The International Database on Longevity

Database structure

The IDL data are classified by country. For each country, there are up to two data files available:

- List of individuals dead at ages 110 and older;
- List of individuals dead at ages 105-109 (if available).

These data files are available <u>in two formats</u>:

1. as comma-separated (csv) text files consisting of the complete country data;
2. as html tables that provide only selected columns from the comma-separated data files.

Each record in a data file contains information about one individual; each person included in the IDL has only one record.

The two csv data files have the corresponding names: idl_XXX_dead_110.csv and idl_XXX_dead_105-109.csv; with XXX representing the country code. The html tables are displayed as a part of the country-specific web page.

In addition to the data files, each country-specific dataset contains one metadata file that provides a general overview of the data, as well as a short description of the age validation process used in a particular country. It is essential to take this information into consideration when performing demographic analyses.

<u>The structure of the comma-separated data files.</u>

Each record in the data files describes an individual and contains the following fields:

IDNUMBER (alphanumeric) is the person's unique IDL personal identification number.

AGEYEARS (numeric) is the person's age in completed years (the age at death).

DAYSSINCEBD (numeric) is the number of days since the person's last birthday.

AGEDAYS (numeric) is the person's age in days (the age at death).

SEX (one-character alphabetic) is a code for the person's sex: M stands for male and F stands for female.

BCOUNTRY (three-character alphabetic) is the person's country of birth. The ISO 3166-1 alpha-3 (or equivalent) codes for countries or areas are used. A description of all of the country codes is provided on the [Codes used in the IDL](#) page of the website.

BREGION (alphanumeric) is the person's region of birth. The NUTS-3, ISO 3166-2, or equivalent codes are used for identifying the principal subdivisions (e.g., provinces or states) of countries. A description of all of the regional codes is provided on the [Codes used in the IDL](#) page of the website.

DCOUNTRY (three-character alphabetic) is the person's country of death. The ISO 3166-1 alpha-3 (or equivalent) codes for countries or areas are used. A description of all of the country codes is provided on the [Codes used in the IDL](#) page of the website.

DREGION (alphanumeric) is the person's region of death. The NUTS-3, ISO 3166-2, or equivalent codes are used for identifying the principal subdivisions (e.g., provinces or states) of countries. A description of all of the regional codes is provided on the [Codes used in the IDL](#) page of the website.

BDATE (DD/MM/YYYY) is the person's date of birth.

DDATE (DD/MM/YYYY) is the person's date of death.

VALIDATION (alphabetic) is a code used to specify whether a record was validated. The possible values are: YES, NO, In progress, and Sample Out.

- YES indicates that a validation procedure was successfully completed. This means that an early life document – preferably a birth record, a baptism record, or an early census record – was available and has been checked. Similarly, the date of death of a fully validated case must be backed up by a document, such as a death certificate or an entry in a death index. For several countries, the validation procedure is not documented as thoroughly as the full validation criteria require, but individual cases have been carefully checked. For example, in some cases the local residence register may have officially confirmed a date of death, but did not issue a personal document on the individual. There may also be cases in which the sequence of entries in national censuses had been carefully checked by the respective national statistical office over the course of the individual's life.
- NO indicates that validation was impossible.
- In progress means that the validation process is in progress, and, thus, that the data have not yet been validated, but will be validated in the future.
- Sample Out indicates that a record that is not in the sample is being validated. This implies that the individual was not validated, but a random sub-sample of the population was validated. All cases selected for sample validation (sample-in) should have a *Validation* equal to *Yes* or *No* (depending on the results of the validation), and a *Validation type* equal to *sample*.

VALIDATIONTYPE (alphabetic) is the type of validation: Exhaustive (complete validation; i.e., each record was validated) or Sample (if only a sample was validated).

VALIDATIONYEAR (numeric) is the year when the validation of a record was completed.

SOURCE1 (alphabetic) is the code that indicates the key source of data (for instance, INSEE - Institut National de la Statistique et des Études Économiques, France; ONS - Office for National Statistics, England and Wales; SSA - Social Security Administration, USA). A detailed description of all of the source codes used in the IDL is provided on the [Codes used in the IDL](#) page of the website.

SOURCE2 (alphabetic) is the code that indicates the secondary source of data (if available). A detailed description of all of the source codes used in the IDL is provided on the [Codes used in the IDL](#) page of the website.

SOURCE3 (alphabetic) is the code that indicates the additional data sources (if available). A detailed description of all the source codes used in the IDL is provided on the [Codes used in the IDL](#) page of the website.

BCERTIF (alphabetic) is the code that indicates whether, and, if so, what kind of birth certificate was used for the record's validation (e.g., COPY stands for a copy of the certificate; REGIST stands for

register, ORIG stands for the original of the certificate, etc.). A detailed description of all of the codes used in the IDL is provided on the [Codes used in the IDL](#) page of the website.

DCERTIF (alphabetic) is the code that indicates whether, and, if so, what kind of death certificate was used for the record's validation (e.g., COPY stands for a copy of the certificate, REGIST stands for register, INDEX stands for death index, etc.). A detailed description of all the codes used in the IDL is provided on the [Codes used in the IDL](#) page of the website.

PROOFS (alphabetic) is the code that indicates whether, and, if so, what kind of additional documents were used for the validation of a record. A detailed description of all of the codes used in the IDL is provided on the [Codes used in the IDL](#) page of the website.

CENSUS (alphabetic) is the code that specifies the type of the census information used for validation: record (with individual characteristics, incl. name, date, and place of birth) or list (incl. name, age, or date of birth).

CENSUSYEAR (numeric) specifies the year of the census from which a record was obtained (if applicable).

NOTE (alphanumeric; XXX_##) is a code that links to a specific note. XXX stands for the country code, while ## is the note number. The numeration of notes is consecutive throughout the entire IDL.

UPDATE (numeric) is a code that indicates the number of the update in which a record was included in the IDL (1 indicates data collected during the first update; 2 indicates data added during the second update; and 3 indicates the most recent available data).

All missing values are coded as dot (".") in the csv data files.

The structure of the html file

Only selected information for users is provided in the html data file:

IDNUMBER (alphanumeric) is the person's unique IDL personal identification number.

AGEYEARS (numeric) is the person's age in completed years (the age at death).

DAYSSINCEBD (numeric) is the number of days since the person's last birthday.

AGEDAYS (numeric) is the person's age in days (the age at death).

SEX (one-character alphabetic) is a code for the person's sex: M stands for male and F stands for female.

BCOUNTRY (three-character alphabetic) is the person's country of birth. The ISO 3166-1 alpha-3 (or equivalent) codes for countries or areas are used. A description of all of the country codes is provided on the [Codes used in the IDL](#) page of the website.

DCOUNTRY (three-character alphabetic) is the person's country of death. The ISO 3166-1 alpha-3 (or equivalent) codes for countries or areas are used. A description of all of the country codes is provided on the [Codes used in the IDL](#) page of the website.

BDATE (DD/MM/YYYY) is the person's date of birth.

DDATE (DD/MM/YYYY) is the person's date of death.

VALIDATION (alphabetic) is a code used to specify whether a record was validated. The possible values are: YES, NO, In progress, Sample Out (for the complete description of each value, see the description of the corresponding field above).

Metadata

Metadata files are available for download for every country in pdf format. These files contain information on the original data sources and the data collection process, including information on the selected birth cohorts, the age range of the data, and the time window in which the information was sampled. When available, information on the age validation procedure is also provided in the metadata file. The name of the metadata file is idl_metadata_XXX.pdf, with XXX representing the country code. On the web, the file is displayed on the corresponding country-specific page.

Pooled data file

In addition, one pooled data file is available on the web, consisting of data for all of the countries combined. The structure of this file is identical to the country-specific data files provided in the csv format (see Structure of the comma-separated data files above).