Contents

[Why is the CSO developing new methods? 1](#_Toc531020828)

[How are these population estimates compiled? 1](#_Toc531020829)

[How is the linking done? 1](#_Toc531020830)

[Why do they differ from existing calculations? 1](#_Toc531020831)

[Which is correct? 2](#_Toc531020832)

[In what ways can I use this data? 2](#_Toc531020833)

# Why is the CSO developing new methods?

From 2024 onwards, it is anticipated that all EU member states will provide Census like information at a detailed geographical level on an annual basis.

It would be too costly for Ireland to conduct a traditional Census every year. Therefore, the CSO is actively exploring new ways to meet this need in a cost-effective manner. The research outputs presented here are generated from new methods being explored by the CSO to compile population estimates from administrative data.

# How are these population estimates compiled?

These population estimates are compiled by first creating a Statistical Population Dataset (SPD) from public administration systems and then adjusting the counts in this list using statistical methods to estimate the population size.

In simple terms, when a person engages with one or more public administration systems in the calendar year, a record will be created in the SPD. This is done using a Signs of Life (SoL) approach. A public administration system can be tax based (Revenue), welfare based (DESAP), education based (DES/HEA) or health based (HSE). Once the SPD is compiled for the year, it is adjusted using Dual System Estimation (DSE) methods.

The DSE method involves linking the SPD with a second list to identify the proportion of records missed by the SPD and then simply adjusting the SPD counts accordingly. For example, if the SPD contains 100 records, and when it is linked with another list of 10 records only 8 records are found, then the SPD counts are adjusted upwards by 25% [(10-8)/8] to estimate the population.

# How is the linking done?

Protected Identifier Keys (PIK) are used to link the data. The PIK is an encrypted or protected version of the PPSN that enables CSO to link data across sources and over time without revealing any identification details to the statisticians undertaking the linking.

# Why do they differ from existing calculations?

The project is a research project to explore new methods and data sources for the purposes of compiling population estimates. With this in mind, differences between the PECADO estimates and the population estimates officially published by CSO can be expected.

Differences between the PECADO annual population estimates and Census results can be attributed to:

* differences in the underlying population concepts between the two measures.
  + The Census counts use the population concept usually resident and present at one point in time of the year, on Census night
  + PECADO annual population estimates use a population concept based on usually resident at any time in the calendar year
* weaknesses in the application of the underlying assumptions in the PECADO methodology
* the possibility of erroneous data being included in the lists from the underlying data sources.

# Which is correct?

The Population & Migration Estimates and Census Reports are the officially published data and the correct version to use. The PECADO research outputs are provided for the sole purpose of demonstrating the potential of new methods and new data sources in this area.

# In what ways can I use this data?

The data should not be used as Official Statistics and is for research use only.