The construction and value of a school database

Ellen Greaves

Motivation:

Changes in school identifiers due to changes in school structure
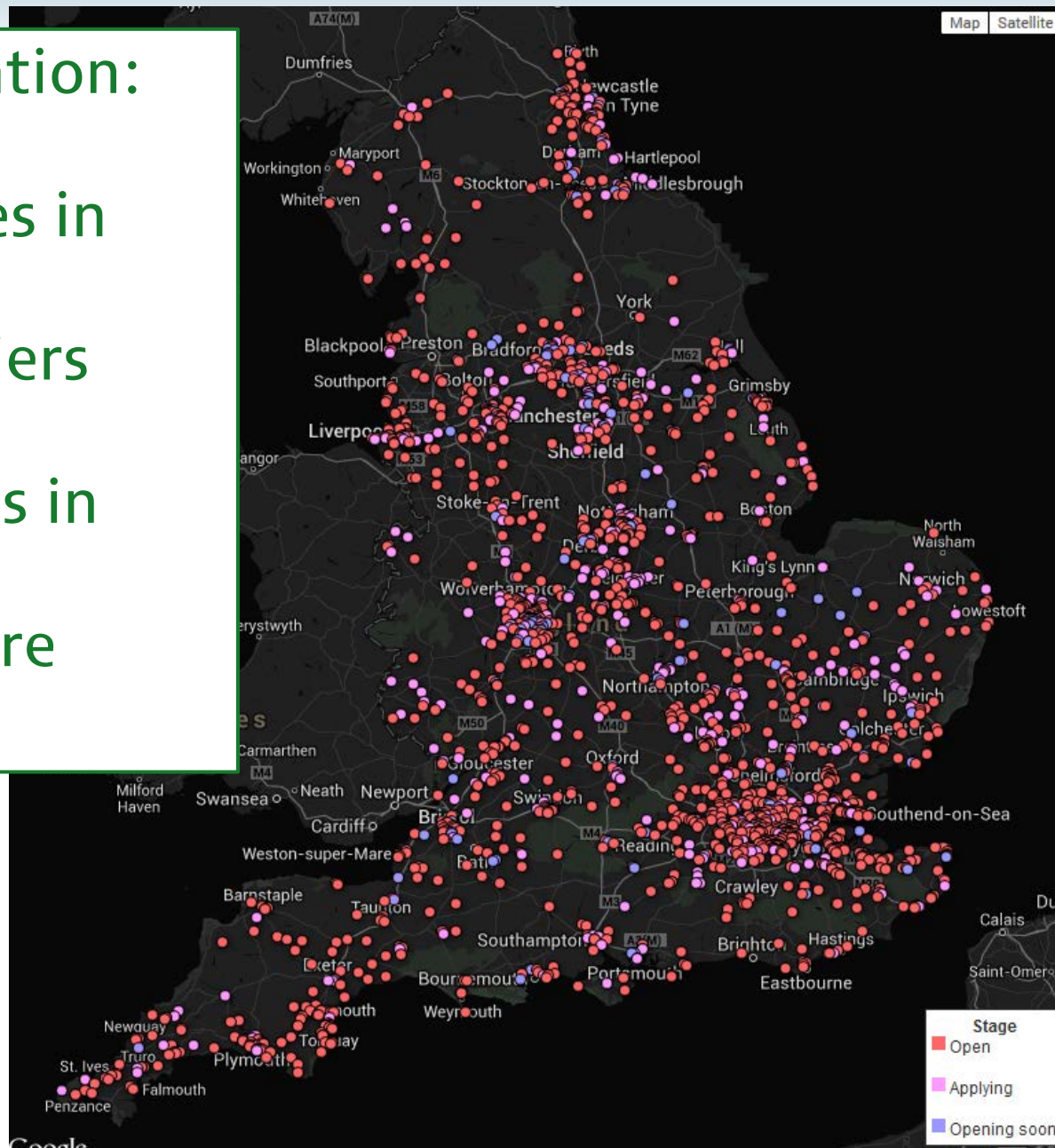

Stage
- Open
- Applying
- Opening soon

Institute for Fiscal Studies

Changing school identifiers: the problem

- Empirical research:

  – How to construct a consistent panel of schools when Unique Reference Numbers (URNs) change?

  – How to adequately control for school-level prior attainment when URNs change?

- Practical work of education charities:

  – How to maintain a list of eligible schools?

  – How track attainment of schools involved in program?

Changing school identifiers: the problem

- Example 1 (we've probably all had...):

```
. use "P:\NCRM MCS NPD\Stata\Data\leasis_2012.dta", clear

. rename urn_2012 urn_2013

. qui merge 1:1 urn_2013 using "P:\NCRM MCS NPD\Stata\Data\leasis_2013.dta"

. tab _merge

           _merge |      Freq.     Percent        Cum.
------------------+-----------------------------------
   master only (1)|      1,253        4.90        4.90
   using only  (2)|      1,236        4.83        9.73
      matched   (3)|     23,092       90.27      100.00
------------------+-----------------------------------
            Total |     25,581      100.00

. list urn_2013 in 1/300 if _me==1, clean

          urn_2013
   78.      100083
   86.      100093
  151.      100188
  239.      100323
  273.      100364
  275.      100367
```

Some URNs only in 2012

Some URNs only in 2013

Only this school actually closes between 2012 and 2013

Institute for Fiscal Studies

Changing school identifiers: the problem

- Example: what is the impact of changing to academy on KS2 Average Points Scores?

- Caveat: simple example - not causal

	Ignore changing URNs	Account for changing URNs
Change to academy between 2010/2011 and 2011/2012	1.054***	0.782***
Standard error	[0.190]	[0.190]
N (schools)	14,413	14,879

Note: OLS regression also accounts for prior APS decile (2010/2011) and the interaction between prior APS decile and change to academy status.

Institute for Fiscal Studies

Changing school identifiers: the problem

- Example 2: linking information from the Millennium Cohort Study (MCS) Teacher Survey (Year 2 and Year 6) to the National Pupil Database:

  - Project currently ongoing to answer research questions:

    1. What is the impact of school practices such as setting/streaming on the distribution of ability in the cohort?

    2. What is the impact of a teacher's qualification and experience on the distribution of ability in the cohort?

  - Must anonymise URN and LAESTAB (Local Authority and Establishment Number) so that schools with MCS cohort members can't be identified

  - Once schools are anonymised, how do we tell whether pupils change schools over time, or whether schools change identifiers over time?

    ➢ Create links between schools before anonymisation, and create binary indicators for whether the pupil changed school in each year

Institute for
Fiscal Studies

Changing school identifiers: a solution (1)

- DfE parent-child linkages (EduBase or SLD - Institution History):

urn	linkurn	linkname	linktype	linkes~e
100016	132245	Kingsgate Primary School	Child	
100017	132245	Kingsgate Primary School	Child	
132245	100017	Kingsgate Infant School	Parent	
132245	100016	Kingsgate Junior School	Parent	

urn	linkurn	linkname	linktype	linkesta~e
100188	138449	Harris Academy Greenwich	Child	31-08-2012
138449	100188	The Eltham Foundation School	Parent	02-05-2013

Two schools merge to become one

School becomes an academy

Institute for
Fiscal Studies

Changing school identifiers: a solution (1)

- DfE parent-child linkages (EduBase or SLD - Institution History):

  1. Reformat DfE parent-child data to become "Child" only (see Appendix 1)

  2. Merge to existing data of interest

  3. Merge to second data of interest, using urn_old

  4. Merge to second data of interest, using urn_new where urn_old does not produce a match

- BUT:

  – Date of parent-child link is often missing (so can't verify when change happened)

  – There can be multiple linkages (in different years or where year isn't known)

  – Very few linkages before 2006 (as recording this data started then)

Institute for
Fiscal Studies

Changing school identifiers: a solution (2)

- Ad-hoc approach:

  1. Look up school link in appropriate year in EduBase

  2. Recode URN manually in syntax

  3. Merge primary and secondary data using recoded URN

- BUT:

  – Feasible for only small numbers of schools

  – Very few linkages before 2006 (as recording this data started then)

Institute for
Fiscal Studies

Changing school identifiers: a solution (3)

- Use **database of schools** to merge URNs in adjacent years:

  – Currently contains school links between 2005 and 2013 (to cover the Millennium Cohort Study cohort)

  – Will update to include previous academic years

  – Will be made publicly available (dataset and syntax)

  – Offers to check syntax are welcome (thank you George!)

Institute for
Fiscal Studies

The database of schools: stable schools

urn	urn_2005	urn_2006	urn_2007	urn_2008	urn_2009	urn_2010	urn_2011	urn_2012	urn_2013
100000	100000	100000	100000	100000	100000	100000	100000	100000	100000
100001	100001	100001	100001	100001	100001	100001	100001	100001	100001
100002	100002	100002	100002	100002	100002	100002	100002	100002	100002
100003	100003	100003	100003	100003	100003	100003	100003	100003	100003
100005	100005	100005	100005	100005	100005	100005	100005	100005	100005
100006	100006	100006	100006	100006	100006	100006	100006	100006	100006
100007	100007	100007	100007	100007	100007	100007	100007	100007	100007
100008	100008	100008	100008	100008	100008	100008	100008	100008	100008
100009	100009	100009	100009	100009	100009	100009	100009	100009	100009
100010	100010	100010	100010	100010	100010	100010	100010	100010	100010
100011	100011	100011	100011	100011	100011	100011	100011	100011	100011
100012	100012	100012	100012	100012	100012	100012	100012	100012	100012
100013	100013	100013	100013	100013	100013	100013	100013	100013	100013
100014	100014	100014	100014	100014	100014	100014	100014	100014	100014
100015	100015	100015	100015	100015	100015	100015	100015	100015	100015
100018	100018	100018	100018	100018	100018	100018	100018	100018	100018
100019	100019	100019	100019	100019	100019	100019	100019	100019	100019
100020	100020	100020	100020	100020	100020	100020	100020	100020	100020
100021	100021	100021	100021	100021	100021	100021	100021	100021	100021
100022	100022	100022	100022	100022	100022	100022	100022	100022	100022
100023	100023	100023	100023	100023	100023	100023	100023	100023	100023
100024	100024	100024	100024	100024	100024	100024	100024	100024	100024
100025	100025	100025	100025	100025	100025	100025	100025	100025	100025
100026	100026	100026	100026	100026	100026	100026	100026	100026	100026
100027	100027	100027	100027	100027	100027	100027	100027	100027	100027
100028	100028	100028	100028	100028	100028	100028	100028	100028	100028
100029	100029	100029	100029	100029	100029	100029	100029	100029	100029
100030	100030	100030	100030	100030	100030	100030	100030	100030	100030
100031	100031	100031	100031	100031	100031	100031	100031	100031	100031

The database of schools: changing schools

urn	urn 2005	urn 2006	urn 2007	urn 2008	urn 2009	urn 2010	urn 2011	urn 2012	urn 2013
130388	126175	126175	130388	130388	130388	130388	130388	130388	130388
126175	126175	126175	130388	130388	130388	130388	130388	130388	130388
126341	126341	126341	130388	130388	130388	130388	130388	130388	130388
130391	.	130391	130391	130391	130391	130391	130391	130391	130391
130826	.	130826	130826	130826	130826	130826	130826	130826	130826
130854	.	130854	130854	130854	130854	130854	130854	130854	130854
130855	.	130855	130855	130855	130855	130855	130855	130855	130855
130856	100669	130856	130856	130856	130856	130856	130856	130856	130856
100670	100670	130856	130856	130856	130856	130856	130856	130856	130856
100669	100669	130856	130856	130856	130856	130856	130856	130856	130856
130857	.	130857	130857	130857	130857	130857	130857	130857	130857
125489	125489	125489	125489	125489	125489	130869	130869	130869	130869
130880	.	130880	130880	130880	130880	130880	130880	130880	130880
102877	102877	130899	130899	130					130899
102880	102880	130899	130899	130					130899
130899	102877	130899	130899	130					130899
114883	114883	130901	130901	130					130901
130901	114880	130901	130901	130					130901
114880	114880	130901	130901	130					130901
130902	.	130902	130902	130					130902
130904	.	130904	130904	130					130904
130907	.	130907	130907	130					130907
130908	111772	130908	130908	130					130908
111772	111772	130908	130908	130					130908
107462	107462	130909	130909	130					130909
130909	107462	130909	130909	130					130909
101150	101150	101150	130912	130					130912
130912	101150	101150	130912	130					130912
130913	.	130913	130913	130913	130913	130913	130913	130913	130913

Note: multiple rows per school, as linked forward and backward and include school mergers.
(Churchfields, the village school formed from Atworth Primary School and Monkton Farleigh & South Wraxall.)

Institute for Fiscal Studies

The database of schools: changing schools

urn	urn_2005	urn_2006	urn_2007	urn_2008	urn_2009	urn_2010	urn_2011	urn_2012	urn_2013
130388	126175	126175	130388	130388	130388	130388	130388	130388	130388
126175	126175	126175	130388	130388	130388	130388	130388	130388	130388
126341	126341	126341	130388	130388	130388	130388	130388	130388	130388
130391	.	130391	130391	130391	130391	130391	130391	130391	130391
130826	.	130826	130826	130826	130826	130826	130826	130826	130826
130854	.	130854	130854	130854	130854	130854	130854	130854	130854
130855	.	130855	130855	130855	130855	130855	130855	130855	130855
130856	100669	130856	130856	130856	130856	130856	130856	130856	130856
100670	100670	130856	130856	130856	130856	130856	130856	130856	130856
100669	100669	130856	130856	130856	130856	130856	130856	130856	130856
130857	.	130857	130857	130857	130857	130857	130857	130857	130857
125489	125489	125489	125489	125489	125489	130869	130869	130869	130869
130880	.	130880	130880	130880	130880	130880	130880	130880	130880
102877	102877	130899	130899	130899	130899	130899	130899	130899	130899
102880	102880	130899	130899	130899	130899	130899	130899	130899	130899
130899	102877	130899	130899	130899	130899	130899	130899	130899	130899
114883	114883	130901	130901	130901	130901	130901	130901	130901	130901
130901	114880	130901	130901	130901	130901	130901	130901	130901	130901
114880	114880	130901	130901	130901	130901	130901	130901	130901	130901
130902	.	130902	130902	130902					130902
130904	.	130904	130904	130904					130904
130907	.	130907	130907	130907					130907
130908	111772	130908	130908	130908	130908	130908	130908	130908	130908
111772	111772	130908	130908	130908	130908	130908	130908	130908	130908
107462	107462	130909	130909	130909	130909	130909	130909	130909	130909
130909	107462	130909	130909	130909	130909	130909	130909	130909	130909
101150	101150	101150	130912	130912	130912	130912	130912	130912	130912
130912	101150	101150	130912	130912	130912	130912	130912	130912	130912
130913	.	130913	130913	130913	130913	130913	130913	130913	130913

Andalusia Academy Bristol opened in September 2005

Institute for Fiscal Studies

The database of schools: changing schools

urn	urn_2005	urn_2006	urn_2007	urn_2008	urn_2009	urn_2010	urn_2011	urn_2012	urn_2013
130388	126175	126175	130388	130388	130388	130388	130388	130388	130388
126175	126175	126175	130388	130388	130388	130388	130388	130388	130388
126341	126341	126341	130388	130388	130388	130388	130388	130388	130388
130391	.	130391	130391	130391	130391	130391	130391	130391	130391
130826	.	130826	130826	130826	130826	130826	130826	130826	130826
130854	.	130854	130854	130854	130854	130854	130854	130854	130854
130855	.	130855	130855	130855	130855	130855	130855	130855	130855
130856	100669	130856	130856	130856	130856	130856	130856	130856	130856
100670	100670	100856	100856	100856	100856	130856	130856	130856	130856
100669						130856	130856	130856	130856
130857						130857	130857	130857	130857
125489						130869	130869	130869	130869
130880						130880	130880	130880	130880
102877						130899	130899	130899	130899
102880						130899	130899	130899	130899
130899						130899	130899	130899	130899
114883						130901	130901	130901	130901
130901	114880	130901	130901	130901	130901	130901	130901	130901	130901
114880	114880	130901	130901	130901	130901	130901	130901	130901	130901
130902	.	130902	130902	130902	130902	130902	130902	130902	130902
130904	.	130904	130904	130904	130904	130904	130904	130904	130904
130907	.	130907	130907	130907	130907	130907	130907	130907	130907
130908	111772	130908	130908	130908	130908	130908	130908	130908	130908
111772	111772	130908	130908	130908	130908	130908	130908	130908	130908
107462	107462	130909	130909	130909	130909	130909	130909	130909	130909
130909	107462	130909	130909	130909	130909	130909	130909	130909	130909
101150	101150	101150	130912	130912	130912	130912	130912	130912	130912
130912	101150	101150	130912	130912	130912	130912	130912	130912	130912
130913	.	130913	130913	130913	130913	130913	130913	130913	130913

One for one change: North Westminster Community School became Paddington Academy

Institute for Fiscal Studies

The database of schools: value in the MCS cohort

Change in Year group / academic year		URN changes		Indicator from school database	
		N	%	N	%
Reception to Year 1	2005/2006-2006/2007	38,633	7.47	34,122	6.59
Year 1 to Year 2	2006/2007-2007/2008	39,197	7.41	33,597	6.35
Year 2 to Year 3	2007/2008-2008/2009	125,418	23.60	121,027	22.77
Year 3 to Year 4	2008/2009-2009/2010	35,289	6.63	32,510	6.11
Year 4 to Year 5	2009/2010-2010/2011	47,732	8.97	43,301	8.13
Year 5 to Year 6	2010/2011-2011/2012	43,210	8.09	23,502	4.40

Institute for
Fiscal Studies

The database of schools: construction

1. Make LEASIS files consistent over time
   a. Create *phase of education* variable based on number of pupils of each age
   b. Make school type consistent over time
   c. Clean school name (eg change "Roman Catholic" to "rc" in all cases)
   d. Add postcode from EduBase

Institute for
Fiscal Studies

The database of schools: construction

1. Make LEASIS files consistent over time

2. Link schools across years, for each year:

   a. Merge using URN

   b. If no link created, merge using LA and ESTAB

   c. If no link created, merge using postcode and clean school name

   d. If no link created, merge using postcode and phase of education

   e. If no link created, merge using postcode and ESTAB

   f. If no link created, merge using DfE parent-child data

Institute for
Fiscal Studies

The database of schools: construction

```
. tab match_2005_2006
```

Match between LEASIS 2005 and 2006	Freq.	Percent	Cum.
Match on URN	28,116	97.46	97.46
Match on postcode and school name id	95	0.33	97.79
Match on postcode and phase	63	0.22	98.01
Match from DfE linked file	211	0.73	98.74
No match in 2006	140	0.49	99.23
No match in 2005	223	0.77	100.00
Total	28,848	100.00	

```
. tab match_2012_2013
```

Match between LEASIS 2012 and 2013	Freq.	Percent	Cum.
Match on URN	26,332	91.03	91.03
Match on LA ESTAB	1,530	5.29	96.31
Match from DfE linked file	441	1.52	97.84
No match in 2013	128	0.44	98.28
No match in 2012	497	1.72	100.00
Total	28,928	100.00	

Institute for
Fiscal Studies

The database of schools: construction

1. Cross-over between DfE linked file and school database
   - Negative: for this project – the cross-over is annoyingly close!
   - Positive: for future projects (that need LEASIS data further back in time) – we are able to match the majority of schools that don't match on URN without the DfE linked file

Match between LEASIS 2011 and 2012	match_2011_2012_dfe 0	1	Total
Match on URN	22,566	3,264	25,830
Match on LA ESTAB	4	2,354	2,358
Match on postcode and	4	45	49
Match on postcode and	4	126	130
Match from DfE linked	0	183	183
No match in 2012	140	0	140
No match in 2011	137	39	176
Total	22,855	6,011	28,866

Institute for Fiscal Studies

The database of schools: construction

1. Make LEASIS files consistent over time

2. Link schools across years, for each year

3. Link backwards and forwards over time

    ➢ Schools that convert have two rows

    ➢ Schools that merge have multiple rows

    ➢ Links can be traced whatever year the researcher starts from

Institute for
Fiscal Studies

Summary

- School identifiers change over time

- This causes empirical researchers and charities trouble

  - Linking schools over time is complicated

  - Difficult to create a consistent panel of schools

- The database of schools provides information on school mergers, school closures, school opening, and changes in school status

- Schools are linked forward and backwards to make linking schools over time straightforward

- Work in progress – feedback and checking code would be welcome!

Institute for
Fiscal Studies

Questions?

ellen_g@ifs.org.uk

Institute for
Fiscal Studies

Appendix 1: Re-format DfE parent-child data

1. Open DfE parent-child data

2. Some manual cleaning to replace parent to child, child to parent etc (email ellen_g@ifs.org.uk)

3. Keep child data only to merge to existing data

```
*    database of links between "old" and "new" schools DfE know about
keep if inlist(linktype,"Child","Merged - change in age range","Merged - expansion of scho

gen urn_new = linkurn
gen school_name_new = linkname
gen urn_old = urn

keep urn_new school_name_new urn_old
duplicates drop
drop if urn_old == 117517                          // The Heathcote School closed and pupils were

save "$data_save\dfe_links.dta", replace
```

Institute for
Fiscal Studies