

OVERVIEW of Inventory of Data Sources on Children's Lives



The authors of this report are:

Anne Marie Brooks

Ruth Geraghty

Sarah Fitzgerald

Gillian Roche

Copyright © Minister for Children and Youth Affairs, 2012

Department of Children and Youth Affairs

43-49 Mespil Road

Dublin 4

Tel: +353 (0)1 647 3000

Fax: +353 (0)1 667 0826

E-mail: contact@dca.gov.ie

Web: www.dca.ie

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission in writing of the copyright holder.

For rights of translation or reproduction, applications should be made to the Head of Communications, Department of Children and Youth Affairs, 43-49 Mespil Road, Dublin 4, Ireland.

Introduction

This Overview document accompanies the *Inventory of Data Sources on Children's Lives*, which is an online inventory of data sources covering multiple dimensions of children's lives. The inventory is available online as a series of meta-data summaries in tabular format at www.dcy.gov.ie.

These resources are aimed at a range of users, including those involved in research, service planning, service evaluation and policy development and monitoring. They are intended to promote awareness and increased use of existing data sources on children's lives.

This Overview sets out the context in which the inventory of data sources was initially compiled, its role in the development of the *National Strategy for Research and Data on Children's Lives, 2011-2016* (DCYA, 2011a), some analyses undertaken on the inventory to date and an overview of the structure and format of the online meta-data summaries.

The summaries themselves are available electronically only, but a print-ready format can be generated from the online resource. Due to the likely need for frequent updates and revisions, hard copies of the summaries were not produced.

Context

The initial impetus for the development of an inventory of data sources arose during the development of the *National Strategy for Research and Data on Children's Lives, 2011-2016* (DCYA, 2011a). This strategy originated in cross-Government initiatives towards more strategic approaches to data and statistics, led by the National Statistics Board (NSB, 2003a and 2003b). However, the approach taken went further – in focusing on research as well as on data, and by taking account of information and knowledge concerning children's lives on a 'whole of Government' basis.

The development process for the strategy was aligned as closely as possible with the *Best Practice Guidelines for the Development and Implementation of formal Data/Statistics Strategies in Government Departments* set out by the NSB (2004). In relation to data sources to meet policy information needs, these guidelines state that:

Departments should identify the data holdings available internally (including agencies and other bodies acting on behalf of the Department) that could contribute to meeting the Department's policy data needs. [Guideline 5]

Departments should identify the external data holdings that could contribute to meeting the Department's unmet policy data needs. [Guideline 6]

Therefore, an inventory of both internal and external data sources was an essential component of the evidence base for the *National Strategy for Research and Data on Children's Lives*. The original inventory was compiled in 2008 and was drawn from both administrative and survey-based datasets that were identified as relevant to the lives of children in Ireland.

Compiling the inventory

A total of 116 data sources were identified for inclusion in the original inventory. The following inclusion criteria were used:

- » data sources that related to children and their families, or services and supports for children and families;
- » data sources that were expected to be updated on an ongoing or regular basis.

Data sources currently being developed and once-off data sources were excluded from the inventory.

The data sources originally selected through use of these criteria are listed in Appendix 6 of the Main Report of the *National Strategy for Research and Data on Children's Lives, 2011-2016* (DCYA, 2011a).

Each variable and the associate response categories from the 116 data sources were entered into a database, resulting in more than 11,000 records. Individual records were then coded according to a number of concepts, including a thematic categorisation (see *Table 1*). This was in line with the analytical approach taken to the consultation element of the *National Strategy's* development (DCYA, 2011b) and the review of policy recommendations, which also formed part of the evidence base for the strategy. Further detail on the structure of all the inputs to the strategy and the analysis conducted is outlined in Chapter 2 of the Main Report (DCYA, 2011a).

Table 1: Thematic coding of inventory

Coding element	Categories
Theme/Domain	Health Education Social/Emotional/Behavioural Financial/Economic Participation Safety Relationships Demographics Environment Justice

Analysis of the inventory

Following the thematic coding of each variable, each data source was then assigned to one or more themes/domains according to the code of the majority of variables within the dataset. These, along with the other categorisations, informed the identification of data gaps in the strategy's development. ***The principal themes/domains for each data source are provided as one of the fields in the meta-data summaries.***

In addition, each data source was examined for the presence of a predefined set of key classificatory variables, including basic socio-demographic information such as age, sex, household structure, income and socio-economic status, and markers for the 9 grounds covered by Equality legislation. The presence or absence of a unique identifier (such as the PPSN in a dataset) was also recorded; this enabled an examination of the extent to which certain key variables were present, but also the extent to which the format of such variables were harmonised across the datasets. This work enabled the identification of issues relating to data processes, such as data linkage, comparison of concepts across datasets and disaggregation by key variables. This also contributed to the identification of data needs, but in terms of data processes rather than specific topics. ***The presence of these variables is identified within the meta-data summaries under the heading 'Key classificatory variables'.*** The terminology used in this field is explained below.

Description of online inventory

The inventory of data sources has evolved since its first development in 2008. Some data sources that were in place at the time have been discontinued. New data sources have been established. Others have been developed or amalgamated in line with changing technologies and information needs. One notable example is the shift from 20 individual reporting forms for notifiable diseases to a Computerised Infectious Disease Reporting (CIDR) system. The content of the current inventory reflects these changes and it is expected that continuous updating will be required as similar developments occur.

The inventory is presented in 3 categories:

- » **Administrative data sources**, i.e. data collected for or through the administration of specific schemes or services
- » **Cross-sectional Surveys and Censuses**
- » **Longitudinal Surveys**

Within these categories, meta-data is supplied for each data source, or in the case of more complex data collections for each element of the data source.

Structure of meta-data summaries

Each summary begins with the title of the dataset and an alphanumeric inventory code for ease of reference (A for Administrative; S for Survey; L for Longitudinal). The meta-data is then set out in tabular format. Different fields are used for the description of survey-based and administrative data due to their differing natures. For example, issues of response rate and sample size are not relevant to administrative data sources. An overview of the structure of each is given below, setting out the different fields used for each type of data source.

ADMINISTRATIVE data sources: Structure of meta-data summary

Data owner	Institute/organisation(s) that generates the data and owns the dataset (may be provided for in legislation)
Brief description	Outline of content, purpose and scope of data collection
Primary function of data	Description of the administrative scheme for which the data are originally collected
Key classificatory variables	Indicates presence of identifier fields or selected socio-demographic and other classificatory variables in the dataset. There are 22 key variables: PPSN; ID Other; Sex; DOB; Age Yr; Age Other; County; Geog Sub County; Geog Other; Marital Status; Family Status; Household; Occupation; Income; Ed Attain Mother; Labour Force Status; Nationality; Disability; Race/Ethnicity; Traveller; Sexual Orientation; Religion
Themes/domains	Indicates main theme(s) addressed by data source from list of 10 categories: Health; Education; Social/Emotional/Behavioural; Financial/Economic; Participation; Safety; Relationships; Demographics; Environment; Justice
Unit of observation	Person or entity who is the subject of data collection, e.g. child, parent, school, household
Start date of data collection	Date on which data collection began
Frequency of data collection	Details on how often data collection is/has been repeated since the commencement of the data source
Data format	How and where data are collected and stored, e.g. computerised <i>versus</i> paper-based, centralised <i>versus</i> localised
Method of data collection	Details of vehicle through which data are collected, e.g. claim form, annual review, medical report
Date of last publication	Details of latest publication(s)/output(s) from data source. <i>Note:</i> Not all of these data sources may result in published outputs.
Access/further information	Contact details for further information or access to the dataset

SURVEY data sources: Structure of meta-data summary

Title of overall study	Title of study on which the dataset/instrument is based
Data creator	Institute/organisation(s) generating the data
Data funder	Institute/organisation(s) funding the data collection/processing/publication
Data publisher	Institute/organisation(s) principally responsible for publishing outputs/reports based on the data
Brief description	Outline of content, purpose and scope of data collection
Key classificatory variables	Indicates presence of identifier fields or selected socio-demographic and other classificatory variables in the dataset. There are 22 key variables: PPSN; ID Other; Sex; DOB; Age Yr; Age Other; County; Geog Sub County; Geog Other; Marital Status; Family Status; Household; Occupation; Income; Ed Attain Mother; Labour Force Status; Nationality; Disability; Race/Ethnicity; Traveller; Sexual Orientation; Religion
Themes/domains	Indicates main theme(s) addressed by data source from list of 10 categories: Health; Education; Social/Emotional/Behavioural; Financial/Economic; Participation; Safety; Relationships; Demographics; Environment; Justice
Unit of observation	Person or entity who is the subject of data collection, e.g. child, parent, school, household
Start date of data collection	Date on which data collection began
Date of last data collection	Date of most recent data collection
Frequency of data collection	Details on how often data collection is/has been repeated since the commencement of the data source
Coverage	Details of target population of data collection, sample size and response rates
Method of data collection	Details methodology used for data capture, e.g. personal interview, computer assisted interview, self-completion
Date of last publication	Details of latest publication(s)/output(s) from data source
Internationally comparable data	Details of any international data collections aligned to the national source, including other countries involved and links to international organisations or networks
Access/further information	Contact details for further information or access to the dataset

LONGITUDINAL data sources: Structure of meta-data summary

Title of overall study	Title of study on which the dataset/instrument is based
Data creator	Institute/organisation(s) generating the data
Data funder	Institute/organisation(s) funding the data collection/processing/publication
Data publisher	Institute/organisation(s) principally responsible for publishing outputs/reports based on the data
Brief description	Outline of content, purpose and scope of data collection
Key classificatory variables	Indicates presence of identifier fields or selected socio-demographic and other classificatory variables in the dataset. There are 22 key variables: PPSN; ID Other; Sex; DOB; Age Yr; Age Other; County; Geog Sub County; Geog Other; Marital Status; Family Status; Household; Occupation; Income; Ed Attain Mother; Labour Force Status; Nationality; Disability; Race/Ethnicity; Traveller; Sexual Orientation; Religion
Themes/domains	Indicates main theme(s) addressed by data source from list of 10 categories: Health; Education; Social/Emotional/Behavioural; Financial/Economic; Participation; Safety; Relationships; Demographics; Environment; Justice
Unit of observation	Person or entity who is the subject of data collection, e.g. child, parent, school, household
Start date	Date on which data collection began
Completion date	Date on which the most recent wave of data collection was completed
Coverage	Details of target population of data collection, sample size and response rates
Method of data collection	Details methodology used for data capture, e.g. personal interview, computer assisted interview, self-completion
Date of last publication	Details of latest publication(s)/output(s) from data source
Internationally comparable data	Details of any international data collections aligned to the national source, including other countries involved and links to international organisations or networks
Access/further information	Contact details for further information or access to the dataset

Key classificatory variables

An extensive analysis of the original 116 data sources was undertaken in 2009-2010. The data sources in the original inventory were examined, firstly for the presence of 22 specified variables which were regarded as key classifications in data sources on children's lives. These 22 key variables were classified into 8 groups and are listed in Table 2.

Table 2: Key classificatory variables in meta-data summaries

Group	Key variables
1. Identifier variables	PPSN ID Other
2. Sex variable	Sex
3. Age variables	DOB Age Yr Age Other
4. Address variables	County Geog Sub County Geog Other
5. Family structure variables	Marital Status Family Status Household
6. Socio-economic variables	Occupation Income Ed Attain Mother Labour Force Status
7. Nationality variable	Nationality
8. Equality variables	Disability Race/Ethnicity Traveller Sexual Orientation Religion

Secondly, the inventory was also examined for the extent of conformity or variation that existed in the collection of each of these key variables (for example, diversity across data sources in question, format and/or response categories to generate similar information). This work provided insight into the level of variation in relation to these key variables that existed across the inventory of data sources. A more detailed report (unpublished) of this analysis has been prepared and is intended for use by an Action Group on key data sources that will be established to progress the Action Plan of the *National Strategy for Research and Data on Children's Lives*. As a starting point, information on the presence of each of these key variables is provided in the current meta-data summaries, including those sources that were not part of the original inventory.

A more detailed explanation of the terminology used for each of the key variables is provided below.

Identifier variables

This group of variables refers to any unique identifier recorded by a data source, e.g. an identification number, a name, or the combination of both. Such variables allow for the elimination of duplication in datasets and other quality controls, and provide scope for data linkage that can further enhance the analytical potential of a data source. The identifiers were divided into 2 categories:

- » **PPSN** (Personal Public Service Number) – an alphanumeric identifier issued by the Department of Social Protection. The PPSN is an element of an individual's Public Service Identity (PSI), which contains a defined set of personal data for use in public service administration and is of particular interest for statistical purposes.
- » **ID Other** includes any form of unique identifier other than PPSN, e.g. a service user number, a specimen or case reference number, or fields such as surname or name of an organisation.

Sex variable

The **Sex** variable refers to any variable that allows identification of the sex of an individual, most frequently as a binary variable of male or female. Some (usually medical) data sources also include a residual category such as 'unknown' or 'indeterminate'. For a small number of data sources, while sex was not directly coded it could be derived from such titles as Mr/Mrs/Miss. Such data sources have been classified as containing information on this variable.

Age variables

This group of variables refers to any measure of the age of the individual to which the data refer and contains 3 categories:

- » **DOB** (Date of Birth) is used where the respondent is prompted to enter his or her date of birth in the format of day, month and year of birth.
- » **Age Yr** (Age in Years) is used where the respondent is prompted to state his or her age in years, either as current age or age at a specified point in time.
- » **Age Other** includes any measurement of age other than DOB and Age Yr, e.g. year of birth collected alone or year of birth collected with month of birth only.

Address variables

This group of variables refers to the collection of details of address of the individual or organisation to which the data refer. Three levels of geographical coding are used:

- » **County** refers to the collection of details of address at the level of county, usually one of the 32 counties in the Republic of Ireland and Northern Ireland.
- » **Geog Sub County** refers to the collection of details of address at a more localised level than county, e.g. Local Health Office (HSE).
- » **Geog Other** refers to any remaining coding for address. This key variable typically contains regional-level coding composed of a group of counties, e.g. Health Service Executive Area.

Family structure variables

There are 3 variable types included in this group:

- » **Marital Status** refers to variables that enable an examination of the current marital arrangement of the data subject(s).
- » **Family Status** refers to variables that identify the number of dependent children within the family, the relationships between the dependent children and their guardians, and the living arrangements between parents.
- » **Household** refers to variables that identify the structure of the household, e.g. the persons who are typically resident within the household and the relationships between the residents.

Socio-economic variables

This group contains 4 categories of variable that can potentially contribute to composite measurements of socio-economic status:

- » **Occupation** refers to the assignment of the individual's occupation group and in some cases may be combined with their labour force status.
- » **Income** involves any assessment of individual or household income, whether from wages/salary or other sources.
- » **Ed Attain Mother** (Educational Attainment of Mother) refers to the highest level of completed education or highest qualification currently held by the mother in the household. This variable is regarded as a reliable predictor of the child's social and health outcomes. In some cases, this information is not collected directly, but may be derived from the dual collection in a single source of (1) highest level of education and (2) the reference person's relationship to the other members of the household – in other words, where mother-child relationship is recorded.
- » **Labour Force Status** refers to the measurement of the individual's current situation in regard to employment (examples include: in employment, unemployed, in education, retired, on home duties or unable to work due to a disability).

Nationality variable

The **Nationality** variable is collected in a range of formats, including free-form questions or coded questions for which response categories vary and which may be as sparse as 'Irish/Other/None'. Questions on related concepts, such as country of birth, citizenship or country of residence, are also used in some datasets.

Equality variables

This group of variables refers to the grounds on the basis of which discrimination in employment, vocational training, advertising, collective agreements and the provision of goods and services is prohibited in Irish law (Employment Equality Acts 1998-2008 and Equal Status Acts 2000-2008). Four of these variables (i.e. Gender, Marital Status, Family Status and Age) are dealt with above under other groups of variables. The remaining categories are as follows:

- » **Disability** is defined in both the Employment Equality Act 1998 and the Equal Status Act 2000 as:
 - (a) *the total or partial absence of a person's bodily or mental functions, including the absence of a part of a person's body,*
 - (b) *the presence in the body of organisms causing, or likely to cause, chronic disease or illness,*
 - (c) *the malfunction, malformation or disfigurement of a part of a person's body,*
 - (d) *a condition or malfunction which results in a person learning differently from a person without the condition or malfunction, or*
 - (e) *a condition, illness or disease which affects a person's thought processes, perception of reality, emotions or judgement or which results in disturbed behaviour, and shall be taken to include a disability which exists at present, or which previously existed but no longer exists, or which may exist in the future or which is imputed to a person.*

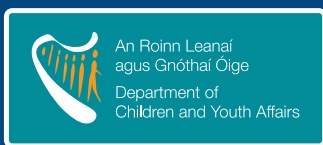
In the current context, this category refers to any variable that elicits information on the presence of a physical, sensory, mental or intellectual disability. It also incorporates measures of multiple or unspecified disability or disabilities.

- » **Race/Ethnicity** is not legally defined in Equality legislation, but in the current exercise is used to refer to the collection of any variable that examines either the racial or ethnic background of the individual.

- » **Traveller** refers to any variable collected with information of membership of the Traveller Community. Equality legislation defines 'Traveller community' as:
The community of people who are commonly called Travellers and who are identified (both by themselves and others) as people with a shared history, culture and traditions, including, historically, a nomadic way of life on the island of Ireland.
- » **Sexual Orientation** is defined in Equality legislation as '*heterosexual, homosexual or bisexual orientation*'. At present, this is collected on only one data source (medical) in the inventory.
- » **Religion** is also defined here in terms of Equality legislation that refers to discrimination on the grounds of 'religious belief', defined in the Equality legislation as a '*religious background or outlook*'. The Census of Population is the only data source in the inventory collecting this information.

References

- DCYA (2011a) *National Strategy for Research and Data on Children's Lives, 2011-2016*, Department of Children and Youth Affairs. Dublin: Government Publications.
Available at: www.dcy.a.ie
- DCYA (2011b) *Report on Public Consultation Processes for National Strategy for Research and Data on Children's Lives 2011-2016*, Department of Children and Youth Affairs.
Available at: www.dcy.a.ie
- NSB (2003a) *Developing Irish Social and Equality Statistics to meet Policy Needs: Report of the Steering Group on Social and Equality Statistics*. Dublin: National Statistics Board.
- NSB (2003b) *Statistical Potential of Administrative Records: An Examination of Data Holdings in Six Government Departments, Working Report*. Dublin: National Statistics Board.
- NSB (2004) *Best Practice Guidelines for the Development and Implementation of formal Data/ Statistics Strategies in Government Departments*. Dublin: National Statistics Board. Available at: www.nsb.ie/pdf_docs/Data_Strategy_Guidelines.pdf



Copyright © Minister for Children and Youth Affairs, 2012

Published by Department of Children and Youth Affairs

43-49 Mespil Road, Dublin 4

Tel: +353 (0)1 647 3000

Fax: +353 (0)1 667 0826

E-mail: contact@dcya.gov.ie

Web: www.dcya.ie

Design: www.penhouse.ie