

Iain Bell

Director for Data Standards Analysis Group
Sanctuary Buildings
Great Smith Street,
SW1P 3BT

Ed Humpherson
National Statistician
1 Drummond Gate
London
SW1V 1QQ

28 November 2016

Dear Ed,

UK Statistics Authority correspondence on Access to Educational Attainment Data and on the National Pupil Database

I thought it worthwhile to update you on the progress the Department for Education has made since your correspondence earlier this year. There have also been a number of reports recently on DfE data collection and processing arrangements which I wanted to update you on.

National Pupil Database – transparency of access arrangements

You wrote to me in April setting out a range of requirements to improve the transparency of the processes around the National Pupil Database (NPD). The material provided to the UK Statistics Authority at this time showed that all these actions were underway. The following documents were published by 6th May:

- Updates to our website guidance on the NPD process, which included clearer text on how we share NPD with external users,
- A simplified process map showing how data flows through the system,
- An updated version of our sample privacy notice for schools to use (making it clearer that NPD data is given out under certain strictly controlled conditions to third parties)
- An update to the list of who we provide NPD extracts to (extending the previous list from the end of 2014 to the end of 2015) and including a status field for each extract since the log began (2012) to show whether the request was finished and data destroyed, or whether it was still live. We then updated the full list on 14th July bringing the list up to date for all requests to end March 2016 and adding more information to the form, adding more recent pipeline requests and publishing the DMAP terms of reference.

The next update of the NPD extracts log will be published shortly which will report on extracts and status to end of June 2016, alongside updates to the pipeline requests document.

DfE take our duties around data protection very seriously and we have been exploring further room for improvement. It is clear from the access list that the establishment of a physical area where users could access the data without DfE needing to produce bespoke data extracts would have improved our service levels. In order to fulfil this user need and reduce the number of extracts, the Department is:

- Working with the Office for National Statistics to make the National Pupil Database available through their Virtual Micro-DataLab (or a DfE-run process which mirrors much of it)
- Exploring the potential for a 'privacy controlling API' to be built on top of NPD. We are currently working in partnership with the Open Data Institute to do discovery work and design how we take this forward. .

Related data access procedures

There has been some concern about the recent communication to NPD users which requested access to research 48 hours in advance of its publication. This was originally thought to be a clarification of existing rules however I understand it caused confusion and I apologise for that.

We have written to all NPD Users to put forward a revised set of terms and conditions which clarifies this requirement. The replacement still provides assurance to the department that the publication is within the permitted uses of the original application - if not a further application will need to be submitted – an essential aim to maintain the trust of those whose data are supplied. This notification of intent to publish will be at 48 hours or, where that is not possible, as early as possible but has no requirement to share the analysis or findings. It also recognises that some users may produce additional analysis at short notice to inform public debate and the revised terms recognise this. We have already begun discussions with interested parties to ensure we get this right in practice. (See Annex A for the Clauses for Access.)

School census and access to data

This year's school census has come under considerable scrutiny due to the new data items being collected on nationality, country of birth and English language proficiency. The purpose of collecting these additional data is so that DfE can better understand how children, with for example, English as an additional language, perform in terms of their broader education, and to assess and monitor the scale and impact immigration may be having on the schools sector.

I would like to assure you, the public and users that these new data items are solely for the use of the Department for Education for research, statistics and

analysis and are not shared beyond this.

The exercise has also shown some areas where we may be able to improve and we will be reviewing all the lessons we can learn from this

Access to Education Attainment data

You sent details of your report on this and I was due to provide an update on these issues. During the process of producing the report, DfE agreed we would:

- make the purpose of the Performance Tables clearer to all users; - explain more clearly the wide range of reasons why schools may have 0 per cent attainment recorded in the Performance Tables;
- Make the scale of the impact of the exclusion of unregulated International GCSEs known to users; - Explore making an alternative version of the National Pupil Database available to users which includes unregulated International GCSEs for research purposes.

Our Performance Tables website is designed with regular, ongoing user testing. Following this testing, we have now updated the tables to include explanations of the purpose of the Performance Tables and the signposting of the reasons why schools may have zero per cent recorded against them. This was released in late September.

In terms of the impact of the exclusion of unregulated international GCSEs, commentary was provided in our statistics release on 13th October on Key Stage 4 Performance which made this clear to users.

Following discussions with users and internally, we will meet our commitment to make an alternative version of the National Pupil Database (NPD) which includes unregulated international GCSEs on a comparable basis no later than the release of the Performance Tables in October 2017. If possible, we will do this earlier. This process takes time as we need to apply the similar discounting rules to the exams we currently include. Failure to do so would mean the data provided a misleading picture for users on pupil attainment. We consulted with users at the September NPD User group event. There was strong agreement that this database should be available through the normal NPD access processes with the condition that data cannot be published at school-level. Users agreed that multiple measures of performance were not helpful at a school level and that there was a need to assess and research the performance of the education system in the round. With the NPD User Group we will review ongoing demand on all issues relating to the NPD.

Later this year we will also publish further guidance on the process for users requesting supplementary data.

I would, as always, be happy to discuss any of the above with you. I am copying this letter to the National Statistician, John Pullinger.

Yours sincerely,

A handwritten signature in cursive script that reads "Iain Bell". The signature is written in black ink and is positioned above the printed name.

Iain Bell

Annex A: Clauses for Access - the revised draft clauses associated with NPD extract access.

These draft clauses will be refined in collaboration with the NPD Steering Group.

The following rules are in place to:

- a) Ensure that NPD data is being used for legitimate purposes by 3rd party users and maintain public confidence in the use of this data.
 - b) Ensure that DfE can better identify inappropriate use which is not related to a permitted use to stop it being placed in the public domain.
 - c) Allow DfE to support users by advising on issues related to use of our data in the public domain.

9.1 Each NPD access request is for a specific purpose. The Requester is not permitted to conduct analysis, nor publish analysis, beyond the scope of the approved Permitted Use. This constitutes a breach of access.

9.2. Should a user wish to conduct analysis on data held for an existing Permitted Use for an additional reason, they must seek permission from DfE via the existing approval process and demonstrate to DfE that the analysis is within permitted use.

9.3. Where research / analysis is going to be published into the public domain, the Requestor should notify DfE a minimum of 2 working days before publication wherever possible. High level description of the research/analysis (relating it back to the original request) is all that is required. The findings/content of that research/analysis are not required.

9.4 DfE recognise that the 2 working day rule will not always be possible without slowing down or reducing impact of analysis. In some such circumstances where:

- a) A permitted use case is already in place relating to the planned analysis, and
- b) Analysis is being produced, or quoted more quickly than 2 working days notice in order to contribute to debate or announcements,

then the Requestor should notify DfE at the earliest opportunity so that DfE is able to advise on any issues on a case by case basis.

9.5 Consent is only valid if you have informed us in writing (via the usual email address) of your intent to publish, in advance of reproduction or publication of the NPD Data.

9.6 The Requester undertakes that whenever the NPD Data is reproduced or used in a publication by it or on its behalf an attribution and caveat are included on behalf of DfE in a form previously approved in writing by DfE. In

any event, the caveat must list the Data Controller whose data has been used in the publication and state that they do not accept responsibility for any inferences or conclusions derived from the NPD Data by third parties.

9.7 For the avoidance of doubt, inclusion of the NPD Data on an internet website will be deemed to constitute publication for which the prior written consent of the Data Controller is required.

9.8 The Requester undertakes to ensure that all statistics published are at a level of anonymisation and aggregation which will ensure that no Personal Data or Sensitive Personal Data are published, and will thereby ensure the confidentiality of individuals. More specifically, the Requester undertakes to adhere to the Standard Disclosure Control set out below in any publication or reproduction of the NPD Data unless an alternative disclosure control mechanism has been agreed in writing in advance of publication by the Data Controller.