An Roinn Caiteachais Phoiblí agus Athchóirithe
Department of Public Expenditure and Reform

# Budget 2022

## Equality Budgeting: Equality Audit of Tusla Data

Noel Howard (DPER), Ciara Murray (Tusla),
Sadhbh Whelan (DCEDIY) and
Emma Hogan (CSO

Prepared by IGEES staff in the Department of Public Expenditure and Reform, and staff in: Tusla; the Department of Children, Equality, Disability, Integration and Youth; and the Central Statistics Office.

**gov.ie/budget | #Budget2022**

IGEES
Irish Government Economic and Evaluation Service

## 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, rather than simply a neutral process of resource allocation.
- Equality budgeting requires equality-disaggregated data. To progress Equality Budgeting, the OECD recommended the development of an equalities data strategy. As a first step in this regard, the CSO completed an equality data audit of 107 data sources from 31 public bodies in 2020.
- In 2021, a *focused* equality data audit was conducted, covering all national data sources held by Tusla (the Child and Family Agency). Four organisations co-operated on this process: CSO, DCEDIY, DPER and Tusla. This report is based on the findings of that audit.
- The aim of this project was to ascertain the degree of equality-disaggregation present in the audited data sources.
- A template was circulated to gather information about the variables (or *types* of data) collected, particularly those that are equality-relevant.
- A total of 22 data sets were included, representing all national data sets held by Tusla. Responsibility for these data sources is spread amongst the various functions of Tusla.
- The audit has nothing to say about the quality or completeness of the underlying data sources. There may be instances where a particular variable is collected for one individual in the data source but not for another. The audit merely highlights the *types* of equality-disaggregated data that are reported for a given data set.
- This report provides an overview and discusses some key findings of the audit. Reading the report alongside the full audit file permits a fuller comprehension of the process and results, and includes specific information about each of the data sources included. The full audit file is freely available to download in Excel format by following this link: https://s3-eu-west-1.amazonaws.com/govieassets/201015/2aaabcd0-dc43-4f56-b766-2cb49bef9920.xlsx

## Key Findings

- Of the nine grounds covered by the Employment Equality Acts 1998-2015 and the Equal Status Acts 2000-2018, *age* (95%) and *gender* (82%) are well represented across the audited data sets. *Race* (including ethnicity and nationality) is captured in 59% of data sets. Each of the remaining six grounds has coverage below 35%.
- Location data is captured in 82% of cases; Eircode is captured in 27% of cases, as is PPSN.
- The proxy measures of socioeconomic status considered by this audit were not widely present in the audited data sources. This is unsurprising considering that many of the individuals represented in the data are children. The same understanding should be applied when considering the poor availability of variables for family status and civil status.
- With a few exceptions, gender is typically captured in a binary manner, offering a choice between male and female.
- There is inconsistency in the classifications used for the collection of both *disability* and *race* data.
- Within Tusla, work is underway to improve the collection and coverage of data for both *race* (in the form of ethnicity) and *disability.*

## Recommendations

- Where possible, classifications used in the collection and reporting of equality-relevant variables should be standardised, as far as possible, across Tusla national data sets. In many cases, the CSO outlines the standard classifications that should be used.
- Continuing to increase the coverage of both PPSN and Eircode across Tusla data sets is of key importance; this facilitates the anonymous linking of data sets, allowing additional equality-disaggregated data to be included in analysis. Data matching activity of this sort should only be carried out by the CSO within the scope of its Protocol for Linking and Integrating Data Holdings for Statistical Purposes.

## 1. Introduction and Background

### *1.1 Equality Budgeting – Overview*

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.

Equality Budgeting began with a pilot initiative introduced for the 2018 budgetary cycle, coordinated by the Department of Public Expenditure and Reform (DPER). Equality objectives and indicators are published every year in the Revised Estimates Volume (REV)[1] in Q4 and the Public Service Performance Report[2] in Q2. The initial focus of Equality Budgeting was on gender, following which the initiative was extended to other dimensions of equality. 12 departments are now reporting Equality Budgeting metrics. An Expert Advisory Group was established to guide development of Equality Budgeting policy, and has met regularly since September 2018.

In March 2021, the Government agreed on several important points to progress Equality Budgeting. Of particular importance was the establishment of an inter-departmental network to facilitate the embedding of the initiative across all of Government. This group has held two meetings since its inception, and is focused on the practical implementation of Equality Budgeting policy in Government departments.

### *1.2 General Equality Data Audit*

The OECD was commissioned by DPER and the Department of Justice and Equality to carry out a Scan of Equality Budgeting in Ireland, with a report published in October 2019.[3] The report provided twelve recommendations for progressing Equality Budgeting; implementation is currently underway. One of these recommendations was that Ireland should develop an equalities data strategy. This recommendation noted

[1] https://www.gov.ie/en/collection/e20037-revised-estimates/
[2] https://www.gov.ie/en/collection/61d3f-public-service-performance-reports/
[3] https://www.oecd.org/gov/budgeting/equality-budgeting-in-ireland.pdf

the importance of equality-disaggregated data for the operational tools of Equality Budgeting, such as Equality Impact Assessments. In a broader sense, the availability of this type of data allows for a better understanding of the circumstances and requirements of individuals and families. In order to ascertain the degree of equality-disaggregation in existing public service data, the Central Statistics Office (CSO) carried out a data audit (hereafter referred to as the 'main 2020 audit'), publishing the audit file with an accompanying report in 2020.[4] A separate report on that audit was simultaneously published by DPER, placing the findings of the audit in the context of Equality Budgeting.[5]

While that audit captured a wide range of data sets from across public service organisations, some organisations were not included. The timeline during which the overall audit was conducted coincided largely with the onset of the pandemic, meaning that the audit could not be as comprehensive as was initially intended. Tusla was not approached at that point to participate, however it was intended to capture Tusla information in the audit at a later stage. This also applies to public service data sources more generally - the overall equality data audit is an ongoing pursuit and will be updated periodically as additional data sources are submitted.

### *1.3 Focused Equality Data Audit – Tusla*

Building on that high-level data audit covering all public bodies, the Equality Budgeting Expert Advisory Group agreed that conducting a *focused* equality data audit for a single organisation would be a useful exercise, particularly for entities that provide services directly to the public and hold a significant degree of individual-level data. Following preliminary engagement between Tusla (the Child and Family Agency); the Department of Children, Equality, Disability, Integration and Youth (DCEDIY); CSO; and DPER, a decision was made to proceed with a focused equality data audit of all national data sources held by Tusla.

This report provides an overview and discusses some key findings of the audit. Reading the report alongside the full audit file allows a fuller comprehension of the

[4] https://www.cso.ie/en/methods/methodologicalresearch/rp-eda/equalitydataaudit2020/
[5] http://budget.gov.ie/Budgets/2021/Documents/Budget/Equality%20Budgeting%20%E2%80%93%20Equality%20Data%20Audit%20Report.docx.pdf

process and results, and includes specific information about each of the data sources included.

The first sheet in the file includes a range of metadata and an indication of the equality-relevant variables being captured in each case. The second sheet lists the classifications used for each variable in each data set. The full audit file is freely available to download in Excel format.[6]

### *1.4 Key Findings*

As with the main 2020 audit, the primary focus of this audit was the disaggregation of data by equality dimension according to the nine grounds of discrimination presented in Box 1. The results show that *gender* and *age* are very well represented.

**Box 1. The nine grounds of discrimination**

Direct and indirect discrimination are prohibited by the Employment Equality Acts 1998-2015 in the field of employment, and by the Equal Status Acts 2000-2018 in the provision of goods and services, accommodation and education. The Acts cover the nine grounds below, as well as the housing assistance ground when seeking accommodation. More information is available at: https://www.ihrec.ie/your-rights/

- Gender;
- Civil Status;
- Family Status;
- Age;
- Disability;
- Sexual Orientation;
- Race;
- Religion; and
- Membership of the Traveller Community.

[6] The Excel file can be downloaded by following this link: https://s3-eu-west-1.amazonaws.com/govieassets/201015/2aaabcd0-dc43-4f56-b766-2cb49bef9920.xlsx

The dimension of *race* also has fairly good coverage (note that data on nationality is present in more data sets than data on ethnicity). The remaining six grounds of discrimination do not have broad coverage. This is in line with the findings of the main 2020 audit.

While *location* is not included under the grounds of discrimination, it is undoubtedly a dimension in which a range of inequalities may exist – *location* data is captured widely in the audited data sets. Neither is *socioeconomic status* included as a ground of discrimination at present, although the 2020 Programme for Government commits to examine the introduction of a new ground to cover this.[7] Although it is clear that socioeconomic inequality is very real, it can be difficult to measure from a data perspective, with imperfect proxy variables such as income or educational attainment typically used. This audit attempted to account for such proxy variables. The poor coverage of these variables in the results should be considered in light of the child-focused nature of many of Tusla's data sets.

The results also find some variation in classifications between data sets, i.e., the way in which a particular variable is collected. Where possible, a standardised approach to classifications could make the data more useful. The unique identifiers of PPSN and Eircode can allow for equality-disaggregated data to be obtained through anonymous linking of data sets. Coverage of these unique identifiers is set to improve as Tusla is increasing its capability to include both Eircode and PPSN in data collection.

### *1.5 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 General Data Protection Regulation (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.

During the conduct of the main 2020 audit, it became clear that some public bodies were unsure about what was changed by GDPR and whether they could collect certain

[7] Programme for Government: Our Shared Future (2020), p.77. Available at https://www.gov.ie/en/publication/7e05d-programme-for-government-our-shared-future/

types of equality data. In some cases, this uncertainty 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. There needs to be a shared understanding of the requirements placed on Public Sector Bodies by the GDPR and, in particular, an understanding of the exemptions that exist when data is being provided for statistical purposes. These exemptions include:

- The further processing and retention of data for statistical purposes, where the data was originally obtained for another purpose. This applies, for example, to the use of administrative data to compile official statistics.
- The allowance for special categories of personal data to be processed for statistical purposes.
- An exemption, in respect of statistical processing, from the obligation to inform the data subject about the use of the data obtained indirectly from other sources.

There may be additional concerns arising from the use of unique identifiers such as PPSN to anonymously link the individual records in two or more data holdings. It must be emphasised that administrative data is strictly controlled, and data matching activity of this sort will only be carried by the CSO within the scope of its Protocol for Linking and Integrating Data Holdings for Statistical Purposes. Within the scope of this protocol, the CSO will assist other public authorities in matching data holdings available to them, provided a certain set of conditions are met. The CSO states that it will maintain a register, which will be publicly accessible through its website, of all compliant data matching activity undertaken. The register will contain, inter alia, details on: the datasets that have been matched; the reasons why the matching was undertaken; and the statistical outputs that have been obtained.[8]

[8] For further information about this Protocol and the requirements and exemptions relating to the GDPR, visit https://www.cso.ie/en/aboutus/lgdp/csodatapolicies/csodataprotocol/

### *1.6 Additional Points*

The audit and this report have nothing to say about the quality or completeness of the underlying data sources. That is to say, there may be instances where a particular variable is collected for one individual in the data source but not for another. The audit merely highlights the *types* of equality-disaggregated data that are reported for a given data set.

The successful completion of this audit provides a methodology that can be adapted for other organisations to conduct their own focused equality data audits. The recently established Equality Budgeting Interdepartmental Network will play a central role in facilitating this in departments, agencies and offices across Government.

The remainder of this paper is structured as follows: Section 2 outlines the methodological approach; Section 3 provides the results; Section 4 discusses the findings of the audit; Section 5 presents recommendations; Section 6 concludes.

## 2. Methodology

A small working group was assembled, with one member each from Tusla, DCEDIY, CSO and DPER. The group met online on several occasions between March and September of 2021. The first step was to draft a template that could capture the necessary metadata for the sources being audited. A template of this sort had previously been designed for the main 2020 audit. That original Microsoft Excel template was used as a starting point, to which were added columns to capture greater detail on unique identifiers, as well as additional variables relating to location and socioeconomic status. As with the original template, a second sheet was also included to determine the classifications used to capture each equality dimension.

The modified template was circulated appropriately within Tusla to ensure that all national data sources held by that organisation were captured. The resulting metadata was then collated by Tusla and entered into one single template. Following some minor cleaning of the data, the analysis was conducted by producing descriptive statistics with the primary aim of understanding each equality-relevant variable as follows:

1. The frequency with which that variable was captured, i.e., how many of the surveyed data sets included a given dimension.
2. The different ways in which that variable might be classified across the surveyed data sets.

Once this primary analysis was completed, this report was drafted with input from all representatives on the working group.

## 3. Results

### *3.1 Overview*

22 Data sets were covered by this audit, accounting for all national data sources held by Tusla (see Appendix A for list). Responsibility for these data sources is spread amongst the various functions of Tusla, as given by Table 1. More detail on each of these sources is available in the full audit file (Excel).

**Table 1**

Breakdown by Section/Unit

| **Section/Unit** | No. of data sources |
|---|---|
| Tusla Operations | 10 |
| Adoption Services | 4 |
| Tusla Education Support Services | 3 |
| Human Resources | 3 |
| Domestic, Sexual and Gender Based Violence (DSGBV) Services | 2 |
| **TOTAL** | **22** |

The findings as given in Table 2 are broken down by equality-relevant variable. *Coverage* in Table 2 is given both for this audit of Tusla data and for the main 2020 audit conducted by CSO that included 107 data sources from 30 public bodies (see Appendix B for list of organisations included in the main 2020 audit). A graphical comparison is presented in Figure 1.

**Table 2**

Breakdown by equality-relevant variable

| | **Variable** | No. of data sources | Percentage coverage in Tusla data sources | Percentage coverage in main 2020 audit[a] |
|---|---|---|---|---|
| Nine Grounds under the Employment Equality Acts 1998-2015 and the Equal Status Acts 2000-2018 | Age | 21 | 95% | 61% |
| | Gender | 18 | 82% | 64% |
| | Race[b] | 13 | 59% | 24% |
| | Disability | 7 | 32% | 24% |
| | Membership of the Traveller Community | 6 | 27% | 11% |
| | Family Status | 4 | 18% | 18% |
| | Civil Status | 3 | 14% | 23% |
| | Religion | 3 | 14% | 4% |
| | Sexual Orientation | 1 | 5% | 6% |
| | Location | 18 | 82% | n/a[c] |
| Unique Identifiers | PPSN | 6 | 27% | n/a[d] |
| | Eircode | 6 | 27% | n/a[d] |
| Proxies for Socioeconomic Status | Own Education | 7 | 32% | n/a[c] |
| | Employment Status[e] | 4 | 18% | 21% |
| | Socioeconomic Group[f] | 0 | 0% | n/a[c] |
| | Income[e] | 0 | 0% | n/a[c] |
| | Parent's Education | 0 | 0% | n/a[c] |
| **TOTAL** | All data sets in audit | 22 | 100% | - |

[a] 107 data sets from 30 public bodies. See Appendix B for details.

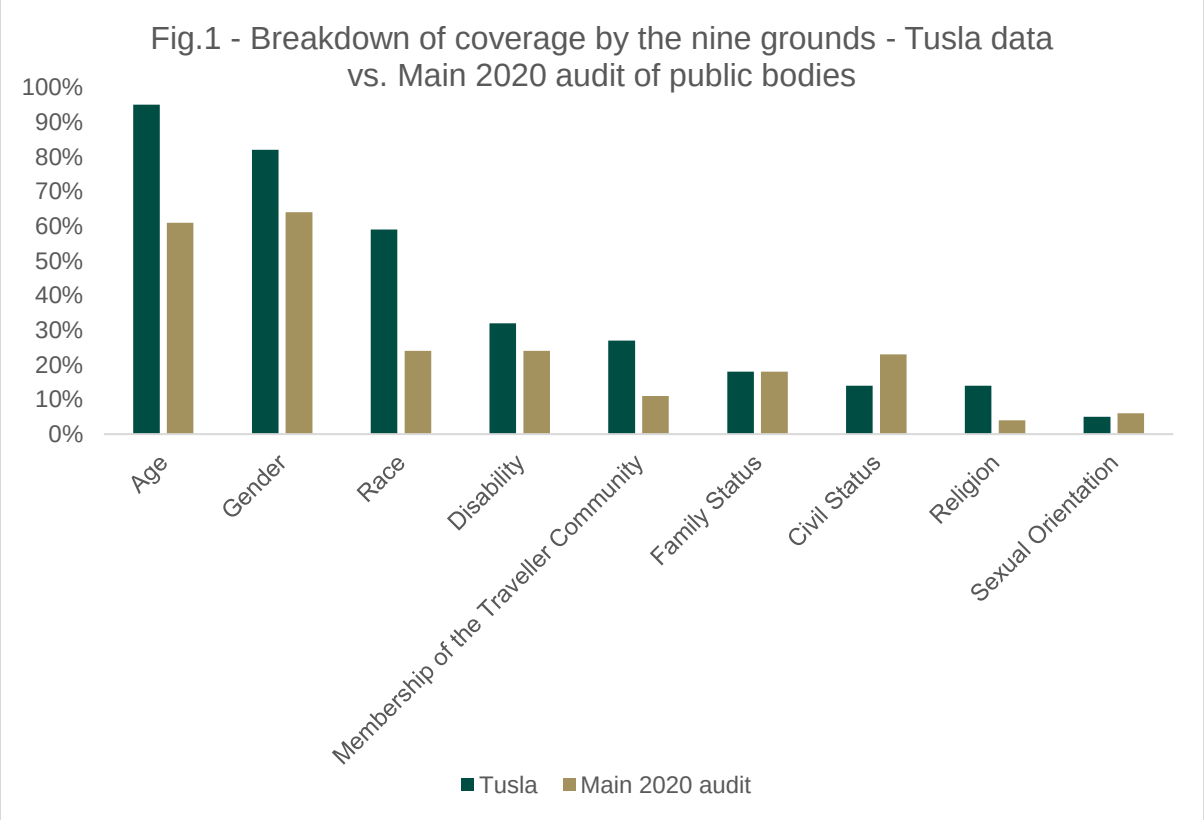
[b] Includes *race. ethnicity* and *nationality/country of origin*

[c] Variable not included in main 2020 audit.

[d] While the presence of a unique identifier (PPSN or Eircode) was captured in the main 2020 audit, it was not specified which identifier was present for each data source.

[e] To account for child-focused data sets, indication was given as to whether this related to parent's *employment status* in the *Notes* column on the Excel audit file. In 3 of the 4 instances, this refers to own *employment status* (adult data sets). The same approach was taken to *income* but this variable was not present in any form in any data set.

[f] Socioeconomic group (SEG) classifies the entire population into one of ten groups based on the level of skill and educational attainment of the occupation (of those at work, unemployed or retired) while all other persons are classified to the socio-economic group of the person in the family on whom they are deemed to be dependent. More information is available at: https://www.cso.ie/en/releasesandpublications/ep/p-cp11eoi/cp11eoi/bgn/




It is important to note that *coverage* in this instance simply refers to the presence of a given variable in a given data set – that there is a "box on the form" for that equality dimension. This says nothing about the completeness of the data sets in respect of that variable, i.e., whether it is collected for most individuals or whether there are a lot of blanks. In many cases it may not be compulsory to collect (or for individuals to provide) information about these characteristics. If one considers the role of (for example) a Tusla social worker, data collection of this sort is not the primary aim of their service provision.

It is also important to point out the unique composition of Tusla data. In a large number of cases, the service user in question is a child. In these instances, the primary characteristics being recorded in data will be those of the children themselves; questions about *civil status* or *family status* are less likely to be of primary concern for those collecting data directly from the child.

One data source which does allow for the collection of a wide range of equality-relevant variables, including *civil status* and *family status,* is the National Child Care Information System (NCCIS). This is an integrated case management system for child protection and welfare cases including children in care. A key objective of NCCIS is to support the management and provision of effective, quality services by providing timely, relevant and reliable information.[9] Information on each of the other data sources is available in the full audit file (Excel).

One final point to note is that some of the data sources included are maintained solely for case management purposes. Data within these is often not collated and in many cases the information is contained within case notes for each individual.

### *3.2 Coverage of the Nine Grounds of Discrimination*

The findings presented in this section relate to the nine grounds of discrimination that are listed in Box 1.

*Age* is captured in 21 of the 22 data sources examined, making it the equality dimension with the greatest coverage. In 17 out of those 21 instances, it is captured as a continuous variable, i.e., the specific age is included (usually as date of birth). This is desirable from a general research perspective, as it allows the analyst to determine any age-related inequalities more accurately than if using data coded in age ranges. One data set captures age as 'years-old' which is almost as strong as capturing age continuously. The remaining three data sets collect age ranges; two of these use the same breakdown and relate to surveys on Domestic, Sexual and Gender Based Violence Services, so it is understandable that date of birth would not be sought.

[9] https://www.tusla.ie/news/tuslas-new-national-integrated-information-system-goes-live/

*Gender* is captured in 18 of the 22 data sources included in this audit. In 11 instances, this dimension is captured as a binary male/female value. Four of the data sources include a specific option to select 'Transgender' or 'Trans'. The remaining three instances are non-standard.

*Disability* is captured in seven of the 22 data sources. In two instances it is included in an unstructured format in case history and notes. In the remaining five cases it is recorded differently in each data source. These different approaches are: a simple yes/no question, an indicator of either mental or physical disability, and three different scales to indicate severity.

*Civil Status* is captured in three of the 22 data sources, although considering the child-focused nature of a large part of Tusla's operations, it is not surprising that this figure is so low. Each of the three data sources captures civil status in a different way, with one offering 10 options, the next offering seven and the other offering four options.

*Family status* is captured in four of the 22 data sources, although, again, considering the child-focused nature of a large part of Tusla's operations, it is not surprising that this figure is so low. This dimension is captured in a variety of ways very specific to each data source, and may require re-coding if conducting research using multiple linked data sets.

*Race* is captured in 13 out of the 22 data sources. The ground of *race* includes discrimination on the basis of colour, nationality or ethnic or national origins.

- Five sources capture data on *ethnicity.* Of these, one uses free text, two use a drop down list (with more than 10 options) and two use the same nine-option list. One additional data source contains information on ethnicity in an unstructured format in case history and notes.
- Eight data sources capture *nationality* or *country of origin*. In two instances, the CSO Standard Country List is used, and in another two instances an unspecified drop-down list of more than 10 countries is employed. Another three data sources capture this information as free text. The final source only captures *nationality* in cases where an interpreter is required.

*Membership of the Traveller Community* is captured in six of the total of 22 data sources. In three of these cases, it is recorded as part of an *ethnicity* drop-down list. In the other three cases, non-standard approaches are used.

*Religion* is captured in three out of the 22 data sources covered. In one instance it is recorded by free text; in another, a large drop-down list is used; in the other, four options are available. Collection of *religion* markers is important to account for many cohorts, not least people of minority faiths in their country of origin, who come to Ireland to avoid persecution on religious grounds.

*Sexual Orientation* is captured in one of the 22 data sources covered by this audit. In this case, the data are used for case management and are not collated or reported at present; the dimension is captured by selection from a list with six discrete categories. Poor availability of data concerning *sexual orientation* was also a finding of the main 2020 audit. In the context of Tusla, it should be noted that the dimension of *sexual orientation* may be less relevant for younger children.

### *3.3 Other Dimensions of Equality*

While *location* is not included under the nine grounds covered by the Equal Status Acts, it is undoubtedly a dimension in which a range of inequalities may exist. These inequalities may persist in that one area is disadvantaged compared to another, or in the sense that there may be a rural-urban divide nationally on certain dimensions of equality. Collecting *location* data makes the assessment of such inequalities possible. *Location* is recorded in some form by 18 of the 22 data sources covered – this is as expected, as, along with *gender* and *age,* location information is one of the key types of administrative data typically collected from service users. *Eircode* is recorded in six data sources, *full address* is recorded in an additional nine data sources, while the remaining three have *area* or *region* as the most granular level of location information.

As with *location*, socioeconomic status is not included under the current discriminatory grounds. This dimension is not straightforward from the perspective of collecting and reporting equality-disaggregated data; socioeconomic inequalities undoubtedly exist but there are different approaches to measurement. One or more related variables are often used as proxy variables (e.g., income, educational attainment, occupation). Work

is ongoing to develop a definition of *socioeconomic status* in light of the Programme for Government 2020 commitment to examine the inclusion of that dimension as a ground of discrimination under legislation.

This audit attempted to determine if any data was routinely being captured in Tusla data sources that could be used by researchers as a proxy for socioeconomic status. The variables considered were: *income, own education, parent's education, socioeconomic group* (a classification included in many administrative data sources), and *employment status*.

*Own education* is included in seven of the 22 data sources covered by this audit; considering the individuals in question are often young children, this variable may not always be a useful indicator of socioeconomic status. *Employment status* is recorded in four data sources; it is captured as free text in two of these, and in different drop-down lists in the other two. In the single child-centred data source that captures the variable, *employment status* refers to that of the parent. Neither *income* nor *parent's education* nor *socioeconomic group* are present in any of the data sources.

In general, there is limited availability of these proxies for socioeconomic status in the data covered by this audit. *Employment status*, *education* and *income level* are among the most useful variables of this sort; given that many of these data sources are child-focused, gaining an understanding of socioeconomic status would require the collection of parents' levels of one or more of these variables. Another potential proxy for socioeconomic status could be obtained by using detailed *location* data in conjunction with the Pobal deprivation index. This index uses census data to measure the relative affluence or disadvantage of a small geographical area (50 – 200 households)[10]. This highlights another benefit to the widespread inclusion of Eircode in data collection.

### *3.4 Unique Identifiers*

It is unlikely that equality-disaggregated data can be collected for all data sources and for all dimensions of equality. However, the absence of disaggregated data from one source can be remedied if that data is captured in another source, if and only if there

[10] https://www.pobal.ie/launch-of-2016-pobal-hp-deprivation-index/

are the means of anonymously linking each record or row by means of a unique identifier. The most commonly used unique identifiers in Ireland are Eircode and the Personal Public Service Number (PPSN). Out of the 22 data sources included in this audit, three include both PPSN *and* Eircode, three include PPSN only, and three include Eircode only. This leaves 13 data sources which do not include either of these two key unique identifiers. Many of these *do* include a unique identifier of some sort, for example a reference number or a Child ID, however, these are generally specific to each data set and may not be useful from a data linkage perspective.

## 4. Discussion

### *4.1 Equality Data Collection*

*Gender* is collected in most instances, however options are typically limited to a binary choice of male/female. Recent years have seen a growing recognition that individuals may have gender identities that are different to their biological sex and that can also change over time. Accounting for these developments requires consideration of new approaches to the collection of *gender*-disaggregated data. The CSO *Take Part* campaign consists of quick and easy online Pulse Surveys about topical issues. These Pulse Surveys include a gender identity question, and request feedback from respondents about the way in which the question is phrased. [11] This process could help to inform future classifications and approaches to the collection of information on gender.

*Disability* can be defined in a variety of ways: the broad definitions included in legislation and case law; definitions employed by CSO; medical definitions used in self-reported health measures or administered by healthcare professionals. Disability data is not widely collected in the data sources under consideration and there is a variety of approaches taken to recording this dimension of equality. This reflects the findings of the overall data audit completed by CSO in 2020.

Data on the ground of *race* is available in the majority of sources included in this audit; nationality (or country of origin) is the most frequently collected variable under this ground. However, while nationality is certainly relevant from an equality perspective,

[11] https://www.cso.ie/en/surveys/householdsurveys/pulsesurveys/pulsesurveyquestionsfaq/

it does not fully account for the ground of race. This might be the case, for example, with an individual that has Irish nationality but has a different ethnic background. For this reason, it is valuable from an equality perspective to collect specific information on an individual's racial or ethnic origin.

In 2017, Irish Travellers were recognised as a distinct ethnic group. In the data sources under consideration, *Membership of the Traveller community* is sometimes collected as an option under the *ethnicity,* and sometimes collected using a separate question. As data on ethnicity is more widely collected, an option for Irish Traveller should always be included.

The fact that only one data set includes information on *sexual orientation* reflects the situation in the wider administrative data landscape. In 2019, the CSO included questions on both gender identity and sexual orientation in its General Household Survey for the first time.[12] This development may provide clarity on best practice in the future collection of this equality dimension.

### *4.2 Ongoing and Recent Developments – Equality Data Collection*

*Ethnicity data collection:* Tusla is committed to effectively planning and delivering culturally appropriate services to diverse populations. Ethnic data collection will facilitate and enable monitoring of the unique needs and outcomes of children and families engaging with its services. In 2019, Tusla commissioned research to support the development of an ethically appropriate, legal ethnic data collection system within Tusla that adheres to a human rights framework[13]. The aims of the study were three-fold:-

- To identify the legislative and policy context within which Tusla can develop an ethnic data collection system;
- To establish best practice guidelines for Tusla in ethnic data collection methods and systems; and

[12] https://fra.europa.eu/en/promising-practices/introducing-questions-sexual-orientation-and-gender-identity-equality-module
[13] https://www.tusla.ie/uploads/content/Ethic-Data-Final-Report-2_230819.pdf

- To ensure that Tusla's information systems (IS) can apply best practice to inform service improvement.

Tusla is currently progressing with the recommendations from that study and will commence a Data Protection Impact Assessment (DPIA) later in the year.

*Education and Children in Care Data Linkage Project:* Officials from DCEDIY, Tusla and the CSO are engaged in a joint project to improve the availability of data on the educational attendance and attainment of children and young people in care and those who are care experienced. Integral to the collection of such data is the need to establish a link between the administrative data collected by the NCCIS on children and young people in care and the information stored in the Administrative Data Centre of the CSO on attendance and attainment.

It is unlikely that data captured by the NCCIS on ethnicity will be used in this project as the accuracy and reliability of that data is under review. It is likely that data on disability will be included.

*Domestic, Sexual and Gender Based Violence Services (DSGBV):* It is planned that future surveys will collect data from service users relating to ethnicity and disability, in line with Istanbul Convention obligations and commitments under the Second National Strategy on DSGBV.

### *4.3 Data Collection – Unique Identifiers*

The presence of a PPSN or Eircode for each individual in a given data source increases the potential for detailed research that can be used to enhance the policy-making process and, ultimately, to deliver better public services and outcomes for citizens. At the least, it is recommended that Eircode is collected wherever an individual's address is requested – in many cases this could be achieved by adding a single line or cell to a form (whether paper or online).

There are, of course, some instances where data may include very sensitive or personal information. In such cases, where it may not be appropriate to include these unique identifiers, additional equality data will not be obtainable through data-linking and so it is advisable to collect as many equality characteristics as is deemed appropriate, in order to understand the service users.

### *4.4 Ongoing and Recent Developments – Unique Identifiers*

*Eircode:* Tusla has acquired a full corporate ECAD (Eircode Address Database) licence which allows for the full integration of Eircode in all Tusla systems. The first system to have Eircode lookup functionality integrated and implemented is NCCIS. This will allow for greater accuracy of address information and address validation for existing addresses. As well as increasing efficiency of data input and quality of address information, the integration of Eircodes into NCCIS will mean better quality data and reduced risk for postal-related issues.

*PPSN:* Tusla already collects PPSN in several of its data sources, including the NCCIS. The organisation has also stated that it will seek to expand its use of the PPSN where applicable to facilitate greater efficiency in delivering its services and improved data management.[14]

## 5. Recommendations

### *5.1 Increasing Data Collection – Pros and Cons*

The collection and reporting of equality-disaggregated data are not straightforward pursuits, requiring consideration of several factors. The availability of such data delivers a better understanding of the needs of public service users from all cohorts, and thus allows for the development and implementation of policy in a more equal way. This must be weighed against concerns around the sensitive and personal nature of gathering some types of data from individuals, and the constraints of data protection more formally in the GDPR. There are also concerns around placing the administrative burden of additional data collection on those individuals who deal directly with service users. The following recommendations are delivered in light of the need to strike a balance between these factors.

### *5.2 Classifications*

One point that emerged in the findings for several equality dimensions was the lack of a standard approach to classification for a given variable. This can be seen in the case of *gender*, where the majority of sources collect only male/female and a small number

[14] https://www.gov.ie/en/publication/82d0c8-tusla-child-and-family-agency-register-of-users/

offer additional options. The lack of a standard approach to classification is also notable for the dimensions of *race, disability, civil status* and *family status*. Where possible, CSO standard classifications should be used as progress is made in improving the collection of equality data. Of course, there may be specific data sets where the use of the standard approach to classification is not be feasible or desirable.

### *5.3 Unique Identifiers*

It is not reasonable to expect that a wide range of equality-disaggregated data will be collected from each public service user at each instance. The collection of PPSN and Eircode is a valuable addition to any data set, potentially offering further insights through the process of anonymous data linking. One such project is already underway with Education and Children in Care Data Linkage Project, as described in section 4.2. Moreover, the developments outlined concerning the increased collection of PPSN and Eircode in section 4.4 are most welcome in this regard. Increasing the coverage of both PPSN and Eircode in Tusla data is in line with the recent recommendation of the National Statistics Board to continue implementation of a National Data Infrastructure. This infrastructure relies on the inclusion of unique identifiers to facilitate the anonymous linking of data across the administrative and CSO systems.[15]

## 6. Conclusion

Developing a corpus of high-quality individual-level data that contains information on key equality characteristics is essential in order to effectively assess the equality impact of new and existing budgetary measures, policy proposals, regulations and legislation. Such high-quality data allows for robust equality impact assessments to be carried out, ensuring the perspectives of all cohorts are included in the business of government. Individual organisations can begin by identifying data gaps where they exist, then addressing those gaps either with the increased collection of equality variables (with standardised classifications) or the widespread coverage of PPSN and Eircode. Of course, both data protection and the potential for increased administrative burden should be considered before deciding to increase data collection.

[15] https://www.nsb.ie/media/nsbie/pdfdocs/NSB_Statement_of_Strategy_2021_2026.pdf

The methodological approach described in this report can be adapted and applied for any organisation collecting individual-level data. In conjunction with the recently established Equality Budgeting Interdepartmental Network, other public service organisations can use the approach outlined in this paper (along with the accompanying template for collecting metadata) to carry out their own focused equality data audits. The full Excel audit file can also a useful tool for researchers seeking an overview of Tusla data from an equality perspective.

Engaging in an equality data audit can be very useful for any public body seeking to fulfil their obligations under the Public Sector Equality and Human Rights Duty to assess, address and report on equality and human rights issues relevant to its purpose and functions.[16] It can also enrich an organisation's understanding of the data that they hold, potentially allowing inequalities among service users to be exposed and tackled. By reducing inequalities, the general level of well-being in society is undoubtedly raised.

[16] https://www.ihrec.ie/our-work/public-sector-duty/

## Appendix A

### Titles of the 22 data sources included in this audit:

- Assessment Consultation Therapy Service (ACTS) Referral Data Base
- Assessment Consultation Therapy Service (ACTS) Clinical Assessment Tool
- National Child Care Information System (NCCIS)
- Educational Welfare Case Management System (eCRIS)
- Out of Hours Crisis Management System
- Tusla Online Portal
- Separated Children Seeking International Protection (SCSIP) - Referrals Database
- Separated Children Seeking International Protection (SCSIP) - Children in Care Register
- National Inter Agency Prevention Programme (NIAPP)
- Aftercare dataset
- Domestic, Sexual and Gender Based Violence Services - Annual Survey Domestic Violence
- Domestic, Sexual and Gender Based Violence Services - Annual Survey Sexual Violence
- Educational welfare services referral database
- Student Absence Report System
- Employee Hire Form
- Annual Staff Disability Census
- Garda Vetting Process
- Child Protection Notification System (CPNS)
- Mother and Baby home database
- Therefore - electronic filing system
- Adoption and Alternative Care Records Sampling Project Report
- National Domestic Adoption Register

More detail on each of these sources is available in the full audit file (Excel); to download, follow this link:

https://s3-eu-west-1.amazonaws.com/govieassets/201015/2aaabcd0-dc43-4f56-b766-2cb49bef9920.xlsx

## Appendix B

List of public bodies included in the main 2020 Equality Data Audit conducted by the CSO:

- Central Bank of Ireland
- CORU (Health and Social Care Professionals Council)
- Dental Council
- Department of Communications, Climate Action and Environment
- Department of Education and Skills
- Department of Employment Affairs and Social Protection
- Department of Foreign Affairs and Trade
- Department of Health
- Department of Housing, Planning and Local Government
- Department of Justice
- Department of Rural and Community Development
- Health Research Board
- Health Insurance Authority
- Higher Education Authority
- Health Products Regulatory Authority
- Health Service Executive
- Irish Blood Transfusion Service
- Local Government Management Agency
- Medical Council
- Pharmaceutical Society of Ireland
- SOLAS
- Revenue
- The Central Statistics Office
- Environmental Protection Agency
- ESB
- Health Research Board
- Health Information and Quality Authority
- Inland Fisheries Ireland
- Irish Prison Service
- National Oil Reserves Agency
- Probation Service