Institute of Education

UCL

Longitudinal Administrative Data and Survey Data linkage: some problems and solutions

Lorraine Dearden

28th January 2016

closer
Cohort & Longitudinal Studies Enhancement Resources

Administrative Data Research Centre England

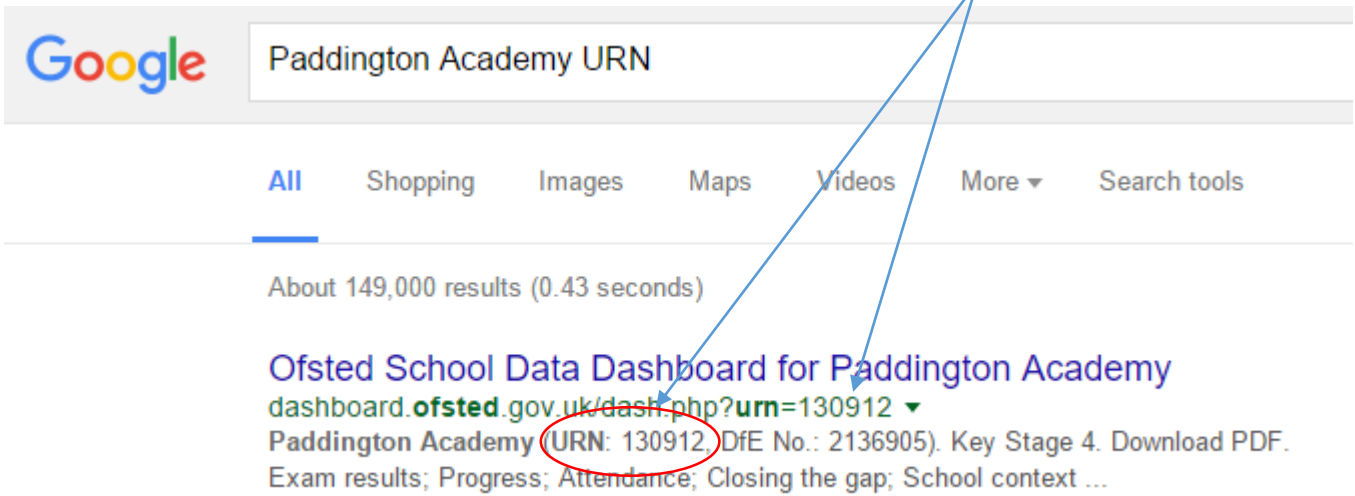
Key issues

- Panel surveys are increasingly obtaining permission to link to administrative data sets
  - Health data e.g. Hospital Episodes Data
  - Crime data held by MoJ
  - Economic data held by DWP/HMRC
  - Education data held by DfE, BIS, HESA, UCAS…..
- So far, it is education data from the National Pupil Database (NPD) which has been most widely linked
  - E.g. to ALSPAC, MCS, Next Steps, Understanding Society

What is the problem?

- For at least some of these longitudinal panels (e.g. MCS) the school identifier in this administrative data needs to be *anonymised* in order to ensure confidentiality of survey participants
  - The school identifier in the NPD – the unique reference number – URN is a publically available school identifier

  - I did a google search for Paddington Academy URN



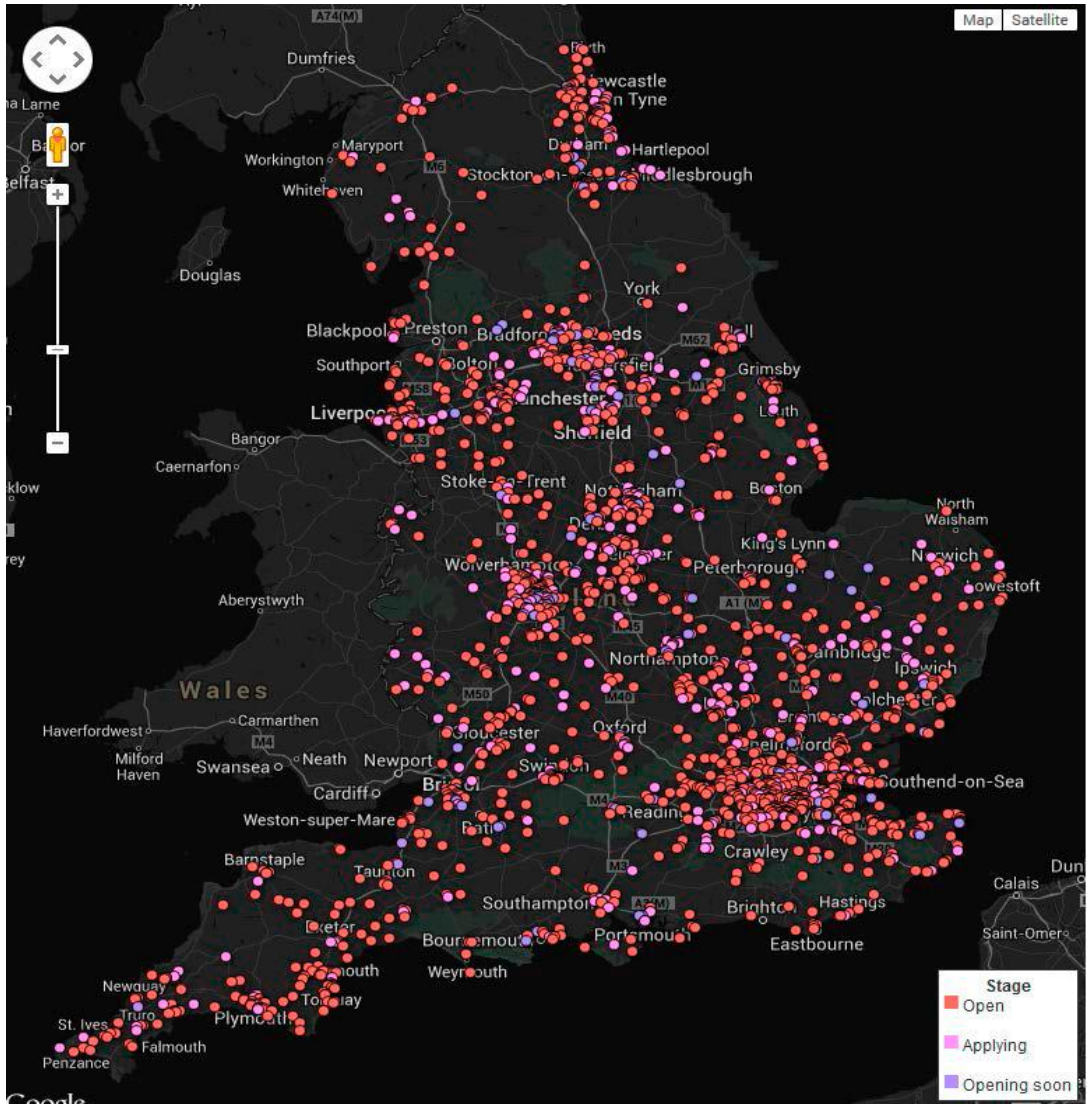
closer
Cohort & Longitudinal Studies Enhancement Resources

Administrative Data Research Centre England

But is it unique?

- The name URN suggests this identifier is *unique BUT* they are far from unique
  - Change if school changes status e.g. from comprehensive to academy
  - If schools merge
  - If boundary changes to local authorities occur
  - If the school moves site
  - ……
- This becomes a huge problem if this identifier is anonymised as a researcher no longer knows if a school identifier changes is a real school move or simply a school changing status etc…..

Changes are large in any one year…



Motivation:

School identifiers (Unique Reference Number – URN) change due to changes in school structure.

closer
Cohort & Longitudinal Studies Enhancement Resources

Administrative Data Research Centre England

Schools Database

Changing school identifiers are a problem!

Empirical research:

- **How to construct a consistent panel of schools?**
- **How to adequately control for school-level prior attainment?**
- **How to control for school level clustering appropriately?**

Education charities:

- **How to maintain list of eligible schools?**
- **How to track attainment of schools involved in program?**

closer
Cohort & Longitudinal Studies Enhancement Resources

Administrative Data Research Centre England

Schools Database

The Problem: Example

Merge LEASIS 2012 with LEASIS 2013 schools data…

_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	

Some URNs only in 2012

Some URNs only in 2013

But only one unmatched school from 2012 actually closes between 2012 and 2013.

It matters….

The Problem: Example

What is the impact of changing to academy on KS2 (age 11) Average Points Scores?

	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 (2010/2011) and interaction between prior APS and change to academy status.

Schools Database

Solution – Database of Schools to merge school identifiers in adjacent years creating a consistent panel of schools:

- We have created a file which links schools between 1994 and 2014
- Will be updated to include 2015 data in next month
- Will be made publicly available (dataset and syntax)
    - Currently talking to DfE about how we can make this publically available

closer
Cohort & Longitudinal Studies Enhancement Resources

Administrative Data Research Centre England

Schools Database

Solution – Database of Schools:

1. **Make LEASIS* files consistent over time:**
   - Create phase of education variable based on number of pupils of each age.
   - Make school type consistent over time (academy, maintained, college etc)
   - Clean school names (e.g. change “Roman Catholic” to “rc”)
   - Add postcode from EduBase

*Note: LEASIS files are data for all school in England (including private schools) that Have been produced since 1994 with. Stopped being produced in 2015.

- Contain pupil numbers by age/gender, %FSM, Pupil Teacher Ratio, %SEN, Ethnicity etc….
- Can be linked to school results files at KS1, KS2, KS4 and KS5

closer
Cohort & Longitudinal Studies Enhancement Resources

Administrative Data Research Centre England

Schools Database

Solution – Database of Schools:

1. **Make LEASIS files consistent over time**
2. **Link schools across years, for each year:**
   - Merge using URN
   - If no link created, merge using LA and ESTAB
   - If no link created, merge using postcode and clean school name
   - If no link created, merge using postcode and phase of education
   - If no link created, merge using postcode and ESTAB
   - If no link created, merge using DfE individual level census data (if available)

closer
Cohort & Longitudinal Studies Enhancement Resources

Administrative Data Research Centre England

Schools Database

Solution – Database of Schools:

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

closer
Cohort & Longitudinal Studies Enhancement Resources

Administrative Data Research Centre England

Schools Database

Solution – Database of Schools:

1. **Make LEASIS files consistent over time**
2. **Link schools across years, for each year**
3. **Link backwards and forwards over time**

- **Database provides information on school mergers, closures, opening and changes in school status.**

closer
Cohort & Longitudinal Studies Enhancement Resources

Administrative Data Research Centre England

Schools Database

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

Stable Schools

closer
Cohort & Longitudinal Studies Enhancement Resources

Administrative Data Research Centre England

Schools Database

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.)

Changing Schools

closer
Cohort & Longitudinal Studies Enhancement Resources

Administrative Data Research Centre England

Schools Database

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
102677	102677	130899	130899	130899	130899	130899	130899	130899	130899
102680	102680	130899	130899	130899	130899	130899	130899	130899	130899
130899	102677	130899	130899	130899	130899	130899	130899	130899	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	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

Changing Schools

closer
Cohort & Longitudinal Studies Enhancement Resources

Administrative Data Research Centre England

Schools Database

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						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

Changing Schools

closer
Cohort & Longitudinal Studies Enhancement Resources

Administrative Data Research Centre England

Schools Database and Data linkage

Database of Schools: Linkage to Surveys

- In principle, any survey capturing school age children post-1993 can be matched to database of schools
  - ALSPAC (1990/1991)
  - BHPS/Understanding Society (1991/2009)
  - MCS (2000/2001)
  - LSYPE (2004).
- Most Importantly we can show from our database, we can say why an individuals school identifier changed before anonymisation in panel survey data
  - Extra variable in survey data:
    - Real school change; change of status of school (and what that change was) school merger; school closure etc – really important for getting analysis right
    - E.g. in MCS can have anonymised school identifier for each year of school and if this changes for individual can say why this change occurred
    - At moment this has NOT been done in MCS (or other surveys)

closer
Cohort & Longitudinal Studies Enhancement Resources

Administrative Data Research Centre England

Conclusions

- Linkage to surveys carried out by data holders
- Data holders generally provide pseudonymised school identifiers – which are consistent over time
- But reasons for changes in identifiers cannot be established once pseudonymised – means longitudinal analysis using education data could be wrong – no problems for cross-sectional work
- Guidelines and approach are currently being established how we grant access to this resource….
- Applications are wider than for education data – similar problems with changes in hospitals, prisons, …….

closer
Cohort & Longitudinal Studies Enhancement Resources

Administrative Data Research Centre England