

Longitudinal Studies Centre - Scotland

Home of the Scottish Longitudinal Study



Linkage of school education data to the Scottish Longitudinal Study

Chris Dibben - cjld@st-andrews.ac.uk

Claire Boag – Claire.Boag@gro-scotland.gsi.gov.uk

linking lives through time

www.lscs.ac.uk

E · S · R · C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

Requirement for education data

- Using the rich information from the census data that we hold to learn more about the household characteristics that influence educational attainment.
- Following pupils from school into higher education and work. We can begin to untangle the effects of schooling versus the effects of the environmental and social background on the subsequent occupational development of young adults.
- To use the educational information as useful explanatory variables in a range of studies. Currently, the census data only include quite limited information about the level of qualification achieved. We might be interested in relating more detailed educational information to outcomes such as health status or life expectancy.

Background – legal basis for sharing data

- **Section 5 of the 1920 Census Act states:**

“It shall be the duty of the Registrar-General from time to time to collect and publish any available statistical information with respect to the number and condition of the population in the interval between one census and another, and otherwise to further the supply and provide for the better co-ordination of such information, and the **Registrar-General may make arrangements with any Government Department or local authority** for the purpose of acquiring any materials or information necessary for the purpose aforesaid.”

- Ability of the Scottish Government to gather education data covered through the 1980 Education Act
- Scottish Government’s ability to use the collected data for research and statistical purposes covered through the Data Protection Act

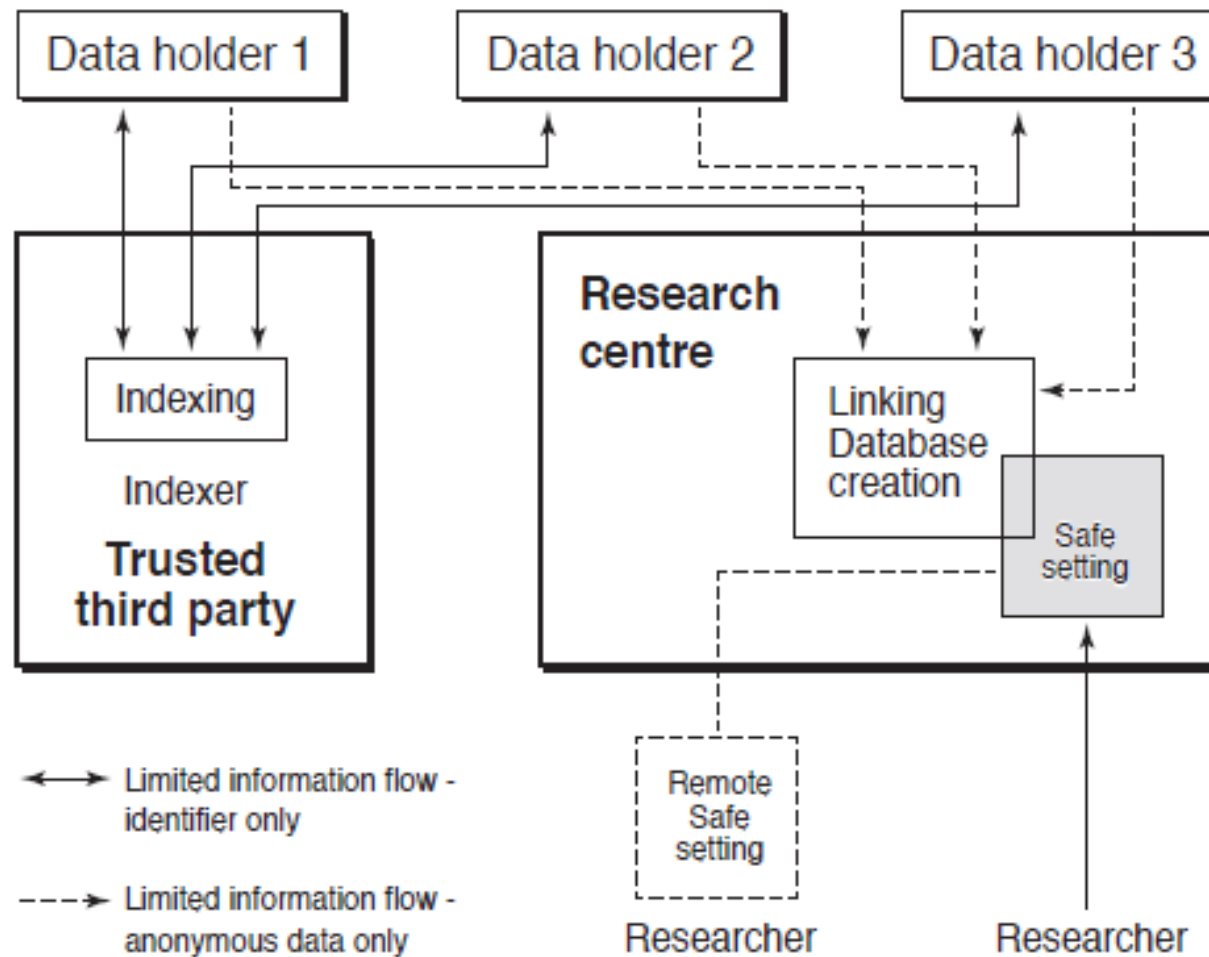
Background - Access procedure

SLS data linkage proposal approved by:

- **Privacy Advisory Committee** - PAC advises NRS on the correct balance between protecting personal data and making data available for research, and helps to ensure that any information releases are carefully controlled.
- **SG Education Analytical Services Data Access Panel.** Approval by the DAP dependent on:
 - Whether EAS Division has the right to share data,
 - Whether the use of the information is consistent with the original purposes for collecting the data,
 - Assurance that no individuals will be publicly identified as a result of the analyses,
 - appropriate safeguards and checks in place to ensure the security of the data throughout its exchange and use.

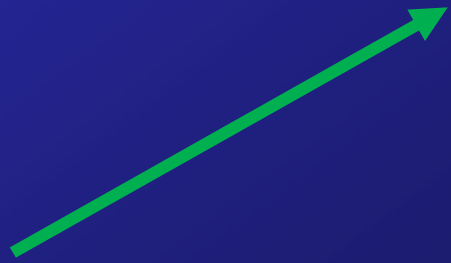
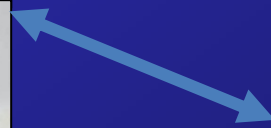
Once approvals in place, Data Access Agreement between Registrar General, Scottish Ministers, and St Andrews University prepared and signed off. This allows for ongoing annual linkages of education data to the SLS

Privacy-ethical advisory bodies



Scottish Longitudinal Study

Research centre



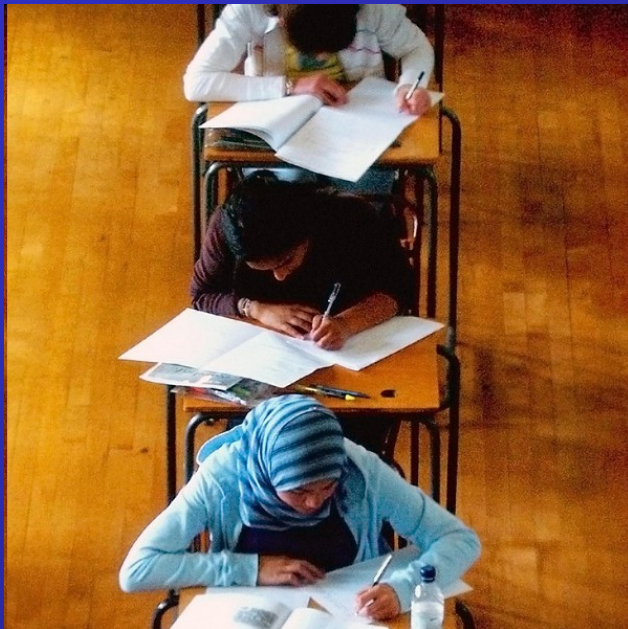
Data holder

TTP indexer

School Education data

– 3 datasets:

- School Census data
- SQA attainment data
- Attendance/ absence & exclusions data available



Education data linked to the SLS

1. School Census variables

- Local Authority Code
- SEED Code (school identifier)
- Scottish Candidate Number (SCN)
- Student postcode
- Student gender
- Student stage
- Student date of birth
- Scottish candidate number
- Admission date
- Ethnic background
- Main difficulty in learning
- National identity
- Free school meals registered
- Individual educational programme (IEP)
- Records of needs (RON)
- Student mainstream integration
- Student attendance – special schools/units
- School type
- Class name
- Class type (size) (primary school only)
- Classes with 2 or more teachers present
- English as a second language

Education data linked to the SLS

2. Attendance/absence & Exclusions

Attendance/Absence

- Scottish candidate number
- Student gender
- Student stage
- Term
- Attendance/absence code
- Value (number of 1/2 days for each code)

Exclusions

- Scottish candidate number
- Student gender
- Student stage
- Student reason for/factor in exclusion
- Removed from register
- Student exclusion start date
- Student exclusion end date
- Student exclusion appeal
- Number of days with no alternative provision of education
- Temporary alternative provision found at ...

Education data linked to the SLS

3. SQA data on attainment and qualifications

Candidate data

- SCN
- Date of birth
- Gender
- Post code

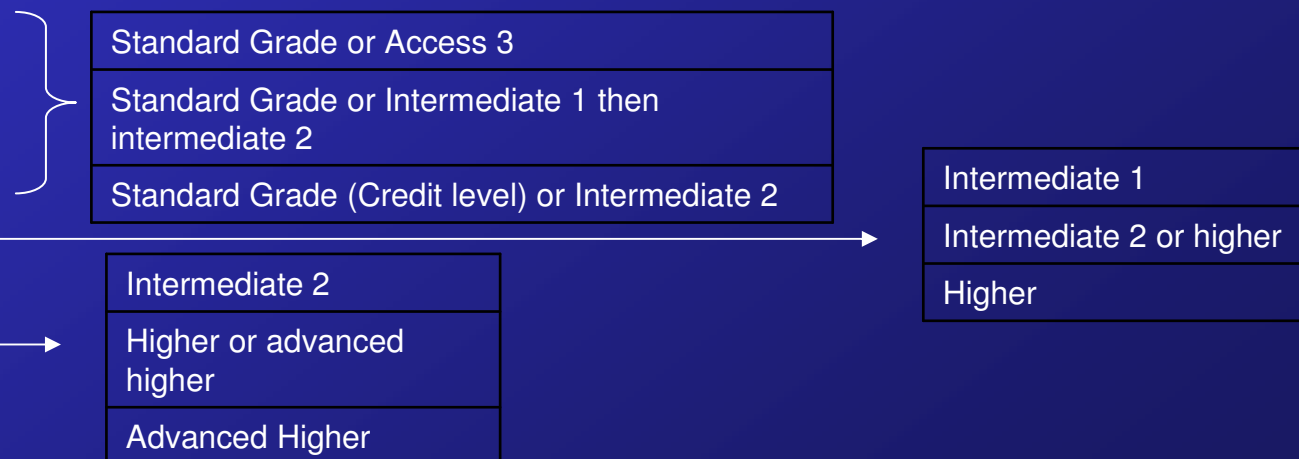
Attainment

- SCN
- Qualification
- Level
- Result
- Result date
- Centre
- Entry status
- Entry stage

Qualifications

- Qualification code
- Qualification name
- Type of qualification
- Level
- Credit value
- Qualification start date
- Qualification end date

| | Age at start of school year |
|-----------|-----------------------------|
| Primary 1 | 4 - 5 |
| Primary 2 | 5 - 6 |
| Primary 3 | 6 - 7 |
| Primary 4 | 7 - 8 |
| Primary 5 | 8 - 9 |
| Primary 6 | 9 - 10 |
| Primary 7 | 10 - 11 |
| S1 | 11 - 12 |
| S2 | 12 - 13 |
| S3 | 13 - 14 |
| S4 | 14 - 15 |
| S5 | 15 - 16 |
| S6 | 16 - 17 |



Linkage methods

1. Provision of data to SLS team

- Names and addresses not held within education dataset (ED).
- The residential postcode, date of birth and gender used for data matching.
- SG EAS provide the SLS with datasets for *all Scotland*
- The SLS team split the datasets into the SLS birth date sample and non-SLS birth date sample.
- Non SLS birth date sample deleted

Linkage methods

2. Matching of data against NHSCR

- Linkage file for SLS sample only passed to NHSCR.
- NHSCR run exact match of linkage file against their database using dob, sex and postcode.
- NHSCR attach SLS numbers to the pupils identified through the exact matching and pass the SLS number and the Scottish Candidate Number (pupil ID used in pilot) back to the SLS.
- SLS check unmatched pupils against the SLS 2001 census data. (This could only link those pupils who did not move between the 2001 census and the school census.)
- Remaining unmatched pupils lost to the next annual run (when another matching attempt would be made).

Linkage methods

3. Addition of data to the SLS database

- SLS link the SLS number returned by NHSCR to the main education datasets using the SCN.
- After the initial run, annual updates of the ED are required:
 - The linkage process will only be re-run for new pupils, and those that failed to match the previous year.
 - Data for pupils already matched would be added to the SLS database (this would include a new year of exam results for each child etc.).
- See handout for diagram of linkage methods (based on pilot exercise using 2007 & 2008 data)

Pilot Exercise

- Methods – as previous slides but only the linkage files provided by SG EAS
- Purpose:
 1. to test the accuracy of matching process using 2 separate years of data (2007 and 2008 school census)
 2. to check the matching consistency over time:
 - Whether those who successfully linked in 2007 also linked in 2008 (and to the same person)
 - Whether those who did not link in 2007 did in 2008 and vice versa

Part 1 – linkage exercise for 2007 and 2008

- Two linkage exercises were carried out, to separately link 2007 and 2008 schools data to the SLS, using dob, sex and postcode as the identifying variables.
- The overall linkage rates were found to be 80.5% in 2007, and 83.0% in 2008.
- i.e. 80.5% of school census records successfully and uniquely matched to an SLS member on the NHSCR database.

Table 1 – Identifying SLS members and passing them to NHSCR for matching

| Description | 2007 | | 2008 | |
|--------------------------------------|---------|---------|---------|---------|
| Records From SG | 695,208 | 100.00% | 684,721 | 100.00% |
| Identified As SLS Members | 38,449 | 5.53% | 37,813 | 5.52% |
| Duplicate Entries | 70 | 0.01% | 88 | 0.01% |
| Records Passed To NHSCR For Matching | 38,379 | 5.52% | 37,725 | 5.51% |

Table 2 – Matching SLS members at NHSCR

| Description | 2007 | | 2008 | |
|-----------------------------------|--------|---------|--------|---------|
| Records From SLS For Matching | 38,379 | 100.00% | 37,725 | 100.00% |
| Matched Records | 32,011 | 83.41% | 32,123 | 85.15% |
| Unmatched Twins | 1,164 | 3.03% | 1,158 | 3.07% |
| Unmatched Triplets | 39 | 0.10% | 30 | 0.08% |
| Unmatched Other | 5,165 | 13.46% | 4,414 | 11.70% |
| Records Returned To SLS | 38,379 | 100.00% | 37,725 | 100.00% |
| ...of which have unique pupil IDs | 38,353 | 99.93% | 37,713 | 99.97% |

Table 3 – Linkage rates and reasons for failure

| Description | 2007 | | 2008 | |
|---|--------|---------|--------|---------|
| Records (with unique pupil ids) from NHSCR after matching | 38,353 | 100.00% | 37,713 | 100.00% |
| STATUS = 1: The pupil was matched at NHSCR and the SLS and their education data is available to researchers for analysis. | 30,890 | 80.54% | 31,300 | 83.00% |
| STATUS = 2: The pupil id has been duplicated and so cannot be used to uniquely link to an SLS member. | 49 | 0.13% | 12 | 0.03% |
| STATUS = 3: The pupil was matched at NHSCR but the SLS number has been duplicated so cannot be used to uniquely link to an SLS member. | 55 | 0.14% | 40 | 0.11% |
| STATUS = 4: The pupil was matched at NHSCR but not the SLS. | 992 | 2.59% | 759 | 2.01% |
| STATUS = 5: The pupil was matched at NHSCR and the SLS but with a different dob on both systems. | 17 | 0.04% | 10 | 0.03% |
| STATUS = 6: The pupil did not match at NHSCR. | 6,350 | 16.56% | 5,592 | 14.83% |

Comments on linkage quality issues

- **Duplicate records from ScotXed (see table 1)**
 - 70 in 2007, 88 in 2008.
 - pupil has more than one entry on the School Census system
 - records appeared as duplicates because only few variables provided

- **Duplicate pupil ID (status = 2)**
 - 70 in 2007, 18 in 2008
 - may represent different pupils who have been allocated the same pupil id

- **Failure to match at NHSCR**
 - This method cannot match same sex multiple births. 3% of records failed to match for this reason
 - The majority of other non-matches caused by postcode differences between NHSCR and the education data

Comments on linkage quality issues

- **Pupils matched at NHSCR but not in the SLS (status = 4)**
 - 992 in 2007, 759 in 2008.
 - **The majority of these are not flagged as an SLS member on NHSCR - could be new entries to the SLS (new births, immigrants, cross border moves) not picked up through VE processing**
 - Small number where SLS number on NHSCR contains typo so cannot be matched to SLS database

- **Bias in unmatched records**
 - the extent to which there is any bias in the unmatched records cannot easily be assessed
 - Known issue with twins and triplets, but are certain groups of pupils, schools etc over or under-represented among the unlinked pupils?

Part 2 - Comparison of 2007 and 2008 linkage results

The comparison covered the following:

- Did pupils who failed to match in 2007 find a match in 2008?
- Did pupils who matched in 2007 fail to match in 2008?
- Did pupils who matched in 2007 and 2008 match to the same SLS member?
- Did pupils who matched successfully in 2007 encounter matching problems in 2008?
- Did pupils who encountered matching problems in 2007 match successfully in 2008?

Pupils only in 2007 or 2008 linkage file

- 4,130 pupils on the 2007 file only and 3,490 pupils on the 2008 file only. These will include pupils entering and exiting the school system. They were matched as follows:

| Description | 2007 | 2008 |
|--|-------|-------|
| The pupil was matched at NHSCR and the SLS and their education data is available for analysis. | 3,021 | 2,566 |
| The pupil id has been duplicated therefore cannot be used to uniquely link to an SLS member. | 28 | 3 |
| The pupil was matched at NHSCR, but the SLS number has been duplicated therefore cannot be used to uniquely link to an SLS member. | 11 | 6 |
| The pupil was matched at NHSCR, but not at the SLS. | 169 | 92 |
| The pupil was matched at NHSCR and the SLS, but with a different dob at the SLS. | 9 | 1 |
| The pupil did not match at NHSCR. | 892 | 822 |

Pupils present in both years

- 26,928 pupils (78.7%) were successfully linked in both years
- 933 (2.7%) linked in 2007, but unmatched at NHSCR in 2008
- 1,642 (4.7%) not matched at NHSCR in 2007, but were successfully linked in 2008
- 152 (0.4%) matched at NHSCR in 2007 but not at the SLS, but were subsequently successfully linked in 2008
- Overall, of the 34,233 pupils in both the 2007 and 2008 linkage files, 81.4% of pupils were successfully linked in 2007, and an additional 5.6% of pupils were successfully linked based on their identifying information for 2008.
- Each successive year of additional linkage information will increase the percentage of linked pupils further
- each pupil has up to 13 chances to be successfully linked, assuming they remain in state education in Scotland from P1 through S6

Next Steps

- 2009 and 2010 linkages have been carried out, and are undergoing quality assurance.
- Data will be made available to external researchers around July 2012.
- Two current SLS users plan to incorporate the data into their projects:
 1. A scoping study investigating low birth weight and its impact on child development, linking maternity, child surveillance and education data.
 2. Extension to existing project, following SLS members born 1992-1995 to 2011 (aged 16-19), exploring whether individual, family, school (truancy, free school meal, exclusion), and local area factors are related to the risk of becoming NEET.
- Investigate potential for linking independent sector education data