



Progress in building a longitudinal framework of linked reference data for England and Wales.

K Loukas¹; P Groom²; S Guggisberg³; B Lisowska⁴

- 1 Office For National Statistics
- 2 Office For National Statistics
- 3 Office For National Statistics
- 4 Office For National Statistics

Abstract:

ONS are developing the methods and data infrastructure to deliver a revised and comprehensive system of population and migration statistics based on administrative data.

We are developing a longitudinal framework to structure and integrate administrative data; indexing data relating to people, business, location and classifications. The core reference data indexes aim to maximise the use and value of administrative data for statistical purposes within ONS where data standards, security and consistency are designed in from the start. Assigning index ID's to data allows personally identifiable information to be removed without impacting the usability of the data. The framework enables analysis on data with links between people, businesses and locations.

The demographic index requires the integration of multiple administrative sources where data need to be standardised, cleaned and structured to enable high quality linkage. As there is no unique person identifier available across the sources of administrative data both deterministic and probabilistic linking methods are used to build the index. The design of the index and how it is updated and maintained over time must meet ONS statistical research needs whilst also adhering to data warehousing standards and keeping personal information safe. Additionally we aim to develop an indexing function enabling users to easily retrieve a wealth of connected reference data through automatic indexing of new data to the framework.

Integrating administrative data has challenges where attributes may not align across sources. We are investigating methods to reconcile the data enabling a single view of some of the key aggregating characteristics of a person. For example this will enable outputs to count a person against only a single geography at any point in time.

Keywords:

administrative data; data linking; timeseries

1. Introduction:

<Introduction>

2. Methodology:

< Methodology>

3. Result:

<Result>

4. Discussion and Conclusion:

<Discussion and Conclusion>

References:

- 1.
- 2.
- 3.

**NOTE: THE MAXIMUM NUMBER OF PAGES
FOR THE PAPER IS SIX PAGES**