Data sources

The list of schools and MATs is derived from edubase data and the list of Southeast MATs provided by SERSC. Performance and characteristic data comes from data made publicly available by the DfE (performance tables, OFSTED's report summary, list of NLEs, etc)

Certain special types of centres are excluded from the analysis (eg special schools, PRUs). Centres are assigned to one or more phases of education usually based upon the PhaseOfEducation field in edubase. Where this is omitted from edubase or ambiguous, the phase(s) are deduced from the age range of the school.

The data sources variously use LAESTAB or URN to identify centres. Where a centre has closed and reopened, changing their LAESTAB, historical data will not be available due to the change of identifier.

MATs

Each MAT includes all centres identified in that MAT either by SERSC, or edubase. It therefore includes some centres outside the Southeast.

Averages for MATs are calculated by aggregating the individual centres. For some fields this is a simple average of the value for each centre. However where there is an obvious denominator this is used to give a weighted average (for example, we use the number of students as the denominator when aggregating points-per-student)

Note: most data fields for MATs are averages rather than totals. This includes fields such as number of pupils which is the average number of pupils per centre.

Performance fields

These are taken directly from the performance tables for each year. The fields which are available change somewhat. There have been substantial changes to the performance tables for summer 2016.

The values for MATs are averages, or weighted averages, of the values for each of their centres. Where a centre has no data reported it is omitted from the average. This usually applies to centres that have just opened and have no results yet, or to centres with fewer than 10 students (and in the latter case such a small cohort is largely irrelevant to the overall performance of the MAT)

Characteristic fields

Fields which report changes over three years use Summer 2013 to the present - since summer 2016 is the last performance tables data that is included.

Total # of pupils

This is derived from the "TOTPUPS" or "TPUP1618" field in the performance tables (depending which phase(s) of education each centre appears in). Although edubase has a total number of pupils, we only had the current edubase extract; whereas performance tables gives us the value each

year. A few new centres may be omitted from performance tables; for an MAT the reported value is the average for all centres appearing in the performance tables.

"Number of special schools", "Number of PRUs"

These fields report the number of special schools (any centre with an TypeOfEstablishment containing "special") or PRUs (TypeOfEstablishment = "Pupil Referral Unit") within the MAT.

Note that these centres are not generally included in the rest of the reporting, not least because they are generally not included in the performance tables data.

"Centres joined in 3 years"

This is based on the "Joined date" in the edubase links file. Centres joining in September 2013 onwards are included.

(The analysis is using August 2016 as the nominal date for the current data, and the cut-off is 3 years earlier. The actual edubase extract used was taken slightly later than this and we chose not to filter out any centres added after August 2016, so the time span for this measure is actually slightly greater than 3 years)

Converter/Sponsored fields

These report as above, separating all centres into either Converter or Sponsored centres. This is derived from the TypeOfEstablishment field in edubase. Sponsored centres includes all types:

- including "Sponsored"
- all variants of Free School
- University Technical College
- Studio Schools

"Total growth in pupil numbers in last 3 years" / "Growth in pupil numbers in 3 years excluding new centres"

These are the difference between the Total # of pupils in summer 2013 performance tables and summer 2016.

"Average distance between" (centres)

The location of each school is taken from the Easting and Northing fields in edubase. If these are omitted the centre is omitted from this calculation. The distance between every possible pair of centres is calculated, and the reported value is the average.

"Cluster: avg distance to nearest 3"

For each centre the distance to the nearest 3 centres is calculated. If there are fewer than 3 other centres in the MAT then all the available centres are used. The value for the MAT is the average of all of these distances.

The purpose of this measure is to measure the distance between each centre and other centres in the MAT that may support it and share resources. If an MAT has 2 groups of centres, one in the north of the country and one in the South the basic "Average distance between" measure will be a large value, because each centre is a long way from those of the other end of the country. However, if there are at least 4 centres in each group, this measure would report a much shorter distance.

"Cluster: %age with 3+ within 20km"

This counts the number of centres in the MAT that have at least 3 other centres within 20 km. However, if the MAT has fewer than four centres in total then the threshold is instead that all other centres within 20 km.

Dissimilarity score

For each care centres within the MAT we calculated a score to reflect how different the nature of the two centres were. The scores translated by looking at a number of factors and adding up "points" where they differ. The factors we included are:

- Gender policy

- Admissions policy (especially a grammar school is considered very different to any other)

- Whether the centre is urban or rural

- Whether the centre contains primary, secondary and KS5 provision

- The difference in number of pupils at the centres (as a ratio: the numbers were divided rather than subtracted)

- The religious nature of the school

- the percentage of FSM students at the school

Each factor adds up to 50 to 150 points for the maximum difference (e.g. urban/rural adds 50 points at most).

The resulting score is never zero (there will always be some difference in number of pupils or FSM pupils); any value under 100 would indicate very similar schools. Values in the high hundreds (eg 500+) would indicate centres of a very different nature.

Average OFSTED score

The OFSTED score for each centre is taken from Ofsted's summary of the latest report for every centre which is published monthly. The overall effectiveness score is used

Avg change in OFSTED from join date to latest where known

To assess the changes in OFSTED score we also use the summary of all inspections from 2005 to 2015 published by OFSTED. The value is the difference between the last available OFSTED score dated before six months after the centre joined the MAT (the joining date is taken from the edubase links file) and the latest OFSTED score. If the date of the last OFSTED report before joining matches the date of the last report then it is disregarded and no value is entered for that centre.

In practice the availability of data for this measure is patchy. It is usually several years between inspections. Centres which have joined an MAT recently may not have another inspection since the last one before joining. Centres which joined an MAT before 2010 might not have a prior inspection to use as the basis.

Furthermore where centres have converted to academies and/or closed and reopened they are typically given a new URN (which is used to identify the centre by OFSTED). We have attempted to track such centres and link the old identifier to the new centre. However in cases where the centre also changed LAESTAB we cannot do this; and for centres which "closed" a long time ago we may not have the data to do so.

The value for the MAT is the average change in OFSTED for centres where a value could be calculated. Therefore for many MATs the average may be based upon only some of their centres.