

TTP DEMOGRAPHIC DATA QUALITY REPORT

DATA SOURCE

DATA SUPPLIER:	Department of Education (DE)
DESCRIPTION:	School registration data for children in Primary and Special Pupil education for academic years 2006/07 to 2013/14, and registrations for Post-Primary education from 2007/08 to 2013/14
COVERAGE:	Northern Ireland
NUMBER OF RECORDS:	2,395,568 records in total made up of the following- Average 146,852 per academic year in Post-Primary Education (total = 1,027,967) Average 166,363 per academic year in Primary Education (total = 1,330,900) Average 4588 per academic year in Special Pupil Education (total = 36,701)

DATA BACKGROUND

Information on pupil enrolments is collected annually from schools, as part of what is commonly known as the schools census. The data collection exercise is an annual snapshot of pupil and school level data for each pre-school centre, nursery, special, primary, post-primary, hospital and independent school in Northern Ireland. As well as data relating to school enrolments and numbers, information in relation to attendance is collected as part of the overall school return. The date of the school census is usually the Friday of the first full working-week in October. For primary, post-primary and special schools, schools upload a version of the data they hold on their administrative system, SIMS, as of census day.

Data is collected via both electronic data interchange and via paper returns and relates mainly to pupil characteristics and attributes such as year group, religion, ethnicity, special needs, etc.

The content of the census is defined by DE information needs. The relevance of this is checked each year against any changes requested by DE branches and the Education and Library Boards and subsequently the information requested can differ with each annual return.

All schools are legally required to return a fully completed school census download under the Education and Libraries (NI) Order 2003 Article 37. A number of validation checks are built into the reporting process in schools, so that they can verify and correct data before it is submitted. Once data is collected, it is subjected to a further set of checks, including verification against ELB information systems for items such as free school meals entitlement.

DEMOGRAPHIC INFORMATION

Information available to the TTP for linkage purposes -

- Unique Pupil Number (UPN)
- Name information of the pupil which includes Forename, Midname and Surname
- Address information of the pupil
- Address information of the school which the pupil attends
- Gender
- Date of birth

DATA QUALITY & PRE-PROCESSING

Record linkage is highly dependent on the quality of the data being linked. Many key variables can be presented quite differently between datasets as the information has been collected and recorded using different methods. This can greatly complicate record linkage unless understood ahead of time. Pre-processing can be used to standardise these variables into a consistent format prior to linkage.

UPN – The Unique Pupil Number is used to link students across multiple academic years. The collection of this information was introduced in 2007 and therefore will not be available for any year prior to this. The coverage is generally very high, however, it should be noted that in 2007, 2008, 2009 and 2011 there were a small number of records with missing UPN (on average 0.06%). Department of Education have attributed this to the late installation of SIMS in new schools at the time of the Census resulting in pupils in those schools being assigned a UPN later in the academic year.

NAME INFORMATION- Name information is not available for pupils in academic years 2006/07, 2007/08 and 2008/09. It is available for academic years 2009/10, 2010/11, 2011/12 and 2012/13 with an average coverage of 96.8% and is fully populated for academic year 2013/14. Midname information is provided where appropriate.

- In some instances multiple names will be recorded as the Forename. This can be edited to leave the first name that appears as the Forename and any additional names as the Midname;
- Punctuation and characters which do not appear in the English alphabet (e.g. Á , % , \$...) can also appear in the name information. A function can be used to replace these with usable characters;

ADDRESS INFORMATION OF THE PUPIL- In total 8% of pupils have a home postcode which is outside Northern Ireland, incomplete or missing.

- Postcode information is standardised prior to linkage.

ADDRESS INFORMATION OF THE SCHOOL WHICH THE PUPIL ATTENDS- All records have address and postcode of the school which the pupil attends

- Postcode information is standardised prior to linkage.

GENDER- All records contain the gender of the pupil.

- Gender information should be standardised prior to linkage.

DATE OF BIRTH- All records contain the Date of Birth of the pupil.

- Date of birth can be parsed to into its constituent parts (day, month and year) prior to linkage

DUPLICATES

The school census data detailed above was analysed by the Census Administrative Data team in NISRA. Through their investigations a number of duplicates were identified-

- 473 UPNs appeared twice within the same academic year with 17 of these occurring after 2009. All but one of these 473 related to the same UPN being used for students with different dates of birth. In many cases name information was not available.
- 1585 students were identified who appeared to have been allocated more than one UPN. In some instances this has been caused by a student attending 2 schools within the same academic year.

MATCHING METHODS

Using the demographic information provided the following matching can be undertaken -

- Person matching of the pupil for academic years 2009/10 onwards
- Address matching of the pupils home address for all academic years
- Address matching of the school which the pupil attends

PREVIOUS LINKAGE MATCH RATES

School Census information for 2009 - 2013 has previously been linked to BSO Health Card Registration Data from 2011 as part of the Census Administrative Data Project and produced an adjusted match rate of 97%. More information can be found at [The Future Provision of Census of Population Information for Northern Ireland](#)

EVALUATION

The demographic data from the school census is generally of good quality and can be used for matching without extensive pre-processing.

It should be noted that when matching to other Northern Ireland specific datasets we will be unable to match a small number of records which relate to students who have a home address outside of Northern Ireland but who attend an educational institute in NI (approximately 0.15%).