CSOs which provided some form of direct service to a group or community tended to use data to both inform and evidence their services. Understanding the needs of individuals and groups helped them to shape their service priorities at a strategic level, to inform service provision and target their own services effectively, and to adequately resource services based on demand. Capturing the outcomes for individual service users enabled CSOs to assess the effectiveness of services provided, and to undertake cost-benefit analyses. Finally, by evidencing the need for, and effectiveness of, services provided, CSOs gathered the evidence base needed to help support funding and grant applications, marketing, and organisational profile-raising.

"As the first generation becomes the second and third, we have a lot of demand, in terms of providing nursing homes or care homes for the elderly people, in terms of the language, culture, and food point of view. It may not be suitable for them to go into the ordinary mainstream care homes and nursing homes, where their life will be shortened for those very reasons. I think we have established that. In my view, we need proper data on how large a population we have, over 70, over 65, which we can't do ourselves." CSO working with Sikh communities

CSOs involved in advocacy collected, collated and utilised data relevant to their organisational priorities to influence public policy, services and legislation. This primarily involved developing an evidence base that could be used (or made available) to help inform local, regional or national services. The evidence could also be used to lobby for changes to legislation or practices, including through responses to public consultations from the government. Data also provides an evidence base that CSOs can use in responding to media stories and to advocate on behalf of specific communities in the media, although some encountered challenges accessing the necessary data to do so. A small number of CSOs produced regular quarterly or annual reports collating and presenting publicly available data on a specific equality area.

"We tend to use our data either to inform what our policy priorities should be and to look at what issues are key children's rights issues we should be pushing to rectify, and also to support our cause in terms of making a strong case for change." CSO working with children and young people

"From a policy and influencing perspective as well, it just provides us with a strong platform on which to challenge government and policy-makers. The more we know about the scale of the issue the easier it is to make the case for change." CSO working with people who are pregnant or on maternity leave

Three primary types of data were utilised by CSOs, across the different equality areas:

1) **Data collected directly from service users or membership** in those CSOs that provide direct support or services to a given community. This included:

    - Administrative-type data to profile service users or members (e.g. age, gender, ethnicity, presenting needs and outcomes as relevant to CSO remit)
    - Surveys of service users or members to gather more comprehensive data on lived experience (i.e. attitudinal or experiential data)
    - Case studies of service users developed from qualitative interviews with service users

2) **'Public data'** collected, collated and publicised predominantly by government departments (often via data.gov.uk or the UK Data Archive) and ONS. This also included local authorities, police authorities, Clinical Commissioning Groups (CCGs) and NHS Trusts, as well as other organisations (such as non-departmental bodies, charities and academic institutions). These data included:
    - Foundational data collated by the ONS, with the Census seen as the gold standard in terms of data collection due to granularity of data analyses afforded by the size of sample and ease of access to aggregate data, microdata and easy-to-read summaries. A small minority of participating CSOs did still highlight that Census data can be challenging to access and navigate for lay users (i.e. people without training or familiarity in using datasets).
    - Wider national survey data or administrative data sets, which varied depending on the relevance to CSO remits and the level of research knowledge and capability that existed within organisations. Commonly referenced data included: Understanding Society, GP Patient Survey, Labour Force Survey and Individualised Learner Record data.
    - Various 'tools' were referenced positively by CSOs as enabling a greater level of data analysis and disaggregation. These included: PHE Fingertips, DWP Stat-Xplore, and ONS Nomis.
    - Freedom of Information (FOI) requests were relatively common within some equality areas (notably in relation to prisoners and ex-prisoners, children and young people, migrants and asylum seekers, and homelessness) to access data that were either not made public or not provided in a format which enabled disaggregation by the key characteristics of interest.
3) **Research commissioned on behalf of CSOs** to fill evidence gaps in data held on service users or gaps in 'official' published data.

The most useful data for CSOs were that which allow for a robust and granular level understanding of a specific group or community to enable targeted actions to be taken in response to an identified need (either at a service level or a policy level). As such, valuable data include:

- **Sufficient sample sizes** and presentation of disaggregated data (i.e. via a tool, dataset or summary) to enable subgroup analysis based on other protected characteristics (notably age, gender, ethnicity, faith and disability) or location. Here it is important that protected characteristics are themselves disaggregated at the most practical, granular level, and not presented in generic groups or bands (e.g. '15-19-year-olds', or 'BAME') which limit CSOs' ability to understand which specific groups are impacted by a given issue.

  "A lot of the large datasets, the size of the sample that's collected for pregnant women is small. I looked at the National Diet and Nutrition Survey last week and that mentioned that the sampling of pregnant women can be difficult because the number of pregnant women present in household-based surveys may be small." CSO working with people who are pregnant or on maternity leave

- **Data which are harmonised** in terms of the way in which they are collected and represented to enable CSOs to have confidence in the quality and reliability of the data, and to allow for accurate comparisons to be made over time.

  "Because of small sample sizes we cannot rely on existing surveys so try to incorporate harmonised questions to allow us to compare with national data sources." CSO working with people of Jewish faith

  "I do a lot of work with local authorities who provide the data and they have asked us very emphatically not to use it because the quality is so poor and they provided us with evidence of, 'This is what we returned to DfE because of how they asked for the information to be returned, these are our actual figures,' and DfE's figures are like 20% out for various reasons." CSO working with children and young people

- **Data which are linked** - or enable links to be made - to other data to add a new layer of understanding of a given issue. For example, this might involve linking datasets across different Departments (e.g. children and young peoples' educational outcome data from DfE and household employment data from DWP) or datasets that allow different units of analysis (e.g. parent and child data within Understanding Society). This was mentioned by a smaller minority of CSOs, largely focused on advocacy, who had an in-house data analytics function or capabilities.

  "Much of the data does exist. It's just getting it to the right place and linking it with other information to make it really useful, which just doesn't happen." CSO working with children and young people

- **Timely data** which presents an up-to-date, representative picture of a given group to enable services to meet actual, rather than presumed, needs, and to enable CSOs to respond to issues that are time-sensitive either due to the policy timetable, political attention and/or media coverage.

  "I know that the raw data will get released in, like, 6 weeks' time or something like that. So, by then, the story's, kind of, gone a bit. You can't get traction." CSO working with disabled people with physical impairments

- **Qualitative and quantitative data,** which are both seen to be integral for understanding more marginalised groups. Quantitative data are recognised as important for assessing the scale of needs/issues and is required for advocacy, influencing and service evaluation activities. Qualitative data are required for understanding intersectionality and lived experiences due to the current limitations of both quantitative methods and achieved sample sizes. Qualitative data are particularly important for fundraising and service development.

  "In terms of policy, the quantitative data is really important in terms of reaching out and having an impact on the wider public. You really need the anecdotal evidence because that's what tells the story. The question of having the both of them. One has the effect on government, and local policy, and local governments, etc. The other one allows you to convince people this is really happening." CSO working for race and ethnic equality

  "I think it's useful to have statistics… People take notice if the figure is high or significant or there's been a change. Sometimes, you need to have that figure to give people an indication of what we're talking about here in terms of the scale. So, I do think it's important, but I do worry if it's particularly accurate or insightful, shall we say, that people may have to just tick the box that best applies, but maybe doesn't actually reflect their status or their situation." CSO working with disabled people with physical impairments

These are further expanded on in Section 3.1. looking at the challenges and opportunities that CSOs see as impacting the inclusivity of data.

The key differences in perspectives of CSOs evidenced through this research related to their remit and resources. Those organisations more heavily involved in service delivery placed value in data that could be tied more closely to a specific locality and beneficiary group (e.g. female Muslims in Glasgow), and therefore had more limited use for national data collected through surveys such as the Census, or the time and resource to invest in trying to disentangle public datasets. Those organisations with more of an overt advocacy role valued access to larger administrative and population survey data that offered the potential for disaggregating data across different protected characteristics to help identify issues or qualify more anecdotal data. This more “robust” data was seen as critical for helping to influence policy and legislation.

2.2. How data are managed by CSOs

Data collected from service users and members are typically collected by service providers as part of the standard process of registering someone to that organisation. Where appropriate, data are also collected to monitor the type of service received by a user or member, and the associated outcome of that service. The actual capability and processes for managing data varied considerably between CSOs, dependent on the characteristics and vulnerabilities of the people they worked with, the types of services provided, and the perceived level of risk involved in the data being collected. Those CSOs providing professional services, often with local or central government contracts, had dedicated resources, processes and platforms for managing service user data (e.g. using software such as Salesforce/In-Form) while those CSOs providing membership services (notably some faith organisations), tended to have less formalised processes, skills and capacity, meaning the data were typically more rudimentary (often limited to contact details). All CSOs reported an awareness of GDPR and of the issues relating to data protection which influence the data collected and shared between organisations. For smaller CSOs and those more involved in direct service delivery, GDPR was seen to hinder data collection and sharing due to concerns around breaching data protection legislation.

Most CSOs involved in advocacy activities were looking to access and utilise publicly available data to inform services and advocate on behalf of individuals/groups. However, this was typically a feature of a Policy role as opposed to a dedicated research or analyst role; of those engaged, relatively few CSOs had access to internal research resources dedicated to collecting and analysing external/public data relating to equality areas. Notable exceptions included organisations such as the Office for the Children’s Commissioner, the Prison Reform Trust, and the Church of England, Church of Scotland and Methodist Homes, all of whom employed individuals that had specific statistical and/or secondary data analysis skillsets.

Quantitative data analysis was recognised as being a specialist skillset, particularly where it involves combining data sets (requiring statistical skills and an understanding of data protection), weighting and extrapolating data. Given the variable resource/skills internally to source and scrutinise the statistics made publicly available, CSOs expressed a desire for access to summative data, though there was recognition that these summaries seldom answer the range of questions CSOs have about the needs of subgroups (typically based on protected characteristics, location/geography and socioeconomic factors).

3. Improving the inclusivity of data on individuals with protected characteristics – the CSO perspective

3.1. The challenges and opportunities for 'inclusive data'

A wide variety of issues were identified by CSOs, impacting their access, management and use of data, inhibiting the extent to which statistics reflect the experiences of everyone in society. There was recognition, with some variation between equality areas, that CSOs potentially had access to a high volume of data but that (i) determining the provenance and reliability of the data available is a challenge, and (ii) deriving actionable insight from the data that are available is problematic, due to data gaps or data aggregation, inhibiting its value.

The primary challenge facing CSOs is the lack of granularity in data to enable an understanding of specific subgroups (e.g. female offenders) and the intersectionality of protected characteristics. The value of data is in providing a reliable understanding of the needs and circumstances of different groups of people, so that information that can be translated into insight that allows for action to be taken. CSOs identified numerous barriers to this, including:

- **Data not being provided (or collected) in such a way as to allow for disaggregation around key characteristics**. Accurate statistics, at either a local or national level dependent on each CSO's remit, is fundamental to CSOs' (and policymakers') understanding of issues impacting the different communities. The current practice of aggregating protected characteristics into broader age, ethnic, faith, disability or sexual orientation categories is seen both to severely inhibit the inclusivity of statistics and to potentially misrepresent the issues and needs of smaller and more marginalised communities. The aggregation of characteristics (e.g. 'BAME', 'Asian' or 'Gypsy, Roma and Traveller'; 'learning disabilities'; or 'LGB+') was seen as misleading, if not potentially "offensive" to members of these groups, due to significant sub-group differences, as well as impacting on the ability of CSOs to understand the needs of the specific sub-groups they worked with.

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

  "If you're looking at policy and trying to effect change, you don't really want to be told that this is not really something that is measurable, it's more anecdotal. Yet again, that data is really important because it's not captured anywhere else, because those categories for those ethnic groups aren't actually mentioned in some of the larger data sets." CSO working for race and ethnic equality

- **Lack of harmonisation in the administrative data collected** from public service use (e.g. schools, police forces, health services) resulting in an inconsistent picture of particular subgroups (notably relating to faith and ethnicity) and misalignment between "official" data and those collected by CSOs on the ground. An example of this is the voluntary nature of responding to faith questions on the Census. This is seen to create a precedent that impacts the inclusion and wording of questions on faith in other surveys and in administrative datasets. There is a desire for ONS to provide more direction around the data collection and questions that should be included, particularly across administrative datasets. There was generally a high level of support for data harmonisation, assuming that this was done sensitively and with thought on the potential implications and harms.

"A lot of organisations are nervous around the discourse around sex and gender identity and they're worried about getting into trouble, they're worried about people attacking them. If they've got an excuse to say, 'We're not going to collect this data,' they will." CSO working to support women's equality

- **Missing data on marginalised and minority communities**, particularly where these are: more recently defined or less well understood groups (e.g. neuro-diversity); "harder-to-reach" (e.g. Gypsy, Roma and Traveller, ex-offenders, asylum seekers); where data are collected in a non-standardised manner (and therefore expensive to collate and publish); or where there is seen to be limited political interest in publishing data (e.g. numbers of asylum seekers or people who are homeless).

  "As a practical challenge, it is hard to get Home Office officials to publish the data." CSO working with migrant and asylum seeker communities

  "I think one of the issues that we find is that obviously, data and evidence is being used by policy makers all the time, but I think sometimes it can be a challenge to know what information the government holds. I think we just don't always know what is being collected." CSO working with people who are pregnant or on maternity leave

- **Statistical data does not provide a comprehensive understanding of the lived experience.** Aligned with a lack of granularity was also an acknowledgement that quantitative data alone can have limited explanatory power, particularly regarding the use of "administrative-type" data. Given the complexity, richness and intersectionality of issues impacting people's lives, CSOs saw it as imperative that qualitative data are used in addition to statistics, enabling a fuller understanding of lived experiences. There was concern expressed at the lack of value placed on qualitative data (and insight) by funders and policymakers. CSOs felt that greater inclusivity could be fostered through developing a more holistic picture of communities through both qualitative and quantitative data.

  "However you do it, I think moving away from forms of categorisation, and a kind of binary statement of religion or non-religion or identity to something that actually has more kind of colour and quality to it and depth I think, but you know it would be more complex of course it would, but it would actually be more accurate." CSO working with people of Christian faith

- **Subjectivity of experience.** CSOs working in areas of ethnicity, gender, sexuality, disability and faith also highlighted that personal identity is subjective, complex and fluid which presents challenges for gathering more quantitative, closed responses. Individuals could be seen to find it challenging to identify where they fit in the categories presented (particularly in respect of fixed/static conceptions of ethnicity, sexual identity, gender identity, disability and faith) which can also lead to non-participation or missing data. Greater consultation with CSOs working within equality areas was seen as helping lead to the development of more appropriate measures and methods for collecting data.

- **Recency of data**. The Census was described as a foundational source of data, however, being undertaken once every 10 years means that data used to understand the population (and to plan policies and services) can be many years out of date. Similarly, the Indices of Multiple Deprivation is only updated every 5 years. The lack of recency means many services are reliant on qualitative research and more anecdotal data to understand the needs and circumstances of groups with protected characteristics. These are valued by CSOs though recognised as less robust data for service planning in terms of estimating absolute levels of service need. Similarly, there is frustration that publicly funded deep-dives on key equality areas (such as sexual orientation) are undertaken infrequently and seemingly with no long-term strategic intent. On the other hand, other CSOs (e.g. trans, non-binary and gender diverse, and some ethnic minority groups) raised issues with their communities being over-consulted around the same topics, which often led to feelings of frustration when no changes were observed as a result of their frequent participation. This links to barriers to research participation, discussed later in this section.

"Sometimes, ONS will have trend data contained in each release. Sometimes, it won't. So, sometimes, if you want trend data, you'll have to go back to each individual ONS release, and basically construct your trend data yourself. Or, they'll only do trend releases every 5 years, or something. Trend can sometimes be a bit of a tricky one." CSO working with older people

- **Lack of data linkages** between different departmental datasets. While there was limited understanding or dialogue across CSOs in relation to what is possible here, a smaller number of CSOs (with analytics capabilities and more of an advocacy focus) felt that the lack of linkages between current administrative datasets is a missed opportunity to better understand under-represented groups. This was seen to result from a lack of a common, unique identifier (such as an NHS number) to allow for individual and household comparisons to be made across different datasets, but also a "siloed mentality" in government policy and services.

  "A lot of the time what you have is data collected in silos by different government departments. What we'd want to see and what we're trying to do is get better linkage between those different datasets, so that we can understand the full journey." CSO working with children and young people

  In some cases, this was felt to be exacerbated by weak or missing legislation (with some specific references to DWP and ONS in respect of legislative limitations to sharing and requesting data). One CSO expressed a view that the framework for sharing personal data introduced under Digital Economy Act was not explicit enough to enable the dataflow into ONS. Given the resource and skills that exist within ONS there was a call, amongst this more data-literate minority of CSOs, for increasing their powers to enable them to play a stronger role in collating and connecting data in a meaningful, GDPR-compliant manner. This was not something discussed consistently across equalities areas and therefore we cannot comment on the extent of consensus around views toward data linkage.

- **Lack of readily available, easily accessible data for CSOs.** Larger national CSOs, particularly those with more of an advocacy remit, tended to have a good understanding of the key data available on the groups which they worked with. However, smaller and more local CSOs, reported greater uncertainty as to exactly what data was collected, and then what data was made available and where, either by government, academia, or other CSOs.

  "I think sometimes it can be a challenge to know what information the government holds. Particularly when we do freedom of information requests or parliamentary questions, for example, data might be released and it doesn't really tell you exactly what is held and what the scope of the evidence is that's there, that we could draw on, if that makes sense. I think we just don't always know what is being collected." CSO working with individuals who are pregnant or on maternity leave

The other key challenge for CSOs is the lack of internal capacity and/or expertise to find and analyse data. As mentioned previously, it was relatively rare that a CSO had dedicated research capacity, or the skills seen to be necessary to source and analyse data. There is a recognised tension, amongst CSOs with more experience working with data, between the desire to have access to more granular data and the implications this then has for data users. Most CSOs do not have research skills in-house and would prefer to have access to more summative data split by key socio-demographics and protected characteristics at a sufficiently granular level. This is particularly the case for those working more directly with service users. At a minimum this would entail data being provided in a way that enabled CSOs to look at the intersectionality of protected characteristics, ideally further allowing the data to be broken into smaller subgroups (e.g. individual ethnicities, ages, disabilities etc.)

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

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

While not explicitly referenced by CSOs participating in this research, it is our view that many of these challenges relate to **data foundations** (data that is fit for purpose, recorded in standardised formats on modern, future-proof systems and held in a condition that means it is findable, accessible, interoperable and reusable) and **data skills** (including data assurance, data processing and data analysis). These challenges have been identified in the recent National Data Strategy (2020) which presents a framework for the action government intends to take on data.

Finally, CSOs also highlighted the barriers that exist for under-represented and vulnerable groups to engage in research. These can broadly be summarised as follows:

- The circumstances and needs of different groups mean there are various practical, cultural and emotional **needs influencing their participation in formal research**, including: access by researchers (e.g. to the prison population, Travellers, asylum seekers); language, literacy or comprehension needs; persecution, discrimination and exploitation resulting from previous disclosure of information and views; exhaustion from over-research; trauma associated with the topics covered; and simply the competing pressures faced in day-to-day life (e.g. managing with a physical disability). A one-size fits all type approach to data collection therefore would not work for many groups, instead there is a need to adopt a more person-centred approach, with alternative data collection approaches developed in collaboration with the communities of interest.

  "If the intention is really wanting to reach out to the Chinese community, at least I think the effort has to be made to have translated to make sure the people can respond quite readily in their own native tongue that they feel most at home in a sense." CSO working in support of race and ethnic equality

- **Trust and trustworthiness of 'authority' and public services**. CSOs mentioned that many of the communities they worked with felt marginalised as a result of negative experiences with public services and government. This in turn influenced the trustworthiness of public authorities and can result in a level of disenfranchisement and unwillingness to disclose information collected on behalf of those public authorities. A breakdown in trusted relationships with government and public services was seen to be a particular issue for certain ethnic groups (notably Black British, Black Caribbean and Gypsy, Roma and Traveller), disabled people, asylum seekers and prisoners or ex-prisoners.

  "If I put down I am from a particular ethnic group, a particular faith, will that impact on the service I am going to be getting? I suppose people are also thinking about that and that is part of the trust issue." CSO working for race and ethnic equality

> *"People want to stay in the country and they fear if they share their circumstances, that officials, Home Office, local authorities, etc, will take negative action against them."* CSO working with migrant and asylum seeker communities

- Individuals from certain protected characteristics groups are seen by CSOs to be relatively over-researched. In particular those where trust barriers exist to sharing information *and* who have engaged with support services are repeatedly asked to provide administrative data which categorises them in a pre-identified, narrow and potentially unrepresentative manner. CSOs working with these groups highlighted that members could feel that **information can be gathered in quite an extractive manner with little to no personal benefit**, or seeming impact on public policy or services, which creates further distrust and lack of desire to participate in research.

> *"I think we quite often come across where people from black minority community will say they're tired. They are tired of being over researched because they participate in research, you have the findings, you have the evidence but then nothing happens in practice."* CSO working for race and ethnic equality

> *"There are lots of issues around collecting data in general, I mean, why are we collecting data? Even the DfE that has data, we know the outcomes are awful in education, the exclusions are really high, are the highest for any ethnic group, and still nothing really changes, so people are a bit cynical about how and when data is used."* CSO working with Gypsy, Roma or Traveller communities

- In tackling these challenges to improve the inclusivity of data, CSOs saw an opportunity for those wanting to collect data to work through trusted intermediaries, **to actively involve minority groups in the development of research approaches**, and to publicise what has changed as a result of needs and issues raised in their responses to questions.

> *"Involving disabled people in how to collect this from the ONS's point of view would be really good, actual people themselves. Obviously, we represent disabled people but perhaps starting up some expert group which would be disabled people with lived experience talking to you guys directly would be good."* CSO working with disabled people with physical impairments

3.2. The impacts of gaps in data and how gaps are currently addressed by CSOs

The challenges highlighted above were seen to result in gaps in data and understanding that, in turn, result in less effective policies and services. Where there is a lack of data CSOs saw a danger that issues are simply not recognised or understood, and instead can be downplayed as marginal or minority concerns. It was felt to prevent important conversations from taking place (e.g. around hate crime, hidden harms, equality) and becomes challenging for CSOs to demonstrate the scale of an issue and therefore a need for services (and to gain funding for delivering these). It also restricts the development of a coherent body of evidence around a given issue or group, including intersectionality, comorbidity, risk factors and prevention. This in turn assures that any work undertaken by government and by CSOs in support of more vulnerable groups remains reactive and siloed rather than proactive and systemic.

> *"Policy is being made without a clear understanding of the population."* CSO working with migrants and asylum seeker communities

“We don't really know how effective our criminal justice system is and particularly our prison system is at stopping people from committing crime in the future.” CSO working with prisoners and ex-prisoners communities

“It gives organisations and individuals who might be apathetic to our plight the wriggle room to suggest that we don't exist, or if we do, then it's an incredibly marginal issue.” CSO working to support equality between sexual orientations

Ultimately, the impact of gaps was seen to be felt by those individuals and communities from specific protected characteristics groups. This can range from discrimination, misrepresentation and reduced life chances, right through to hidden harm and loss of life for those in highly vulnerable circumstances.

Moderator: *“What do you think is the impact of these gaps in knowledge and understanding on both polices and services, and on outcomes for children and young people?”*

Participant: *“I don't want to sound over dramatic, but children die. I think the outcomes can be the absolute worst. It's beyond they don’t get 5 GCSEs. That is the very worst end of the spectrum but it's happening.”* CSO working with children and young people

A lack of inclusive data was also seen to have the potential to disenfranchise people through not accounting for their circumstances, views and needs in services and policies designed to serve different communities, resulting in an absence of positive role models. This was said to contribute to a cyclical reduction in trust, further disengagement and a lack of data, further reducing understanding and effective policies and services.

The challenges identified through this research result in gaps in data which CSOs typically attempt to address through a combination of:

- collecting their own data through bespoke primary research delivered internally or through an external provider (ranging from individual researchers to academic partners)
- deploying internal staff resource to undertake additional data analyses, or looking to access analyses of other partners (CSOs and academics)
- submitting Freedom of Information requests to access data which are not otherwise publicised
- cross-referencing the data that are publicised with other current/historical data held in a given equality area to look at trends and key differences/changes
- guesstimating prevalence of groups/issues by weighting service-level data using national-level data sets, or using anecdotal information based on professional judgement or more qualitative forms of data (which may be accurate but can lack credibility with external audiences, policy makers and funders)

However, dedicating time to undertake research or seek evidence to fill the gaps in data presents a challenge for nearly all the CSOs we spoke to, as staff resources need to be invested in collecting, organising and presenting data rather than in delivering services or focusing on advocacy activities. Only in a minority of cases did CSOs have the capacity, skills and remit to actively invest time in this. For the majority of CSOs, regardless of their focus on service provision or advocacy, looking to fill gaps in data was done out of operational necessity rather than being a strategic priority for the CSO.

“There may be quite a wealth of data in these big, longitudinal surveys, but it doesn't really make it to the light of day in terms of public statistics. You have to go in there and find it yourself. If you're a smaller

charity it can be quite difficult and quite frustrating, because you're like, 'The data is there, we just can't get it.'" CSO working with older people

3.3. The risks of collecting and making available more inclusive data

While a wide range of benefits coming from more inclusive data were recognised by CSOs, there were also actual or potential risks that could come from more inclusive data (which would often involve more people being asked for more data). People are becoming more aware of their rights in respect of their data, but a small number of CSOs highlighted a danger that more inclusive data collection could occur without individuals being aware of their rights. Members of the general public will vary in their ability to provide truly informed consent (e.g. for reasons of literacy or mental capacity) and, as such CSOs consistently highlighted a need for consent and data protection to be part of a considered ethics process for any data collection, as is the case with national population surveys.

"I think the major risk is to the individuals taking part. People come with a lot of emotional baggage. They've gone through an awful lot in their lives. You want to get their consent, first of all, and make sure it's ethically appropriate, whatever research is happening and that they can leave the research at any point they want. The last thing you want is for them to feel even worse, or less comfortable, as a result of that research process." CSO working with individuals with learning disabilities, neurodiversity or dementia

Improper use, sharing and publication of data (even aggregated data) was seen as having the potential to be disclosive, in breach of GDPR and potentially presenting risk of harm to individuals. To be effective custodians of data, CSOs need the resources (both the time, capacity and expertise) to adhere to GDPR; for example, on the protection of data, writing data protection impact assessments, and data sharing agreements. Enabling access to funding and grants for CSOs, as well as a credible source of information and guidance, were identified as ways to support CSOs in better managing data should they have a desire to be more substantively involved in data collection.

CSOs, across equalities areas, highlighted that more inclusive data have the potential to be used to discriminate or persecute communities, both at an individual and group level. For example, discrimination of Roma groups on planning applications or ex-prisoners in employment processes, or cherry-picking statistics to promote damaging narratives in the media. Some groups are very conscious of being "counted" due to historic persecution (e.g. Jewish communities, Gypsy, Roma and Traveller communities, or refugee and migrant communities), which can also impact their willingness to participate in research. In looking to make data more inclusive, CSOs felt it was important to also account for the past experiences of individuals and communities with data collection to ensure that processes do not inadvertently present barriers for certain groups. One way of achieving this is through greater consultation with CSOs working with these communities early in the research process, "from start to finish".

"That failure to understand what people are doing when people are disclosing, and for Black men, for Black women, their visibility can be about danger as much as it can be about empowerment." CSO working to support equality between sexual orientations

Collecting more data (either in scale, frequency or form) needs to account for the additional burden placed on individuals, and to be clear on the personal/community benefits that collecting this 'more inclusive' data confers. For more under-represented groups, CSOs identified a high degree of fatigue in providing their data and a concern that trust is further compromised where situations do not change in response to the views provided. This also impacts the ability of CSOs to work with these communities. CSOs felt it important to take a participatory approach to developing research and data collection methods for there to be a very clear rationale for collecting data. In addition, evidence of the value of data collection needs to be relayed to those participating.

“*I think we quite often come across where people from black minority community will say they're tired. They are tired of being over researched because they participate in research, you have the findings, you have the evidence but then nothing happens in practice.”* CSO working for race and ethnic equality

4. Data use, challenges and opportunities disaggregated by equality area – the CSO perspective

4.1. Disabilities associated with physical impairments

The key data for CSOs working with disabled people with physical impairments are data around the prevalence and comorbidity of different disabilities, including prevalence data structured using more actionable framing (such as the Washington Group question set on functioning[7]). These were seen to be particularly actionable as they help to identify and address the structural barriers faced by disabled people. Qualitative data on lived experiences (often in the form of case studies) was also important to help generate a more comprehensive understanding of the needs of disabled people, and how structural barriers are overcome (either by services or individuals themselves).

"We tend to focus more on collecting qualitative data and I think that's because the sort of campaigns that we run tend to focus on the impacts of being disabled, on people's lives. We often want to know people's stories." CSO working with disabled people with physical impairments

The primary sources of data utilised by CSOs engaged in this research included data available from ONS and data.gov.uk that allowed for disaggregation of disability, for example through the Labour Force Survey, Family Resource Survey and Understanding Society.

The key data challenges for CSOs working with disabled people with physical impairments included:

- A lack of granularity in data available to allow for (i) disaggregation of disabled people with physical impairments (ideally by location), or (ii) disaggregation of disability data by other characteristics (such as age) to understand intersectionality. A related issue is the huge degree of variation within some forms of disability (such as hearing loss) which limits the value of aggregated data on certain disabilities.

  "Getting hold of raw data, I find, is often tricky and you have to apply for special access often, particularly with disability data sets. And we often are reliant on the standard press releases which have the data accompany them to start looking at things a bit, rather than the bigger data sets." CSO working with disabled people with physical impairments

  "We often end up having to go to, say, polling companies that have panel-based people to get some of the level of detail that we're interested in a survey scale sort of size, as well as the qual work that we do as well. And that's because the ONS isn't satisfying what we want to find out." CSO working with disabled people with physical impairments

- Missing data, either because it is seemingly not collected (for example, the number of wheelchair users in the UK), or because individuals are not participating in research (due to more pressing issues impacting their ability to participate, distrust of the governments intentions, or greater use of more temporary forms of accommodation). Some individuals will also not self-identify as disabled as they may not consider themselves to have a disability.

---

[7] Note that the Washington Group was referenced by CSOs participating in this research, though other measures of disability are available, such as the WHO-DAS scale and the Government Statistical Service harmonised impairment questions which also focus on areas of functioning affected

> *"You're obviously relying on people reporting the figures at the time. Do people always declare that they are disabled or consider themselves to be disabled?"* CSO working with disabled people with physical impairments
>
> *"I think trust is a big issue. Disabled people are often framed as taking the piss and claiming benefits when they shouldn't be, and there's this real distrust with participating in surveys."* CSO working with disabled people with physical impairments

- Questions around physical disability which are seen to be outdated and based on a deficit model for conceptualising disability that does not sufficiently capture the lived experiences of individuals (including needs, structural barriers encountered and overcoming barriers).

> *"I have thought about how disability is captured in this country because it's based on the Equality Act definition on the whole, which is basically, 'Do you have something that affects your day-to-day ability?' And that's obviously medical model based anyway. I did wonder whether the approach might well be about access needs and requirements of society to support you? So, it's much more framed around not what your condition is, but more how that manifests in a way which you could have support structures around you. Which I personally think people might be more willing to share because it's more personalised."* CSO working with disabled people with physical impairments

The key opportunity for improving inclusivity for CSOs working with disabled people with physical impairments would be the collection and representation of data that presents the lived experience of disabled people and the structural barriers they face. The availability of statistics is helpful but there is a need to supplement this with qualitative research to understand the stories behind the statistics. This requires building a trusted relationship with communities of disabled people which was seen as best achieved by working through CSOs.

"What we're interested in really is, as a disabled person, what's your experience of employment? Do you want to work? Do you have a job? Is the housing where you live, is it accessible to you? Is it sufficient? Is it of a standard that's acceptable to everybody? They're the sort of questions we should be asking. What are the impacts on your day-to-day life and what's missing? What is it you need? Not so much, 'Tell us the name of your condition or conditions that you have." CSO working with disabled people with physical impairments CSO

A secondary priority is enabling CSOs greater access and use of data which will result from streamlining applications/access to datasets, but also from collecting and presenting data that allows for easy disaggregation of different physical disabilities alongside other protected characteristics.

4.2. Learning disability, neurodiversity or dementia

The key data for CSOs working with individuals with learning disabilities, neurodiversity or dementia are data around the prevalence and comorbidity of different learning disabilities. CSOs also value qualitative data on the lived experience of individuals as the complexity of conditions (in terms of associated conditions and impacts on people's capacity to develop skills and relationships) can require both statistics and personal stories.

There is seen to be a dearth of publicly available data published by ONS or other government departments such as DHSC. In theory, the data provided by PHE, and at more of a local level the Joint Strategic Needs Assessments produced by local authorities and Clinical Commissioning Groups (CCGs), should be helpful. However, learning disabilities are often not identified in these data. Instead organisations rely on ad-hoc studies and prevalence data identified via Google, and through undertaking their own research (e.g. Westminster AchieveAbility Commission) to provide data on prevalence. Mixed-method research with service users is a key source of data for CSOs, as is data resulting from partnerships with universities in which academics access funding to undertake research that CSOs can then feed into.

"We had to invent our own way of sourcing the data because the data wasn't there, in order to get the neurodivergent voice truly heard. We did it through social media and inclusive surveys and interviews." CSO working with individuals with learning disabilities, neurodiversity or dementia

"You have to have inside knowledge to know where to go. There are pots of data at different charities and different universities." CSO working with individuals with learning disabilities, neurodiversity or dementia

CSOs referenced research undertaken on specific issues, such as neurodiversity, which have helped both to inform the work of other charities and ensure that less well represented issues were are accounted for in inclusive policymaking. An example was referenced in relation to ensuring the Welsh Government was aware of the support needs of parents of children with learning disabilities as part of looking at how to mitigate against children being taken into the care of the local authority.

"Our report did have quite a massive effect on several small charities, and it supported the whole lobby on 'Nothing about us without us' and making sure the neurodivergent voice is central to all lobbying and discussions over policy." CSO working with individuals with learning disabilities, neurodiversity or dementia

The key data challenges for CSOs working with individuals with learning disabilities, neurodiversity or dementia included:

- A lack of data, due to learning disabilities and neurodiversity information either not being collected, or not being reported on, in national datasets. Where it is published, there is a further issue with (i) a lack of harmonisation between local authorities in how information is represented, and (ii) a lack of granularity or ability to allow for disaggregation by locality or other protected characteristics. CSOs in the devolved nations are often forced to use data from England to try and assess prevalence or need in their own region. This often leads to inaccuracies in the conclusions drawn around prevalence or need of those with learning disabilities.

  "The data situation for learning disabilities in Wales is pretty dire. We end up having to rely on data from England because we just don't have data in Wales." CSO working with individuals with learning disabilities, neurodiversity or dementia

- The accessibility and dissemination of data collected as part of studies within academia. While a range of relevant research may be taking place, CSOs working with disabled people may not be aware of it or have access to the findings.

  "There seems to be a disconnect between academia and the third sector. We struggle to get that. It gets published in academic journals, but it doesn't necessarily get shared wider than that. It almost feels like research for research's sake. We're crying out for this stuff, and yet we don't know it's happening, or we would have like to been involved in that research, whether that's just signposting researchers to groups of people that we think would be fantastic to be involved, and then also seeing the findings of that." CSO working with individuals with learning disabilities, neurodiversity or dementia

- Due to the specific needs of individuals with learning disabilities and neurodiversity (which will differ dependent on individual ability), the data collection methods used for the general population may not be appropriate. This may lead to issues with trust in the researcher, informed consent, negative experiences of participating and the validity of the data collected.

  "One thing I realised early on when I was working with parents is that it was important what I wore. There was a lot of suspicion when I first started. Parents wouldn't talk to me, because they believed I was going to feedback to their social worker and I was there to spy." CSO working with individuals with learning disabilities, neurodiversity or dementia

  "There is no one fit. So, you have to have a range of research tools that will reach a range of neurodivergent people. You can have a survey. You can be dealing with interviews. You can be doing online work, telephone stuff. You have to make sure the same questions are happening for consistency throughout the research." CSO working with individuals with learning disabilities, neurodiversity or dementia

  "People with a learning disability are very good at saying what they think people want to hear, and hiding and masking their learning disability, because they have learned that's often what people want. If someone says, 'Do you understand?' they know they're supposed to say yes, and that bad things might happen if they admit they don't understand what's going on, they will take my children away, rather than saying, 'No, I don't understand. I need some support with this'." CSO working with individuals with learning disabilities, neurodiversity or dementia

The key opportunity for improving inclusivity for CSOs working with learning disabilities, neurodiversity or dementia would be the collection of data on learning disability and neurodiversity to fill the current gap in the evidence base around prevalence. These data need to be collected sensitively (i.e. adopting a person-centred approach to thinking about methods and questions) and focus on capturing the lived experience of individuals. Once data has been collected it will be valuable to provide these data in such a way that enables CSOs to look at the intersectionality of learning disabilities and neurodiversity with different protected characteristics such as age, gender and ethnicity.

4.3. Race or Ethnic equality advocacy and support groups

The key data for CSOs working for race and ethnic equality are ethnicity data split by key life outcomes of interest, where equality of outcomes may be impacted by ethnicity (e.g. health, social care, education and employment). Most recently this has included COVID-19 infections, treatment, deaths and vaccinations. These data can also include specific questions around discrimination due to ethnicity, nationality or caste (for example, in terms of hate incidents, employment or housing decisions); greater specificity enables a stronger depth of understanding.

CSOs are accessing a wide range of local data (via local authorities and CCGs) and more national population level data through ONS, government departments and non-departmental bodies (e.g. English and Scottish Housing Surveys, National Pupil Database, PHE Fingertips). The Census was referenced as a highly important resource for CSOs, particularly amongst those working in regions of the UK where samples on other studies are too small to allow for meaningful analysis due to repression of data and associated wide confidence intervals.

"I'd agree the census is a really rich and useful source of information specifically because it systematically collects information on ethnic minorities. Other data might not even be desegregated in the same way.

They might use different categories. That makes it extremely complex." CSO working in support of race and ethnic equality

Specific reference was made to an attempt to have caste recognised as a protected characteristic. While the legal case was, ultimately, unsuccessful, the data that CSOs supplied to indicate the number of people that may be impacted by caste discrimination was seen to be integral in getting as far down the legislative process as they did.

The key data challenges for CSOs working for race and ethnic equality included:

- A lack of granularity in majority of data available to allow for (i) disaggregation of ethnicities beyond 'BAME' around specific issues (e.g. housing, health, education) and locations, or (ii) disaggregation of ethnicity data by other characteristics (such as gender) to understand intersectionality. These are particularly prevalent for devolved nations where there is a smaller overall raw population number of ethnic minorities, and for those who identify themselves in the 'Other ethnic group' category for whom data are rarely disaggregated. Issues of data quality relate in part to public services using different systems for managing data, which can prevent sharing in a standardised format.

  "While it might be possible for me to say the Asian communities within our local area are X or Y or Z, it becomes more difficult then to say 'Okay, what about just the Asian women?' or 'What about Asian women with a disability?'" CSO working in support of race and ethnic equality

- Missing data, either due to ethnicity data not being collected (potentially due to political sensitivity) or not being provided by individuals (for example due to questions over the trustworthiness of public services and concerns around how that data may be used and/or fear of persecution/discrimination). One example provided of missing data was ethnicity not being reported on death certificates in England pre-COVID-19.

  "A lot of people are thinking about repercussions. If I put down I am from a particular ethnic group, a particular faith, will that impact on the service I am going to be getting? … that is part of the trust issue." CSO working in support of race and ethnic equality

- Outdated questions and definitions around ethnicity which conflate ethnicity and nationality (e.g. Black British or Chinese, which can often refer to other Chinese language speaking countries), provide broad categories (e.g. Black African) that are not fully reflective or inclusive of how people see themselves.

  "It's always surprising to see ethnicity and nationality categories on different data collections forms! for example Black British - not quite sure how this is collating ethnic data." CSO working in support of race and ethnic equality

- Language is a significant challenge to CSOs working in support of race and ethnicity. Whilst the efforts of the 2021 Census was acknowledged for provision in alternative languages, other data collection tools and methods, alongside the distribution of communications pre and post-data collection, are considered barriers to inclusivity when they are not provided in a range of languages. This included qualitative data collection, such as focus groups.

  "Main correspondence language being in English itself is a huge barrier for a lot of community members. So, whenever they receive letters, say from ONS or from the Census, they do not instantly know what that is. The situation in Wales is like where they've written the address, and the first thing is like in Welsh. Like it's the word for 'householder' in Welsh. And then people thought it's actually

addressing someone else, and they were so afraid to open the letter because they've thought oh it's not for me." CSO working with the Chinese community

- There are significant sensitivities around what can be asked around caste which mean that data are often not collected from communities where this is relevant and impacts lives.

  "Many people don't want to officially name themselves for fear of repercussions [so] it's very difficult to reach out to them" CSO working in support of race and ethnic equality

The key opportunity for improving inclusivity for CSOs working for race and ethnic equality would be a revision of the ethnic classification to allow for a more nuanced understanding of people's ethnicity (e.g adopting categories of East Asian and Southeast Asian, although this differed notably to comments made by individuals). This would require consultation with the range of organisations working with ethnic minorities in the UK and would have implications for the harmonised collection and provision of ethnicity in more administrative datasets.

"Our issue is much more with sourcing it or accessing it. Knowing the right source for a particular piece of information. We would really benefit from some guidance on that. Even if ONS perhaps did an ethnic minority specific digest or something." CSO working in support of race and ethnic equality

Alongside this, CSOs are looking for greater access to these data in a form which allows for disaggregation by other characteristics, or for the production of a summative data digest for CSOs with more limited resources to access and analyse data themselves, as well as better provision of research and data in other languages than English.

"We're often asked to conduct surveys and consultations and then as soon as you ask, so how much budget do you have for translation and interpretation? And so many turn round, and these are public bodies as well, and say well we haven't got a budget. So how on earth do you think you can consult with people in a community that do not speak English as their first language? How does anyone?" CSO working with the Chinese community

4.4. Gypsy, Roma and Traveller advocacy groups

The key data for CSOs working with Gypsy, Roma and Traveller communities are data around the number and location of Gypsy, Roma and Traveller communities, the types of accommodation they are living in, and data on presenting needs and life outcomes of interest where equality may be impacted by ethnicity (e.g. health, housing, education and employment).

CSOs highlighted the importance of census data available from ONS which allows for disaggregation of Gypsy, Roma and Traveller groups, though it is now recognised as being very out of date. Other sources included data collected by CSOs themselves, data and reports collated by the Race Disparity Unit and MoJ, and research specifically on Gypsy, Roma, Gypsy and Traveller communities undertaken by academics at universities (such as the London School of Economics).

"We've now started to see data come down through the MOJ in relation to the justice state. That's also enormously helpful because what it's showing us is that gypsy traveller young people are being imprisoned and given custodial sentences for crimes where other groups would be given community sentences. That's how we use it effectively and we do lobbying work. 'Why is that happening?' That's where we get to deeper dives, so that's why it's useful for us." CSO working with Gypsy, Roma or Traveller communities

Several examples were given of data being used to influence service provision, such as population data being used to help identify the need for new sites in various areas, and of 'mystery shopper'-style research undertaken by one CSO (Friends, Families and Travellers (FFT)) into issues relating to GP registration that has since resulted in changes to practice.

"One of the things we did at FFT was mystery shopped GPs, and that triggered NHS England to look more seriously at the issue of registration for gypsy and traveller people but also people experience homelessness, vulnerable migrants and other communities, and we have seen them respond to that". CSO working with Gypsy, Roma or Traveller communities

The key data challenges for CSOs working with Gypsy, Roma and Traveller communities included:

- Gypsy, Roma and Traveller communities are a relatively small ethnic minority and, for various reasons, tend to be underrepresented in the statistics. Small sample sizes can undermine the credibility and actionability of the data gathered but is also seen to be used as an 'excuse' against publishing data that disaggregates the Gypsy, Roma and Traveller community from 'White' or 'Other White'. As the outcomes are likely to be worse for these specific groups, CSOs highlighted that disaggregation would reflect badly on authorities. This results in challenges for (i) understanding what issues are impacting this community and life outcomes (notably health and education), and (ii) in looking at protected characteristics within this, such as age, sexual orientation or children in care.

  "The numbers who took part in the Census and self-identified in the last Census massively underrepresents people who would identify as Gypsy or Traveller or Gypsy Roma." CSO working with Gypsy, Roma or Traveller communities

- 'Gypsy, Gypsy, Roma and Traveller' is used as a generic category for presenting data which hides distinctive ethnicities and cultures that CSOs view as important to disaggregate. This is felt to be offensive to some and the equivalent of grouping Asians into a generic ethnic category.

  "The ticktick box on the Census was gypsy/Irish traveller. The two communities have got completely different origins. Something like that's a bit problematic, because it skews it to start with." CSO working with Gypsy, Roma or Traveller communities

- A lack of trust in 'authority'. Gypsy, Roma and Traveller communities were reported as feeling marginalised due to negative experiences engaging with public services, perceived discrimination by government and representation in the media. This is also seen to link to a fear of being counted and historic persecution. This can result in a level of disenfranchisement and unwillingness to disclose information collected on behalf of public authorities.

  "The last time somebody started getting all these figures on a big grand scale, half of the people ended up in the concentration camps." CSO working with Gypsy, Roma or Traveller communities

- Given they are such a under-represented group with greater levels of inequality compared to the general population (in terms of outcomes in key areas such as health and education), CSOs felt a lot of research is undertaken within a relatively small population in quite an extractive manner. Where issues and concerns raised do not lead to action, it can create barriers to willingness to participate in additional research and on relationships both with government and with charities working with Gypsy, Roma and Traveller communities.

  "It risks alienating the community we work with who are filling out these questionnaires with us because we work alongside them and because they trust us. Are we risking them not trusting us anymore by being part of something that didn't yield any results last time? It's a difficult one. There's

no point just gathering the data and having the figures. There needs to be steps taken to actually achieve." CSO working with Gypsy, Roma or Traveller communities

- For Gypsy, Roma and Traveller communities that move around, it can lead to challenges for engaging people in research. Given other demands on their time, participating in research may simply not be a priority for those who are travelling. Also, where Gypsy, Roma and Traveller communities move location it can mean that local authorities fail to take responsibility or ownership for engaging in administrative data collection exercises, as they are not officially seen as residing within the authority.
- Finally, not accommodating for literacy capabilities can present a barrier for a number of Gypsy, Roma and Travellers participating in a written survey requiring self-completion.

The key opportunity for improving inclusivity for CSOs working with Gypsy, Roma and Traveller communities would be the collection and presentation of more representative data relating specifically to these groups (in terms of characteristics, location, needs and outcomes), which will predominantly come through national survey approaches such as the Census. To achieve this, there is a need for public services to build trust of the Gypsy, Roma and Traveller communities. CSOs are attempting to support this via campaigns to encourage their participation in the Census, but it will be important for researchers to demonstrate that voices have been heard and concrete actions taken as a result of their input into research. This presents greater opportunities for more inclusive engagement in the future.

4.5. Homelessness

The key data for CSOs working with people experiencing homelessness are data around the number and prevalence of different forms of homelessness across different parts of the country, ideally collected over time to enable more longitudinal analysis of trends. CSOs also valued data on presenting needs and outcomes which may be impacted by an individual's housing status (e.g. health, housing and employment), and any data enabling an understanding of intersectionality (e.g. in relation to gender or mental health) and triggers to homelessness. This tends to be more qualitative in nature due to the complex circumstances of those with experience of homelessness (i.e. homelessness was often a result of a long chain of other life events and structural factors such as relationship breakdowns, unemployment or poverty).

CSOs utilise data available from ONS (such as data on deaths of homeless people) as well as data from PHE, MHCLG and DWP. ONS data was specifically identified as being "independent" and "robust". Reference". R was also made to the Combined Homelessness and Information Network (CHAIN) as an excellent source of data on homelessness due to being a multi-agency database, but this is limited to London at present. There was reported to be data sharing occurring between CSOs working to tackle homelessness and those working across other sectors and equality areas (e.g. children and young people, offending, mental health etc.). Building effective relationships with local authorities is helpful in enabling better data sharing and, as an example, has been done well in Northern Ireland. However, in other parts of the UK, CSOs often have to raise FOI requests to access data from local authorities.

A range of examples of good practice in applying data to influence services and policies were referenced by CSOs. One example was current work with academics using CHAIN data to investigate relationships between mental health, young people and homelessness to influence policy work and fundraising. Another example was using DWP data to lobby for reversing a decision to remove housing benefit from 18 to 21 year olds by highlighting the minimal cost benefits the action would result in.

The key data challenges for CSOs working in support of people experiencing homelessness included:

- A lack of granularity in the majority of data available to allow for (i) disaggregation of socio-demographics within the homelessness group (notably age, gender, ethnicity), or (ii) understanding the relationship with other outcomes such as health, offending and employment. To best design and deliver preventative programmes, it is important to understand how combinations of different issues intersect with homelessness (for example ethnicity and mental health). This depth of understanding was not seen to be possible currently and CSOs felt it difficult to investigate the relationship between different data sets due to the differences in measurement and reporting.

  “We don't usually see the data sets so we will see a kind of, interpretation or an analysis of what that data means, so it's very hard for us. We tend then to get into a lot of back and forth about, you know, who does that apply to, where did you get that information, what was the research methodology?” CSO working with people experiencing homelessness

- There is believed to be a high level of missing data on homeless populations. In part this is due to them being a transient and “hard-to-reach” group, but also because - politically - there is felt to be a desire to play down the scale of homelessness at a local and national level. As a result, this is seen to result in accurate data not being collected. Services do collect data on their service users, but CSOs recognised that this is limited by their reach and lacks the robustness needed to reliably assess prevalence. Missing data also means there is not an accurate longitudinal picture of the scale of homelessness in the UK according to CSOs involved in this research.

  “Young people are not well represented in MHCLG data. So, for example, we have what we call our data bank project. We try and ascertain the numbers of homeless young people out there, and that data isn't easily available. So, we do an FOI of all local authorities saying how many young people are approaching you who are homeless. We really shouldn't be having to do that.” CSO working with people experiencing homelessness

- Definitional issues mean that certain groups (notably women and younger people) who are not accessing services or are not overtly homeless (i.e. rough sleepers) are seen to be missed from data collection. There are others who are in vulnerable situations, such as those in hostels and sofa surfing, who are also typically missed in the data. This leads to a lack of data on what CSOs referred to as hidden homelessness.

  “If you're looking at young people who are sleeping rough, they won't be counted in data unless they're sitting up or something, which is just ridiculous. Night buses are now counted but actually there are other places where people are sleeping rough, and they're just not counted as sleeping rough.” CSO working with people experiencing homelessness

- CSOs acknowledged that individuals experiencing homelessness can be very challenging to engage in research due to: (i) high levels of transience, (ii) distrust in sharing personal information, (iii) low levels of literacy which can impact on survey participation, and (iv) various issues impacting their ability and willingness to consent and participate in research (including trauma, alcohol and substance misuse).

  “We did some work with other youth organisations, around young people sleeping rough, and out of about ten organisations, only two were able to share their information because when young people come to us, they sign something saying that, actually, we can share their data anonymously. If you can't do that, it's very hard to track their progress.” CSO working with people experiencing homelessness

The key opportunity for improving inclusivity for CSOs working with people experiencing homelessness is collecting, and making accessible, a more comprehensive set of data relating to homelessness. This would be with the primary aim of enabling a better understanding of the homeless population and their living circumstances (including those in hostels). A change in legislation to broaden the definition of homelessness, so it was more inclusive of those who are currently hidden or not accounted for despite the vulnerability of their accommodation, was seen by CSOs as helping facilitate improvements in the accuracy of statistics collected.

By collecting more comprehensive data, CSOs could better understand the intersectionality of different characteristics, particularly between health and mental health issues and homelessness, to support preventative action.

Finally, CSOs highlighted that one of the key challenges at present is missing data from local authorities. Looking forward, CSOs working in this area feel it would be beneficial for there to be common data standards for local authorities recording homelessness (similar to those used in CHAIN), and for these data to be shared more openly with CSOs.

4.6. Buddhist faith organisations

The key data for CSOs working with people of Buddhist faith are contact details held on mailing lists, which is a measure of their active membership, and some record of their more formally committed membership. These data are collected and collated at an individual centre level, for their own organisational purposes, which mainly comprise advertising residential and non-residential teaching events, and communal practice.

Based on the interviews undertaken, Buddhist CSOs appear to play a limited role in relation to serving community need beyond teaching about Buddha, which they claimed reduces their need for data.

"Buddhism generally throughout the whole of history has been very decentralised. We don't operate in an equivalent way like the Church of England or the Roman Catholic Church within a hugely centralised organisation. And this is certainly true of Buddhists in the UK." CSO working with people of Buddhist faith

"We're primarily a teaching organisation, we don't provide pastoral care or community support, so looking out for health needs isn't a part of what we do." CSO working with people of Buddhist faith

There are no key data challenges recognised by CSOs working with people of Buddhist faith, however there is recognition that there is limited data on the number of Buddhists in the UK, what number are practising or connected to temples. This was not seen as a concern given the remit of Buddhist CSOs.

"But when it comes to UK, we are not sure how many Sri Lankans are here, how many Buddhists are here. So, what we do is normally we go to the Embassy and tell them, but my experience is even the Embassy is yet to find out or they don't have the data for how many people are here." CSO working with people of Buddhist faith

The key opportunity for improving inclusivity for CSOs working with people of Buddhist faith would be the provision of guidance from ONS, or another body, on how they could better collect, access and use data. In addition to this, information on GDPR-compliant data collection should be provided, so there is clarity around what approaches could be taken to collect and share data on Buddhist communities amongst networks of individual centres and temples. As mentioned, those CSOs engaged as part of this research hadn't considered that data could be used to advocate for their needs at a policy level (for example, in advocating for spiritual needs in relation to death and burial).

4.7. Christian faith organisations

The key data used by CSOs working with Christian communities and within the Churches of England and Scotland include their own surveys, as well as census data, from both England and Wales census and Scottish census. The data collected by CSOs includes the Statistics for Mission (English and Scottish versions), collating information from churches on, for example, church attendance, clergy and congregation demographics, church finances, and building and land possession within and across parishes. British Religion in Numbers was also seen as a useful data resource.

CSOs use data to allocate ministry resource to parishes for provision of services, which can include assistance with food banks, running toddler groups and helping or running homeless shelters. Those interviewed for this research worked closely with data, collecting and analysing it on behalf of the Church of England and Church of Scotland.

Key data challenges for CSOs working with Christian communities included:

- Representativeness of data was of greatest concern to those interviewed, regarding both data collected by their own organisations and other Christian faith organisations, as well as England and Wales census data. Issues of representativeness related to a lack of granularity, in that data available may not provide a detailed picture of issues such as deprivation experienced by those attending a particular parish church, even if deprivation statistics are available for the local area.

> *"Some of this information gets used in decisions about deployment of people and where the money goes. I guess to put it crudely... the deprivation stats that I produce mapped onto parish boundaries in the same way as P1 does in Scotland get used in the similar way that there is a weighting of money towards places that are more deprived"* CSO working with people of Christian faith

- The potential for response bias within data collected by parish churches was highlighted by those interviewed. They felt data may be more comprehensively collected for congregations with certain demographics than others, such as older populations, or those with more extreme opinions, leading to further unrepresentativeness of data. Concerns were also shared that clergy, who are ultimately responsible for collecting the data within parishes, may introduce a level of researcher bias to data collection.

> *"I mean we used to until very recently... asked the chaplains to generate religious data, kind of census data about the residents within homes and retirement living schemes. We've actually abandoned it because we felt it was unduly influenced by the denomination or faith of the chaplain"* CSO working with people of Christian faith

- Complexity and subjectivity of religious identity and faith was also reported as a barrier to collecting inclusive data. Quantification of religious views and activities was viewed as problematic, due to the complexities involved in religious belief and identity, and the very personal nature of faith. Concerns were raised that current data is collected based on binary and potentially outdated ways of thinking about religion and collecting data through 'tick-box exercises' limited respondents' ability to express their views. Asking people to identify themselves as 'religious vs not-religious' does not reflect modern thinking about faith.

> *"I tend to feel if you offer people boxes, they get into them. It kind of prevented contradictory identities, and there's quite a bit out there in the research literature, are people who, when they go into hospital, say I'm not religious. But then they'll say, oh yes, I pray"* CSO working with people of Christian faith

- Participants noted that they had moved to a more digital form of data collection more recently, but many in the community still preferred traditional forms of data collection such as paper surveys. In some cases, this may be helpful as it was noted that older people may be more likely to experience digital poverty which could have an impact on online survey response rates.

Key opportunities for improving inclusivity for Christian faith CSOs were identified as firstly, the use of census data for lower-level geographical areas. Collection of more granular data through the census and other sources was suggested as a means of collecting more inclusive data about Christian communities, including lower-level geographical areas and information about different denominations and forms of practice. The data which collected by the Scottish census was offered as a good example of how this is done.

"I think what one of our other challenges is getting the sufficient granularity with datasets. So because there are, whatever it is, 12 and a half thousand churches and parishes, some of which are tiny in terms of their population. Information that's available at local authority level is often not granular enough for our purposes" CSO working with people of Christian faith

Secondly, a move away from binary ways of thinking about religion and faith. When asked about top priorities, another strategy suggested for producing more inclusive data, was to use methods other than traditional quantitative approaches to collect data on religious beliefs and identity. This would enable people to better express their views and could encourage wider engagement by moving away from thinking of faith in the binary 'religious vs. non-religious' sense.

"However you do it, I think moving away from forms of categorisation, and a kind of binary statement of religion or non-religion or identity to something that actually has more kind of colour and quality to it and depth I think, but you know it would be more complex of course it would, but it would actually be more accurate." CSO working with people of Christian faith

4.8. Hindu faith organisations

The key data for CSOs working with people of Hindu faith are that which enables an accurate representation of Hindu people at a more localised level, to understand the profile of Hindu communities (importantly the numbers of families and children), needs and issues (including experience of hate crime). This helps with accessing funding, planning and targeting services, and for engaging with communities on important issues (e.g. encouraging COVID-19 vaccine uptake). Local level member data are also key to keep in contact with the community and to invite people to events.

ONS Census data was seen as providing the most robust understanding of the Hindu community at a national level, though this was not particularly relevant for many of the CSOs involved in this research. Instead, the primary source of actionable data comes via surveys and administrative data collected from CSOs about their members. While some CSOs accessed data made accessible by local authorities, this was seen to be of variable quality in terms of differentiating the Hindu community from other ethnic or religious groups, limiting the extent to which it was used.

The key data challenges for CSOs working with people of Hindu faith included:

- A perceived lack of standardisation and consistency in collecting faith data, and in the way in which data on faith is collected between different surveys and different public services. This results in a lack of understanding the profile, needs and circumstances of Hindu communities at a local, regional and national level, in turn impacting service provision (such as food options within hospitals). It also means that Hindu communities can be grouped under broader ethnic groups, such as 'Asian', and misrepresented in the media where generic labels are used in negative news reports.

"We are saying everybody should put British, other, and Indian, and in the box for the religion they should put whatever their religion is. You could be Indian Hindu, Indian Muslim, Indian Christian, whatever. Whatever your religion is." CSO working with people of Hindu faith

- 'Hindu' is an umbrella term under which there are different denominations, sects and schools of philosophy that differ in their beliefs. Capturing these differences, at least at the level of the four denominations (Vaishnavism, Shaivism, Shaktism, and Smartism) would help provide more granularity to any data on Hindu communities.

  "Hinduism is an umbrella thing which comes under a lot of little things, like people might call themselves Hindus but they might say that they follow a particular priest or particular guru or particular sect. All that needs to be properly recognised under the umbrella of Hinduism." CSO working with people of Hindu faith

- Older people in the Hindu community are not always able or comfortable in communicating in Urdu (as opposed to Hindi or Gujarati) which can be a barrier to their involvement in research. Engaging members of the Hindu community to support with Census administration was seen to be a positive step by the ONS.

  "The important thing is that the ONS has employed these 12 Hindu community representatives, or Indian representatives, and that was a good thing they did because then they can go and participate in various webinars and they can actually explain about it. That was good." CSO working with people of Hindu faith

The key opportunity for improving inclusivity for CSOs working with people of Hindu faith is the inclusion of a harmonised religion question across all national data sets, including population surveys and administrative data, alongside the collection of ethnicity data. This should present Hindu as a main faith option, rather than as part of an 'Other' category. A second, aligned priority is to ensure that questions are asked, and responses can be provided, in Gujarati and Hindi to promote greater inclusivity amongst those members of the community that do not speak Urdu or English.

4.9. Islamic faith organisations

The key data for CSOs working with people of Islamic faith are data that (i) enables an accurate representation of Muslims in the UK national statistics (so any data that allows for people to identify as Muslim in terms of their religion), and (ii) data that enables a detailed understanding of issues impacting particular subgroups at a more local level (notably by gender and age). Dependent on the remit of the CSO these data help with identifying needs to support fundraising, planning and targeting of community services, for advocating on behalf of the Muslim community as a whole (in national policy and legislation), and for understanding civic engagement and issues impacting Muslims (such as experience of hate crime).

At a national level, ONS Census data was seen to be a key source of data as it allows for key information on employment and education to be broken out by ethnicity and religion, providing a robust source of data against which more localised needs (at least in England) or behaviours (such as voting patterns) could be assessed against. Aside from the Census, there was limited use or awareness of other national survey data and administrative datasets, except for some use of statistics on hate crime from the Crime Survey for England and Wales, and reports from the Race Disparity Unit. Instead, there was greater use of data gathered by CSOs working with Muslim communities at more of a localised and issue-specific level, which tended to be shared informally within CSOs networked with one another.

The Muslim Council of Britain (MCB) produced a report 'British Muslims in Numbers' in 2015, summarising data from the 2011 Census, and is intending to produce something similar, if not more expansive, following the 2021 Census, including their own analysis and interpretation of the data. MCB has historically had the benefit of individuals within their organisation with the remit and skills to review publicised data as well as to undertake bespoke research, sometimes in collaboration with other organisations. This has enabled them to use or generate data to respond to consultations on issues ranging from women's equality within the labour market and the two-child limit on welfare benefits through to the government's COVID-19 response.

The key data challenges for CSOs working with people of Islamic faith included:

- A perceived lack of standardisation and consistency in collecting faith data, and in the way in which data on faith is collected between different surveys and different public services. Recording of religion is missing from key administrative datasets (e.g. births and deaths, experience of crime) which presents issues with understanding key outcomes such as life expectancy or experiences of religiously motivated hate crime broken down by faith.

  "When we're trying to show the level of Islamophobia up here in Scotland we're not able to show it with the police information; also the data that we get from the police quite often is slightly far behind, so it's maybe on disaggregated every year or every other year and it's not on a regular basis." CSO working with people of Islamic faith

- A lack of granularity in the national survey data and administrative datasets to profile people of Islamic faith across the range of data collected. In respect of the Census, there was some frustration that data on each question was not broken out by faith and ethnicity consistently, while across wider datasets the data was often not presented in such a way as to allow for disaggregation of faith. Even where data are broken out by faith, the Muslim community is very diverse and Muslims from different ethnic backgrounds are felt to have very different cultures, backgrounds, challenges and outcomes. A person's immigration status or where their parents have migrated from will have a key impact on their outcomes. At present these data were not collected or collated in a way that allowed for this level of analysis.

  "For example, about 10%-15% of the Indian population are Muslim. How do they perform on education and employment compared with Muslims from Pakistan, and Bangladesh or African Caribbean or Middle Eastern countries? And same thing with employment? Do we have any statistics? That would be very, very helpful to find out how Muslims from different ethnicities, you know, fare on different indicators of life. Sometimes we make very generic and very general remarks about the Muslim community, without understanding the diversity that exists within the community." CSO working with people of Islamic faith

  "How do you make sure that the statistics reflect for example, Southern Glasgow, or even [location] alone, 23 languages are spoken. So how are you ensuring that that is captured within the surveys that you do?" CSO working with people of Islamic faith

- There are certain types of data which CSOs highlighted would be particularly valuable to obtain broken down by religion, to better support their work. Examples included: standard occupational classification to understand the proportion of Muslims in senior leadership roles, experience of racially motivated hate crime, and proportion of people that have reverted to Islam. Aside from this it was life outcomes disaggregated by both religion and ethnicity or nationality.

"I think one of the issues most Muslims have been facing, and data will really be powerful in advocating for this, is the lack of Muslims in senior leadership within private and public sector organisations." CSO working with people of Islamic faith

- While the Census data was seen as being useful for looking at the broad socio-demographic profile of Muslims in the UK, it offered less value in the devolved nations – notably Northern Ireland - where the smaller sample size meant the results were often not disaggregated by the Islamic faith. Furthermore, when looking to understand intersectionality of different protected characteristics - specifically age and gender - this becomes unfeasible with Census data. For this reason, many organisations working with subgroups of the Muslim community will capture their own administrative data or undertake bespoke research which is then used to advocate on very specific issues such as Islamophobia in relation to mosque attacks.

  "While ONS data or Census-type data is very useful at a headline level, especially for England, it can get more challenging for the devolved nations especially when you then look to break it down by particular smaller subgroups, and you're talking about not just people of Islamic faith but women in Scotland." CSO working with people of Islamic faith

- A significant proportion of the CSOs working with Muslim communities operate at a grassroots level, managed and delivered by small numbers of volunteers. These organisations often do not have the skills and capacity to source relevant research findings, let alone analyse and draw implications from published data. While there is a degree of information sharing between Islamic CSOs (in the form of CSO-produced reports), this was seen as primarily happening in pockets, with not enough easily accessible information filtering down to the larger body of small voluntary organisations.

  "I don't think they [reports] have circulated widely enough to, to the grassroot community organisations which tend to be very small organisations. They don't have much resources for administration skills. I think a lot more needs to be done to really disseminate the information to as many organisations at grassroots as possible." CSO working with people of Islamic faith

- Finally, language needs amongst both older generations in Muslim communities and for first generation migrants, many of whom come to the UK without being able to communicate in English, could impact participation. This was seen to be a particular challenge for reading and writing, with implications for consent and self-complete surveys.

The key opportunity for improving inclusivity for CSOs working with people of Islamic faith is to ensure that all data collected via population studies such as the Census, and administrative data, is disaggregated both by faith and by ethnicity. Particular emphasis and interest here were in looking at employment and education outcomes, which were seen to be predictive of a wider range of longer-term outcomes. A secondary priority was to acknowledge the diversity within the Muslim community, and the need to make the research process more inclusive for sub-groups (such as Muslim women) who may face other challenges to participating in research. This will require working through the network of grassroots CSOs rather than just the mosques or umbrella membership bodies and ensuring a greater diversity in the research field force.

A final recommendation was that ONS, or another body, take responsibility for ensuring that all relevant data, not just from the Census but from other population surveys and administrative datasets, are centralised to increase their findability.

4.10. Jewish faith organisations

The key data for CSOs working with people of Jewish faith are data that enables an accurate understanding of the Jewish people in the context of wider statistics, so any data that allows for people to identify as Jewish in terms of their religion or their ethnicity. These data were then seen as most valuable when viewed together with outcomes of interest, in particular educational outcomes, socio-economic status and health outcomes (COVID-19 deaths, mental health, health comorbidity). Data helps in planning services for the Jewish community (e.g. school places, care home provision) and to ensure the safety of the British Jewish community (e.g. capturing experiences of antisemitism and crime victimisation data).

ONS Census data was viewed as foundational, and the inclusion of the religion question in the 2011 census was reported as having generated a much more robust picture of the Jewish population than had previously been possible. A wide range of other national survey data and administrative datasets were used by Jewish CSOs including: School Census, Individualised Learner Record data, Annual Population Survey, GP Patient Survey and British Crime Survey for England and Wales. The latter two surveys now include religious classification which was seen to be helpful for viewing data by faith. CSOs also seek to access administrative data from local authorities around education and health outcomes to fill gaps in the data or improve the recency of data.

The Institute for Jewish Policy Research (IJPR) produces a range of policy-focused reports and outputs that provide much of the evidence on education and health outcomes. They also undertake their own research on the Jewish population (e.g. the National Jewish Community Survey, 2013) and have established a research panel of Jews in the UK which aims to be a resource for future data collection. The IJPR also collect primary data on births, deaths, marriages, and divorces, as well as community data such as attendance at school, and synagogues. These data have variously been used to:

- determine how many elderly care places are likely to be needed over the coming years (for Jewish Care)
- determine whether or not new Jewish secondary schools need to be built-in North-West London to accommodate demand (for PaJeS)
- ascertain how many Jews of different ages in each region of the UK have different levels of learning disabilities – from severe to mild (for Langdon)
- provide evidence about whether existing facilities are in the right places and scale to support people with specialised housing needs because of disability (for Jewish Blind and Disabled)
- establish where low-cost housing units should be built to support disadvantaged individuals and families to live in close proximity to community (for Industrial Dwellings Society)

The key data challenges for CSOs working with people of Jewish faith included:

- A perceived lack of standardisation and consistency in collecting faith data, and in the way in which data on faith is collected between different surveys and different public services. Recording of religion is missing from key administrative datasets (e.g. deaths, marriages, hospital admissions, vaccinations, housing, victims of crime etc.) or is recorded inconsistently (e.g. within state schools) which means key data relating to education and health outcomes is either missing or contains gaps. A related issue is that ‘Jewish’ is no longer collected as part of the ethnicity question in the Census, which is seen as then influencing the questions used in other population-level surveys. This lack of harmonisation leads to gaps in data and understanding of issues impacting the Jewish community (and other faith groups) and impacts the longitudinal picture of trends.

"In 2017, we chose as a community, not to have Jewish amongst the ethnic questions. It was presented to us at the time that this was about the census and nothing else. However, immediately afterwards, ONS then puts out recommendations that all ethnic questionnaires should use the census classifications. On police questionnaires, I am now conscious that you report a crime, you're not asked a religious question, you're asked an ethnic question, which doesn't include us. We've gone invisible." CSO working with people of Jewish faith

"The continued provision of a full Census is vital for a minority community which accounts for less than 0.5% of the population. We are interested in how the alternatives to the Census can be developed to support continuing research about ethnic and religious minorities but want the Census to continue in the future until there is absolute clarity that ethnic and religious minority groups will be able to conduct research on their target populations if this vital source comes to an end." CSO working with people of Jewish faith

- The UK Data Archive makes some data readily accessible, but other data sources are seen to be much more difficult to access (e.g. data held by the NHS, CCGs and Trusts). Furthermore, the valuable linkages of data such as mortality records, GP Patient Survey data and Hospital Episode Statistics involve a level of skill and resource that CSOs are not able to provide without access to additional funding and training. While there are certain Jewish CSOs specialising in data, many others do not have the ability or capacity to access or analyse data, and many synagogues do not share data for fear of breaching GDPR.
- There are certain cultural and religious influences on research participation, notably the fear of being counted and that counting Jewish people could be viewed as prohibited under Halakha (the collective body of Jewish religious laws). While the Census provides a comprehensive picture of the population, there was a feeling amongst CSOs that it was likely to present a partial, and under-representative picture of Jewish people. It also happens too infrequently to capture the mobility of people, which can be a feature of the Jewish community.

  "We don't count who is a Jew, traditionally. And that has been overlaid by, of course, the memories of the Holocaust." CSO working with people of Jewish faith

Challenges with gaps in data are typically met through a combination of cross referencing and manually supplementing administrative data with that held by the Jewish community, and re-weighting local data back to Census data. These are both time and resource intensive exercises that more research focused-literate CSOs recognised as introducing biases and inaccuracies that ideally would not be present (e.g. around determining what constitutes a race crime). The IJPR has invested in developing their own panel to meet some of the current perceived shortcomings in administrative data.

"I think one of the things that you do have in the Jewish community, you do have a network of organisations that have an excellent understanding of its local community. We could present you with an application that literally maps and reach every Jew in the country that wants to be identified as a Jew within 97% accuracy, which you could never find, probably, in any other community." CSO working with people of Jewish faith

The key opportunity for improving inclusivity for CSOs working with people of Jewish faith is prioritising the inclusion of a harmonised religion question across all national data sets, including population surveys and administrative data, in addition to the collection of ethnicity data (which therefore would not need to also capture Jewish). Data should then be made available in such a way to allow for the integration of ethnicity and religion data; one option here is to produce a ‘long-form’ ethnicity variable which incorporates religion. CSOs would welcome the opportunity to be involved in consultation on the inclusion of religion in administrative data sources and national surveys beyond those led by ONS.

“If administrative data collected or contained religion, this would provide us with important additional information, for example about health, education and poverty. We would like administrative sources to be reviewed in this regard and to be consulted on this important issue.” CSO working with people of Jewish faith

During the COVID-19 pandemic, ONS demonstrated the potential for data linkage by linking Census to mortality data. CSOs would value support in increasing access and use of linked administrative data sources, particularly if there were funding and training support was available to help building the capacity of CSOs to undertake data analysis.

4.11. Sikh faith organisations

The key data for CSOs working with people of Sikh faith are data that enables an accurate representation of Sikh people in the UK national statistics; this includes any data that allows for people to identify as Sikh in terms of their religion or their ethnicity. This helps with planning and targeting services (e.g. assessing likely localised demand for language teaching, or understanding likely dietary requirements in response to pandemic and providing food handouts), for advocating on behalf of key groups (such as care homes for elderly Sikh populations), for understanding access to services (healthcare and benefits) and tracking outcomes such as health or employment. Local level data are also key to keep in contact with the volunteer base and to invite people to events.

At a national level, ONS Census data is seen to be a key source of data, as is the ‘British Sikh Report’ (produced by City Sikhs) which has historically been limited to England but is due to cover Scotland and Wales from 2021. This annual report is viewed as presenting the most comprehensive picture of the Sikh community and is a valued source of information, but there is also a perception that it may be more skewed toward middle-class Sikhs and therefore not entirely representative of the community. Other key data sources are membership data collected by Gurdwaras and data collected by CSOs amongst members and service users either in the form of surveys (such as that administered by Sikh Sanjog focused on their target audience of Sikh women in Scotland) or qualitative case studies.

“City Sikhs, is a London-based organisation which actually prepares the data for the Sikh community on a UK base, and I think they have done a wonderful job over the last 10 years or so. That report actually is based on England but as I see it, they launched in Scotland, they launched last year in Wales as well, so hopefully over a period they will become a comprehensive UK-wide report.” CSO working with people of Sikh faith

The key data challenges for CSOs working with people of Sikh faith included:

- A perceived lack of standardisation and consistency in the way in which data on religion is collected between different surveys and different public services. This means that the only robust data that exists for the Sikh community is at a national level (England or UK), which limits the ability to act on identified issues. Recording of faith is missing from key administrative datasets or is recorded inconsistently as it isn’t seen to be mandatory to collect, reinforced by the Census format. The absence of Sikh from the ethnicity Census question presents challenges for CSOs in getting a comprehensive picture of the Sikh community, as not all Sikhs are practicing or would identify themselves as Sikh in questions around religion. As a result of this, it is quite easy for Sikhs to become subsumed within a broader category of Indian ethnicity, restricting the ability of CSOs to understand the specific needs of the Sikh community.

"The reason why I'm here is because we're saying we want our own tick box. That's ultimately why I'm here. We should have it. The other countries in the world have got it. I think it would help us in the future." CSO working with people of Sikh faith

- For certain communities where Sikhs are less well represented in the general population, such as Sikhs in Wales, there is very limited understanding of what issues they are experiencing (such as experience of hate crime) or intersectionality of issues (notably by gender and SEG) due to a lack of disaggregated data. More marginalised, lower-SEG Sikh groups are also felt to be under-represented in the data. The Sikh Report is seen to be relatively comprehensive and is used by devolved nations to estimate prevalence of various issues at a devolved level.
- Levels of trust and knowledge of how and why their data needs to be collected by government or other bodies can limit participation in research, particularly for older members of the Sikh community. Additionally, CSOs mentioned that, culturally, Sikhs are considered to be a relatively reserved community, which can also limit research participation. CSOs typically look to address these challenges through collection of data from local hubs (typically located outside of temples), though again this may overlook those in the Sikh community who do not attend places of worship or community events.

"A lot of women we work with experience different forms of abuse, coercive control, and that sort of data cannot be picked up by standard data collection efforts. That's where much of our qualitative analysis comes in." CSO working with people of Sikh faith

"There's not a lot of data to access, unless you stand outside the temple with a clipboard and ask people questions, you'll get some information, but then that's not all of it. The only people you're going to get there are the temple goers. What about the 90% that don't go to the temple?" CSO working with people of Sikh faith

The key opportunity for improving inclusivity for CSOs working with people of Sikh faith is the inclusion of a harmonised religion question across all national data sets, including population surveys and administrative data. Alongside this, CSOs felt the nationwide collection of ethnicity data should also enable people to identify their ethnicity as Sikh. Gaps in current data could be filled through much greater involvement of the Sikh community in publicising data collection efforts and collecting the data (e.g. through peer data collection and at community venues/events). A second, aligned priority for CSOs is to ensure that questions are asked, and responses can be provided, in Punjabi to promote greater inclusivity amongst those members of the community who do not speak English.

"*My recommendation would be to create a culturally sensitive census and methodology, where that data collection system recognises that the people can be bilingual and they can be monolingual, but as bilingual speakers, they will still interpret that very differently from somebody who's monolingual, even though they're familiar with the language."* CSO working with people of Sikh faith

4.12. Migrants, asylum-seeking and modern slavery

The key data for CSOs working with migrants and asylum seekers and those impacted by modern slavery are granular data around the number, location and status of migrants and asylum seekers. Important supplementary information included: (i) socio-demographics (including nationality, age and gender to enable identification of vulnerabilities); (ii) access to support services (including benefits); and (iii) life outcomes (notably employment, housing and health). Data on language skills (including English proficiency) was also helpful in understanding the likely challenges of assimilation. The other data that organisations use was typically more anecdotal and relates to experiences of discrimination. Qualitative data on lived experience, often in the form of case studies collected from service users, is important given the complexity of circumstances surrounding their migration and the challenges in collecting survey data from migrants and asylum seekers (relating to both language and perceived trustworthiness of organisations or individuals collecting the data).

Official data are available from the Home Office who produce quarterly releases of immigration statistics detailing asylum applications, decisions, resettlement, appeals and support data. This tends to be supplemented by other data collected and collated by CSOs working with asylum seekers and refugees, and universities (e.g. the University of Nottingham) undertaking research amongst these groups. Another example source of data is the Anti-Trafficking Monitoring Group, which is comprised of a number of CSOs, and which undertakes and collates research around trafficking; this was seen as a good practice example of collaboration and data sharing among various NGOs and service providers. CSOs have also recently drawn on data relating to COVID-19 infections undertaken by the Centre on Dynamics of Ethnicity at the University of Manchester.

CSOs reported using data on the impact of COVID-19 on ethnic minorities to make a successful funding case for undertaking research with frontline support staff working with migrants to understand the specific impact on refugees and asylum seekers. This has highlighted disparities in risk on groups such as Filipinos working in the health service, and of ethnic minority security guards in shopping centres. Another example was a recent success in making a legal case for extending the long-term support offered to victims of trafficking based on an individual’s needs, rather than cut it off at 45 days. This was underpinned by data that highlighted the significant differences in the nature and length of individual need.

The key data challenges for CSOs working with migrants and asylum seekers and those impacted by modern slavery included:

- Missing data. There is a perception that official statistics as reported by the Home Office significantly underrepresent the number of migrants, refugees and asylums seekers, and presents an unreliable picture of these groups. It is possible that these data exists, and some CSOs are even commissioned to collate these data (though not to share or publicise these data). It is possible that these data exist, and some CSOs are even commissioned to collate these data (though not to share or publicise these data). However, CSOs felt that there is minimal incentive for government to share data on migrants, refugees and asylums seekers as the figures will not “look good”. The Home Office publish numbers of asylum seekers in each local authority, but several things remain unclear with this data, for example: are the people included the same each time; if people have moved, where have they moved to; and what are people’s outcomes or experiences (including how different types of asylum seekers fare).

  “Sometimes it feels as though we might be drip-fed data. That data exists and they give us a vague picture, but we want the actual details, and you submit FOI requests and you get some information back, and it's just more questions from the data they provide.” CSO working with migrant and asylum seeker communities

> *"Government can't and won't – go away, compose data, quantitative or qualitative, which will paint a picture – because it will condemn their policies. Let's be honest it will show the government…in a very bad light."* CSO working with migrant and asylum seeker communities

- Issues with data quality and standardisation. The systems that are used by local and national public services working with migrants, refugees and asylums seekers are seen to be designed for casework, as opposed to the reporting of data. As a result, there is a lack of standardisation and complexity in providing data that limits what is shared by departments such as the Home Office (and what can be shared cost-effectively). There is also seen to be optional standards for data collection at a local authority level in relation to documenting migrants, refugees and asylums seekers. This leads to differences in the extent to which local authorities prioritise the collection of data. CSOs manage this by estimating the prevalence of populations and issues based on service and organisation-level data.

> *"If you're looking at local, regional data, some local councils or local authorities are much better than others at providing the data that you need. So, when you try to find comparative data for another area, you either can't find it or it's not as detailed. So, there's a real disparity between authorities."* CSO working with migrant and asylum seeker communities

- A lack of granularity in data available to allow for (i) disaggregation and comparisons to be made by locality, or (ii) disaggregation of groups by other characteristics (notably how socio-demographics, English language capabilities, and other presenting needs such as disability intersect with life outcomes) to understand intersectionality.
- Given individual experiences leading to their making asylum claims within the UK, there can be low levels of trust and engagement in research by refugees and asylum seekers, and a wariness of authority and the potential repercussions of disclosing views and circumstances. As such, the information shared during first encounters often does not provide a complete picture of their circumstances, with more comprehensive information shared following the development of trusted relationships. There are also compounding language and literacy needs impacting participation in research, which highlights the importance of working through CSOs and undertaking peer research in this area.

> *"People want to stay in the country, and they fear if they share their circumstances, that officials, Home Office, local authorities, etc, will take negative action against them."* CSO working with migrant and asylum seeker communities

The key opportunity for improving inclusivity for CSOs working with migrants and asylum seekers and those impacted by modern slavery is collecting, and making accessible, a more comprehensive set of data relating to these groups to enable a better local and national picture to be developed. CSOs feel that these data are available, just not publicised. Having access to data would enable the needs of these communities, and the issues impacting them, to be better represented and understood by services and policy makers.

4.13. Children and young people

The key data for CSOs working with children and young people ranges across datasets which cover life outcomes, rights and inequalities. This includes data to understanding the prevalence of: poverty and homelessness; educational exclusion; care experience, adoption and guardianship; stop and search, incarceration and police use of force; immigration and asylum seeking children and young people; substance and alcohol misuse; and COVID-19 infections amongst children and young people. CSOs draw on a variety of both qualitative and quantitative data, with qualitative case study data most useful for fundraising and service development.

Data published by ONS is regarded as robust and of good quality, with references made to Census data and the Crime Survey for England and Wales. Understanding Society was also seen to be a particularly valuable source of data due to the ability to look at linked parent and child level data across a robust, high quality sample frame. Also, at a national level, a range of data was drawn on from across different Government departments dependent on the remit of CSOs including DWP, DfE, PHE, MoJ and MHCLG. CSOs also utilised data from other bodies including the Child Poverty Action Group and the Office of the Children’s Commissioner (whose Childhood Local Data on Risks and Needs app was seen to provide a comprehensive resource on indicators of children at risk). At a more localised level, data from local authorities, NHS trusts and CCGs, and police authorities was also seen to be key for addressing gaps in the wider datasets reported at a national level.

Several examples were mentioned of data being used to either hold government to account in respect of youth justice and policing actions, and to inform policy in relation to school closures and plans for opening in respect of COVID transmission.

“One example was we took a legal case recently around the new COVID regulations relating to remand, which extended the remand time. We were challenging that it shouldn't apply to children because it would have meant an extra 2 months in custody on remand. We did a pre-action letter to the government before we took them to judicial review. We included quite a lot of the government's own statistics in terms of the impact that would have on disproportionality of children in custody and in prison. We used that to challenge that their equality impact assessment was robust enough. We obviously used that in a lot of our submissions to the court as well. We actually won that case.” CSO working with children and young people

The key data challenges for CSOs working with children and young people included:

- CSOs can work to varying different definitions of what constitutes a child or young person. Data are collated and provided by organisations differently, typically grouping children and young people into various age categories that may not align fully to CSOs remits. This is particularly problematic at the 16-19-year-old range where CSOs held different definitions of what constitutes the transition from young person to adult. The use of various groupings can also hide nuances in the data that can limit understanding of which ages are most impacted by a given issue.

  “We were looking at the Home Office immigration statistics on the routes to citizenship, and they publish for under 18s but as under 18 as a whole group. And we got them to give us the data for each individual age group. It wasn't that difficult for them to do; they just choose not to.” CSO working with children and young people

- Issues with a lack of standardisation also arise due to (i) differences in definitions (e.g. what constitutes a ‘vulnerable child’ between CSOs and public services and (ii) agencies (notably police authorities and health services) seen as using different processes/systems at a local level, leading to issues with standardising the way in which data are recorded and subsequently aggregated. In many cases CSOs will access funding or have it as part of their remit to conduct their own research amongst children and young people due to issues with different definitions and quality of data. Examples of administrative data mentioned that are currently missing or subject to substantive gaps included statistics on:

  - children that committed a crime when they were aged under 18 but didn't go to court, or were sentenced when they turned 18 entered the adult justice system
  - police use of force for under 11-year-olds
  - homeless care-leavers over and under 21 years old
  - overnight detention of children in police cells broken down by ethnicity

- A lack of granularity in the data. There is a desire to look at issues, outcomes and intersectionality with other protected characteristics and vulnerabilities (notably looked after children, young carers, LGBTQ+ young people, Gypsy, Roma and Travellers, children informally excluded from education, young asylum seekers) broken out by age and location in a way that currently is rarely possible in the published data due to the above issues. This makes it very challenging to get a more actionable understanding of marginalised subgroups who otherwise get lost in the data. The quantitative data available generally also lacks explanatory power, meaning that it typically needs to be supported by qualitative research or by drawing on practitioner/professional understanding.

  "The fact that they just choose not to publish this data baffles me and I don't understand who makes that decision, why it gets made, but it seems that children's data is always an afterthought, and they don't think, 'Actually, we should be as transparent as possible and publishing as much as possible.' They want to restrict what they release to just a single, national figure which isn't that helpful." CSO working with children and young people

- Where CSOs submit FOI requests to access data from local services and agencies (notably for youth justice, policing and offending data) these are felt not to be treated seriously, requiring significant time and resource to draft and follow-up. There is a view among CSOs that there is little incentive for organisations to share data on issues and outcomes relating to children and young people that will make them look bad.
- There is a lack of data linkage between different datasets published at a national level, in particular between child and household level data due to a lack of any consistent unique identifiers that would enable these linkages to be made.

  "A lot of the time what you have is data collected in silos by different government departments. What we'd want to see and what we're trying to do is get better linkage between those different datasets, so that we can understand the full journey, and also being able to look at parental experience related to the child." CSO working with children and young people

- A greater range of approaches are required to engage and conduct research with children and young people than with adults (accounting for differences in language comprehension and preferences for expressing views). This isn't necessarily more resource intensive, but it does require consideration of what approaches are most appropriate and having the skills to implement these.

  "I don't think it's something you can just ask any researcher to do. There's a huge amount of training. A lot of my job is not so much doing the work but writing the organisational policy that protects the

> *work that we do and protects the children that we're working with. The amount of administrative work we have to do to make sure that all staff are trained appropriately, safeguarding training, engagement policies, data protection policies, even thinking about how you are writing notes during an interview and how is that data being protected, that's a really important thing. It takes quite a lot of effort to make sure that you're GDPR-compliant and protecting the children that you're working with."* CSO working with children and young people

The key opportunity for improving inclusivity for CSOs working with children and young people is the publication of disaggregated data to allow for analysis of the intersectionality of issues and protected characteristics in respect of different age groups. There is a desire for ONS to provide direction and a framework for this, in particular to ensure that standardised data are produced and published by local authorities and police authorities. Allied to this is the need for summative reporting of data by age (not age bands) and different categories of vulnerability (related to protected characteristics). This needs to be publicised to ensure smaller CSOs are able to access and use these data.

Some CSOs in this group highlighted that greater value will come in future from linking data, at an individual and household level, to look at the relationships between parent/household factors and outcomes for children and young people in those households. This requires agreement over unique identifiers, such as the use of an NHS number.

Finally, there needs to be recognition that the methods used to engage children and young people need to be age and ability appropriate. There are a wide range of CSOs working with children and young people that have this expertise and can advise or undertake research.

4.14. Pregnancy and maternity

The key data for CSOs working with people who are pregnant or on maternity leave includes any data which helps to build a better profile of individuals that make up this audience (for example family structure and support networks) and data on equalities-related outcomes (specifically regarding health and employment) to better understand the prevalence of discrimination.

ONS Census data and the Labour Force Survey were both highlighted as key sources of data, as was information from NHS England and PHE on maternal health, and data collated by local authorities on pregnancy and maternity. These data are largely used to understand the number and socio-demographic profiles of pregnant people and mothers. Data on equalities related issues tended to be gathered from ad-hoc research undertaken by CSOs or academics, with several CSOs reporting collecting data on the impact of COVID-19 on women to generate evidence to advocate on behalf of support for pregnant people and vulnerable groups such as single parents.

The EHRC and BEIS report into Pregnancy and Discrimination (2016) was identified as a foundational piece of evidence for understanding the scale of pregnancy and maternity discrimination, underpinning the work of many CSOs working for the equality of pregnant people and mothers.

The key data challenges for CSOs working with people who are pregnant or on maternity leave included:

- Aside from the broad administrative data on numbers of conceptions and births there is a lack of data captured on pregnant people in government statistics. This extends to a lack of data on socio-demographic characteristics (such as ethnicity or age) or on the intersectionality with outcomes of interest (such as the numbers of pregnant people in work). This presents challenges in building up a profile of this group or to look at longitudinal trends in data.

"One of the questions that we can't answer is the number of pregnant women in the Labour Force Survey at any one time. The Labour Force Survey questionnaire for 2020 is over 200 pages long and there's one reference to being pregnant or caring for other children as a reason for why people are out of work. It would be great if the Labour Force Survey could capture information about the number of pregnant women in the workforce." CSO working with people who are pregnant or on maternity leave

- The quality of administrative data provided by health care practitioners is felt to be extremely variable, with different processes used for capturing and recording data across different services. Some CSOs with closer links to health care midwifery services reported a high degree of handwritten notes used in healthcare settings meaning that data isn't fully digitised and reported on consistently. CSOs may make FOI requests to try and gather hospital data or look to raise parliamentary questions to fill gaps in primary data, though in some cases gaps are filled by undertaking research with practitioners and professionals rather than with service users themselves.

"The data is so poorly recorded, collected during pregnancy, basically by midwifery teams, it's all paper-based in I'd say 90% of the UK." CSO working with people who are pregnant or on maternity leave

- Research undertaken by CSOs and academics typically involves relatively small sample sizes which creates issues with building up a robust, granular picture of subgroups. There is some concern that the research undertaken specifically around pregnancy and maternity is not fully representative of mothers in the UK, and is more skewed toward white, middle class participants.

"The women that we would be interested in, and I think certainly a lot of the other national charities, that are women where there's been inequality or discrimination, I think there's definitely an issue there about bias towards certain types of women." CSO working with people who are pregnant or on maternity leave

"I do feel that getting an opt-in consent leads to a big bias, you're going to get only people that are either very good at communicating, probably English-speaking, pleased with the service, to talk about it, or really the opposite where they're very negative about something." CSO working with people who are pregnant or on maternity leave

- Pregnancy and maternity leave are relatively time-limited states, meaning that there is a challenge (both for CSOs and others undertaking research) in accessing and keeping in touch with a population who may not be accessing support services for longer than 6-12 months.
- Quantitative research is seen by CSOs as not presenting the full picture of complex issues such as discrimination and needs to be supported by qualitative research to help understand the lived experience and real impacts this has on individuals.

The key opportunity for improving inclusivity for CSOs working on issues relating to pregnancy and maternity is to prioritise improving the collection and publication of more data about pregnant people: their characteristics, needs, experiences and outcomes, and how these intersect with other protected characteristics. These data need to be more findable and accessible via summary papers.

4.15. People in prison and people who have previously been in prison[8]

The key data for CSOs working with prisoners and ex-prisoners include official statistics produced by: (i) MoJ (annual Offender Management Statistics, Safety in Custody, Proven Reoffending and Criminal Justice statistics, and then more biannual publications relating to ethnicity and gender); (ii) Home Office (arrest data) and (iii) ONS and other government departments. These help CSOs understand the profile of the prison population, people that have left prison, the issues they face, their experiences and outcomes of prison leavers (reoffending data, housing, employment etc.). CSOs working with prisoners will have access to more detailed personally identifiable information on release dates and offence data via the Prison-NOMIS database.

CSOs use parliamentary questions and FOI requests to uncover data that isn't routinely published, and utilise their own administrative data on service users and conduct primary research to improve understanding of complex issues and to uncover inequalities faced by ex-prisoners (e.g. in respect of employment processes, and challenges faced by particular groups such as women). This provides a richer understanding of the key issues that may impact prison leavers ability to resettle (such as the proportions leaving prison without a bank account). Qualitative data collected by CSOs is also particularly important in understanding intersectionality with other protected characteristics (such as disability, mental health, gender and sexuality) due to the complex nature of individual circumstances.

Particular CSOs, such as the Prison Reform Trust, play a key role in the sector by collating published data in summative form (e.g. Bromley Briefings and Prisoner Facts covering sentencing through to release and resettlement) and have invested in developing the skills and capabilities internally to bring together different data sets for more detailed analysis. The Prison Reform Trust is working with The Prisoner's Advice Service and the Prison Advice and Care Trust to combine data gathered from service user enquiries into a shared database to get insight into the challenges that prisoners are facing.

The key data challenges for CSOs working with prisoners and ex-prisoners included:

- Missing data or lack of publicised data. For example, where certain age groups of prisoners are being held in the UK, the number of foreign national women in prison, or data on homelessness of people leaving prison during the COVID-19 pandemic. Over one-third of the data on prisoners' accommodation circumstances post-release is unknown. Some data are also excluded from some releases, seemingly without any reason (e.g. reconviction rate for women who have been to prison and served a sentence of fewer than 12 months, which is the majority of women going to prison). There is a lack of clarity as to what data is collected and by who, with a perception that a lot of relevant data is collected and then not published. FOI requests are also reportedly often turned down on account of a 'lack of resource capacity'.

  "The big problem with those data sets is usually around a third or even sometimes more is in the unknown category, there is obviously a big gap there." CSO working with prisoners and ex-prisoners

- A lack of granularity in the data that are available. While national data on the prison population may be disaggregated by gender, or by ethnicity, CSOs report that it is rarely presented in a way that enables them to look at the intersectionality of different protected characteristics (such as ethnicity, gender or sexuality). Similarly, it is rare to have access to data looking at the reoffending rate of

---

[8] Note that, as compared to 'offenders and ex-offenders', this is a preferred and more accurate term for describing the groups represented by CSOs engaged in this research.

people released from custody by region. Issues with disaggregation undermines the ability of organisations to draw actionable insight from the data.

"Maybe we can now get people from BAME backgrounds or maybe we can look at women, but can we look at both? Sometimes that data is just missing, so it's really hard then because you can't paint a picture of what's happening." CSO working with prisoners and ex-prisoners CSO

- Organisations working with the prison population and those that have left prison only have access to a very small subset who are engaging with their services. Research with these groups is therefore acknowledged by CSOs as being with a small, and potentially unrepresentative, sub-sample which creates challenges with generalising the findings and influencing policy. While there is seen to be a high degree of consistency in their findings over time, there is a lack of opportunity to triangulate with wider datasets and opportunities to do larger surveys are limited by resources and access. CSOs can struggle to access the prison population due to the approval processes required, or to reach those who have left prison and are not in receipt of some form of service from the CSO. The Prison Reform Trust has established the Prisoner Policy Network, working with prisoners, ex-prisoners, relatives and other supporting organisations, to help address these identified challenges.
- It can be difficult to get people to tell their story due to a lack of any perceived benefits from their participation. This can also impact the accuracy of administrative data collected, with a view that many prisoners question the trustworthiness of 'the system', are fatigued by bureaucracy, and protecting themselves against potential harms by not disclosing certain information (e.g. relating to their ethnicity or sexual orientation, or in their needs assessments). It can require building a trusted relationship with people to enable them to provide personal information, and to allow them to frame their responses to questions rather than simply providing closed survey options. Having this greater depth of understanding enables a much stronger understanding of the risk factors that influence offending/re-offending, and protective factors that support re-integration and decrease chances of re-offending.

  "You're talking about a demographic that on the whole feel mistrust and resentment towards a system that, as they see it, is setting them up to fail, and not necessarily there to help them particularly." CSO working with prisoners and ex-prisoners CSO

- There are also identified challenges surrounding literacy, digital literacy, and the prevalence of learning difficulties (e.g. dyslexia) and learning disability in the prison population, all of which were felt to impact on people's ability and willingness to participate in research.
- Longer-term outcome data for prison leavers is not collected as standard which makes it very difficult to understand their needs and outcomes, and as a result to understand whether the criminal justice system is effective. A small number of projects were cited in regard to this, including the linking of MoJ and DWP/DfE data to look at short and medium-term outcomes.

  "I know that there is now regularly published data on how many people are in employment, I think it's 6 weeks and up to 12 months after release. But beyond that, it's really difficult and so for those things we rely on other pieces of information." CSO working with prisoners and ex-prisoners CSO

The key opportunity for improving inclusivity for CSOs working with prisoners and ex-prisoners is to capture and publicise data in a way that is more transparent, consistent over time and enables a more granular level of analysis to understand the prevalence of issues and outcomes by different protected characteristics. This requires thought around the ways in which data can be collected that is a more authentic and accurate reflection of lived experiences, which also requires improvements to prisoners and ex-prisoners trust in, and access to, research. CSOs recognise the importance of feedback and demonstrating the impact that this feedback has in building trust with those they support.

"[There is a need for] the longer-term outcomes around key things like employment, housing, health, etc. But also the basics like reoffending." CSO working with prisoners and ex-prisoners

The other key priority is thinking more inclusively about the way in which questions are framed so they help address the underlying issues being explored rather than exacerbating them. Work by Transform Justice was cited which showed the importance of how data was presented, including the use of qualitative data that illustrated individual stories and contexts, in influencing public attitudes and perceptions toward offenders and ex-offenders. A related example was provided in the form of a YouGov poll, released by MoJ, that focussed on public support for employers who recruited people with criminal convictions, which helped reframe people's views about those leaving prison. Similarly, care should be taken around the labelling of groups, so rather than 'offenders and ex-offenders' it may be more appropriate to use terms such as 'people in prison' and 'people that have left prison'.

4.16. Older people (aged 70+)

The key data for CSOs working with older people relates to profiling the population by age and getting an accurate picture of presenting needs (including disabilities), service use (e.g. health and social care) and life outcomes (e.g. health and employment).

Census data published by ONS is the primary source of information and regarded as largely accessible and usable, without requiring additional analysis. In addition to this, CSOs drew on various administrative data sources published by both government departments and non-departmental bodies including HMRC, DWP, NHS Digital and PHE, and other organisations such as the Dementia Intelligence Network, Blueteq (health data) and the UK MS Register. Understanding Society and The English Longitudinal Study of Ageing were also referenced as key sources of national survey data. Aside from these sources, CSOs were also collating data on service users and commissioning bespoke pieces of research to fill gaps in the wider evidence base.

The key data challenges for CSOs working with older people included:

- A lack of granularity in the data available on older people. While there was acknowledged to be a wealth of data collected about older people, there was a view that this was not always presented or disaggregated in a way that was most useful in understanding the needs and experiences of specific sub-groups. Notably these included breaking data on older people out by other protected characteristics (such as ethnicity and sexual orientation), and by issues such as dementia types, social care needs, and whether they were themselves carers. Disaggregating data can lead to issues with reliability due to the relatively small sample sizes in the published data. Gaps in data are currently filled through CSOs commissioning their own research, often in collaboration with academic institutions, and then linking their own data with published data.

  "While it's relatively easy to get data on different things for older people as a homogeneous group, and maybe even older people in ten- or fifteen-year age bands, it's much harder to get data on BAME older people …. Even within that you'd want to get data on black, Asian, and so on, older people, LGBT older people, older people who are carers, older people living alone. There is a gap there." CSO working with older people

"Any data that exists is just on dementia. Don't even get into the fact that there's 100 different sub-types of dementia, which can have radically different symptoms and levels of need and progression. There is no data on most that aren't the main couple, at best. Very, very frustrating". CSO working with older people

- While ONS data was seen to be of high quality, there has been some issues reported with trend data not being presented in a consistent manner, leading CSOs to have to manually construct trend data.
- Social care data was seen to be particularly problematic due to the wide variety of different organisations involved in providing care services across public, private and voluntary sectors, and therefore the different systems used to record data.
- There are also some specific considerations CSOs raised in conducting research amongst older people. These included the need for greater diversity in the field force both in age and gender to better reflect the profile of respondents, and care required around engaging with vulnerable groups such as those suffering from dementia. CSOs also highlighted the considerable challenges in collecting accurate data from self-completed surveys of older people, as they may not see themselves as being in poverty or acting as a carer.

"We found that the terminology used in surveys doesn't always work for different audiences. For example, when we were conducting interviews of older people who were in poverty by various definitions, they were, but they wouldn't have recognised themselves as being as such. They would use completely different language to talk about it." CSO working with older people

The key opportunity for improving inclusivity for CSOs working with older people is improving accessibility of more granular data, through providing tools that enable organisations to explore data collected on older people by different categories of interest. Reference was made to Stat-Xplore and PHE Fingertips as examples of the type of tools/functionality desired. This would help to empower CSOs to develop and present a more inclusive picture of older people in the UK.

"One of the top things I'd like them to think about is how they can empower consumers to create the data they need, rather than being too prescriptive in how that's presented. I'd really like to look at constructing tools similar to what other data producers are starting to do, like DWP, and Public Health England. I think there should be a focus on trying to empower that kind of data consumer to be able to easily piece together what they need, rather than having to work with the tables that ONS are giving them." CSO working with older people

4.17. Sexual orientation

The key data for CSOs working to support equality between sexual orientations are granular data on sexual orientation (i.e. disaggregating lesbian, gay, bisexual and heterosexual) and the intersection of these data with other protected characteristics (such as ethnicity), as well as data on life outcomes (including mental health, homelessness, digital exclusion and financial circumstances).

While some data was drawn from NHS Patient Surveys, CSOs predominantly gathered data from research undertaken with their own service users or from bespoke studies undertaken with the LGBTQ+ community. These included Queer Futures, Count Me In Too (Browne, 2010), research from organisations such as Stonewall and BlackOut, and a national LGBT survey conducted by DCMS (2018). Reference was also made to a systematic review of mental health conditions, suicide, and deliberate self-harm in lesbian, gay and bisexual people (King, 2008). These publications were each seen to help provide a more nuanced understanding of the specific needs and challenges faced by those identifying as lesbian, gay or bisexual.

Data from Count Me In Too was reported as being used to make the case for suicide prevention services, and was particularly useful because it enabled a better understanding of intersectionality. ‘akt’, an LGBTQ+ youth homelessness charity working with young people aged 16 – 25, were also seen to have run a successful campaign about mandatory data collection on sexual orientation, with the message: 'If you're not counted, we don't count.' The effectiveness of this was linked to the consultation with the community when the campaign was run, to overcome controversies, mistrust and sensitivities.

The key data challenges for CSOs working to support sexual orientation equality included:

- Sexual orientation is not mandatory to collect in administrative data and therefore there is felt to be variation in which services collect and report on sexual orientation, leading to missing data. Allied to this is a lack of knowledge amongst CSOs around which services have access to data on sexual orientation.

  “Sexual orientation and gender identity is not a mandatory thing that needs to be filled out when you access a housing or homelessness service. So, what you find is, some local authorities will fill this information out and some won't. For us, we don't really know, really accurately, what the levels of queer youth homelessness are.” CSO working to support equality between sexual orientations

- A lack of granularity in the data available on sexual orientation. Even some of the reportedly ‘better’ surveys were felt by CSOs to have relatively small sample sizes, which reduces the extent to which they are considered reliable and presents challenges when looking to disaggregate data by other protected characteristics or subgroups of interest (e.g. lower income and ethnic minorities).

  “What struck me is the complete paucity of evidence. If you're an intersectional group, intersection between protected categories, never mind beyond that, there is no data that tells you anything universal. So, I'm seeking to understand where particular inequalities are for this group, and yet, we're using data that's really just quite dodgy.” CSO working to support equality between sexual orientations

- Sexual orientation is both a sensitive and complex construct to measure. It is something that can be fluid and changes over time, and is seen to be closely tied to gender, which has implications for the need to capture data on gender identity as part of understanding the lesbian, gay and bisexual picture. Sexual orientation can also require a degree of trust to be comfortable disclosing (particularly for vulnerable groups such as children and young people, refugees or prisoners) which needs to be accounted for in the measures and methods used for data collection.

  “I know you're talking about sexual orientation, but for us, it's very much about gender identity, as well. For me, that's inseparable, really. Obviously, there are some things that I can talk about that affect particular groups within that heading, but I can't leave out the trans and non-binary experience.” CSO working to support equality between sexual orientations

  “I think when you're looking at a survey, sometimes it feels like there's really defined type categories around someone's gender identity, or sexual orientation. It's about being able to expand that as much as you can whilst also still getting the data that you need.” CSO working to support equality between sexual orientations

The key opportunity for improving inclusivity for CSOs working to support sexual orientation equality is prioritising the inclusion of questions around sexual orientation as standard practice in administrative data and national surveys. It is recognised by CSOs as a sensitive issue, so there is a need for those involved in data collection to have training on what to ask and the best methods for doing this. In surveys there needs to be a mechanism to enable more fluidity on sexual orientation and gender in survey answer codes (e.g. 'questioning') and to always provide a self-describe option. Aligned to this, there was a view that there needs to be a much greater level of engagement between those involved in collecting and collating UK statistics and the LGBTQ+ community to improve the perceived trustworthiness of official data collection and promote participation in largescale data gathering exercises. Part of this engagement should involve discussions around the relationship (and decision to decouple) gender and sexual identity.

"There are many young people whose sexual identify is a journey, a process. It's not just a straight line from, 'I thought was straight and now I'm gay.' There's a back and forwards, self-understanding. That fluidity element as being part of that LGBTQ experience is an important thing to integrate into surveys." CSO working to support equality between sexual orientations

4.18. Trans, non-binary and gender-diverse communities

The key data for CSOs working with trans, non-binary and gender-diverse communities relates to data collected by and for those within these communities.

CSOs were working with data from the Northern Irish Life and Times Survey and the Equalities and Human Rights Commission (EHRC) commissioned NatCen British Social Attitudes Survey (BSAS), which collects information on attitudes towards transgender people. CSOs also reported using data produced by the LGBT Consortium, as well as work from the Trans Learning Partnership, Stonewalls' Transforming Futures project, and the Integrated Care for Trans Adults Foundation.

CSOs use data to target their resources in the areas where they are required most, such as working with health service providers to bring care closer to their members, educating others about issues faced within the community, and advocating for and supporting trans, non-binary and gender-diverse communities in different areas of the UK.

"The most useful and most prominent bits of research that's been done in the past while has been trans people inclusion in the Northern Ireland's Life and Time[s] Survey. Basically, the Statistics and Research Agency do this yearly kind of public attitudes survey that showed a blanket validity in terms of trans people using gender services like toilets and changing facilities. Trans people using domestic and sexual violence refuges and things like that. It was all net positive, people were generally approving of this so, that was pretty good data to come out of that and we were quite able to use that." CSO that works with trans, non-binary and gender-diverse individuals

The key data challenges for CSOs working with trans, non-binary and gender-diverse communities included:

- Underrepresentation. When asked about the key data challenges, participants from CSOs working with trans, non-binary and gender-diverse individuals replied with "*where to start?".*
  CSOs emphasised throughout the discussion their belief that current data and research is largely unrepresentative of the trans, non-binary and gender-diverse communities due to problems with underreporting.

  "We find statutory research and statutory data collection is almost entirely useless to us. In the past month the PSNI [Police Service of Northern Ireland] has recorded one transphobia hate crime which is just, we know there's been more than one transphobia hate crime!" CSO that works with trans, non-binary and gender-diverse individuals

- A lack of awareness of existing trans-led research. CSOs reported that duplication commonly occurs within the field of gender identity research. This is because researchers may not be aware of work already undertaken by trans, non-binary and gender-diverse researchers into relevant issues and topics. It was felt that such work is not taken seriously, and any recommendations which emerge from such projects are overlooked, in favour of research designed and undertaken by researchers who are cisgender (which means their gender identity is the same as the sex they were assigned at birth).

  "There's tendency among researchers to objectify the trans population as opposed to involve them." CSO that works with trans, non-binary and gender-diverse individuals

- Research fatigue occurring due to duplication. CSOs pointed out that members of trans, non-binary and gender-diverse communities often experience unnecessary trauma or distress from having to recount life experiences due to participating in such research, when often similar work has already been undertaken within the trans, non-binary and gender-diverse community.

  "I do think that the kind of issue of research fatigue and people just consistently being asked the same question, 'Okay, we have another researcher coming in to talk about mental health. Great, excellent.' Because the trans community aren't sick of talking about our bloody mental health." CSO that works with trans, non-binary and gender-diverse individuals

- A lack of involvement in research planning, with CSOs reporting that when new research is being commissioned and undertaken, there is generally very little involvement with the trans, non-binary and gender-diverse community in its planning. They highlighted that this results in research being undertaken which overlooks important issues and fails to include relevant community members and stakeholders in trans, non-binary and gender-diversity research and policy.

  "I think that's a really important thing to think about when you're designing research whenever you're thinking about what research to do, you think about who you are and what you have to bring to that? Because I think genuinely, if every research project that is currently being done on trans communities was being led by a trans person, I think we would just a field of incredible, amazing research and data and statistics to use but because it's not, we don't." CSO that works with trans, non-binary and gender-diverse individuals

- A lack of disaggregation, with an emphasis on the lack of disaggregation of transgender people from wider LGBTQ+ data. CSOs reported that in the rare instances that transgender data are disaggregated, sample size are often too small to obtain useful information. CSOs also recognised that small sample sizes put trans people participating in research at risk of being identified.

  "One of the things we find there's a lot of the surveys and research we look at is done on the LGBT population and they either don't disaggregate the T data, or they have a token T person who you could almost if you know the person, you could almost identify them from the information that's given and that's really breaching confidentiality." CSO that works with trans, non-binary and gender-diverse individuals

- Accessibility and a lack of co-ordination. CSOs reported that accessibility of existing data proves a considerable challenge, with a great deal of research being hidden behind paywalls. It was mentioned that CSOs working with the trans, non-binary and gender-diverse communities often lack sufficient funding to enable members, staff and volunteers to access key pieces of research and analysis. Furthermore, a perceived lack of coordination and oversight of research into gender-diversity which is undertaken by a wide variety of institutions and organisations was reported, which results in fragmented, incohesive and difficult to find research spread across the internet and other media sources.

"I think the biggest point for us would definitely be around the paywall for research papers because we're a small organisation... I didn't go to university. I don't have these log in details and I think it is a class barrier as well in terms of like there's a class of people that can access this kind of data, that can access this kind of research and there's a class of people who can't." CSO that works with trans, non-binary and gender-diverse individuals.

- A lack of research-led policy or service amendments. CSOs emphasised a lack of action resulting from existing data and research. They felt there is a lack of follow-up on issues reported by the trans, non-binary and gender-diverse communities, leading to the perception that even if good quality research is undertaken, it is unlikely to lead to any useful changes in policy. The situation surrounding repeated consultation on the Gender Recognition Act was provided as an example of this.

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

The key opportunity for improving inclusivity outlined by CSOs working with trans, non-binary and gender diverse communities is co-production of research between institutions and members of the communities in order to produce more sensitive and relevant research. Ensuring that stakeholders within the trans, non-binary and gender-diverse community have the funding to conduct research themselves was said to be preferable, but failing this, participants suggested that ensuring they are consulted throughout the research process would lead to more inclusive research. It was felt that this approach would result in less power imbalance between researchers and participants, ensuring the most appropriate methods and terminology are being used and that the most pertinent issues are considered, rather than ones which are presumed to be important.

"*We want to work with other people but having trans in real co-production makes things work properly. That's why the Welsh gender service works well because it has been, and it still is run by co-production.*" CSO that works with trans, non-binary and gender-diverse communities

4.19. Women's equality

The key data for CSOs working to support women's equality included any data relating to experiences and life outcomes that enables a better understanding of equality issues for women.

CSOs were working with data from a variety of sources including from the ONS (Census, Labour Force Survey and The Annual Survey of Hours and Earnings), and national administrative data reported by various government departments. This was supplemented by academic studies such as Understanding Society and the University of Essex Tax Benefit Model, data collated by other CSOs (e.g. Nuffield, Joseph Rowntree Foundation, New Economics Foundation) and their own research undertaken with members or administrative data.

The Domestic Abuse Bill was cited as an example of where data was used successfully to influence policy.

The key data challenges for CSOs working to support women's equality included:

- Gaps in data due to a lack of clarity in the standardised format and mandatory nature of questions relating to sex and gender. ONS, through the voluntary positioning of the gender identity question in the Census, was seen to have contributed to a situation where public services do not feel they need to ask people about both their sex and gender, resulting in gaps in data. There are some specific needs and audiences for which there is notable lack of data, for example on the mental health of young women, and on the numbers of: transgender prisoners, homeless women, pregnant women, and Gypsy, Roma and Traveller women. Challenges with gaps in data are currently met through CSOs undertaking their own research, with examples including ongoing research undertaken by the Young Women's Trust into the impact of COVID-19 on young women, with booster samples for ethnic minority respondents (to increase the range of subgroup analyses that can be undertaken).

  "Our issue is more that we no longer trust the data that we see and that the work we've done gives us reason not to trust that data. How can we inform good policy and how can other people be making good policy if they don't have data that's true?" CSO working to support women's equality

  The conflation of sex and gender identity, which are seen to be different things: the sex observed and recorded at birth (which can include intersex) and the gender you identify with (which can include a range of transgender identities). Sex and gender reassignment are both protected characteristics, and it was felt to be detrimental to the understanding of both if sex and gender data collection are not treated separately. This has been a key issue for CSOs working for women's equality and one which has recently gone to court regarding guidance to be used in the 2021 Census.

- A lack of granularity in the way in which data is presented and made available to understand intersectionality of sex or gender with other issues (such as rough sleeping). This is even more challenging where CSOs are looking to understand issues impacting different age or ethnic groups by gender, or to disaggregate data by the devolved nations. The resulting lack of evidence impacts CSO's ability to access funding to provide support services for those in more vulnerable circumstances.

  "One of the big issues that we have is that often when data is presented, where you get breakdowns by certain characteristics each protected characteristic is treated separately and in turn, so you'll have sex and then you'll have race and then you'll disability, and it's very different to see intersections between them, so that's a primary issue." CSO working to support women's equality

- The collection of data on sexuality at a group level for the 'LGBTQ+' community, but not each specific one, leads to issues for CSOs working for women's equality when trying to advocate on behalf of lesbians as a distinct group.

- Issues relating to the mode of collection and presentation of data. CSOs identified that data collected at a household level makes it hard to understand how public services can better serve women. For women impacted by domestic abuse, it may also reduce willingness to disclose if these data are being collected in the home. Furthermore, the push toward conducting research online means that research can miss more deprived women without access to Wi-Fi or laptops/smartphones which leads some CSOs to feel that the data available (e.g. about employment) is unrepresentative of the true picture.

- Participating in research, especially when disclosing traumatic experiences, can be distressing. CSOs highlighted the importance of data collection being undertaken sensitively by appropriately trained researchers and organisations.

The key opportunity for improving inclusivity for CSOs working to support women's equality is access to data in a format which enables more granularity in analysis to understand the intersectionality of issues impacting women, and women with other protected characteristics. Where the data doesn't allow for this (e.g. due to small sample sizes), it would be beneficial to take this into account for future sampling strategies to boost the sample size where needed. To achieve this, CSOs believe there needs to be clear guidance for all organisations running surveys and collecting administrative data on collecting both sex and gender data, and to present this in a standardised way in public datasets. CSOs working in these areas would be happy to be consulted on the most appropriate wording and format for these questions.

"I think if the dual discrimination and the provisions in the Equality Act were enacted it would put more pressure on public bodies to look at intersectional data because they would be potentially liable for employment tribunal cases on more than one characteristic, and obviously you'd then need that data in order to ensure that you're not laying yourself open to a potential tribunal case, or a potential challenge under the Public Sector Equality Duty." CSO working to support women's equality

5. Preferences and principles for the collection of more inclusive data

As part of this research, we spoke with over 90 members of the public that have different protected characteristics or are under-represented in the UK statistics, to understand their views toward the collection of their data and making data more inclusive. This section presents their views on inclusive data, including their understanding and experiences of sharing personal data, the benefits and risks of sharing data, preferences for sharing data and principles for those collecting and using their data.

5.1. Understanding of personal data

People's awareness of what constitutes their personal data varied considerably both across, and within equality areas and under-represented groups. Definitions were tied more to their individual understanding of 'data' and 'personal data', the types of services and organisations they interacted with, and how much thought they had previously given to their data as something which held value (for themselves or for others).

"Definitely your date of birth, where you were born, name, age, where you live, all the things that can be identifiable to you, to exactly where you live." Individual of Muslim faith

"My name, telephone number, email address, bank details, social media profile, marital status, political views, members of societies and organisations. Also things like debit cards, supermarket loyalty cards, public transport smart cards, council tax, utility bills, that's to name but a few." Individual with a learning disability

In discussions with participants it was clear that they differentiated between types of personal data according to (i) how sensitive or personal they considered it to be, and, allied to this, (ii) personal experiences of providing these data to companies or organisations. There were broadly five categories of data that were commonly referenced by individuals participating in this research, though we have named and differentiated these categories as part of our analyses. These included:

- Basic administrative data including your name, address, telephone number, email address, place of work, and date of birth which is provided on request to receive a service or through national surveys such as the Census
- Sensitive administrative data including your religion, ethnicity, gender, sexuality, marital status, political views and health, which is provided on request to receive a service or through national surveys such as the Census
- Transactional data, including online search behaviour and purchasing behaviours, which people largely consent to as part of receiving a service
- Security or financial data including account details, biometric data and passwords, which are carefully managed and rarely shared, even with relevant/affiliated services
- Private personal data, including photos/videos, information about friends and family, details of conversations, information posted on social media and general attitudes, experiences and beliefs toward different subjects. People typically make an explicit, conscious decision as to what they choose to share, when and with whom

The extent to which people could identify instances of sharing personal data varied and related to what types of data they recognised as forming their personal data. In most cases participants identified instances of providing data which were tied to the receipt of a service: a public service (such as GPs, hospitals, the police, school or university, libraries, the DVLA and HMRC); a voluntary service (e.g. accessing support from a charity); or a commercial service (such as a search engine, social media account, email provider, comparison site, online retailer, leisure centre or bank).

While most participants did not talk about the data provided or held by commercial organisations, for those who did there was an appreciation that private companies are likely to know a large amount about people as a result of their purchasing decisions. Similarly, a small proportion of participants highlighted that government will hold a lot of information on people across the variety of services they are responsible for. People were generally uncomfortable with organisations holding a lot of identifiable data on them, though typically recognised this was needed for them to either access a service, or to receive more personalised service (i.e. adjusted to their circumstances and needs). Overall, it wasn't something that many people reflected on prior to these discussions.

"I think companies such as Amazon, Argos, all of those commercial companies probably know more about me than government." Individual of Sikh faith

"[Government will know] *whether or not I'm on state benefits, whether or not I'm solvent, who I live with, whether I rent or own the flat I live in, next of kin. They would have my bank details in most cases, and also potentially my political beliefs as well."* Individual with a learning disability

A small minority of participants reported experience of providing their data to contribute to a specific, ad-hoc information gathering exercise (which could range from employer equal opportunities surveys through to participating in market and social research activities).

5.2. Willingness to share personal data

People's willingness to share their data is less about exactly what data are being collected, and more about why those specific data are being collected, who by, and how they will be used. People engaged through this research reported that they are typically happy to provide their personal data, if there is a clear and logical rationale for why they are being asked to share that information. Where an individual, service or organisation asks for information, people reported typically making a fairly instinctive judgement as to whether the request makes sense based on the perceived rationale for the information being needed and whether the individual or organisation requesting their data is trustworthy. Providing finance-related information on a tax-return is understandable and comfortable but providing the same information to an estate agent to rent a property can be less understandable and more uncomfortable for people.

"I think just capturing data for the sake of data, that's a breach of civil liberties for me, but if you said to me, 'We're going to capture this data this year, on this data we manage to do this but because we had a data gap and needed to identify your faith, we could have been able to offer you this,' then I would be prepared to." Individual of Sikh faith

There is an innate desire amongst most people for a certain level of privacy, particularly in relation to aspects of personal lives that they wish to remain confidential. Across those participants we spoke with, there was a wide degree of variation in levels of comfort in sharing the same types information dependent on circumstances, experiences and attitudes to their data and to those who might collect their data. As such, there was no desire amongst people for their lives (and their 'data') to be an open-book. Rather data were seen as something that should be sought, and granted access to, on a case-by-case basis. People were least comfortable with sharing financial information (outside of the context of sharing with relevant services such as their bank), or personal photos and private conversations, due to the risks associated with that data being used to harm them. While many also highlighted concerns over sharing health-related information, some noted the importance of sharing this information with GPs and medical research organisations to improve service delivery and prevent unnecessary risk to their community. Coupled with this were calls for health care professionals to be educated about the issues affecting specific groups, to improve the experiences of these groups with public services.

"I suppose it's a good thing. I can see two sides to it. I'm quite a private person. I can see that they want to help you, but then I can see that the big brother is watching you-type thing. I need all the help I can get, to be honest, so they've got to know, but I like to keep some things private." Individual with a learning disability

"I've heard lots of anecdotes about practitioners and how they speak to trans people and there just needs to be more education overall on trans issues. Specifically, in the medical field." Trans, non-binary or gender-diverse individual

Most people were not consciously thinking about how their data is being used, or could be used, by organisations (i.e. outside of this research, most people were typically not thinking about how their data was being used or collated). Sharing information in exchange for a service seen to have become a routine transaction and there is an implicit expectation that data shared is treated confidentially, stored securely and required in order to provide them with a specific service or to comply with some form of rule or regulation (e.g. equality and diversity legislation). People appear to rarely be questioning this and, given the circumstantial urgency some individuals face (e.g. recent migrants or those experiencing homelessness), they may feel that they have little choice in the matter.

"As a society we've got to share a certain amount of data for us to get a service. It's like me saying to Amazon, 'I need a jigsaw,' 'How old's your kid?', 'I'm not telling you, it's personal.' You've got to be able to divulge a certain amount to actually get the service." Individual of the Hindu faith

Where data was being collected for the public good (i.e. there is a clear link made between the request for information and an outcome at a group, community or population level), people were typically happy to provide their data. This is most obvious in national surveys such as the Census but also for more detailed surveys such as Understanding Society. Many people would also be happy to provide biological data, though this was seen as more intrusive, controversial and likely to be best administered through GPs and doctors; it was also least acceptable for some marginalised communities that had more strained relationships with authorities. Where people were less happy with sharing their data for public good, they tended to refuse to provide that information.

"I don't mind my personal story, my personal information can go anywhere as long as it can be used for a positive purpose." Individual migrant or asylum seeker

"I like the Census. When I was unemployed, every few months, they used to give the unemployment rate, I can then see how the area is in Cardiff. There's so much grey to it but it's nice to know how many people and how the money is spent, your taxes." Individual with a learning disability

Views on the sharing of personal data did vary dependent on the organisation collecting and using that data, and the associated (often assumed) rationale for their use of that data:

- People may have a limited understanding of how the **government or public services** make use of their data in informing policies or services, but there is a general assumption that it should be used for both personal and public good, and therefore there is usually a high degree of willingness to share their information. A recent, high-profile example of this was around using data to help tackle to COVID-19 pandemic. However, there was a widely acknowledged disconnect between individuals providing information, and actions being taken that are seen to align with, or result from, that data. This leads to some uncertainty over how people's data is being used by authorities.

"I'm honest with you, it's a bit difficult to answer why I think they need it, but I just think they do. You've just grown up with it, you're just so used to people taking this information, and so I'm like, fair enough, if that's what they require, there must be some reason behind it, but I don't know what the proper answer." Individual of the Islamic faith

There was an expectation that data provided to local or national government would be treated confidentially, kept secure and not passed on to any third parties.

While this was the majority view, there were notable differences between different under-represented groups, with those individuals from Black British, Black Caribbean and Roma, Gypsy and Traveller ethnicities, and those from the trans, non-binary and gender-diverse communities most likely to question the authenticity and trustworthiness of those collecting their data. Individuals from trans, non-binary and gender-diverse community in particular questioned whether there is an necessary justifiable purpose for gender markers on legal documents, including passports, driving licenses and HMRC documents.

"Pretty much the only reason HMRC require data on somebody's gender is to do with archaic details on how much national insurance you paid when you were married...there's no need for gender markers on legal documents." Trans, non-binary or gender-diverse individual

- People, with a high degree of individual variation, tended to be a little more cautious in sharing their data with **private companies**. There was a general uneasiness about sharing personal data with companies who were using that data to make profit, though many people recognised that the trade-off was that they received benefits in relation to a more bespoke and valuable service.

  "I've been onto ClearScore, and it's horrifying how much information people can see about you. If I can see them, other people can see them, and there doesn't seem to be a way to delete that … I'm not really consenting for other people to see that but clearly, if I want a credit card, they're going to have to get that information." Woman with experience of inequality

  A small minority of people actively reviewed the terms of service and/or terminated services where data being accessed was deemed to be private personal data that was not seen to be necessary to receive that service (e.g. certain communications services access to photos and conversations). Typically larger companies were seen as more trusted to collect and hold data due to assumptions around the policies and processes in place to govern the use of their data, though there was some degree of reluctant acknowledgement (and frustration) that their data was shared with third parties.

- People were largely positive about providing their data to **universities or charities** to help provide a better, more detailed understanding of a given issue or community. Where usage was seen to be used benevolently for public good, participants trusted that the data they provided would be treated confidentially, kept secure and not passed on to any third parties.

  "Medical's the key one, scientific research, government budgeting, pensions, that side of things, and again, universities, university hospitals, these are the folk who discover stuff every day and are able to make our lives better going forward." Older person

- A small number of participants specifically referenced GDPR and awareness of legislation around data protection and were very positive about the level of privacy protection these were seen to afford people. These were typically individuals that worked (or volunteered) in an environment where they had received training on data protection.

5.3. Benefits and risks of sharing personal data

The benefits of sharing personal data tended to revolve around access to a service, and the quality and personalisation of that service. Personal information around circumstances, needs and preferences were seen to enable companies and organisations to provide services, or make reasonable adjustments to services or expectations, that better addressed individual needs (e.g. accounting for religious observances). Another personal benefit, which was mentioned much less frequently, was of "having your voice heard". This was particularly important for those individuals that felt (or had previously felt) marginalised (e.g. due to their ethnicity).

"It is really important to me because most times I feel as a black person, most times when you bring up some issues or you try to talk about something, it gets ignored or just played down." Individual migrant or asylum seeker

Only a small minority of participants talked about personal benefits of sharing data, specifically in respect of their protected characteristic, in relation to affirmative action. Examples here were around inclusion and diversity within the workplace, and access to specific support or initiatives (e.g. business grants for LGB+ people).

The perceived wider public benefits of sharing personal data were in respect of using data to better understand people and communities, to develop national and local policies and subsequently deliver more effective public services. Examples here varied across individuals but included:

- Police use of crime data to protect a locality or group
- Providing a more representative picture of more marginalised groups (for example Muslim communities, or Gypsy, Roma and Traveller communities)
- Understanding representation of different groups within organisations to tackle under-representation
- Tackling social inequalities
- Providing education, leading to a better understanding of minority groups, with the aim of improving social cohesion and cultural competency
- Monitoring inclusivity and diversity in the workplace
- Ensuring services were accounting for religious practices and observances
- Allowing researchers to determine links between protected characteristics, or intersectional populations, and health. For example, using employment data and health data to determine occupational hazards

"Discrimination is already there in society anyway. I think that, again, the clearer we are, the more open we are, even sharing our personal information, to understand other people's views, we will reach a common base." Individual of Buddhist faith

"Obviously, there are some inequalities and in order for anything to be done about those you usually do need some kind of proof that those inequalities exist in the first place, so data is one way to prove that." Individual with a learning disability

The risks of sharing personal data were generally less well understood in abstract form. However, many under-represented groups, such as the trans, non-binary and gender-diverse communities, were better able to identify risks. Others were able to identify some potential risks to sharing of data, which were typically seen to impact them at an individual level. These included:

- Sales and marketing 'spamming' by commercial organisations or fundraising requests
- Stereotyping individuals on the basis of certain identifiable characteristics
- Discrimination (e.g. in respect of what services you are offered, or the price of services) on the basis of certain identifiable characteristics
- Facing harm or abuse
- Facing prejudice from employers or potential employers
- Risk of being defrauded or identity theft
- Risk of reductions or withdrawal of government support (e.g. around benefits) as a result of changes to status or criteria for assessing needs/circumstances
- Risk of identification, even where confidentiality is assured, due to small sample sizes in research
- Risk of harm to livelihoods to those whose work is considered illegal or non-regulatory e.g. sex workers
- Risk of misinterpretation and misuse of data, leading to negative media portrayal

"It's all so often we are horrendously represented in media" Individual from the trans, non-binary and gender-diverse community

"The only risk and harm I see is if certain individuals, I know some people don't do it and they don't have it on purpose but some people already have a view of, 'You're a Hindu,' or, 'You're Indian so you're-,' and that's the only worry I have about it is people having these stereotypical views. Sometimes mixing religions up, I know some people don't know the difference between maybe Hinduism, Sikhism, Islam. A lot of these religion they mix them up and they're all very different". Individual of Hindu faith

"I'd want to know the actual reasons. Say, for example, if I was looking for employment, are they going to use that information to actually genuinely help me look for work and use that information to make reasonable adjustments in that place of work, or are they going to use that just to bully me to actually get off of state benefits?" Individual with a learning disability

"The worst case scenario is providing if they're using all of that data to create, I don't know, a fake passport or something, fake ID card in your details and your name." Individual of Islamic faith

The majority of participants, across equality areas, saw minimal risks of sharing data with public services. Participants perceived these organisations to be using data for providing a personal or public service, and as having greater accountability than commercial organisations. Related to this was a view among several participants that in respect to providing information to the government, while they valued their privacy, they were law-abiding and so had "nothing to hide". Nonetheless, while individuals were largely willing to share their data with public authorities, there was an expectation that, in return, actions were taken which accounted for their views and experiences.

5.4. Importance of more inclusive personal data in providing a more representative picture of minority groups

It was felt that the administrative data people provide does not always give an accurate or fully representative picture. While it may enable services, and government, an understanding of broad socio-demographic information, it was seen as aggregating people into broader groups which felt disconnected from them as individuals. Most importantly, people felt that administrative data doesn't convey what makes them an individual: (i) their experiences, attitudes and beliefs; (ii) their hopes, fears and dreams; (iii) what they like and dislike; (iv) what they need and what they have to offer; and (v) how they feel about their circumstances. If organisations want to build a more inclusive picture of people, either as individuals or as a group, these factors are crucial. It was further noted that household data collection may exclude groups such as the trans, non-binary and gender-diverse community, as a higher proportion of these groups are homeless than in the general population. Consequently, there was reported to often be a sampling bias preventing data from being generalisable to the whole community.

"I don't think they ask very deep questions. They just want to make a proper study of society, so how many people have got such and such a faith, how many people have no faith at all. Who is practicing religion, who is not practicing. And I think it's more for this tick box basis, but I don't think they want to know information about your faith in that deeper sense." Individual of Buddhist faith

"I think with the data that they get, they get a lot of quantitative data but they don't get any qualitative data. Do you know what I mean? They get X amount of people use a wheelchair or are partially sighted, or whatever, but they don't get a life experience. They don't get day-to-day challenges. It's purely numbers. All we are numbers sat on a page and that doesn't explain people's struggles and people's story and what they need help with." Individual with a physical disability

"You read the files and you put numbers on them, those are people's lives, hopes, dreams, everything. It's not just a number, it's more than just that data. It's a person." Migrant or asylum seeker

"It's definitely not representative. Where I go, who I'm with, what kind of foods I like, where I go abroad, what I do with my family in my personal time, what my relationship is like with my husband, with my children, who my friends are." Individual of Islamic faith

"I'm just going to mention filling the census online or even in person is very difficult if you don't have a set place to live. A friend of mine is currently homeless and trans and obviously with libraries not being opened, COVID and stuff like that, makes it very difficult and obviously there are much high rates of being homeless in the trans community makes things like filling in a form really difficult." Trans, non-binary or gender-diverse individual

More sensitive administrative data - in the form of gender identity, ethnicity and religion – was also seen as important for people to be aware of so that they can better understand, respect and support people. It was reported that often administrative data does not account for the intersectionality of different characteristics, and therefore cannot provide a representative picture of individuals. Again, people were largely comfortable providing this information to organisations where this was requested, provided that there was a clear rationale as to how it would be used to affect positive change. Some participants were less comfortable providing this information as it was seen to be sensitive and private, and there was a lack of understanding as to why it was of relevance for an organisation to have that information. Participants in the trans, non-binary and gender-diverse discussions highlighted an additional risk of disclosing sensitive personal data, as it could result in revealing an individual's gender identity before they are ready to share this information themselves.

"I'm of Islamic faith and during the Ramadan period, where I'm fasting, there was nobody else in my department, my manager didn't understand what that meant, so, you know, I had declared it and I spoke to HR and they were fully aware of my situation, etc. They supported me in terms of being able to start later. I wasn't asking for special treatment, it was just so people could understand where I was from. So it does help me to share that information, and it does gain knowledge in terms of your workplace." Individual with experience of homelessness

"I think it's helpful to let people know because obviously you want as much help as you can get with it. I've recently been in a school actually volunteering to finish off a teaching assistant course and it's really important they know I'm deaf rather than me standing there pretending I've heard them when I really haven't." Individual with a physical disability

"DBS checks and for a credit checks, both of those will require details of previous names. Both of those things are things employers usually want at least one of, and most of them will want a DBS, a disclosure barring service check, and there's no way really to conceal it from an employer and unfortunately it makes things really difficult." Trans, non-binary or gender-diverse individual

Unless they identified themselves as being more "politically-engaged", people tended not to have a strong understanding of what data was collected around protected characteristics or how it was used. This made it challenging for people to understand if it enabled an inclusive picture to be presented of their specific 'group'. However, based on the categories under which participants were familiar with data being presented (i.e. by broad protected characteristics such as being of Hindu faith, or where distinct groups are grouped together under broader categories such as 'Black and Minority Ethnic'), there was an assumption that the current 'picture' of different equality areas was likely to be limited. There was an expectation that data would present a very homogenised picture of groups that contain a large degree of variation and nuance, which undermines understanding.

5.5. Preferences for data collection

The ways in which people wanted to contribute their data varied largely independent of their protected characteristics, and related to their individual preferences, confidence and literacy, and on the type of information being shared and purpose of providing that information. Ultimately, the method for data collection/provision came down to personal preferences.

"How they ascertain that information doesn't really matter to me. I think it's how they use that information and how that information is shared on, that's more of a concern to me." Individual with experience of homelessness

"If you knocked on someone's door and they're unconventional, they might have poor eye-to-eye contact. They might not be very good socially. They might be anxious. They might not want to be scrutinised and asked a lot of questions about their personal life. Some people don't like these calls because they find it anxiety-triggering. They'd rather probably do it on a computer, where they feel safe." Neurodiverse individual

Standard, administrative-type data was seen to best be collected through surveys and forms, which are largely accessible for people if the appropriate support is given to enable completion (e.g. accounting for language capabilities). Data which are seen to contribute to a more inclusive and holistic understanding of an individual, or a community, was felt to require more of a qualitative approach (or at minimum to allow for more open-ended responses to be given). Discussions, at an individual or small group-level, were often recommended here as enabling people to develop the personal connection required to open-up about more personal opinions or experiences.

In principle, there was a preference for more active forms of individuals providing data and the processing of these data rather than passive, automated collection. I.e. participants preferred the idea of consenting to each data transaction. However, those individuals that reflected on this recognised that the terms of service (and associated benefits) often mean consenting to commercial organisations accessing and processing data in a more automated manner. This was largely accepted as a necessary trade-off for receiving services though there was concern at how personal data is subsequently used for the targeting of advertising.

"They all seem to know what you're thinking these days when you don't even say it. I always get adverts come up about things you've thought about. There's got to be something listening, you need milk in Asda and then you get adverts for milk, you know?" Individual with a physical disability

There was a high degree of variation in participant views as to who is best placed to collect data. Collection from within a 'community' was seen as leveraging existing, trusted relationships to help access people within those communities. However, it was recognised that this could lead individuals to withhold information on views that were perhaps perceived as more controversial, or went against perceived community norms, for fear of being identified. This was particularly the case within faith communities where 'outside' researchers were seen as offering the greatest potential for more objective data collection.

"Oh this is for the government', I will respond to those questions in a different way. Depending on how I feel about the government and authorities, that will impact on how I respond. As if it is the Catholic Church doing some sort of research, my own bias about these other organisations plays in the way on how I respond to those questions." Individual of Christian faith

The majority of participants felt that researchers working on behalf of the government were best placed to conduct research with all but the most under-represented groups (e.g. asylum seekers, homeless people, and Gypsy, Roma and Traveller communities), but that the process for engaging communities would be facilitated through trusted CSOs.

5.6. Principles for data collection

Participants were asked as part of this research what rules, or principles, should be adhered to by those organisations that collect, process and use people's personal data. There was a large degree of consensus around a common set of principles which align with a number of the cross-cutting data principles underpinning the ONS Data Strategy, namely:

- **Transparency and Proportionality**. Participants emphasised that data should only be collected if it is proportional to use. Researchers need to be transparent and upfront about their need for collecting data, and there must be explicit statements, in plain English (and other languages to meet individual needs), on what data are being collected, the purpose for collecting it, how it will be used and/or shared, and around confidentiality. Greater transparency was seen to be important in building trust and the trustworthiness of those organisations collecting and using data.

“Just to be clear on where it's going, and why it's all needed. Who is going to have access to look at it, and will it be used in the future.” Child or young person

“It needs to be proportionate access. Do not need to know how much there is outstanding on my mortgage if you are not going to use that information. If it is not beneficial to your service or to me even. Needs to be clear about what is going to happen to that data.” Individual of Christian faith

- **Inclusive approaches to data collection.** A one-size-fits all approach, while cost efficient, was recognised as neither desirable nor inclusive in enabling people to participate in providing their data. Instead it was seen to be important to think carefully about not just the methods for collecting data, but who was collecting their data.

  “Whether you get a paper version of something, whether you can do it online, some people may need a face-to-face appointment to provide all that. Someone may want to do it over the phone, they may not want to use a computer. There are so many different reasons why someone may need an alternative way of providing that.” Individual with a learning disability

- **Categorisation and Classification.** LGB, trans, non-binary and gender diverse individuals highlighted issues surrounding inaccurate or inadequate classifications within data collection. Categories such as ‘other’ and ‘prefer not to say’ were mentioned as being neither inclusive nor helpful for intersex and non-binary people, who often feel most invisible in data collection. Concerns around a lack of coherence and harmonisation in the categories used for collecting data on these individuals were also raised, as this was reported to contribute to inaccurate and unrepresentative data.

  “I prefer putting trans woman for gender but like there is a lot of inconsistency because on forms I go between trans woman and woman and other...Often you will see just transgender as its own option, which is true in the same way that LGBT can be its own option, but it doesn’t tell you specifically what’s going on with somebody.” Trans, non-binary and gender-diverse individual

- **Informed consent**. The current challenge identified by many participants is that information on how data are used is often lost in the small print, which people either do not read or cannot understand. There is a need for this to be simplified and alternative levels of consent provided to not unduly restrict access to services. For some individuals, being able to speak to people about how their data is to be used is reassuring.

  “I don't really think there are many risks but it's just the thought of not knowing who's going to access it or who could access it or who has accessed it if I'm not explicitly consenting to people who can do it each time.” Individual with experience of homelessness

  “Consent. What part of the information I would like to be shared, what part of the information I would not want to be shared.” Migrant or asylum seeker

  “I thought they kept basic records about your health, it probably protects the company in case you need to claim against them. But some of the things, I thought well, some of the things on that record, I wasn’t aware of.” Individual of Christian faith

- **Anonymity**. People expected that data shared with organisations would be collated and used in such a way as to maintain individual anonymity so that they could not be identified by users of a given dataset. The majority of people felt that this was achieved through aggregating their data with data gathered from others and removing any personally identifiable information. Trans, non-binary and gender-diverse individuals, however, noted that they are particularly conscious about the anonymity of data for individuals within their community, due to the prejudice and potential harm they could face if their information fell into the wrong hands.

“If you're part of that pool, you'll take your age range, your gender, your religion, you're pooled in such a large category, you'll be unidentifiable. It's not going to cause me any harm. I don't think my data will be misused in any way. I'm happy to share it if it's going to benefit somebody.” Individual of Islamic faith

“I think that's a fair use of the data because it would probably be anonymised. It's not got anything to do with you anymore, it's just group data.” Individual with a learning disability

- **Data security**. Where personally identifiable information was kept in more of a raw data form, there was an expectation that this should be kept securely and encrypted to protect against data breaches.

“I think it would be more of a concern of how the information is stored than collected, especially in the current climate, we've heard a lot of big organisations, the databases being hacked and being sold on the black market for fraud purposes.” Pakistani individual

- **Consultation feedback**. Beyond administrative data that is provided in exchange for a specific service, there was a desire for feedback in response to providing more extensive or ad-hoc personal information. Many under-represented groups feel that they are consulted regularly but are rarely informed of the research outcomes. Additionally, there is often a belief that their participation in research is not used to meaningfully change policy or improve their daily lives, do not see their feedback as resulting in any changes to policies or their lived experience, leading to frustrations that their voices are not being heard.

“I haven't been presented with any evidence that tells me that them having that kind of information about me as an individual benefits me in any way, shape or form”. Black British individual

“I think the big thing around data for me is transparency, what are we doing with that data? It's that you've said, we did, type of thing. If you're telling me, Mr Singh, if you divulge this information to us, it's protected, what we're using it for is this, and based on what you gave us in the last 10 years, we've given this to your area based on that data. That makes a massive difference.” Individual of Sikh faith

“There does seem to be like a big disconnect between data and policy acting upon it.” Trans, non-binary or gender-diverse individual

“I mean it was nice for everybody to receive a reminder to do the Census. I would love to see the actual document sent out to every household saying wow isn't this amazing? Do you know a nation of people are 1% trans, 1% intersex and of that, this much is non binary people. And we're also 34% from an ethnic minority. That kind of information going out to households... It feels more engaging as well you think I'm doing the census now and they'll tell me what the results are. That would be my biggest suggestion. Tell everyone what the results are not only is a great way to publicise it but you are giving the people that answered it some real facts about what they're doing and also dispelling any beliefs they might have about diversity in the country.” Trans, non-binary or gender-diverse individual

- **Access**. This was only mentioned by a small number of participants, but the point raised was that it would be a positive step for people to be able to see the data held about them, rather than needing to make a request. Currently there is felt to be a power imbalance between the researcher or data collector and participants, particularly if any personal data is being shared with third parties. As the data subject, people felt they should have access to the administrative and wider sensitive data held on them.

6. Building a more inclusive picture of individuals with protected characteristics or from underrepresented groups, disaggregated by equality areas

6.1. Equality area: Disability and Neurodivergence

Amongst those people we spoke with that identified as having some form of physical disability, learning disability or neurodiversity, there was generally a high level of willingness to provide data on their circumstances and needs as this was seen as enabling services and organisations to take this into account, helping identify individuals in vulnerable circumstances and supporting service planning.

"How many people in Brent have Down's syndrome? The powers that be could see there were a lot of disabled people in Brent and they needed help, more facilities." Individual with a learning disability

"The local authority knows all about my daughter being disabled, because I have a social worker and I'm involved with them, like getting her into day centres and getting respite for me so she can go off somewhere. That she's got Down's syndrome, and that I've got a disability myself, a chronic depression it is. I suppose we're listed on a list somewhere that we're vulnerable." Individual with a learning disability

However, there was also a relatively low level of trust in central government due to the perception that disabled people were treated as a low priority (as evidenced by austerity policies reducing benefits entitlements and prioritisation of COVID-19 roll-out). Participants felt that government collect lots of data from and about disabled people (notably around physical disabilities) but did not act on this data as policies and services were not felt to meet individual needs. Across the individuals involved in group discussions and interviews there was a high degree of disillusionment in how data was used by government.

"There's lots and lots of things I'm sure we could all complain that you don't get help with or they don't do, and I think the government knows about these but often doesn't want to address them because they're not a high priority. They've got the data but having the data and doing something about it is an entirely different thing." Individual with a physical disability

"It's what they do with the data that matters. They can say so many people are registered disabled, or what have you, but if they never actually do anything to provide services. That's the important thing. So what?" Individual with a physical disability

Participants reported feeling that administrative data, and data used to represent disabled people at a national level, is not representative or inclusive in terms of understanding the circumstances of disabled people. Just the terms 'disability' and 'learning disability' are seen to mask a wide variety of conditions and experiences, leading to confusion both for those classifying themselves and for those who need to use that data.

"The disability question, there's a lot of variation on how they put it. Sometimes they want you to specify what kind and then the groups that they write are always different for each form. Sometimes it can be hard to think, 'Which one does it fall under?' Disability is such a broad spectrum and there's so many things within that. Sometimes I don't know which box to tick and sometimes I don't know whether I want to share that information. Yes, I think it depends on the form and the context I think whether I want to. Sometimes I might put the 'did not want to answer' option." Individual with a learning disability

"I've got no field of vision in one eye and my other eye is trying to work but is deteriorating fast. I've got to try and do school work with a 9-year-old, a nearly 13-year-old as well as be a carer, as well as try and

look after myself. You can't put that on a piece of paper and get them to understand it." Individual with a physical disability

The benefits of sharing data on personal circumstances were typically seen as relating to personal benefits in terms of improving access to a public service or reducing the potential for embarrassment in situations where disabilities are not understood or accounted for. There was a high degree of frustration and exhaustion expressed by participants at having to provide the same administrative data to multiple public services, and a desire for data to be linked or shared between public services to reduce fatigue.

"If they could all collate all the information together I wouldn't have to repeat myself every couple of years on a DLA or PIP form. That would be nice. Why do they send that form out? Cerebral palsy doesn't get better in a year and neither does the spinal cord. It drives me mad that. Don't send me the paperwork. It drives me mad." Individual with a physical disability

"The benefits, it's good for the user's point of view because it means the services are joined up, all linked in together, which for me, and particularly as someone one the autism spectrum, I get tired of having to repeat myself all the time. When I repeat myself I get tongue-tied and quite agitated." Individual with a learning disability

When asked to think about their specific disability, individuals felt that important information to collect to aid inclusivity (specific to people identifying as having some form of physical disability, learning disability or neurodiversity) included:

- The lived experience of people, including support needs
- Life outcomes, including housing, income and employment situations
- The differences between neurodivergent and neurotypical people

"I think an important one for neurodivergence would be like 'who do you live with? Do you rent your own house or flat or are you still living with your parents?' I think that can be a big thing for neurodivergent people, being able to take the step of moving out and not relying on parents." Learning Disability

Risks of sharing data related primarily to concerns around the potential for discrimination by businesses (e.g. in respect of loans, insurance or employment) and by government in relation to assessing and/or checking on benefit claims. To mitigate against these risks there is a need for people to have real clarity on the purpose and both current and future uses of the data being requested.

"Somebody that was offering you a job might not want you because they'd think, 'Oh, well we're going to have to make all these adjustments if she comes to work for us and things so we're just not going to pick her.' That's really wrong but if they got hold of your data it would be really easy for them to do that." Individual with a physical disability

"I think the negative side, what would concern me with certain aspects of my information is how the agencies would actually use it without my permission. Whether or not there would be caveats, or whether or not I would be reassured or explained what circumstances that my information would be used. For example, say I was on state benefits, I would worry that the government might use that information to spy on me, to actually check that I wasn't swinging the lead." Individual with a learning disability

A wide range of individual needs were identified relating to physical disabilities – from hearing or visual impairment through to tetraplegia – and for learning disabilities which can also vary considerably by time and circumstance, so participants recommended that methods for data collection should flex according to preferences and needs. Face-to-face data collection may be needed for some people to enable their participation, though for others it may promote stress. For online data collection people with learning disabilities reported they could get confused by the structuring of questions, the terminology used and options available, which can also create a high degree of anxiety. There was a desire for ensuring that the processes (questions and approaches) used to collate data are inclusive, such that people don't feel excluded or stressed as a result of the approaches taken.

"The only problem I come across personally is that there has been instances where he may have been contacted directly by these government bodies where it's inappropriate. It's inappropriate because he could say the wrong thing, he could not understand. I'm happy for them to have his information as long as they know there's somebody else to contact before they contact him." Carer of individual with a learning disability

"I never have any idea on what to pick on the ethnicity question. That's the thing I hate most about taking a form. Every time I fill out a form, I have an existential crisis. The way they're divided it's like, 'Are you mixed white and Asian? Are you mixed white British and Irish?' I'm half British, quarter Irish and quarter Nepali." Individual with a learning disability

"I think neurodivergent people are more likely to, for example, be home-schooled and that kind of thing and be less easily reached by outside things and have less connection with the rest of society and that kind of stuff. If someone can't have a job or hasn't managed to get a job yet or if they're home-schooled rather than normal school then they're less reachable and less in society." Neurodiverse individual

6.2. Equality area: Ethnicity

Group discussions were undertaken with a variety of different ethnic minority groups, including people identifying as Black British, Black Caribbean, Somali, Pakistani, Indian, Chinese and Gypsy, Roma or Traveller. Across these groups there were a number of key similarities and differences in views toward the inclusivity of data. Key in understanding the perceptions of different ethnic minority groups is to recognise the relationship they currently have with UK institutions and government, and the extent to which, as an ethnic group they feel well represented and supported by policies, services and in society.

Certain ethnic minority communities, including Black British, Black Caribbean, Gypsy, Roma and Traveller ethnic groups, were more questioning over the trustworthiness of government and certain public institutions (notably police authorities). Participants within these communities often described feeling disadvantaged, persecuted, and disempowered, all of which reduced their willingness to share data with 'authorities'.

"One thing I have noticed about the black community is that it's based on trust. I feel because there's no trust, they're not really willing to give their information so therefore there's not much that can be done in terms of helping and using information." Black British individual

Amongst Gypsy, Roma and Traveller participants there was a very high degree of suspicion of sharing any data with government or public services as it was seen to be an invasion of privacy. If there was a perception that the data was being collected specifically on Gypsy, Roma and Travellers this raised further suspicion as to the purpose, and a degree of anger at being categorised and "tracked".

"*It's like big brother really, isn't it? It's like they're keeping an eye on you from all angles. They're just keeping an eye on you from all angles as if you're doing something or you're stepping out of line."* Gypsy, Roma or Traveller

“We're private people. We don't really want to be giving out what we're doing, where we're going, the next campsite we're going to. You know what I mean? It's our business, it's nobody else's business.” Gypsy, Roma or Traveller

Experience of stereotyping and discrimination (e.g. with local authority planning permissions for Gypsy, Roma and Traveller communities) led to concerns amongst certain ethnic minority groups that their personal data will be used to label them, to discriminate against them and reinforce negative perceptions of their ethnic group. Amongst some Black British, Black Caribbean and Pakistani participants there was a view that data linked to ethnicity could be used for racial profiling, and Chinese participants specifically noted that collecting data on political views in conjunction with data on ethnicity was unacceptable as they believed this had the potential to be “*weaponised*”. This was also referenced by Roma participants in relation to the holocaust. This feeling of persecution for their ethnic identity and culture led some Gypsy, Roma and Traveller participants to not disclose ethnic status, and holding a general aversion to sharing their information.

“When my kids are at school, I don't put they're a gypsy or traveller, I just put them down that they're white because I don't want them singled out in any way at school.” Gypsy, Roma or Traveller

A key concern for ethnic minority individuals is how data collected by government are used to influence policy and practice. Amongst Bangladeshi, Chinese and Pakistani individuals there was a general feeling of positivity toward data collected by public authorities, with links made to the potential for this data to establish more targeted community services (for example, the COVID-19 response) based on ethnicity. However, amongst Black participants, there was the view that whatever data are being collected is not used well, as the policies and decisions made by government are not favourable toward ethnic minorities. Black participants highlighted their concern that UK institutions and public services do not fully understand nor appreciate the complexity of the lives of Black people in the UK, and the social inequalities faced as a result of public policy. A common view amongst all ethnic minority groups was that data is being collected, just not always used effectively to address needs, or research outcomes communicated.

“This is the problem; they've got the information already. It's how they perceive it and how they work it out.” Black Caribbean individual

“I think it's collected okay but I'm not really sure how it's used, so I can't really judge if it's used well or not. You get that message don’t you, this will be used for research purposes, but we don’t actually know what the definition of that is, and we don't really see the benefits of it. It's just sometimes just goes into the ether.” Pakistani individual

For most ethnic minority participants, the question of improving inclusivity was less around what data was being collected, and more around who was collecting their data and for what purpose. There was positivity around the potential that better use of data could have for communities, in highlighting both disadvantages and areas of strength within communities, in building understanding, empathy and appreciation.

“I think that's actually a good thing if it's to try and help everybody, like for schools and whatever else and houses and if it was to do with travellers then maybe thinking about putting, 'Well, we need camp sites there.' I don't think that's a bad thing.” Gypsy, Roma or Traveller

“I think it [better use of data] would encourage more support for our community. Whether it is at work or in the health service or just generally how people interact with us, I think there would be more compassion.” Black British individual

"I'll say the most important thing is to explain to them how important it is to give different information and how it's going to benefit the community, yes, it's about explaining and making them understand. Nobody does anything if they don't know what it is for and why." Somali individual

Chinese participants expressed the importance of considering data ethics when asking people to share personal data. Participants noted that data collection should not only be proportional to use but should also be transparent in its purpose, stored in accordance with data protection regulations, and outcomes clearly communicated to participants.

"The important thing is to get our consent to share the data. I would mind if they shared the data with another public agency and they didn't seek my consent. If they seek my consent explaining why they're sharing my data with another public agency, then I'll be quite prepared for that to happen." Chinese individual

"They're really good at how they conduct this [Understanding Society] at Essex because they send you this newsletter a couple of times a year saying what they've done with the data. And you think you're contributing, and they do some amazing things with the data in terms of looking at the population and that feeds into" Chinese individual

However, many individuals, notably in the Black British group discussion, felt there was very limited personal benefit that could come from sharing their data. Instead there was a perception that their data would be used for racial profiling and to potentially disadvantage them.

"I think there is an awful lot of profiling that is going on in collection of people's information. I am not 100% sure that some of it is not being gathered for nefarious reasons." Black British individual

When asked to think about their ethnicity, individuals felt that important information to collect, specific to people's ethnic identities, to aid inclusivity included:

- Experience of racism, hate crimes and inequalities, and how this impacts physical and mental health
- Health issues amongst ethnic minorities, in particular mental wellbeing and blood disorders
- Languages and dialects spoken by the population in the UK, to inform service provision such as translation services
- Experiences of school and education
- How people understand and feel about their personal identities
- How people of different ethnicities socialise, different family and social structures, and which social platforms are used to communicate
- Experiences of social care
- Experiences of poverty and income levels
- Age, and age-related services accessed

"I think it's really important that the people have an understanding of how that [lived experience of being Black British] really affects someone's health, from the suppression of emotions and the suppression of identity as well, in some cases." Black British individual

"Regarding mental health I think some Somali people are war-torn and they've got bad PTSD or other mental health conditions, so they might have to ask individually or as a collective how they've been affected or how their mental state has been affected by what they've been through." Somali individual

"I think it's a case of, first of all, just asking people about the problems that they face today. It could be doing that kind of research. I would imagine you will identify some common themes that will particularly affect the Chinese community. You know, the community getting older. People dying off and there'll be a lot of lonely elderly Chinese people in especially my part of London. Knowing that through the official data, through the official stats. You know, that could be useful for our community leaders to do something about it. So, basically asking people and then do it in like questions around those problems in order to collect data and statistics on that." Chinese individual

Participants also highlighted the importance of collecting nuanced data in relation to ethnicity, that respected the wide variety of ways in which people understood and articulated their ethnic identity. There was a view that current questions on ethnicity used a categorisation which failed to capture the nuances and complexity of their identities (e.g. a hybrid Black/Caribbean/British heritage, or Pakistani Kashmiri and Indian Kashmiri), or introduce confusion on the basis of overlapping categories such as African and Somali, or Black British.

"I like to say, 'I'm from an Afro-Caribbean, my mum's Ghanaian and my dad's from Barbados,' but there's no form that can get me to say that. Then I just feel outcasted." Black Caribbean individual

In contrast to what several CSOs working with Chinese people in Stage 1 research stated, Chinese participants felt that a broader Asian categorisation in data collection and analysis caused distress and did not feel representative of the Chinese population.

"Chinese category has been taken out and substituted into Asian. So, I think there's been some complaints about that by people because our Chinese experience compared to, for example, people from Manila or Pakistan is very different." Chinese individual

Amongst some communities there was more of an active pushback on providing data relating to sexual orientation and political views, which were seen to be private matters not for collection in research.

"Your sexuality and stuff, I'm not comfortable with that. Not really, no. That's your own personal taste, I feel." Roma, Gypsy or Traveller

Individual participants from across the ethnic minority group discussions highlighted the importance of seeing themselves in those people that were collecting their data. For some, having researchers from the same ethnic background would be a key facilitator for encouraging more inclusive data collection. Others highlighted that it was important to recognise that ethnic identity is just one part of how they saw themselves, with shared localities and life experiences also critically important.

"We trust our own folks so we would definitely listen better to somebody that we know, somebody from your own background. Definitely." Gypsy, Roma or Traveller

"I think the reason why the majority of Somalis don't partake [in the Census], it is because they don't understand what the use is for, even though the information is in the letter, it's not widely discussed or widely well-known within our community of this." Somali individual

"I think it's to do with the type of people that collect the information in the first place…you need more people with different life experiences that can understand this situation." Black British individual

Finally, dependent on the community and related factors such as age, there were also other needs relating to language, literacy, digital literacy, cultural preferences for communications, the research environment and competing pressures on people's time, all of which impacted participation in data gathering exercises.

"A lot of travellers, I don't know if I can speak for everybody but my father, he can't read and write. A lot of people that will be a massive thing." Gypsy, Roma or Traveller

"I live in quite a small town in Manchester, it's full of people from my race, people who have come over from Bangladesh. Their English isn't that good so they wouldn't understand anything anyway, they'd rather not participate in anything." Bangladeshi individual

"I think the key thing is the language. Basically, the focus group has to be conducted in one of the Chinese languages. Cantonese, Mandarin… Hakka, Hokkien" Chinese individual

6.3. Equality area: Homelessness

Amongst those people we spoke with that had experience of homelessness there was a high degree of trust in government[9]. There was a belief that data collected by government is used to provide better services at an individual and whole society level. While people with experience of homelessness were used to providing administrative data to services, they were often not used to providing information on their thoughts, feelings and experiences related to homelessness, and had little sense that by not providing that data it led to any material advantages or disadvantages for them. This lack of experience and consideration of using data in this way led to challenges with people identifying potential benefits of more inclusive data outside of the context of a transactional service. Their day-to-day concerns were much more pressing than concerns around providing data and how this data was being used.

"Whenever I go to the doctor they don't ask me how am I doing for surviving, they don't ask me that question. I don't tell them because they don't ask me that." Individual with experience of homelessness

People with experience of homelessness reported experiences of "being heard, but not listened to". While their temporary accommodation status was accounted for in their interactions with local public services, their experience of this accommodation was not felt to be fully understood. There was also frustration expressed by participants at having to provide the same administrative data to multiple public services, and why their administrative data cannot be centralised.

"It would be easier if there was just one massive system where everyone could access it rather than constantly having to repeat it. It would be easier if multiple people could access the same points". Individual with experience of homelessness

The level of comfort amongst participants in sharing information on their living circumstances varied depending on their individual circumstances. Those in temporary accommodation but otherwise retaining (or attempting to retain) independent lives were concerned at disclosing their circumstances due to the potential for discrimination and negatively influencing the way they are perceived. They may not identify as homeless and may not want to identify as homeless in any way that could be personally identifiable for "fear of the unknown and a fear of uncertainty of who is seeing my data". There is a level of awkwardness and shame in the circumstances of some homeless participants that may prevent voluntary disclosure of their situation, notably to employers.

"I'm quite conscious about what information I share at the moment. I just don't want it to hinder me. For example, if I was to apply for another position, would they ask for my address? I mean, it is something that I've thought about, whether they'd ask for my address, and place of residence and provide proof of

---

[9] While this contrasted with the view of CSOs, this may reflect the circumstances of those individuals that we spoke with as part of this qualitative research

all that. So I'm kind of staying put at my current location in terms of work, just to make sure that I don't have to provide that information at the moment." Individual with experience of homelessness

The benefits of sharing data on personal circumstances are typically seen as related to personal benefits rather than benefits for the broader group (i.e. people experiencing homelessness), and relatively little in the way of personal benefits were identified beyond access to a given service (often health services). Participants were often unclear how sharing data on their specific homelessness circumstances can help others.

There was an appreciation amongst those involved in the research that Census participation may be hindered through housing circumstances and transience, which was a disadvantage in terms of representation. Dependent on whether you are homeless on the street or homeless and in temporary accommodation has a large impact on participant views as to the best processes for engaging and collecting information.

If looking to collect data from homeless people on the street, then this is best collected in person, via CSOs that are already working with those people, as a trusted relationship already exists. Due to the insecurity of circumstances and potential risks of data sharing there is a degree of wariness that can be addressed through more of a qualitative approach (via phone or face-to-face).

Finally, it was clear through discussions that questions around circumstances needed to be framed in an empathetic and supportive manner to promote the provision of more honest and inclusive data.

6.4. Equality area: Faith

Amongst those individuals interviewed from Buddhist, Christian, Hindu, Jewish, Muslim and Sikh faiths, there was a high degree of trust in government, with data collected on behalf of local or national government typically associated with being used for public good.

"The government, at the end of the day, is trying to improve people's lives so I think when you say, 'I'm from the government and I would like to collect your data to do X, Y, Z and improve your life,' then it's completely fine." Individual of Hindu faith

The data held by government was expected to present an unrepresentative picture of older members of faith communities due to language, digital access and cultural needs impacting participation in research. When asked to think of their faith group, individuals felt that important information to collect, specific to the faith equality area, to aid inclusivity included:

- Better data on the local faith populations in terms of socio-demographics to enable them to be accounted for in the planning of public services and in funding community facilities
- How people understand and practice their faith, how much time is spent within places of worship, how that time is spent (i.e., time practicing religion vs more social and communal activity) and generational differences
- Dietary requirements and preferences
- How people have (and how they felt they have had to) compromise their faith/culture to fit in to the UK
- Issues faced across different aspects of their lives. E.g., finances, homes, living costs, work, health, culture, sports

- Prevalence of more sensitive issues impacting community, such as mental health needs and alcohol misuse, which are often not discussed openly and therefore not properly addressed by services
- Experience of persecution or discrimination based on faith and ethnicity. It should be noted here that individuals of Christian faith engaged in this research did not consider experience of persecution in relation to their faith and did not anticipate risk of harm from disclosing faith information. However, they did acknowledge that this may not be the case for other groups, including other faith groups, in UK society.
- More data and research are needed on underrepresented groups within faith groups, so that their voices are heard.

"I would only give that type of information to an established organization. Government, or local government are a neutral body, they're not going to be discriminatory, or try and persuade you to move to another religion." Individual of Jewish faith

Better data on faith groups was seen as having the potential to ensure that public services are better able to meet religious needs (e.g. ingredients in medicines, food provided in public services, or faith support services in hospitals) at more of a localised level. It was also felt that a lack of general understanding of minority religions in the UK, and mis-representation of minority religions by the mainstream media and government (e.g. in terms of generalising practices of one sect or ethnic group across an entire religion) contributed to issues of discrimination, racism and stereotyping. More inclusive data was seen as having the potential to address this.

"The way the government portrays people like me are based on one particular group that do not paint a picture of everyone who are in our Muslim faith. So, it would benefit us, in a way, that we share how we're really like, how we're like with others, how we interpret life in that way, for them to understand, for them to represent us in a good way. It actually brings anger like, 'Why haven't you guys picked up on the fact that we are not like this?'" Individual of Muslim faith

Within the data collected, participants felt it would be important to break out findings by age and by different religious schools, sects or denominations (noted specifically for Christian, Hindu, Muslim and Sikh communities) to allow for more of nuanced understanding of minority faith groups. Also important was to ensure that population-level data was broken out both by faith, ethnicity and nationality as there was concern that some smaller faith groups are lost within a broad and culturally blunt 'BAME' grouping which was seen to be "*offensive*".

"Even though we're a small religion and a small size in this country, we're not all the same, we all have different needs and behaviours. Different demands and again, there's this issue that globally amongst Asians that you don't talk about things." Individual of Sikh faith

"The Catholic Church is totally different to the Church of England. And what is expected of people in Catholic religions, so it's all different and people have their own ways and beliefs and thoughts, so they've all got to be respected. But they've all got to be reflected in society." Individual of Christian faith

There was acknowledgement that language and cultural needs can make it hard to access information about older generations and those who are more recent migrants; and there is consequently a need to ensure questions/responses can be provided in relevant languages (e.g. Hindi, Punjabi and English for Sikh community) to facilitate participation. Data would ideally be collected via spaces of religious connection (particularly if seeking to engage older groups), by an engaging researcher who uses appropriate language to communicate with participants. While collecting data from 'within' communities would facilitate access, this was also seen to compromise openness on more problematic issues and reduce participation of those members of a faith group who did not attend places of worship.

"No community will ever admit it but there is this hidden agenda within communities to protect communities. Whether it's honour-based violence, alcohol abuse, arranged marriages, we see that struggle all the time. Through that protection of the wider image is where you have things going wrong, so I think for the government local authorities to say, 'We're collecting this information.'" Individual of Sikh faith

"I think it [the researcher] should be somebody impartial, as unbiased as possible, just so it does not reflect whatever judgement or beliefs or whatever the Catholics might have. I feel it is really hard to tell without knowing what you need these data for, but for example, if we would think about views of Catholics on pregnancy termination let's imagine, and you get the local priest to collect these data, you would get biased data for sure . Whereas if you get somebody external, somebody impartial, probably it would be more accurate." Individual of Christian faith

When asked whether they feel data represents them, some participants recognised that quantitative data or administrative data cannot fully represent a person and their views in relation to their faith or religion, especially as these can change over time.

"No, I do not think [it fully represents me] because if I think about these admin data, it is concrete data, isn't it? But a human being is much more than that and people change as well, you know? Life is dynamic, so today I'm married, tomorrow maybe I won't be married. Today I am Catholic, tomorrow I might be Agnostic. So yes, it is not a static sort of concrete thing." Individual of Christian faith

"I don't think they ask very deep questions. They just want to make a proper study of society, so how many people have got such and such a faith, how many people have no faith at all. Who is practicing religion, who is not practising. And I think it's more for this tick box basis, but I don't think they want to know information about your faith in a deeper sense." Individual of Buddhist faith

Finally, it was suggested that people should be given a choice as to how they engage in research. This included giving people the choice over survey mode e.g., responding online, in person, over the telephone or via physical copies, to enable people with diverse characteristics and circumstances to participate fully. It also included giving people the choice between answering via 'tick boxes' or free text responses.

"I think that's inclusive because that's, so long as people have the choice to do either [respond digitally or physically], that's important...I like doing it digitally because I'm dyslexic to start off with, so I can fill in things easier on computer than I can on a form." Individual of Christian faith

6.5. Equality area: Migrants and asylum seekers

Amongst those people we spoke with that had experience of migrating to the UK there was a high level of trust[10] and acceptance of sharing data with the government and public services. Those with experience of seeking asylum in the UK recalled comprehensive interviews undertaken with the Home Office as part of the immigration process and accepted that it was necessary for the Government to collect personal data to assess the validity of people's claims to asylum. Provision of personal data to public services was

---

[10] While this contrasted with the view of CSOs, this may reflect the circumstances of those individuals that we spoke with as part of this qualitative research

also typically linked to positive personal outcomes, in particular for more recent asylum seekers who associate sharing of data with receiving a service.

"When I give my information in hospital they give me papers, I can go to hospital for my check-up. If I give to the Home Office my information for my life, I came here to save my life. It's good for me to give my information, because they help me after I give my information, they know about me." Migrant or asylum seeker

Administrative data was seen as providing an accurate, but incomplete picture of people. In particular, there was felt to be a richness and complexity of circumstances that have led to people's arrival in the UK that is not captured in most administrative data.

"They don't ask very personal story, feelings, attitude, anything that's in my mind that I would like to say. My own voice as a person of a refugee background. They don't lend themselves to all these kinds of things." Migrant or asylum seeker

Newly arrived migrants and asylum seekers were seen to be a highly vulnerable group as they often lacked social connections within the UK or English-language skills, both of which make them vulnerable to exploitation (often by those 'communities' they feel they can trust because of a shared nationality or language).

"I think that because the newly arrived refugees, asylum seekers, immigrant community are very, very vulnerable. Once they come they are confused. It's a different language, different culture. And then lots of negativity about government policy for refugees. People have very little support. Only a few lucky ones who know a little bit of the language, let's say 30%, they can get support from the related agencies and organisations. The problem is you must speak English a little bit, if you don't speak English then you don't have confidence and then you just rely on your environment around you. If you are lucky you get good people, if you are unlucky then bad people use you, which is the most likely." Migrant or asylum seeker

When asked to think about their experience of migrating to the UK, individuals felt that important information to collect, specific to migrants, asylum seekers and those impacted by modern slavery, to aid inclusivity included:

- Information relating to their home country situation and reasons for migrating to the UK (for purposes of asylum claims)
- Their current situation and experience of integrating into UK society (specifically housing, employment and benefits)
- Skills and qualifications on entry to the UK
- Sources of support, support needs (to enable better signposting of support and information provision) and wellbeing
- How migrants are contributing to UK society (to tackle more negative narratives in the media)

"I would have liked people to know about my qualities, my ability, my experience, my qualifications that I have come here with. Unfortunately, I find myself to be an unknown quantity." Migrant or asylum seeker

"What kind of social life are you doing? Are you in a club, a group? Do you do volunteering? Are you just sitting in a flat the whole day, out at work then back in your flat, not interacting with the community? This is important because it gives you the sense of belonging. You've come in as a stranger, you feel segregated, you want to feel part of the community that you're being introduced to." Migrant or asylum seeker

"Some people think that asylum seekers, some communities, are coming here just to reap the benefits, and that's not the case. They work hard. There is some kind of wrong narrative here." Migrant or asylum seeker

In spite of the perceived benefits of sharing data, there was also awareness of a number of potential risks, notably fraud. Some migrants who had been here for longer periods of time reported dangers of discrimination due to immigration status, and encountering problems accessing services. The Biometric Residence Permit was seen by migrants to the UK as being a sensitive form of administrative data as it conveys their status as a refugee, which could be used to discriminate or exploit them.

"When you present your ID as a proof of identification. On it now it says you are a refugee, that makes people look at you in an undesirable way, to me, that's how it comes across when people see that status. To avoid that I applied for a driver's licence, I use that most of the time instead of my resident permit card." Migrant or asylum seeker

Dependent on the circumstances under which some migrants and refugees came to the UK, and the country they were migrating from, some participants reported being less comfortable in sharing personal information. This impacted what they would be willing to share openly in administrative data or survey responses.

"You maybe have got equality here, but I still consider that from where I come from, to get into how you guys are so open and so responsive. I'm not comfortable yet, because I came from a society that was very judgemental." Migrant or asylum seeker

Participants felt that migrants and asylum seekers would be very willing to share data with the government and with public services, though greatest reach within refugee and asylum seekers would be achieved working through local CSOs who have established trusted relationships with these communities. There was a strong preference expressed for data to be gathered through more qualitative research methods, which were seen as more personable and enabled participants to better articulate themselves.

"You have already dealt with these people, interacted with them, so you've got no qualms on giving over further information because you've already done it before." Migrant or asylum seeker

There was recognition amongst those interviewed that there will be a subset of refugees and asylum seekers, in particular those in the UK illegally and experiencing modern slavery, who will not be reached in the vast majority of government-led research as they are not engaged with any public services or CSOs.

6.6. Equality area: Children and young people

This research involved speaking with six children and young people, of varying ages between 9-17 years old, with those aged under 10 years old also interviewed alongside their parent or carer. These children, regardless of age, could articulate what they considered to be their data, though younger children typically articulated this as their "personal information" and this was largely related to personal identifiers such as name, date of birth and contact details. Experience of providing data increased with age, so those younger children engaged in this research had very minimal experience, and data was very much viewed from the lens of online gaming and safety online.

Children and young people were very clear on what they saw as being appropriate and inappropriate data to collect, with younger children more likely to mention parents as being a key source of information and direction here. There was generally a high level of willingness to provide data to public services, and an awareness that parents typically did this on their behalf which was seen to be appropriate for more administrative data, at least until they were 16 years old. There was an expectation that services collecting these data required it to provide access to services, and participants were largely comfortable with this. Younger children recounted experiences of being asked for personal information when accessing apps or games online, and the need to be wary about who to give their data to follow guidance from parents.

"I think I'd more trust the government, or someone working in the government than a private organisation. I wouldn't know what's in place to keep it safe. The government already have access to a lot of personal information probably, so I think they're probably best with it." Child or young person

Some forms of data were however seen to be highly personal – notably health, care-status, sexual orientation and faith-related data - and there were more sensitivities around whether that data was collected and how that data was treated. As an example, amongst those children and young people who had care experience, there was discomfort at how information relating to their care or adoption status was requested repeatedly across public services, and publicly discussed in open environments. There was a desire amongst children and young people for privacy around potentially sensitive information, and to feel in control of what they are asked and who their information is shared with. There was generally a higher level of concern amongst younger children of their data ending up with "someone random" and an expectation that data collection should be kept to a minimum

"*They have this information so they shouldn't make it come up. It's not hidden but it means that other patients hear, and he gets upset each time we say it."* Adoptive parent

"Only professionals should know my information and they need to ask permission to share it. I'll give my permission if it helps." Child or young person

"Sexual orientation, and even your religion, it's sort of fading out now and it's becoming a bit more personal. It seems as though it's too personal, it's too identifying for you." Child or young person

For other children and young people, there were limited concerns around sharing data, as long as it was collected and used for clear purposes that they had consented to. Of particular importance for public services to have access to was data relating to experience of school, health conditions, mental health, care experience and faith. These were seen to be important in understanding individual needs and situations to help contextualise behaviour and account for in the services provided.

"I think I'd ask about mental health, and if you've been seeing a counsellor, or anything like that. Is there anything that they can do to help." Child or young person

"If you come from a background where you are in care or you were previously in care, medical history, obviously that's common sense, really. To some extent, your religious background. Obviously, teachers want to know if you're a Jehovah's Witness at school, they'll also want to know because they need to cater their lessons for you." Child or young person

One older young person interviewed as part of this research highlighted the importance of educating children in school about the importance of data gathering exercises such as the Census, to explain about why questions are being asked, how data is stored, and what difference taking part in the survey makes.

"I'd forgot it even existed until this year, so they could do an informative couple of sessions in the curriculum in schools, so young people feel as though they understand what it is and how it represents them, even if they're not filling it out themselves." Child or young person

Children and young people highlighted a wide range of different ways in which their information could be collected, from different methods to different platforms for engaging with them such as through schools, youth and community groups. There wasn't seen to be one 'right' approach, rather a range of different approaches that could be taken dependent on the nature of the questions being asked and who the research needed to involve. For example, older young people would be most comfortable with online self-completion options. Particular consideration should be made to ensure that the approaches taken were inclusive and didn't exclude people due to age, learning or physical disabilities.

"Don't put it in adult language." Child or young person

6.7. Equality area: Older people (aged 70+)

Amongst those individuals we spoke with that were above the age of 70 years old, there was generally a high level of experience and willingness to provide their personal data to government and public services, and some individuals even participated in initiatives like the UK biobank. It was generally seen to be positive that public services linked relevant data (e.g. on disabilities and financial benefits) and some participants were aware that this is done using their NHS number as a linked identifier.

"The National Health, I'm quite comfortable with them having my personal information, I mean, any government department that are dealing with my needs or what I'm entitled to, DWP, HMRC. Maybe also whoever deals with public transport, I wouldn't mind them having details about how often I use public transport or whether I can use it, that sort of thing, whether I'm able to." Older person

The benefits of sharing data on personal circumstances were typically seen as relating to personal benefits in terms of personal health and welfare, and for improving the public services received (including to tackle issues like loneliness). There was a high degree of frustration and exhaustion expressed by participants at having to provide the same administrative data to multiple public services, and a desire for data to be linked or shared between public services to reduce fatigue.

Participants reported feeling that administrative data, and data used to represent older people at a national level, is not representative or inclusive in terms of conceptualising older people as a homogenous group. This is particularly the case for those who are only recently retired or just over the age of 70 years old, and who don't identify (or don't want to identify) with the label of 'older person'.

"*I'm fitter now than before I was retired, but they make assumptions because you are retired. It offends me being called 'age rich'."* Older person

Where older participants were least comfortable in completing forms, participating in research and sharing their data was where it was unclear what the information was needed for and how that data might be used. This was in part driven by experiences and/or awareness of fraud and older people being taken advantage of for financial gain. As a result, there is felt to be a relatively higher level of caution in sharing even basic administrative data, except in direct contact with public services.

"This year's census adverts are poor, it didn't feel relevant or specific, they [another older person] were worried and asked me." Older person

There was a high level of concern at sharing data relating to health and financial circumstances for fear that (i) disclosing changes to circumstances could be used as evidence to remove certain types of support that had previously been available, or (ii) that health data could be used against them (e.g. forcing them into care homes). In this respect there was a desire for control over what information is shared and with whom, particularly in respect of disabilities. Making reference to GDPR, data protection processes and gaining explicit consent for data sharing was seen to help build trust.

"[People are scared] that their information will get into the wrong hands, or get them in a muddle later down the line, like stopping income support etc." Older person

While individual participants had different preferences for how to provide their data, there was acknowledgement that older people in general were less likely to feel comfortable online, or in providing data face-to-face with a stranger. Key to ensuring inclusivity in collecting data from older people is developing their trust in the data collection process. The most effective routes through which to engage older people would be through public services (such as GP practices) and well-known charities such as Age UK or Help the Aged. One person also recommended including information on written or online questionnaires as to who people could speak to, to assess the validity of a given data collection exercise.

"If it's face-to-face they can see who they're talking to, you can judge a bit of body language, and people have a trusting face or don't have a trusting face. In the older generation we're really quite cautious about who we deal with, so I think face-to-face is the best way really." Older person

"In my age group and plus, our brains don't work so well, so being sensitive to the fact that we are slowing down. And some people, me included, it makes you feel a bit sad that you are in that category. So, I think it's good to handle it sensitively." Older person

6.8. Equality area: Sexual orientation

Amongst those individuals we spoke with that identified as lesbian, gay or bisexual, there was generally a high level of willingness to provide data relating to their sexual orientation with government, public services and, depending on the intended usage, with private companies. There was an expectation that organisations collecting these data would only have positive intentions behind its use. Participants did not see there as being any significant risk of providing data, though some believed that there is more stigma attached to bi-sexual, polysexual and transexual orientations, and that sexual orientation was not something that many people would openly identify in a more public forum such as social media.

"I think for a gay woman, I can't say for gay men, but I feel like we're more accepted now, and a lot of public figures out there are gay, whereas other groups may feel more vulnerable and not as widely accepted, and maybe open to problems by giving that information." LGB+ individual

Participants recognised there as being a range of personal benefits to sharing data, ranging from more appropriate and inclusive services through to affirmative action. Specific references were made by participants to healthcare and tourism as two particular areas where providing information around sexual orientation can be extremely helpful in minimising the potential for discrimination and ensuring that organisations have the necessary information to have informed conversations (e.g. around fertility options).

Collecting data on sexual orientation was seen as having the potential to highlight areas/sectors where there was under-representation of LGBT people as compared to the general population, helping to combat issues of bias and discrimination, and improving inclusivity and diversity. There was also acknowledgement that being able to see diversity of sexual orientation represented in statistics can help people to feel less alone, particularly for those who are still exploring their own sexual identity.

"There is a certain level of comfort there, knowing that you're more accepted, and there are more people out there that feel and are the way that you are." LGB+ individual

When asked to think about their sexual orientation, individuals felt that important information to collect, specific to people identifying as lesbian, gay, bisexual or transgender, to aid inclusivity included:

- Preferred pronoun
- Lifestyles, health and housing situations
- Relationship structures (including marriage)
- Family structures (including adoption and fostering of children)

Participants often highlighted that sexual orientation can be a fluid part of people's identities, and that the general categories often used in surveys and administrative forms don't always reflect how people would like to identify their sexual orientation.

"You have the option of putting your own categories, possibly within that, I think that would probably encourage more people to do it, where they didn't feel necessarily like they were being channelled down a specific path, but they were given a bit more voice." LGB+ individual

While personal preferences for modes of data collection varied across participants, there was a general view that data on sexual orientation is a private and sensitive subject, and therefore having the option to provide that data in confidence (i.e. not through direct questioning) was important. However, data collected at events such as Pride are likely to get high and very open response rates and lend themselves better to more face-to-face collection methods.

"*If it were on my doorstep, and somebody were filling out the census form in front of me, and they asked if I were gay, then I probably wouldn't answer that question, if I'm honest, and I'm really chilled out about it, because I would feel that was quite personal.*" LGB+ individual

"If you were to ask this information at a gay Pride for example, you'd get loads of people filling it in, because you'd get all like-minded people together, and they'd happily fill in and give opinions, because you're together as a group, and you'd feel comfortable." LGB+ individual

6.9. Equality area: Trans, non-binary and gender-diverse individuals

Trans, non-binary and gender-diverse individuals reported that their willingness to share personal information is dependent on the purpose of the data collection. Participants noted they feel more comfortable sharing information relating to their gender identity for health or education related purposes, as a lack of data sharing could result in significant harms for this community. It was additionally raised that medical professionals must be adequately educated in the specific issues pertaining to trans, non-binary and gender-diverse individuals to improve their experiences in accessing and using public services.

"There's plenty of instances where it would be harmful [to not share information]. For example, it depends on what level of information someone has got, knowing for example I was transgender and I was born as female at birth, somebody might assume then that once I finish my transition I never need to be invited for a smear test again. That's one instance where something has a health risk and people not having the right depth of information of you." Trans, non-binary and gender-diverse individual.

Several concerns around data sharing were raised amongst trans, non-binary and gender-diverse individuals, including:

- Unintentionally revealing an individual's gender identity before they feel ready to do so.
- Safety risks of harm or abuse to individuals if their data falls into the wrong hands, and specific harm to sex workers for whom data sharing could threaten their livelihood.

- Facing prejudice from employers or potential employers.
- Misgendering of individuals due to poorly considered data collection categories.
- Misuse of data and statistics in the media and by the public to promote an ‘anti-trans’ rhetoric.

“It’s all so often we are horrendously represented in the media.” Trans, non-binary or gender-diverse individual

Several key data challenges were identified specific to trans, non-binary and gender-diverse individuals, particularly around underrepresentation and inaccuracy of this group in data. It was highlighted that household data collection may disproportionately exclude people in these groups, as a higher proportion are thought to be homeless than in the general population. It was further noted that the classifications used in data collection pathways are often inaccurate or inadequate and may contribute to the exclusion and perceived invisibility of trans, non-binary and gender-diverse individuals in data. A lack of action or progress following data collection consultations with trans, non-binary or gender-diverse individuals was additionally highlighted as problematic, as not seeing progress discourages further participation in research and reduces trust in researchers and decision-makers.

“If I put non-binary in there, legally the government doesn’t recognise that non-binary people are a thing.” Trans, non-binary or gender-diverse individual

“...part of the problem is that on the one hand you want your government to know that there are more trans people out there and that there are more people who want to access services, but...if the government wants more information on trans people, our history says that’s just going into further oppressions, not to more help.” Trans, non-binary or gender-diverse individual

Trans, non-binary or gender-diverse individuals made multiple recommendations for how to improve inclusive data for this community moving forward. The use of longitudinal studies to track the experiences of these individuals was proposed, to produce a more insightful understanding of their experiences of treatment and potential discrimination within the health system, but also within education environments moving forward. Improving the communication of research outcomes was also regarded as being crucial in increasing trust and respect between researchers and the community. Trans, non-binary and gender-diverse individuals also recognised that while paper-based data collection methods will slow down data sharing practices and therefore service delivery, it would allow these individuals to express their views and expand their answers more easily and could be a safer option for data storage to avoid cyber-attacks.

Additional suggestions for improving data inclusivity included:

- Asking qualitative research questions to investigate views, attitudes and experiences of the trans, non-binary and gender-diverse community, and subsequently build better policy. This could include questions surrounding relationships, household income, or personal wellbeing.
- Filling basic demographic data gaps, for example numbers for the intersex and non-binary populations, and data on deaths and causes of death among trans, non-binary and gender-diverse individuals.
- Guaranteeing anonymity in research to represent a true reflection of the issues people face.
- Using media and social media to advertise research opportunities, especially paid opportunities, to increase uptake amongst this group.
- Undertaking outreach work to facilitate the participation of a more representative group of transgender, non-binary and intersex people.

“The best way to sort of reach trans people who are out there who will have these experiences, specifically of living their lives and having to deal with these things, is through service provides like education system, universities, colleges, schools, through GP offices, through local authority.” Trans, non-binary or gender-diverse individual

6.10. Equality area: Women’s equality

Amongst those women we spoke with that identified as having experienced some form of discrimination, there was a high level of willingness to provide data relating to their experiences, with an associated expectation that these data were used to address inequalities. More generally, participants were most comfortable sharing data with government and public services, where data was seen to be used for personal or public good.

The main issue raised by women that had experienced inequality was where data on gender – most commonly in relation to employment - was used as a data point against which to discriminate against women in respect of job offers and career progression. There was a view that stereotypes about women and traditional gender roles still influence expectations of what women’s role should be in society and in the home, with implications on opportunities within the workplace. Participants also highlighted examples of being made to feel uncomfortable explaining employment gaps on job applications or CVs which could relate to maternity leave and raising children, but equally may relate to health or other personal issues. Gender discrimination was often felt to go hand-in-hand with age discrimination, with older women seen to be doubly disadvantaged by supplying their age and gender data.

“When you think about internal promotions as well, and internal recruitment campaigns, where you're just asked the most inappropriate and bizarre questions. Things like 'You haven't got any kids have you?' Or 'You're not thinking of starting a family any time soon, are you?' And those aren't recorded, because obviously they would never dream of saying that to you in an interview…” Woman with experience of inequality

When asked to think about women’s equality specifically, individuals felt that important information to collect to aid inclusivity would be data from women on experiences of inequality and discrimination, ensuring the ability to look at this by age and ethnicity. The challenge here, that would require thought around the way in which questions were posed, is that people from diverse cultures have very different understanding and experiences of inequality dependent on social and cultural norms which may lead to under-reporting of inequalities.

Aside from data on individual experience of inequalities, participants saw there as being wider administrative data relating to employment patterns and outcomes, and the equality of provisions/facilities made by employers and educational establishments, that could be assessed to get an objective understanding of inequalities.

“You should collect some data about history of women up to that point in their careers as well. That's sometimes why women are disadvantaged, is because they haven't had the same introductions and training and networking opportunities.” Woman with experience of inequality

Participants felt that the best way of exploring views of people relating to their experience of inequalities, regardless of gender, were through more qualitative research approaches that (i) allowed for more open responses, led by the respondent, and (ii) could be undertaken ‘off-the-record’ where needed. These might include interviews and group discussions.

“It's quite difficult to understand the subjectivity of people's opinions. It is really difficult. I think focus groups are the way but maybe through community groups, so you're not actually having a focus group on gender inequality, you're just having a women's group, for example, or a men's group.” Woman with experience of inequality

7. Implications for improving inclusivity in data amongst individuals with protected characteristics

ONS commissioned Basis Social to undertake research with individuals and CSOs working with individuals with protected characteristics across a wide range of equality areas to understand how to improve inclusivity in the data and evidence which is then used to determine public policy in the UK. Our conversations with people have illustrated that data inclusivity can be viewed through multiple lenses:

- Inclusivity in the ways in which individuals, organisations and communities are engaged, and supported to engage, in providing and accessing data
- Inclusivity in respect of the subject matter which individuals are asked to provide their data and views on and around, and the language, concepts and categories that are provided for closed-response options
- Inclusivity in the methods used for collecting data from people, and the degree to which these meet individual needs and preferences
- Inclusivity in the types of data that are collected *and* publicised for public scrutiny, enabling individuals to ‘see’ themselves as represented in the data and enabling CSOs to advocate on behalf of specific communities
- Inclusivity in the extent to which people who provide their data have control over their data, understand how that data has been used and the impacts this has had or will have

Each of these lenses was represented in the way in which participants in this research discussed improving the inclusivity of data in the UK, regardless of whether they were participating as an individual or an organisation. This report contains a wealth of evidence for ways in which CSOs and individuals can better be supported to ensure that our data better reflects the rich tapestry of lived experiences across our multi-faceted society. Key opportunities that emerge from the conversations held with CSOs and individuals include:

- Harmonising the types of data collected on protected characteristics to mitigate against the danger that the needs and issues impacting vulnerable individuals and specific minority groups are not identifiable in administrative and national survey data. This could involve a relevant organisation, such as the ONS, consulting with bodies representing different equality areas to update the definitions, categories and forms of questions that are asked around protected characteristics to make these more inclusive, informative and harmonised. It would also require public services and agencies adopting standards in their (digitised) data collection nationwide, and making data publicly available in a consistent, interoperable format to support secondary analyses, and to improve subsequent decision-making
- Recognition, especially if there is a move toward using more administrative data, that many protected characteristics are not static and can change over time. For example, this could include an individual’s faith or sexual orientation, but it could also include other personal information which is not disclosed because of a lack of trusted relationships between the individual and the organisation or individual collecting the data. It will be important to work with relevant CSOs in determining the best ways of capturing data to ensure it minimises burden (and distress) but also maintains a high degree of accuracy

- Researchers providing clear communications around the purposes of data collection, with suitable reassurances around how data will be stored, managed and shared to improve public trust in data. Allied to this is ensuring researchers – and policy makers commissioning research - are more aware of the need to be explicit in how data is to be used, including highlighting the benefits of research/data to individuals during recruitment (with associated proof points)
- Researchers to better match the channel and form through which data is gathered to the audience (including face-to-face, telephone, online, peer-to-peer, qualitative and quantitative). It is our opinion, based on the implications of the evidence gathered, that this should be determined less by looking to standardise the mode of data collection and more on the purposes for which data is being gathered. This will involve greater collaboration with CSOs to understand the needs and preferences of different audiences, and how to tailor approaches to maximise participation. This would also involve greater awareness amongst both CSOs and policymakers as to the opportunities for qualitative research, and ways to increase the robustness and representation of qualitative research (e.g. through appropriate sampling, methodological, analysis and reporting considerations)
- Researchers and research commissioners publicising the way in which data is being used, the decisions taken in response to data gathered, and what this means for changes to policies and services. This is particularly important for those communities/individuals who feel ‘over-researched’ and/or unclear of the benefits of sharing their information
- Researchers and research commissioners providing administrative and population survey data in such a way as to enable (GDPR compliant) manipulation of disaggregated data by different protected characteristic to allow for more contextualised, intersectional and actionable understandings of groups, at a localised level
- There is a need to actively support greater inclusivity of data access and use by CSOs (and other organisations), accounting for their needs, skills and resources. Increasing the accessibility of data is important, however there needs to be recognition in doing this that CSOs are not operating on a level playing field, either within or across equality areas. CSOs (and other organisations, or groups of organisations) could be supported in leveraging data for the benefit of those with protected characteristics by making training, resources and funding available for CSOs to undertake their own collection (where there is a desire to do so). Similarly, data or analyses of data relating to specific equality areas and/or protected characteristics could be made more accessible (via centralising or signposting content)
- ONS (or another relevant organisation such as the Information Commissioners Office) should also provide information, advice and guidance for CSOs on GDPR and best practices in collecting and collating information on service users or members. This could include templates and tools
- Researchers and those using research minimising, wherever possible, the use of homogenous terminology for grouping individuals into meta-categories which may misrepresent or conceal the needs of specific groups within these categories (e.g. ‘BAME’, ‘Asian’ or ‘LGBTQ+’)
- ONS to help facilitate discussions for exploring the use of a unique identifier and data linkages in administrative data sets, both for the purposes of research but also more effective public services. Related to this would be putting in place legislation that strengthens ONS powers to request and receive equalities-related data from public authorities, for appropriate data linking, analysis and reporting, helping to mitigate concerns around data being withheld due to political sensitivities

- Allied to data linkage, ONS to help facilitate discussions with CSOs and members of communities where a balance needs to be struck between capturing data on protected characteristics and respecting the wishes of people not to be identifiable by these characteristics due to concerns around the potential for discrimination. Key examples here relate to faith (e.g. Jewish), ethnicity (e.g. Black and Gypsy, Roma and Traveller), and disability
- Engaging members of the public, CSOs, public sector authorities and private organisations in discussions around the principles of ethical data collection and collation. Greater clarity and adherence to simplified (and harmonised) forms of consent and data access would help improve people's understanding and comfort in how their personal data is used
- For ONS, and other public bodies to work toward the principles set out in the National Data Strategy, where data is 'findable, accessible, interoperable and reusable'

None of these opportunities or 'lenses' however sufficiently encapsulate the primary challenge to inclusivity, which is not just a barrier to members of vulnerable and marginalised communities providing their 'data', but to their feeling valued and heard. A consistent theme throughout this research with members of the general public, was a lack of understanding as to why certain pieces of personal data are collected, how these data is used or what benefit providing these data has either at an individual or community level. There was an assumption amongst most people that it was used for public good, though some individuals we spoke with saw less in the way of public or private benefit having come from sharing their data. For these individuals (often from ethnic minority groups), longstanding inequalities, discrimination and poor outcomes were evidence that their data, and their voices, held little weight with decision makers.

A related theme which emerged through our conversations with both CSOs and individuals from under-represented groups was around the importance of trust, and the trustworthiness of those organisations and individuals that are both collecting and using their data. People are largely willing to share their data, as long as they (i) have cause to trust in how it will be used, and (ii) believe it will benefit either themselves individually, or members of their community (however this is defined). Question marks around the authenticity and trustworthiness of organisations representing 'authority' – which often refers to central government – currently undermine the extent to which under-represented groups feel willing to share their data.

If there is a real desire to make a step-change in the inclusivity of UK data and evidence this needs to be taken in-step with members of those communities that are currently feeling under-represented and mis-represented, not just in UK data, but in UK society. To ensure everyone counts in UK statistics and evidence it will be important to increase people's ownership and control of their data, raise their participation in shaping the agenda for collecting data and in influencing strategic decisions made in response to the data being collected.

'Nothing about us without us' is not just about having a voice, but also about having an element of control over decisions taken for and about you.

List of Annexes

Annex A Master copy topic guide - CSO focus groups

Annex B Master copy topic guide - individual focus groups

Please note, both Annexes are the master copies of the topic guides used by Basis and ONS in Stage 1 (Annex A) and Stage 2 (Annex B) research. These master copies were adapted accordingly to fit an interview or focus group format and language was tailored for each equality area.

Annex A Master copy topic guide - CSO focus groups

ONS Inclusivity Research

Final Topic Guide (CSOs)

FEBRUARY 2021

Note to researcher

The Office for National Statistics (ONS), have commissioned Basis Social to undertake research to help explore how data and evidence can be more inclusive, and what the requirements are for making this possible. To ensure that statistics represent everyone in society, ONS established a new Inclusive Data Task Force. The taskforce would like to hear from people with different perspectives including Civil Society organisations (CSOs), local government, academics, think tanks and the public.

This will help understand and account for the experiences of more vulnerable and protected groups in policy and service development. This research will be undertaken in two stages. The first stage will involve engaging with CSOs across the UK and working across a variety of protected characteristic groups (i.e. those at greater risk of disadvantage where there are also particular data challenges), including:

- Physical disability or impairment
- Learning disability, neurodiversity or dementia
- Race or Ethnic equality advocacy and support groups
- Migrants, asylum-seeking and modern slavery
- Homelessness
- Faith and religious groups
- Children and young people
- Older people (aged 70+)
- Sexual orientation
- Trans, non-binary and gender diversity
- Gender equality

The second stage will involve speaking with members of the public across the UK from each of these groups.

This discussion guide is for use in discussions with CSOs. Discussions will last 120 minutes.

Basis final topic guide (CSO focus groups)

Timings	Content
0-5 minutes	Moderator to introduce the session. *Welcome everyone. Great to meet you all. My name is XXX. I work on behalf of Basis Social. We are an independent research agency who have been asked to carry out this research with you on behalf of the Office for National Statistics (or ONS), an independent government department and the UK's largest provider of official statistics.* *To complement a wider consultation on inclusive data the ONS have commissioned this research to specifically understand how facts and figures can be more inclusive, and what the requirements are for making this possible. This means how we could improve the way information is collected for certain groups in society, so it better represents their views and circumstances. This research is to involve individuals and organisations from across the UK and across a wide variety of equality areas. Those selected will represent - or advocate for - communities who are potentially less represented in current national statistics, who are more vulnerable or who have protected characteristics.* *The session will last 2 hours. We know this is quite a long time to spend online so there is a comfort break midway through. If you need to take calls, respond to emails or excuse yourself for whatever reason please do feel free. We want this to be as natural and comfortable a conversation as possible.* *You'll see that there are a number of people on this call and some of you may know each other already. You all work in similar areas, in this case in supporting [NAMED EQUALITY AREA]. We have purposefully brought people together to help develop as comprehensive a picture as possible of the needs and opportunities that you see existing for collecting data that can help you, and wider public bodies, to have a more granular understanding of the needs and lived experiences of [NAMED EQUALITY AREA].* *We do have a lot to get through so please excuse me if I move the conversation on at times. If there are things you want to say please do use the chat function to express these. This is one of the benefits of doing this online rather than face-to-face!* *As mentioned in our information sheet, this session is being video recorded and we would like to be able to reference your participation, including using attributable quotes in our reporting. If people have any points which they would like to be treated confidentially then please indicate this during the discussion. We will share with you any material where you have been quoted before any outputs are shared publicly.* *Basis are a company partner of the Market Research Society (MRS) and abide by their code of conduct. Participation in these discussions is completely voluntary and you are able to withdraw your consent to participate at any point in the process.* *Does anyone have any questions before we start?*
5-15 minutes	• Can you each briefly tell me about the remit of your respective organisations, including the groups that you work with or on behalf of, and your organizational objectives? Probe on:

	○ Organisational aims and objectives ○ Range of services/programmes provided ○ Issues and challenges faced by [NAMED GROUP]
15-55 minutes	• How is data and evidence currently used in your organization? Explore: ○ What types of data organisations use ○ Sources of data, how data is collected and managed (both that generated internally and externally) ○ Perceived accuracy, credibility and reliability of different data sources ○ What questions is it used to answer ○ Different uses of data (including for organizational strategy, service provision, advocacy and lobbying, partnerships, funding etc.) ○ Perceived representativeness of the ways in which data relating to [NAMED GROUP] is analysed and presented • What data do you currently believe provides the greatest value in understanding the circumstances and lived experiences of [NAMED GROUP]. Explore rationale for value judgement and how this data is used. • How aware are you of the wider uses of data which relate to [NAMED GROUP] – for example in wider policy, service development and in academic research? Probe on: ○ Examples of known data and evidence collection on [NAMED GROUP] ○ Examples of how data on [NAMED GROUP] is used outside of the CSOs participating here, including views toward this data use ○ Extent of data sharing between institutions, and what facilitates or hinders this • Thinking specifically about [NAMED GROUP] are there any examples you would highlight of data or evidence being used to good effect in shaping current public policy or services? Specifically, explore: ○ What underpins the positive effect of this data (e.g. whether it relates to quality of data, timeliness, granularity, openness, interoperability, linkages between datasets etc.) ○ How could it be improved further
55-60 minutes	BREAK
60-100 minutes	• Thinking about [NAMED GROUP] and your organizational remit, how easy is to: ○ find relevant data ○ collect relevant data ○ access relevant data (i.e. collected or managed by other organisations) ○ link relevant data (i.e. the extent to which different data sets can be linked given differences in definitions and questions) ○ compare data (e.g. over different time periods, or between different locations/regions) • What are the main challenges in collecting, analysing and presenting data on [NAMED GROUP]? ○ Does survey design create any barriers to collecting data from [NAMED GROUP]? ○ Are questions asked in a way that enables people's participation? ○ Are there examples of ways in which organisations have attempted to overcome data collection challenges?

	• Where are the main gaps or issues in the data and evidence on [NAMED GROUP]? Please think here about the type or quality of data that you think would help you, and wider organisations working with [NAMED GROUP] to gain a more accurate, inclusive and useful understanding of their circumstances and lived experiences. Moderator to prompt on: o types of data o measurement issues o intersectionality o level of detail o access o representation o presentation Specifically, prompt on the harmonization of measures and any issues this may cause in comparability of [NAMED GROUP] across time, geographies, data sources and with other population groups. Generate longlist of key issues and desired improvements, and then explore whether this list can be prioritized. • What value would addressing these gaps provide to: o the work of your organization (i.e. what utility might this provide)? o the planning and deliver of wider support/services for [NAMED GROUP]? • How (if at all) are these gaps currently managed/addressed in: o your own organization? o wider public policy (nationally or locally)? • What conditions do you think need to be met for the gaps in the data relating to [NAMED GROUP] to be addressed? For each of the key issues and improvements identified prompt what would be needed to enable data to be collected and used effectively (for example, engagement with gatekeepers, changes to data collection methodologies, improving trust in data etc.)
100-115 minutes	• What do you think the impact of these gaps in knowledge and understanding have on: o policy and services for [NAMED GROUP] o outcomes for [NAMED GROUP] • Are there any key actual or potential risks relating to more inclusive data collection on [NAMED GROUP]? Explore: o Risks or harm to individuals resulting directly from providing data or participation in research (for example this may include distress or persecution) o Risks or harm to [NAMED GROUP] resulting from inaccurate or unrepresentative data o Potential negative spillover effects or longer-term impacts on [NAMED GROUP] o How might these risks be mitigated against? • Thinking about the desired improvements to data on [NAMED GROUP], what role, if any, does legislation or regulation have to play in enabling these improvements while also protecting the rights of individuals? I.e. is there a need for changes to

	legislation or regulation to facilitate more inclusive data collection, management or sharing?
115-120 minutes	**Close** • Any final thoughts that anyone would like to share? • What is the top issue they would like to be considered by the ONS in respect to collecting, analysing and presenting data on [NAMED GROUP] • Explain that the next steps in the process will involve engaging with members of the public themselves to understand their views toward making data more inclusive. We will produce a report, and a standalone summary for CSOs in early April. • Thank and close

Annex B Master copy topic guide - individual focus groups

ONS Inclusivity Research

Topic Guide (individuals)

FEBRUARY 2021

Note to researcher

The Office for National Statistics (ONS) have commissioned Basis Social to undertake research to help explore how data and evidence can be more inclusive, and what the requirements are for making this possible. To ensure that statistics represent everyone in society, ONS established a new Inclusive Data Task Force. The taskforce would like to hear from people with different perspectives including Civil Society organisations (CSOs), local government, academics, think tanks and the public.

This will help understand and account for the experiences of more vulnerable and protected groups in policy and service development. This research will be undertaken in two stages. The first stage will involve engaging with CSOs across the UK and working across a variety of protected characteristic groups (i.e. those at greater risk of disadvantage where there are also particular data challenges), including:

- Physical disability or impairment
- Learning disability, neurodiversity or dementia
- Race or Ethnic equality advocacy and support groups
- Migrants, asylum-seeking and modern slavery
- Homelessness
- Faith and religious groups
- Children and young people
- Older people (aged 70+)
- Sexual orientation
- Trans, non-binary and gender diversity
- Gender equality

The second stage will involve speaking with members of the public across the UK from each of these equality areas.

This discussion guide is for use in discussions with participants in group discussions. Group discussions will last 90 minutes.

Basis topic guide – (individuals)

Timings	Content
0-5 minutes	Moderator to introduce the session. *Welcome everyone. Great to meet you all. My name is XXX. I work on behalf of Basis Social. We are an independent research agency who have been asked to carry out this research with you on behalf of the Office for National Statistics (or ONS). The ONS help the government find out what people think about important things like the environment and health. They help them to do this by collecting information on people around the country and asking questions to different groups of people and listening to their answers. Then they write a report, summarising what people say and publish the results, often as statistics – for example percentage of children who own a smartphone.* *ONS have commissioned this research to understand how the information that is collected about people can be more 'inclusive'. This means how we could improve the way information is collected for certain groups in society, so it better represents your views and circumstances. In this case we are interested in hearing from people from XXX.* *The session will last 1.5 hours. We know this is quite a long time to spend online so there is a comfort break midway through. If you need to take calls, respond to texts or excuse yourself for whatever reason please do feel free. We want this to be as natural and comfortable a conversation as possible.* *You'll see that there are a number of people on this call. You have all been asked to join so we can have a discussion that takes account of different views and experiences. You all have a shared XXX identity but many other aspects of your lives may be different.* *We are keen to hear from everyone so please do be respectful of the views of others. We also have a lot to get through so please excuse me if I move the conversation on at times. If there are things you wanted to say but haven't had the chance, please do use the chat function.* *There are no right or wrong answers to the questions we will be asking. If any questions seem obvious, it's because we don't want to assume what you are thinking, we want to hear it in your own words. You don't have to answer any question if you don't want to, you can just ask us to move on."* *As mentioned during recruitment, we would like to record this discussion but only if everyone is happy with this. This recording would only be viewed by me and other members of the Basis Social research team. Does anyone have any concerns or worries about this? If so, please don't hesitate to raise these now – it won't affect your participation in this research as I can take notes. If everyone is happy, I'll start recording the session now.* *Finally, Basis are a company partner of a body called the Market Research Society and abide by their code of conduct. Participation in these discussions is completely voluntary and you are able to withdraw your consent to participate at any point in the process. This includes during this discussion, or even afterwards.*

	Does anyone have any questions before we start?
5-20 minutes	• As a bit of a warm-up and to help us get to know one another, it would be great if we could introduce ourselves and each say what is your favourite time of the day, and why? I'll start! [Moderator to introduce themselves] Moderator to explain: Personal data is any data about an individual that can be linked directly back to them. You could also think of it as personal information or information about you. • If I were to ask you for some examples of what you would consider 'personal data', what kind of examples come to mind? Feel free to shout these out – there are no right or wrong answers here. [Moderator to encourage participants to spend a minute generating a longlist of different types of data] Moderator to explain: Special categories of personal data include information on a person's race or ethnic origin, sexual orientation, health, political opinions and religious beliefs among other things. This research is about your data, which could include any information to do with your life situation, attitudes and experiences. This could include information you provide which is not directly linked to you as an individual (for example through your name, date of birth, address, phone number etc.). • Can you think of times where your information has been collected by the government or local authority? Leave for spontaneous mentions, which may include surveys and Census, then prompt on 'administrative data', this is when an individual interacts with a public service, such as making an appointment with a GP, starting a new job or enrolling at a school – a record is made to enable service providers to carry out day-to-day activities. When these data are combined or linked, they can provide a deeper understanding of our society. For example, linking data on ethnicity and hospital admissions was important in identifying the higher risk of ethnic minorities to COVID-19. Or linking data on GCSE/A-level results with sex or ethnicity to understand if there are differences between groups. Explore: o What kinds of uses do you find more or less acceptable, and why? o Do you feel this type of administrative data presents a complete picture of you and your circumstances? Why / why not? What do you think is missing to present the whole picture of your situation? o How well, or badly, do you think administrative data is currently used? What makes you say that? Can you think of any examples? o Does the government collect the right type or level of data to ensure that everyone counts, and is counted, and no one is forgotten? What should it collect more or less of?
20-40 minutes	• What are some of the potential <u>benefits to you personally</u> of organisations having access to your personal information? o Have you experienced any of these benefits yourselves? Or seen them benefit others who you consider to be similar to you? o Thinking about your own personal circumstances and views, specifically in relation to XXX are there any types of personal information that you think are

	important for organisations planning public services and policies to be aware of? Probe: ▪ Are there areas where people from [NAMED EQUALITY AREA] may be disadvantaged compared to the general population (e.g. thinking about transport, housing, employment, education or health services). o Do you think people having a better understanding of these could lead to personal benefits for you? If so, how? • What about the potential <u>risks or harms to you</u> of organisations accessing your personal information? By harm, this means you may be put in a worse position than you were initially if your information had not been collected o Do you perceive there to be any risks or harms to organisations having access to your personal data? If so, what do you perceive these risks to be? Examples might for instance include risk of bias or discrimination, financial loss, damage to reputation or character. o Have you experienced any of these harms yourselves? Or seen them harm others who you consider to be similar to you? o Again, thinking about your own personal circumstances and views, specifically in relation to your [NAMED EQUALITY AREA], are there any types of personal information that you are less comfortable sharing with organisations planning public services and policies because of their potential risk to you personally? • Do you ever consider the personal or community benefits or risks of sharing information about yourself? By community-level, I mean other people from the same [NAMED EQUALITY AREA] as yourself. Explore: o Under what circumstances they weigh up benefits and risks o How do people judge when and with whom they are happy to share their personal information? • Are there some organisations you are more comfortable collecting or using your personal information than others? Probe on different use of data: o For use by businesses delivering profit-making services and products o For use by charities, national and local government for public good o For use by academics in universities for purposes of research
40-45 minutes	BREAK
45-70 minutes	Moderator to explain: there are a range of ways in which your information could be collected, from surveys/polls, interviews and focus groups through to application forms, monitoring of services and products, and administrative data (examples of which we mentioned earlier). It could be information you provide, information others provide about you (for example credit checks), or information that is collected through your activities (e.g. shopping, mobile phone use or internet). • Thinking about some of the examples of personal information you gave earlier. How would you prefer this personal information to be collected? Gage here whether prefer for it to be more ‘manual’ (i.e. they are asked for it) vs more automatic (i.e. organisations collect and share data more automatically). Explore preferences and reasoning.

	• I'd like to explore some specific examples of data that is currently collected by the Office for National Statistics (ONS) and other public bodies to understand your views toward these. For each of the following examples, the moderator should explore: o Awareness of the data sources and any experiences of participating o Perceived value/benefits and risks o Barriers to participation for people [FROM NAMED EQUALITY AREA] (and themselves personally), probe on modes of collection o How might these barriers be overcome? o Facilitators to participation, probe as to whether who collects information influences participation o Whether they themselves would participate 1. The Census. The Census is a count of all people and households in the UK, undertaken every 10 years by the ONS. It is the only data source which provides a detailed picture of the entire population, and is unique because it covers everyone at the same time and asks the same core questions everywhere. This makes it easy to compare different parts of the country. The information the Census provides allows central and local government, health authorities and many other organisations to target their resources more effectively and to plan future housing, education, health and transport services. It can be conducted online or on paper, and contains questions relating your demographics (age, sex, gender, ethnicity etc.) and personal circumstances (including work and education). It takes 15-20 minutes per household. 2. Understanding Society. This is an annual survey, led by a team at the Institute for Social and Economic Research at the University of Essex, involving 40,000 households. An interviewer visits the same house annually to complete a 40-60-minute questionnaire on subjects such as health, work, education, income, family and relationships, social life, and attitudes and behaviours (e.g. political views, attitudes to gender roles and the environment). The content is much more detailed than the census and people are paid £10 as a thank you for participating. The data is used by government, charities and academics. 3. Health Survey for England. The Health Survey for England, collected by the NHS, monitors trends in the nation's health and care. It provides information about adults and children living in private households in England. The survey consists of an interview, followed by a visit from a nurse who takes some measurements and blood and saliva samples. This is undertaken annually with around 9,000 households.
70-90 minutes	• The ONS is interested in how we can promote more inclusive data; where everyone feels represented in the statistics, policies and services in this country. If you were tasked to build a more inclusive picture of people from [NAMED EQUALITY AREA], with a richer and more detailed understanding of their lives and views, how would you go about doing this? Probe: o How would you build their trust? Or how could organisations be more trustworthy? o What approaches would work best for collecting information? o What types of questions would you ask of people from [NAMED EQUALITY AREA] that you believe are important for informing services and policies? o Are there any groups that would still be excluded through this process? How

	might we reach them? • Thinking about your personal data, what rules do you think there should be for organisations who have access to your data? Working as a group I'd like you to think about: o Who should collect data that helps promote the participation and interests of people from [NAMED EQUALITY AREA]? o Who should be allowed to access to this data? o The different reasons why they should be allowed to collect and use this information? o What kind of data should and should not be collected? • Are there any other principles or criteria against which you think the collection and use of data which relates to ensuring better representation of [NAMED EQUALITY AREA] should be assessed? Probe for example on transparency of uses, control over data/consent, anonymity (being identifiable in data), relevance of data being collected for specific purposes and potential impacts (positive and negative)
90 minutes	**Close** • Any final thoughts that anyone would like to share? • What is the top issue you would like to be considered by the ONS in respect to collecting, analysing and presenting data on people from [NAMED EQUALITY AREA]? • Thank and explain that everyone will receive their incentive shortly (incentives are issued on the Monday following a discussion and could take up to three days to be processed). We will be producing a report for ONS in the next few weeks and will be happy to send them a one-page summary if they have consented to us sending them an email with this.