
Working under the [data] Clamp-down

**Sean Demack, Sheffield Institute of Education,
Sheffield Hallam University.**

Introduction

Joe Strummer's mumbled first line of Clampdown (Strummer & Jones, 1979) poetically captures recent movements in the access to educational data in England

"The kingdom is ransacked, the jewels all taken back"

In the context of this paper, the 'jewels' are education data. Perhaps equating data as 'jewels' is overblown given glaring imperfections of education data in England but the data infrastructure is internationally impressive. Until recently, I had worked through times of increasing access to education data. This paper reflects on the reining in of this access; a data clampdown. Written at the end of the year in which new European data protection legislation became law with General Data Protection Regulations (GDPR). The clampdown began before GDPR became law but the two are clearly entwined. The paper begins by briefly discussing educational data sources before using experiences in accessing pupil level data from the National Pupil Database (NPD) to illustrate a data access saga. Finally, implications of the clampdown for a critical statistical future are reflected on.

National Pupil Database (NPD)

In the first two decades of the 21st century, the NPD has become to be the main quantitative educational data source in England. Within the context of neoliberal policy interventions, state schools in England are legally required to regularly submit pupil-level data to DfE for the school census. This data is compiled into the NPD and this data can then be accessed for research and evaluation purposes. Prior to NPD, educational surveys were the main quantitative evidence base (e.g. the Youth Cohort Study, LSYPE). The switch from surveys to NPD can be viewed positively and

negatively. First, one positive is improved coverage; NPD provides data on (nearly) all pupils whilst surveys only provided data on a responding subsample from a random sample of pupils. NPD coverage is not universal (for example, data on pupils in private schools is rather limited) but it does enable analyses of attainment for pupil cohorts (with around $N > 600K$ pupils per pupil cohort). The switch from survey to NPD took place in a time when schools were taking centre stage as the heroes and villains of the English education system. Because of sample size, surveys could only be used to provide a statistical view on the education system as a whole but NPD made it possible to 'drill down' to a geographical or school level and this fitted closely with the policy zeitgeist. The breadth of coverage enabled an increasing statistical spotlight on schools but, in NPD, the perspective has always been limited by poor measures of pupil socioeconomic background. NPD provides pupil-level binary variables that identify whether a pupil is socioeconomically disadvantaged (in terms of eligibility and uptake of free school meals and/or being 'looked after') or not.

Additionally, the NPD includes the socio-geographic income deprivation affecting children index (IDACI). This uses pupil home postcodes to show the proportion of children under the age of 16 that live in them who come from low income families. Low income families are defined as families that either receive certain benefits or Working Tax Credit or Child Tax Credit with an income (excluding housing benefit) below 60 per cent of the national median before housing costs. In summary, the focus of the two NPD variables is on socioeconomic disadvantage. IDACI might be used to identify pupils who live in areas with low concentrations of child poverty but also brings analytical ecological headaches. IDACI might be a valuable socio-geographical context variable if used alongside more precise individual (and/or family/household) level measures. However, the only such measure(s) currently included are binary indicators of pupil social disadvantage. This leads to analyses that can only look down on a 'problem group' that have been officially identified as poor and to compare their educational experiences to a group not deemed to be officially poor and (therefore) not a problem. This happens to reflect a neoliberal political culture of policy that

espoused an "intense relaxation" about social advantage, wealth and privilege but is a glaring methodological weakness in NPD. The focus on social disadvantage might enable the 'performance' of pupils in different schools to be compared more fairly than if this difference was ignored but the lack of any detail on social advantage serves to limit this . NPD does include a very disaggregated measure of pupil ethnicity (except for private schools). This enables an examination of the (measured) educational experiences for different pupil ethnic groups as they progress through the system. However, given the strong statistical association between ethnicity and socioeconomic position, the depth of analyses here is also limited by the poor measurement noted earlier.

Early warning

Prior to 2016, key problems for NPD were measurement and the limited data on private schools (both of these might be addressed with some effort and political will). Then, in October 2016, the DfE sent an email to all researchers who used NPD data:

Figure 1: DfE email to NPD users from October 2016

Subject: DfE change to conditions of use of data authorised for research and analysis

Following queries from users, DfE are making a minor change to conditions of use of data authorised for research and analysis for all future applications. This is to ensure consistency across all our user community. The additional condition of use states:

All research and analysis must be shared with DfE 48 hours before publication.

Many of our data requestors already operate in this manner, but this condition will create a consistent approach. It will ensure that relevant policy teams and press office are aware of the valuable research that you carry out based on DfE data. This will reduce the risk that DfE are caught off guard by being asked to provide statements about research the appropriate people have not seen.

To be clear, it is not DfE's role to check or approve the outcomes of

research and/or analysis during that 48 hour period, merely to ensure the right people have had time to digest it.

It would be appreciated if all current requests also follow this new condition.

The email signalled a new layer of scrutiny that researchers were required to undergo; 48 hours prior to publication but not for checking/approving outcomes of analyses and research. This email was met by a flurry of disgruntlement amongst NPD users and was followed by a number of updates to the required process (and forms) for accessing NPD data.

Prior to 2016, I had accessed NPD data for many projects and had grown to expect a wait of around two to three months for the process to be completed and data delivered ready for analyses. A fairly lengthy form was submitted that set out the analysis plans, reasons and funders. This form was quickly checked and acknowledged as being OK before being placed in a queue. When the request reached the top of this queue, NPD looked more closely at the request and some clarification emails were exchanged before a final schedule was written and a data file prepared. Shortly afterwards, the data would be encrypted and transferred onto a secure University server. University IT security was evaluated through a second form completed by the IT department. The length of time (and forms) for accessing NPD crept up but generally a two to three month wait with around 10 email exchanges for each request was expected.

Data Access Saga

In November 2017 as part of a DfE funded evaluation research project with a team of five academic colleagues, I submitted a request for NPD data. The evaluation ran for just over a year and the aim was for all analyses to be completed by summer and the final report submitted at the end of September 2018. Following the submission of the NPD request, I received an acknowledgement email which stated to expect a wait of three months. I then began to chase the data in February 2018 and gradually entered an increasingly curious world that was filled with new requirements in

the shape of forms, impromptu training and analytical surveillance. This was during the lead up to the arrival of GDPR in May 2018 and in the wake it left. Trawling through the emails, the saga resulted in over 100 emails between the research team, NPD, ONS, DfE and the client and the completion of nine forms (commonly for each of the research team). The data access timeline below provides a summary of this data access saga (I started it whilst "Dance of the Cuckoos" (Hatley, 1928) was playing on the radio, it seemed appropriate).

Figure 2: Data Access Timeline

<p>Nov 2017:</p> <p>Submit NPD application & received acknowledgement</p>
<p>Apr 2018:</p> <p>NPD staff allocated to application & request details of consent obtained to match data to NPD (this detail is supplied)</p> <p>All new NPD applications paused (but current applications will progress)</p>
<p>May 2018:</p> <p>New requirement from NPD to identify all specific variables / data fields we plan to use in the analyses (from a spreadsheet with over 20 sheets each with up to 1,200 fields listed) - and update original application</p>
<p>Jun 2018:</p> <p>Submitted updated application to NPD</p>
<p>Jul 2018:</p> <p>Ongoing discussion with funder (DfE) resulted in some hopeful words via email "We've escalated your request regarding access to the NPD data, so hopefully things should start moving soon. We will of course update you as soon as we hear more, but in the meantime please rest assured that it's in hand"</p>

Aug 2018:

NPD state that all secure data transfers have been paused but access will still be possible via one of the ONS physical secure lab or the ONS virtual environment [from September].

ONS secure labs are located around 200 miles from our location and so we consult with the client to ask the funder (DfE) whether they will provide funds to cover time, travel and accommodation for access via a ONS secure lab (in face of this new NPD/DfE requirement)

New requirement that all staff who access the data will require ONS 'safe researcher' training.

NPD notified us that data access is agreed ...once all staff complete training and pass follow on test.

Sept 2018:

Attend ONS safe researcher training [rumour in the training room was that the ONS virtual environment is nowhere near ready to use - therefore it seems that secure physical labs are the only route to access the data]

NPD request further details on the consent we obtained

Upload a file for NPD to match data to.

DfE extend the evaluation contract to end of December.

Oct 2018:

Pass ONS safe researcher test

Client confirms that DfE will not provide additional funds to cover the additional expenses for using ONS secure physical labs

NPD confirm that the data has been transferred out of DfE/NPD to ONS.

Nov 2018:

After some confusion and liaison between NPD, ONS and ourselves,

ONS confirm that they have received the data.

ONS supply a number of forms for completion (it is unclear whether we will be able to use the virtual environment). These forms relate to the people who will access the data (again) and the IT security of the university.

Sign form to agree that "All outputs from the controlled environment will be checked before they are made available to the researcher"

Dec 2018:

ONS send details on how to access the virtual environment using the computer specified in the ONS forms.

We access the virtual environment! - but find no software or data.

ONS confirm that the virtual environment is not functioning

ONS discuss supplying a laptop which would be able to link directly to the ONS secure labs.

ONS laptop delivered - however university WiFi is not compatible.

Wifi problem circumnavigated using an Iphone and access to the data is finally achieved!!

Client contacts us to say DfE will not be extending contract beyond December. We respond to underline that DfE (as evaluator funders) have specified this deadline but this cannot be met because of the length of time that DfE (as data gatekeepers) have taken to deal with the application for NPD data - and allude to the Terry Gilliam film 'Brazil'.

DfE agree to extend contract

Jan 2019:

Acclimatize to the security hurdles and startling warning pop-ups met whilst setting up the ONS laptop to access the NPD data.

Become familiar with the three stages of security needed to access the laptop.

pop-up message from ONS employee appears on laptop asking me to ensure that I log off after each session. Become aware that ONS employ people to check people accessing the secure environment using regular screenshots and collected key strokes.

It should be highlighted that the saga was not a result of any individuals from NPD or ONS, both organisations were extremely helpful in providing guidance to navigate data access requirements as they emerged. Similarly, individuals in DfE attempted to help smooth or speed up access. It seems that individuals in ONS, NPD and DfE found themselves in this administrative data access maze and did their human best to find a way out. At a systems level, the DfE as an organisation is more open to criticism. This is perhaps most humorously illustrated with DfE (as funder) deciding to extend the contract deadline after 'finding out' that it was the DfE (as NPD data gatekeepers) who had the main responsibility for the delay to data access. Whilst access was finally achieved after 13 months, the result is quite dramatic changes in working practices. For this project, I now work with an awareness/expectation of regular surveillance (via recorded key strokes and screenshots) and that no output will be released to me until it has been approved by ONS. It seems I have moved from a position external to (and independent of) Government to one where my work requires approval of Government (via the ONS wing).

ONS researchers or statisticians within Government ministries might find this unremarkable but the security hoops and Government surveillance practices will be new and disconcerting to other researchers. Scrutiny is expected and built in to academic practices such as ethics committees, data management and storage plans and peer review of publications. However, recording screenshots and key strokes are steps beyond these. Further, this results in notable imposed changes to working practices (e.g. time, length and location of planned analyses). It seems that big Government got bigger under the guise of data protection in the age of GDPR. This echoes the accountability/power shift at the school level away from local authority and towards central Government.

Systems are now being put in place to ensure Governments will be unsurprised by research findings either by not approving output or by providing time to prepare/spin a response to output that has been approved.

"What are we going to do now?"

Increased barriers to data access will ensure that data is held in the hands of narrower pool of researchers with sufficient resource, motivation and tenacity to complete a data access saga. Future data access sagas will probably be less epic in length but, to paraphrase Scott (2018 p5) the legacy of the clampdown will ensure that the 'correct people' have access to the data and that they make the 'correct decisions' in their statistical analyses.

Whilst clearly positive, Hans Rosling was keen to stress that his 'things are better than you think' perspective was not because he was a naïve 'optimist' (Rosling et al., 2018, p69). Instead, Rosling (ibid) identified as a 'possibilist' to highlight that his global perspective was based on empirical evidence in the form of socio-geographic statistics using open-access data (www.gapminder.org). This open-access data is commonly collected via censuses and household surveys with considerable variation in population coverage (Carr-Hill, 2013). Specifically, Carr-Hill (ibid) estimates that 250 million of the 'poorest poor' are missed worldwide. Wider methodological criticism on global poverty measurement has been expressed by others (Amand et al., 2010). This suggests that Roslings perspective is likely to have been rose-tinted. Of course, data and statistics do not provide objective, neutral 'truths'; they are socially constructs built on methodological foundations. Understanding methodological flaws (such as missing data and analytical assumptions) is important. The apparent objectivity of methodological 'quality-control' checks can result in overlooking the subjective (social construction) decisions behind data collection, measurement, analyses and interpretation. The nature of this social construction determines how useful data and statistics are in helping to challenge and change current structural inequities, i.e. whether they have critical value. With open access data, statistics are open to scrutiny by all people with the necessary technology

and understanding. Each barrier of access will result in narrowing this pool of people.

Scott (2018) highlights how the critique of statistics has tended to come from within the statistical paradigm (Huff, 1976; Dilnot & Blastland, 2008). This within-paradigm critical gaze has focused on statistical literacy and/or the deceitful use of statistics, but is more limited on the social construction and impact of statistics. In other words, for data and statistics to be of critical social value, they need to be open to the critical gaze both inside and outside their paradigm. Data and statistics are open to manipulation and misinterpretation that might be thwarted through improved statistical literacy. However, the social value and impact of statistics will be assessed qualitatively, philosophically and externally to the statistical paradigm. For example, from Critical Race Theory, Gillborn et al (2018) develop principles for the analysis of quantitative data; QuantCrit. This is important because of the complicity of statistics as empirical foundations for racist beliefs including eugenics (Zuberi & Bonilla-Silva 2008). Without an engagement with an external theoretical framework, statistics will always be limited in their critical (transformative) potential. Statistics can be clear, accurate and unambiguous whilst also being of little critical value or worse, used to justify or defend wealth, privilege and power. Scott (2018) discussed how the statistics discipline colluded with power to become complicit in genocide, surveillance and inhuman workplace practices. This contemporary and historic collusion underlines how important it is that statistics is open to an external critical gaze. The clampdown of access to educational data in England risks weakening the external critical gaze. The data access barriers raised around the arrival of GDPR are likely to result in a diminished, less diverse pool of people who access educational data. Additionally, this diminished pool of people will be required to agree to conduct the analyses under the scrutiny of Government statisticians. This means that the external gaze is likely to focus more on accessible published tables and less on the inaccessible data and measurement behind these tables. Therefore, the data clampdown is a new obstacle for a critical statistical future. For example, where statistics might be used to

critique and challenge structural status quo(s) whilst emphasising the qualitative importance of lived human experience. Where reification is avoided, the social constructions of statistics are acknowledged and beliefs of statistical objectivity and neutrality are rejected.

In the first decade of the 21st Century, the availability of educational data hinted at a possible future where constructions, patterns and meanings of educational success were open to scrutiny by citizens. Perhaps future Governments would be expected to account for and change structural inequalities engrained into the English education system? Perhaps this would help to construct an education system that moved away from reproducing socioeconomic inequity and towards equity, meritocracy and enjoyment of learning? Before the data clampdown, improvement in terms of NPD data coverage and validity of measurement were clearly needed, but ease of data access provided the possibility of a critical future. The clampdown serves to dampen such optimism and will limit the extent to which future Governments are subjected to independent critical scrutiny.

Later in Clampdown, Joe sings of teaching twisted speech and training blue eyed men "to be young believers" before the final line:

"And I'll give away no secrets" (Strummer and Jones, 1979)

The clampdown on access to educational data in England will ensure that in the future, patterns of structural educational inequality will be examined through Government approved analyses. This is an obstacle for the future critical use and role of education statistics in England. A few secrets may still emerge, but not before Governments have had sufficient time to develop a counter-narrative.

References

- Anand, S., Segal, P. & Stiglitz, J.E. (2010) *Debates on the Measurement of Global Poverty*. Oxford University Press
- Carr-Hill, R. (2013) *Missing millions and measuring development progress*. World Development.
- Dilnot, A. & BlastLand, M. (2008) *The Tiger that Isn't: Seeing Through a World of Numbers*. Profile Books.
- Gillborn, D., Warmington, P. & Demack, S. (2018) QuantCrit: education, policy, 'Big Data' and principles for a critical race theory of statistics. *Race, Ethnicity & Education* Race Ethnicity and Education, 21 (2), pp158-179
- Hatley, T.M. (1928) *The Dance of the Cuckoos* (Laurel & Hardy theme tune).
- Huff, D. (1976) *How to Lie with Statistics*. Harmondsworth: Penguin.
- Lord, A., Easby, J. & Evans, H. (2013) *Pupils not claiming Free School Meals*. DfE Research Report RR319.
- Rosling, H., Rosling, O. & Rosling Ronnlund, A. (2018) *Factfulness: Ten reasons we're wrong about the world and why things are better than you think*. London: Sceptre
- Scott, S. (2018) *Statistics as Reification*. *RadStats* 121 pp4-25
- Strummer, J. & Jones, M. (1979) *The Clampdown*. Song on London Calling Album, London: CBS records.
- Zuberi, T. & Bonilla-Silva, E. (editors) (2008). *White Logic, White Methods: Racism and Methodology*. Lanham, MD: Rowman & Littlefield.