This paper has been prepared by staff in the Central Statistics Office and IGEES staff in the Department of Public Expenditure and Reform. The views presented in this paper do not represent the official views of the Department or Minister for Public Expenditure and Reform.
Executive Summary

- Equality budgeting is a way of approaching and understanding the budget as a process that embodies long-standing societal choices about how resources are used, not simply as a neutral process of resource allocation.
- The OECD conducted a Scan of Equality Budgeting in Ireland, publishing their report in 2019 with twelve recommendations. One key recommendation was the development of an equalities data strategy focused on improving the availability of data disaggregated by equality dimension.
- The CSO, as a member of the Equality Budgeting Expert Advisory Group, made a statistician available to conduct an audit of existing data from an equality perspective, an exercise that complemented the CSO's plans for a “SPAR 2” initiative. The audit was also motivated by the data collection guidelines published by The European Commission’s Subgroup on Equality Data in 2018.
- The aim of the project was to examine all data sets in public bodies and determine the extent of data available on the one or more of nine dimensions of equality: gender, marital status, family status, age, disability, sexual orientation, race, religion and membership of the Traveller community.
- Data providers were asked for a range of information for each dataset, such as the equality dimensions covered, the organisation responsible for that data set, timeliness, theme (aligned to themes of well-being), and an indication of whether a unique identifier (such as PPSN or Eircode) was captured. Further information was requested relating to classifications and harmonisation.
- The audit itself can be viewed and downloaded at this address: https://www.cso.ie/en/methods/methodologicalresearch/rp-eda/equalitydataaudit2020

Key Findings

- Of 41 public sector organisations contacted, there was an estimated response rate of 85% in total (75% responding with relevant data; 10% responding with no relevant data available). This resulted in 107 data sets for audit.
- Roughly 55% of the audited data sets report regular updates, at minimum yearly, with many having daily or weekly updates. Around 15% of data sets are more than yearly or a one-off survey.
Key Findings

- Information was sought on variable classifications used; this was reported for less than 30% of data sets. However, even from this small sample, there was significant variability in classifications. Standardised classifications make comparing and matching different data sources much easier, allowing for more effective analysis.

- It is important to note that while many data sets include nationality as a marker, only the Census captures race. Similarly, sexual orientation is not an actual marker included in the audited data sets – instead this finding indicates a question on marriage or civil partnership that has a same sex option.

- In some cases, public bodies have removed variables from collected data, due to concerns about compliance with GDPR. While acknowledging responsibility to protect personal data and comply with regulations, it is important to ensure that data sets capture information on the dimensions of equality, ensuring that statistical analysis can fully inform policy-making, legislation, and public service provision.

- This conflict between data protection and availability of disaggregated data could potentially be resolved with more widespread use of a unique identifier such as PPSN, allowing data sources to be matched.

- The audit uncovered data gaps for many dimensions of equality, with varying levels of disaggregation across the 107 data sets. Table 1 outlines the coverage of 9 dimensions of equality across the audited data sets.

<table>
<thead>
<tr>
<th>Dimension of Equality</th>
<th>Number of Data Sets</th>
<th>% Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>68</td>
<td>60.7</td>
</tr>
<tr>
<td>Age</td>
<td>65</td>
<td>63.6</td>
</tr>
<tr>
<td>Race¹</td>
<td>26</td>
<td>22.4</td>
</tr>
<tr>
<td>Marital Status</td>
<td>25</td>
<td>23.4</td>
</tr>
<tr>
<td>Disability</td>
<td>24</td>
<td>17.8</td>
</tr>
<tr>
<td>Family Status</td>
<td>19</td>
<td>24.3</td>
</tr>
<tr>
<td>Member of the Traveller Community</td>
<td>12</td>
<td>3.7</td>
</tr>
<tr>
<td>Sexual Orientation²</td>
<td>6</td>
<td>5.6</td>
</tr>
<tr>
<td>Religion</td>
<td>4</td>
<td>11.2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>107</strong></td>
<td><strong>100.0</strong></td>
</tr>
</tbody>
</table>

¹Includes Nationality as well as Race
²Relates to questions on same sex marriage/civil partnerships
1. Introduction

Equality budgeting is a way of approaching and understanding the budget as a process that embodies long-standing societal choices about how resources are used, rather than a neutral process of resource allocation. In 2019, the OECD conducted a Scan of Equality Budgeting in Ireland; the final report contained twelve recommendations to drive this work forward. One key recommendation was the development of an equalities data strategy focused on improving the availability of data disaggregated by equality dimension. The first step in implementing that recommendation was to carry out a comprehensive survey of the existing public service data landscape, from the equality perspective. To achieve this objective, the Equality Data Audit was proposed; it is a new initiative to highlight the availability of equality-relevant data, and expose data gaps where they may occur.

The CSO, as a member of the Equality Budgeting Expert Advisory Group, made a statistician available to conduct the audit, an exercise that complemented the CSO’s plans for a “SPAR 2” initiative. The European Commission’s Subgroup on Equality Data published their guidelines concerning the collection of equality data in 2018. The guidelines provide a series of steps and actions to improve the collection and use of Equality data. One of these steps includes a data audit, along with a range of institutional, structural and operational activities. These guidelines were a valuable resource and used to inform the progress of this audit.

Put simply, the primary aim of the project was to examine all data sets in public bodies and determine the extent of data available on the one or more of nine dimensions of equality: gender, marital status, family status, age, disability, sexual orientation, race, religion and membership of the Traveller community. The process commenced by reaching out to public bodies and requesting them to fill out an audit template. Once the returned audits were collated and reviewed, analysis was carried out to expose those dimensions that suffer from low data availability (e.g., relating to Race or Sexual Orientation) and those captured by a wide variety of data sets, such as Age and Gender. Respondents were also asked to provide other characteristics of each data source, such as the theme and whether a unique identifier (PPSN or Eircode) was captured or not.

The audit itself is available online, and allows the user to access more detailed information about each of these data sets than is presented in this report. The Equality Data Audit is planned to be a living document that will be periodically updated. To view and download the audit, and for information on how to submit updates please visit [https://www.cso.ie/en/methods/methodologicalresearch/rp-eda/equalitydataaudit2020](https://www.cso.ie/en/methods/methodologicalresearch/rp-eda/equalitydataaudit2020).

More generally, queries, feedback and updates can be emailed to: [sscu@cso.ie](mailto:sscu@cso.ie)

---

1. [Statistical Potential of Administrative Records (CSO, 2003)]
2. [Guidelines on improving the collection and use of equality data (EC, 2018)]
This paper is set out in sections as follows:

Section 2 is a summary of results;
Section 3 reports findings for each dimension of equality;
Section 4 outlines data issues and some recommendations;
Section 5 discusses the implications of the audit for Equality Budgeting.

2. Summary of Results

2.1 Responses and Availability

This audit commenced by contacting all members of the National Data Infrastructure (NDI) champions group, covering 41 public sector groups. CSO attendance at events relating to equality data resulted in additional contacts being made; in total 48 groups were contacted during this audit. In total, 85% of organisations responded to this audit; 75% responding with relevant data and 10% responding that they held no relevant data. The results of the audit outlined in this report cover 107 data sets in total.

It is important to note that the number of groups contacted, and therefore the number of data sets audited, was limited by the size of the NDI champions group. While the audit aimed to be as comprehensive as possible, it is likely that not all relevant data sets have been covered – as the audit is designed to be a living document, continued updates will allow additional sources to be captured over time. It must also be pointed out that not all of these data sets are accessible by the public, or may only be available for use in an aggregate form.

2.2 Coverage of the Dimensions of Equality

Once data providers had returned the requested information, these returns were analysed to ascertain whether variables relating to each dimension of equality were being captured. This section of the report will present the results of this analysis; a more in-depth discussion of the implications of these results for Equality Budgeting is included in Section 5.

Table 1 provides a breakdown of the number and percentage of audited data sets that capture information on each of the nine grounds of discrimination covered by the Equal Status Acts 2000-2018. The number of sources listing age and gender as a data point was higher than for those identifying any of the other dimensions of equality. This is likely to reflect the fact that age and gender are routinely captured in data collection and are used as standard breakdowns in most statistical releases. In contrast, some

---

3 See Appendix for more information on the NDI
of the other characteristics, most notably sexual orientation and religion, are not routinely included in data collection and so are listed in fewer audited data sources.

There are two other particular points to note, concerning the reporting of the dimensions of race and sexual orientation. Race was captured in only two data sets: the Census of Population and the Higher Education Authority (HEA) Equal Access Survey. However, race is not available for analysis from the HEA Equal Access Survey due to the low response rate for this variable. As the audit progressed, it became evident that many data sets include nationality or ethnicity but not race; nationality and ethnicity are both covered by “the ground of race” according to the Equal Status Acts 2000-2018, and so these characteristics are grouped together under the race (see sections 3.7 and 4.1 for more). Data on sexual orientation is rarely collected; all references to sexual orientation in the audit indicate that a question on marriage or civil partnership had a same sex option, and does not reflect an actual marker on sexual orientation in the data set.

Out of responses for a total of 107 data sets, only 34 included a further breakdown of each dimension of equality, with details on how the data is collected. For example, in the case of age, the further breakdown reports whether age is collected as date of birth, as age in years, or as an age grouping/range. Furthermore, the data collected in the audit only shows high-level information on the data sets. It does not show the overall coverage of data related to a particular cohort. In some data sets there may not be a large enough number of data points to allow for meaningful analysis of a particular minority group.

Table 1: Breakdown of Coverage by Dimension of Equality

<table>
<thead>
<tr>
<th>Dimension of Equality</th>
<th>Number of Data Sets</th>
<th>% Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>68</td>
<td>60.7</td>
</tr>
<tr>
<td>Age</td>
<td>65</td>
<td>63.6</td>
</tr>
<tr>
<td>Race¹</td>
<td>26</td>
<td>22.4</td>
</tr>
<tr>
<td>Marital Status</td>
<td>25</td>
<td>23.4</td>
</tr>
<tr>
<td>Disability</td>
<td>24</td>
<td>17.8</td>
</tr>
<tr>
<td>Family Status</td>
<td>19</td>
<td>24.3</td>
</tr>
<tr>
<td>Member of the Traveller Community</td>
<td>12</td>
<td>3.7</td>
</tr>
<tr>
<td>Sexual Orientation²</td>
<td>6</td>
<td>5.6</td>
</tr>
<tr>
<td>Religion</td>
<td>4</td>
<td>11.2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>107</strong></td>
<td><strong>100.0</strong></td>
</tr>
</tbody>
</table>

¹Includes Nationality as well as Race

²Relates to questions on same sex marriage/civil partnerships
2.3 Coverage of Well-being Themes

In addition to analysing the coverage of equality dimensions, the audit also assigned an overall theme to each data set, where possible, covering 12 domains of well-being. All 12 themes are included in Table 2 below, including those themes that were not covered by any reported data set. It is important to note that some data sets were allocated multiple themes, so that the total of this table will be greater than the number of data sets – for this reason, percentage coverage is not reported in Table 2.

The most frequently occurring theme was Health, covered in nearly half of all the data sets. Education & Skills was covered in 26 data sets while Jobs & Earnings and Living Standards were each covered in 21. The remaining themes were covered by very few data sets in the audit.

Table 2: Breakdown of Coverage by Well-being Theme

<table>
<thead>
<tr>
<th>Theme</th>
<th>Number of Data Sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health</td>
<td>51</td>
</tr>
<tr>
<td>Education &amp; Skills</td>
<td>26</td>
</tr>
<tr>
<td>Jobs &amp; Earnings</td>
<td>21</td>
</tr>
<tr>
<td>Living Standards</td>
<td>21</td>
</tr>
<tr>
<td>Culture &amp; Identity</td>
<td>5</td>
</tr>
<tr>
<td>Housing</td>
<td>5</td>
</tr>
<tr>
<td>Personal Security</td>
<td>2</td>
</tr>
<tr>
<td>Income &amp; Wealth</td>
<td>1</td>
</tr>
<tr>
<td>Work-Life Balance</td>
<td>0</td>
</tr>
<tr>
<td>Civic Engagement &amp; Governance</td>
<td>0</td>
</tr>
<tr>
<td>Ireland’s Economy</td>
<td>0</td>
</tr>
<tr>
<td>Environment &amp; Sustainability</td>
<td>0</td>
</tr>
</tbody>
</table>

2.4 Timeliness

Just over half (55%) of the data sets have regular updates, at least once a year, with many having daily or weekly updates. About one in six (15%) data sets had an update of less often than once a year, or were based on a one-off survey.
3. Findings for each Dimension of Equality

3.1 Gender

The Equal Status Acts 2000-2018 defines the ground of gender as follows:

(2) As between any two persons, the discriminatory grounds (and the descriptions of those grounds for the purposes of this Act) are:

(a) that one is male and the other is female (the “gender ground”)

Data on gender was contained in 68 of the 107 data sets, making it the most commonly collected dimension of equality. A detailed breakdown of the gender variable was provided for 25 of the 68 data sets with 23 of these providing a breakdown of Male/Female/Prefer not to say. Just two data sets had other options – one contained ‘Transgender and Non-conforming’ as an option while the other contained an ‘Other gender’ option.

The ground of gender in Irish equality legislation primarily reflects EU sex discrimination law. EU case law, and its reflection in the recitals of the Recast Gender Directive, recognises that transgender people may experience discrimination on the ground of sex – specifically with reference to ‘gender reassignment’ or transition. However, EU equality law does not define sex as referring to male/female. Relevant Irish case law is based on EU law, as domestic legislation does not treat these issues. In practice, therefore, the ground of gender is interpreted by the courts and the Workplace Relations Commission to include transgender persons, as required under the case law of the Court of Justice of the EU (CJEU).

The National LGBTI+ Inclusion Strategy includes an action to review this legislation to ensure that transgender, non-conforming and intersex people have explicit protection within the equality grounds. The new Programme for Government has a similar commitment. Such changes will not necessarily change the definition of sex in law, or for the purposes of data collection. In the UK for example, an additional ground (called ‘gender reassignment’), has been added to equality legislation alongside sex and the other grounds.

Gender was captured across all eight of the themes reported as being covered in Table 2, with 20 instances each for Education and for Health, and more than 15 each for Jobs & Earnings and for Living Standards.

3.2 Age

The Equal Status Acts 2000-2018 defines the ground of age as follows:

(2) As between any two persons, the discriminatory grounds (and the descriptions of those grounds for the purposes of this Act) are:

(f) subject to subsection (3), that they are of different ages (the “age ground”)
(3) Treating a person who has not attained the age of 18 years less favourably or more favourably than another, whatever that other person’s age, shall not be regarded as discrimination on the age ground.

Age is one of the most frequently collected dimensions and was present in 65 data sets, albeit with some variability in the way it is collected and reported. A detailed breakdown of the age variable was provided in 26 of these data sets. Age was collected in a continuous way in 21 data sets (either as date of birth or years of age), which is most useful for analytical purposes. The remaining five data sets gathered data in age groupings, negatively impacting capacity for meaningful analysis on people of different ages. In most cases, these approaches constructed age groups of about ten years (e.g. 10-19, 20-29), with poor granularity for age particularly evident amongst older and younger age groups. Of the five data sets collecting age groupings, one data set has its eldest grouping as “55+”, three have “60+” and two have “65+”, while three have young people grouped as “Under 18”, “Under 25” and “15-24”.

Age was captured across all eight of the themes reported as being covered in Table 2, with 20 or more data sets covering Education and Health, followed by Jobs & Earnings and Living Standards covered by more than 15 data sets each.

3.3 Civil Status

The Equal Status Acts 2000-2018 defines the ground of civil status as follows:

(2) As between any two persons, the discriminatory grounds (and the descriptions of those grounds for the purposes of this Act) are:

(b) that they are of different civil status (the “civil status ground”)

(‘civil status’ means being single, married, separated, divorced, widowed, in a civil partnership within the meaning of the Civil Partnership and Certain Rights and Obligations of Cohabitants Act 2010 or being a former civil partner in a civil partnership that has ended by death or been dissolved)

Roughly one in four (23.4%) of the 107 audited data sets contain civil status and, of those, 15 have also included a breakdown of the variable. The level of detail collected for civil status is very good. Where the breakdown was provided, all the data sets with this variable have at least five levels of civil status (single, married, widowed, divorced, civil partnership), with some containing up to 15 options. Same sex marriage was included as an option in some cases. Civil status was captured across all eight of the themes reported as being covered in Table 2, with 11 instances covering Health 8 instances each for Education, Jobs & Earnings, and Living Standards.
3.4 Family Status

The Equal Status Acts 2000-2018 defines the ground of family status as follows:

(2) As between any two persons, the discriminatory grounds (and the descriptions of those grounds for the purposes of this Act) are:

(c) that one has family status and the other does not or that one has a different family status from the other (the “family status ground”)

Fifteen datasets, or 17.8% of all those audited, had an indicator for family status, with 9 providing a further breakdown of the family status variable. Most of the data sets have three main breakdowns of family status – lone parent, married parent or no children. Family status was captured across all eight of the themes reported as being covered in Table 2. Eight instances covered Living Standards, while there were seven each covering Education and Health.

3.5 Disability

The Equal Status Acts 2000-2018 defines the ground of disability as follows:

(2) As between any two persons, the discriminatory grounds (and the descriptions of those grounds for the purposes of this Act) are:

(g) that one is a person with a disability and the other either is not or is a person with a different disability (the “disability ground”)

Just under a quarter (22.4%) of the audited data sets included a marker on disability. Eight of these provided a breakdown of the variable but it varied widely. In some cases, the severity of a disability is collected, in others the category/type of disability is collected and in yet other cases, a yes/no question is used. It is important to note that while a variable for disability may not be included in a given health-related data set, patient histories, or other notes on the same system, may record a disability variable for a given individual. Disability was captured across all eight of the themes reported as being covered in Table 2. For each of Education and Jobs & Earnings, there were 8 data sets capturing disability; there were 7 data sets for each of Health and Living Standards that captured disability.

3.6 Sexual orientation

The Equal Status Acts 2000-2018 defines the ground of sexual orientation as follows:

(2) As between any two persons, the discriminatory grounds (and the descriptions of those grounds for the purposes of this Act) are:

(d) that they are of different sexual orientation (the “sexual orientation ground”)

Sexual orientation is not directly captured in any of the audited data. In some cases, it can be deduced where same sex marriage is included as an option in questions relating to civil status.

3.7 Race

The Equal Status Acts 2000-2018 defines the ground of race as follows:

(2) As between any two persons, the discriminatory grounds (and the descriptions of those grounds for the purposes of this Act) are:

(h) that they are of different race, colour, nationality or ethnic or national origins (the “ground of race”)

Race was included in only two data sets – the Census of Population from the CSO and the Social Inclusion and Community Activation Programme (SICAP) data set. Another 24 cases in the audit are marked as covering race, but these actually capture nationality or ethnicity; while this is not the same as race, it is covered by the Equal Status Acts definition of the “ground of race”.

In the case of nationality, all data sets collected this variable as free text. For race, there are at least seven options given in the breakdown, in every case where it is collected. In other instances, ethnicity or country of birth are collected. These usually have options for the most frequent countries/ethnicities, with a free text option also included. Race was captured across all eight of the themes reported as being covered in Table 2, with more than ten instances for each of the themes of Education, Jobs & Earnings and Living Standards.

3.8 Religion

The Equal Status Acts 2000-2018 defines religion as:

(2) As between any two persons, the discriminatory grounds (and the descriptions of those grounds for the purposes of this Act) are:

(e) that one has a different religious belief from the other, or that one has a religious belief and the other has not (the “religion ground”)

Religion appears in only three data sets, covering the themes of Education & Skills, Health, Jobs & Earnings, Living Standards and Personal Security.
3.9 Membership of the Traveller Community

The Equal Status Acts 2000-2018 defines the ground of membership of the Traveller community as:

(2) As between any two persons, the discriminatory grounds (and the descriptions of those grounds for the purposes of this Act) are:

(i) that one is a member of the Traveller community and the other is not (the “Traveller community ground”)

Membership of the Traveller Community was captured by 12 of the audited data sets, four of which covered the theme of Education and Skills; overall these data sets reflected seven of the eight themes reported as covered in Table 2. Where Membership of the Traveller Community is included in a data set, it is captured either as a yes/no question, or included in a question about race or ethnicity.

4. Data Issues and Recommendations

This section highlights four key issues that have emerged as a result of the audit, and provides some recommendations.

4.1 Data collection

Section 3 has outlined this audit’s findings regarding the availability of data disaggregated by equality dimension. While some dimensions are reasonably well covered, there are clearly gaps in the data when considered from an equality perspective. In order to produce robust and representative analysis on any minority group, or to study equality in any sector, high quality data must be collected, with all the necessary breakdowns to provide visibility on all cohorts. What follows is an outline of some specific data collection issues that emerged.

4.1.1 Sex and Gender Identity

As noted earlier, while the term gender is used in the equality legislation - and appears in many of the data sets - it is principally used as a synonym for sex and is generally collected in male/female (binary) format in the data sets covered in the audit.

The topic of gender identity has gained prominence in recent years, with a growing recognition that not all people identify with their current sex or indeed with any sex, and that these identifications may change over time. This means that public bodies must consider whether they are accurately capturing this gender identity, allowing for more robust analysis. The United Nations Economic Commission for Europe (UNECE) discussed the potential impact of a gender identity question⁴, advocating a rigorous testing programme before such a question is included in a census.

⁴ Recommendations for the 2020 Censuses of Population and Housing (UNECE, 2016)
The CSO introduced new questions on both gender identity and sexual orientation in its General Household Survey in the first quarter of 2019. The first release from this iteration of the survey was published in July 2019. The CSO will monitor the collection and production of data on both gender identity and sexual orientation. This will allow the CSO to develop questions which are well understood and acceptable to users, stakeholders and respondents, producing confidential, robust and consistent data.

From a legal perspective, the EU requires that all National Statistical Institutes collect census data on sex with a binary classification scheme, but there is no requirement as yet to collect data on gender identity.

4.1.2 Sexual Orientation
There are no data sets in this audit with a specific question on Sexual Orientation. It can be sometimes deduced from a marker for same-sex marriage, however this also creates issues as it does not account for the broad spectrum of sexual orientations. A program to test the impact of any new questions on this topic should be completed before including it in any survey or data collection.

4.1.3 Nationality vs Ethnicity
Under the ground of race, the Equal Status Acts 2000-2018 includes race, colour, nationality and ethnicity. Only two data sets collected race, (see section 2.7), with nationality, and sometimes ethnicity, collected in another 24 data sets. Race, ethnicity and nationality need to be clearly distinguished. For example, an immigrant with Irish citizenship may have a different ethnic background and could be more vulnerable to inequality or discrimination but this may not be visible if only nationality is captured. In the data sets audited, nationality was collected more frequently than ethnicity or race.

4.1.4 Membership of the Traveller Community
Irish Travellers were recognised as a distinct ethnic group in March 2017. Membership of the traveller community was ascertained with a question on ethnicity in some audited cases, but was a standalone question in other data sources. This can make it difficult to know if a data set contains data relating to Irish Travellers. As discussed above, nationality is collected more frequently than ethnicity, therefore data on the Irish Traveller community may be dependent on the availability of ethnicity data. It is recommended that ‘Irish Traveller’ be included in ethnicity questions, but it is essential to collect this data separately if there is no ethnicity question.

4.1.5 Disability
Data on disability is often dependent on self-disclosure. It can be defined using either the very broad definition in equality legislation, the more detailed definitions used by the CSO, (in the Census of Population and the Labour Force Survey (LFS)), or inferred from self-reported health surveys, which offer a more medical definition. It can also be reliant on the degree and nature of self-disclosure, (e.g. it is often optional and based on a very basic tick-box on a form, which provides little information). Given the

---

5 General Household Survey Module on Equality (CSO, 2019)
particular complexity of disability data considerations, this presents unique challenges that will need to be examined further to generate improved levels and quality of disability-related data.

4.1.6 Statistical Disclosure
Questions on all of the above topics may also give rise to statistical disclosure control issues, due to small population cohorts. These often arise for National Statistical Institutes when considering publishing data on relatively small population cohorts at small area levels, which is a defining geography for census data. Similar confidentiality issues may also arise when preparing cross tabular outputs. Therefore, care should be taken when collecting and disseminating any data on minority groups. The EU Equality Guidelines have some advice on confidentiality, building trust, and the use of booster samples when publishing data on small population cohorts.

4.1.7 Well-being Themes
This audit has also shown that some themes of well-being are well covered in data, while others are covered poorly or not at all. Following the publication of this report, the owners of the various data sets can contact the CSO if details need to be amended. It is possible that many themes appear less well-represented because the CSO did not have all of the necessary contacts in each area. If the next iteration of this data audit continues to find that some themes are lacking in data, then this topic will be revisited by the CSO.

4.2 Harmonisation and Classifications
This audit contains a breakdown of the classifications used in about 30% of the audited data sets. It is clear from this small sample that there are many different classifications for each variable. Standardised classifications make comparisons and matching of different surveys and data sources much easier and allows a broader analysis to be made.

The use of standard classifications at both the collection and dissemination phase of statistical production should be promoted and supported across the Irish national statistical system. By improving the use of common standards, this will lead to an improvement in the consistency and coherence of outputs. Standard classifications will also ensure that data is comparable over time and will provide a common link between different data collections.

4.3 Intersectionality
Intersectional data needs to be available so that the impact of policy interventions for various sub-groups in larger cohorts can be assessed, and ensuring that information can be cross-tabulated across different cohorts. This allows for measurement of any disproportionate impacts based on the multiple identities of any individual whose data
is captured, e.g.: women with disabilities as a sub-set of data on gender, Travellers with disabilities as a sub-set of data on Travellers or ethnicity, etc.

4.4 Data Protection and GDPR

The legal basis now exists to allow public bodies to process equality data. The identification of an appropriate legal basis for processing equality data under Article 6, and a permissible condition under Article 9, of the GDPR is a matter for each public body as a data controller. However, the collection and processing of equality data using section 51 of the Data Protection Act, 2018 is also legally permitted for public bodies.

It is clear to the CSO, from discussions related to the EU Equality Data Guidelines and this audit exercise, that some public bodies are unsure about what was changed by GDPR and whether they can collect certain types of equality data. In some cases, this uncertainty has led to variables being removed from the data collected.

While there is a responsibility to protect personal data and comply with regulations, it is very important to ensure that data is collected on all populations, including minority cohorts, to ensure that statistical analysis can be used to inform future legislation, policies and services.

5. Discussion: Implications for Equality Budgeting

The OECD Scan of Equality Budgeting in Ireland™ listed 12 key recommendations for the path forward; among these was a recommendation to develop an equalities data strategy. Improved availability of disaggregated data is essential to allow government to assess where the equality gaps lie, and to assess the effectiveness of government activity across all dimensions of equality. This type of data is a prerequisite if high-quality outcome and impact indicators are to be constructed; the reporting of these indicators allows operational tools to function, be it in the form of ex post or ex ante analysis. The Irish Government Economic and Evaluation Service (IGEES) regularly produces ex post analysis in the form of Spending Review and Social Impact Assessment papers. In 2019, DPER published a review of these papers from the perspective of Equality Budgeting™. That review found that while equality perspectives were not consistently included in ex post analysis, this was very often due to the poor availability of data disaggregated by equality dimension.

Before the availability of disaggregated data can be improved, it is first necessary to ascertain the level of disaggregation that exists at present in data sets across the

---

6 OECD Scan: Equality Budgeting in Ireland (OECD, 2019)
7 Equality Budgeting – Relevant Findings from Ex Post Evaluation (DPER, 2019)
This audit is a first step in exposing data gaps and will continue to be updated with detail from data sets not yet captured. While it is not yet a complete picture of the data landscape, the audit has shown that while there is generally good availability of disaggregated data for certain dimensions, there are also significant gaps to be found.

5.1 Dimensions of Equality

It is no surprise that gender is the most commonly captured dimension of equality, featuring in 63.6% of the audited data sets, closely followed by age which is captured in 60.7% of data sets. Gender or sex, and date of birth are commonly included in the basic personal information collected from public service users. However, even though coverage is good, that does not mean that there are no issues with these two dimensions. While a breakdown of gender classification was provided for only 25 out of 68 data sets, 23 out of that 25 capture gender in terms of male/female/prefer not to say. Only 2 data sets provide additional options; this suggests that, for the majority of data sets collecting gender information, consideration needs to be given to the implications of the National LGBTI+ Inclusion Strategy.

Similarly there is some variability in the manner in which the age variable is collected and reported. When age is collected as a continuous variable, i.e., as date of birth or age in years, it is most useful for the purposes of analysis. However, when it is collected in ranges or groupings, e.g., 18-29, 65+, etc. it is less useful, especially when those groups are not standardised across data providers. Date of birth or age in years was collected in 21 out of 26 data sets providing a breakdown, indicating that this more useful approach is the most widespread.

Beyond gender and age, the other dimensions of equality are included in a more sporadic fashion across the public service data sets covered by this audit. Sexual orientation and religion are two dimensions that were very rarely captured. This is not altogether surprising considering that these characteristics are often considered to be personal matters, with public service providers unlikely to request that information except for very specific purposes.

Under the Equal Status Acts 2000-2015, the “ground of race” includes: race, colour, nationality, and ethnic or national origins. This audit revealed a variety of approaches to collecting this category of information. This makes a real difference to analysis for this dimension, e.g., nationality does not necessarily capture ethnicity – this could be a vital distinction in the case of those born in Ireland to non-Irish parents, who could face particular equality issues due to their race or ethnicity. Membership of the Traveller Community is not captured widely in the audited data sets, and is sometimes included with a question on race/ethnicity.

In the case of disability, there also appears to be a variety of approaches to data collection. While only a small number of data providers gave a breakdown of how this
variable is collected, there seem to be three broad approaches: a yes/no question, a question gauging the severity of the disability, or information on the category or type of disability. Again, this variability can make analysis difficult, and therefore can be an obstacle to improving service provision. This finding is no surprise, and has often been raised as a real obstacle to a fuller understanding of the public service requirements of people with a disability. As with other dimensions of equality, a harmonised approach to collecting and classifying disability in public service data sets should be considered.

Roughly one quarter of audited data sets contain civil status; where it is captured, it is consistent and detailed, with all data sets having at least five levels (single, married, widowed, divorced, civil partnership) and some having additional options, including same sex marriage. The clear, and commonly understood, definition of categories in this case might help to explain why it is so consistent. Family status is less widely collected (17.8% of data sets) and typically classified as follows: lone parent, married parent, or no children. Lone parents in particular have been identified as a group facing significant equality issues; the improved availability of family status information could allow for a better understanding of these challenges.

5.2 Data protection vs. analytical capacity

It is important to recognise that this type of information is personal and can often be sensitive, hence there is no implication that data on all dimensions of equality should be collected for every public service data set. This fact has become even more pertinent over the last decade, as increased connectedness due to technological developments has led to high profile data breaches and demands to improve data protection, as evidenced by the adoption by the EU of the General Data Protection Regulation (GDPR). As mentioned in section 4, some public bodies seem unsure of the implications of GDPR and, in some cases, have removed some of the variables they collect. This has a negative implication for the progress of Equality Budgeting, by reducing the availability of disaggregated data, therefore impacting the capacity for analysis across the various dimensions of equality.

This points to a conflict between, on one side, the desire for rich data sources that allow for comprehensive analysis and, on the other, the need to comply with data protection requirements and limit data collection from public service users to that which is completely necessary. In the conduct of this data audit, there was one additional piece of information captured that may offer some solution to this problem. Data providers were asked whether a given data set included a unique identifier, specifically PPSN or Eircode. Of these two, the presence of PPSN is particularly useful from an equality perspective, as it relates to a specific individual, whereas Eircode alone only allows for household-level analysis (although it is itself still very useful, particularly for analysing regional inequalities).
The presence of a unique identifier allows for the possibility of data matching exercises, such as that carried out for a Spending Review paper concerning Carer’s Supports that was published by DPER in 2019. This exercise shows that, if two data sets contain PPSN, the data in both can be anonymised and then linked. Thus, for example, if a researcher wishes to include marital status as a dimension of analysis, but the data source they are using doesn’t capture that dimension, they can obtain this information from another data source that does capture marital status by using the PPSN as a link. As all administrative data sources that contain sensitive information are strictly controlled, it is essential that both data sources are anonymised. It is not feasible or desirable that every public body would try to collect information on every dimension of equality, for every data source. A solution may lie in the more widespread collection of PPSN and Eircode which could, through data matching, enable researchers to access a rich vein of data, disaggregated by equality dimension.

This audit of equality data provides a foundation for the development of an equalities data strategy, in line with the recommendations of the OECD. The implementation of these recommendations is currently underway, guided by the Equality Budgeting Expert Advisory Group, with representatives from both within and outside of government. This consultative approach will be maintained in all efforts to improve the availability of disaggregated data. The availability of such data can empower the proofing tools that will ensure that equality perspectives are fully integrated into the budgetary cycle and the process of resource allocation.

---

8 Review of Carer's Supports (DPER, 2019)
Appendix

Background

CSO is supporting the development of the National Data Infrastructure (NDI). The core concept of the NDI involves the collection, maintenance and storage on all public sector data holdings of the associated PPSN, Eircode and Unique Business Identifier (UBI, to be developed) whenever they are relevant to Public Sector Body transactions with customers and support the development of targeted policy interventions. To this end the CSO has established a public Sector “NDI champions” group of public sector data holders and has developed a dashboard to monitor the use of Eircode and PPSN across the public sector. This dashboard is produced for the Civil Service Management Board (CSMB) and the Public Service Leadership Board (PSLB).

One of the main benefits of the NDI is to enable administrative data to be used to provide insight for policy makers. The CSO has planned to build on the NDI dashboard and repeat the Statistical Potential of Administrative Records exercise first completed in the 2000’s.

Section 42 of The Irish Human Rights and Equality Commission Act 2014 established a Public Sector Equality and Human Rights Duty. This places a statutory obligation on public bodies to eliminate discrimination, promote equality of opportunity and protect the human rights of those to whom they provide services and staff when carrying out their daily work. The law requires public bodies to implement the Public Sector Equality and Human Rights Duty through a three-step process in the context of strategic planning and reporting. This provides an important framework to systematically consider and reflect the particular needs of staff and service users at risk of inequality, discrimination or disproportionate impact, and helps to mitigate and avoid unintended consequences.

Equality budgeting is a way of approaching and understanding the budget as a process that embodies long-standing societal choices about how resources are used, rather than simply a neutral process of resource allocation. In practice, this means that equality budgeting attempts to provide greater information on how proposed or ongoing budgetary decisions impact on particular groups in society, thereby integrating equality concerns into the budgetary process. Ireland’s Equality Budgeting pilot was initiated by DPER in 2017, with dedicated equality indicators included in the REV; the 2017 and 2018 Public Service Performance Reports included an Equality Budgeting Update. While the initial focus of equality budgeting has been on gender, the initiative is being incrementally extended to other equality dimensions. This process has been guided by the Equality Budgeting Expert Advisory Group, which first met in September 2018.

Moreover, the OECD has recently conducted a Policy Scan of Ireland’s approach to Equality Budgeting programme; their report was published on 8 October 2019 and delivered 12 recommendations. One key recommendation was the development of an
equalities data strategy which would rely on improved availability of disaggregated data. The CSO made a statistician available to conduct an audit of existing data from an equality perspective, an exercise that complemented the CSO’s plans for a “SPAR 2” initiative.

The European Commission’s Subgroup on Equality Data published their ‘Guidelines on the Collection of Equality Data’ in 2019. The guidelines provide a series of steps and actions to improve the collection and use of Equality data. One of these steps includes a data audit, along with a range of institutional, structural and operational activities.

**Approach**

The aim of the audit project is to examine all data sets in public bodies and determine the extent of data available on the one or more of nine dimensions of equality in the equal status act: gender, marital status, family status, age, disability, sexual orientation, race, religion and membership of the Traveller community.

Socio-Economic status was discussed by the audit team and the other stakeholders and identified as an important indicator of deprivation, poverty and inequality. It is not included in this version of the audit, as further discussion is needed to identify what variables in data sets could be used as indicators of socio-economic status.

This audit is intended to be used identify data sets that can be analysed to inform policy.

This process was broken into several stages:

1. Meta data in the Administrative Data Centre (ADC) of the CSO was analysed to identify datasets relevant variables.
2. CSO survey teams were contacted to establish if their surveys had relevant variables.
3. Data providers to the NDI were contacted about relevant variables in NDI data sets.
4. The NDI Data Champions were asked to find any other relevant data sets within their own departments which are not included in the NDI data sets, (which are defined as data sets which contain PPSN or Eircode).
5. Contact was made with public sector groups, beyond the NDI Data Champions, for other available data sources.
6. This report on the process and results of the audit was prepared.

In the data gathering/auditing stage, for each of the data sets listed, respondents were asked to provide information on the following:

- the equality dimension(s) covered;
- the organisation responsible for the data set;
- the underlying dataset – the survey source or administrative dataset;
- a unique identifier – such as PPSN or Eircode;
- the theme – a free text field to capture the broad subject matter of the source;
- how often it is released;
- the time period for which the data are available;
- any other relevant notes – free text.

Information was requested related to harmonisation in the second page of the audit. Details of the data collected on the equality dimension was included to show similarities and differences in the data sets collected. For example, age can be collected as date of birth, age in years or as an age grouping which would have to be adjusted when analysing multiple data sets.

Once the audits were all returned, further work was done to refine the results for cohesion and ease of use. Fields that were left blank by respondents have been completed where possible. For example, if the title of a data set included the type of data set (Survey, Administrative etc), this would then be filled in. In other cases, reaching out to respondents for clarity helped to fill in blanks and if there was no solution, the field was left blank.

Similarly, where possible, themes were allocated to each data set. These were chosen to align with the 11 domains of well-being from the well-being budgeting team.

As far as possible, the themes reflect was included in the free text field by respondents. Where the free text response did not align with any theme then the field was left unchanged. In contrast, where the response indicated more than one theme, duplicate records were created for that source, each of which captured a different theme. This helps users to easily analyse by theme.

**Useful Resource:**

In 2003, the National Statistics Board (NSB) and the CSO released a report from a steering group. The report had two major aims: to identify data within existing administrative records that could be used to build social statistics; and to ask those directly involved in policy making in government departments and agencies to identify their precise data needs in the context of the growing importance of evidence-based policy making. They performed a similar audit to the one discussed in this report and this can be seen in Appendix F at the following link:

Quality Assurance process

To ensure accuracy and methodological rigour, the authors engaged in the following quality assurance process.

- Internal/Departmental
  - Line management
  - Spending Review Steering group
  - Other divisions/sections
  - Peer review (IGEES network, seminars, conferences etc.)

- External
  - Other Government Department
  - Other Steering group
  - Quality Assurance Group (QAG)
  - Peer review (IGEES network, seminars, conferences etc.)
  - External expert(s)

- Other (relevant details)