

# BBC Newsnight

“this just strikes me as being worth examination”

I would like the following sensitive fields (most of which are in tier 2 releases):

Gender\_{term}[yy]  
AgeAtStartOfAcademicYear\_{term}[yy]  
MonthPartOfAgeAtStartOfAcademicYear\_{term}[yy]  
Ethnicity\_{term}[yy]  
EthnicGroup\_{term}[yy]  
EthnicGroupMinor\_{term}[yy]  
EthnicGroupMajor\_{term}[yy]  
FSMeligible\_{term}[yy]  
EVERFSM\_ALL\_{term}[yy]  
COA\_{term}[yy]  
LSOA01\_{term}[yy]  
IDACIScore\_{term}[yy]  
FirstLanguage\_{term}[yy]  
EVERFSM\_6\_{term}[yy]  
InCare\_{term}[yy]

ServiceChild\_{term}[yy]  
GandTindicator\_{term}[yy]  
ModeOfTravel\_{term}[yy]  
Disability\_{term}[yy]  
SENprovision\_{term}[yy]  
Primary SEN  
DistCurrSch\_{term}[yy]  
DistNearSch\_{term}[yy]

The tier 1 variables I have requested are

- **Output area** - This is important for analysing the effect of policies such as academy conversion. Using output areas, I can get the child's neighbourhood with greater precision. When analysing competition effects, for example, knowing the nearest schools with greater precision is important. LSOAs are slightly too large for comfort, so the child may not be allocated to the right network of local schools for analysis.

- **Ethnicity and mother tongue** – this is an addition to the request: I would like these variables because, as [redacted] has argued this week, I don't think we quite understand the ethnic and cultural components of the London school improvement. I can't spot the effect of, say, A8 immigration on schools because I can't spot Poles, say, as opposed to Germans. Nor can I distinguish Somali and Nigerian children (who are rather different).

- **Service children** – They now attract extra money, so I would like to know if they have any recognisable differences. This just strikes me as being worth examination.

- **Gifted and talented** – These pupils have particular needs whose experiences, again, would bear examining.

- **Distance from schools & mode of travel** – School travel bears examining, especially since the DfE has sought to change traffic planning rules. I'd also like to know whether the decision to eschew nearer schools/drive to school is predictive of anything. (From the output areas, I will be able to ascertain the distances anyway. But since you have calculated them...!)

- **Primary SEN** If areas display dramatically different treatments for the same recorded disadvantage, that will help decisions about whether to permit SEN as a factor. In general, I try to avoid it because I think it is endogenous to the schools. If there's consistency, I might rethink that.

Among the other sensitive data, I am requesting the usual variables which can help to explain differences in school performance. I appreciate that this :

- Broad ethnic groups
- EAL status
- FSM eligibility
- SEN
- Age
- Gender

**- In care** – this is a crucial group, who suffer multiple disadvantages; assuming I can meet the anonymisation rules (which might mean not passing comment on lots of LAs to avoid breaching the small number rules), I would like to take a look at differentials in care. Given that some of these children physically get moved around the country a lot, this is particularly important (if we shipping children to schools with weak LAs.)

5 years' worth of data (2009-2013) from the NPD were released including children's sensitive personal data, in care and service children, SEN and FSM indicators, ethnicity and language.

These identifying and sensitive items, or identifying data items were matched at individual pupil level with census data for KS2, KS4 and KS5 attainment datasets before individual level release.

**There is no clear legal basis for passing journalists individual level, sensitive data meeting Schedule 3 of the Data Protection Act 1998:**

*“ we did not require [any] requestor to inform us of the conditions for processing that they relied on.”*  
Department for Education (click to view original FOI)

**Tier 1:** pupil level, identifying and highly sensitive

**Risk:** No small numbers suppression, fully identifying

**Download original requests :**

1. [Data application August 2014 .pdf \[download\]](#)
2. Further request since data granted in 2010-2012