

---

**Director General for Regulation**

Jen Persson  
(By email)

21 March 2018

Dear Jen,

**NATIONAL PUPIL DATABASE DATA ACCESS**

Thank you for your email of 4 November 2017 raising various concerns about the Department for Education (DfE) and their handling of the National Pupil Database (NPD). The concerns highlight important issues about public confidence in the NPD.

The NPD is one of the UK's most valuable data assets. It provides statistical insights that help to improve children's lives that would not otherwise be possible. We spoke to many data users in the course of this investigation who highlighted the need for more data to help understand vulnerable children's needs and improve their outcomes.

The specific actions we have asked DfE to take in response are set out in the annex to this letter, as are more details about our investigation. We have also written separately to DfE.<sup>1</sup> Some of the concerns raised were about Data Protection Act compliance which is for the Information Commissioner's Office (ICO) to determine. Our primary concern is the use of the NPD to support statistical analyses. However, we consulted the ICO during our investigation and will continue to work closely with them about NPD issues in future.

We recognise that access to this resource must be appropriately governed to safeguard children's privacy and maintain public confidence. The steps required to safeguard data privacy are constantly evolving. We are confident that DfE recognises this and we support the work they are undertaking to modernise data access procedures. DfE is currently consulting with NPD users and a wide range of external ethics and data experts to ensure that any proposed solutions strike an appropriate balance between data access and data privacy. We welcome DfE's commitment to involving users in this process.

However, the transparency and communication of this work could be better and we have asked DfE to address this. We are also recommending DfE do more to demonstrate how the National Pupil Census and NPD safeguard data privacy and comply with data protection legislation by: using the Five Safes framework<sup>2</sup> in all NPD outputs; conducting a voluntary Data Protection Impact Assessment<sup>3</sup> of the NPD; and proactively involving ICO when making major decisions about areas in their remit. Actions like this are particularly helpful when new developments happen, such as the recent changes to the Alternative Provision Census.

Maintaining public confidence is vital when governments collect and use data about citizens to deliver services and improve outcomes. More public engagement could contribute valuable insights and perspectives about DfE's practices and ongoing proposals for NPD developments. We have recommended that DfE explore options to consult pupils and parents on a more formal basis, and

---

<sup>1</sup> <https://www.statisticsauthority.gov.uk/correspondence/national-pupil-database-access/>

<sup>2</sup> See information here about the Five Safes - and the section at end of annex

<sup>3</sup> ICO guide to data protection impact assessments

that they test the privacy notice wording with pupils and parents directly. They have started work on the privacy notice testing.

We assessed DfE's response to Parliamentary Questions (PQs) 109065<sup>4</sup> and 106644<sup>5</sup> about NPD data. The PQ response was correct to state that no "personal confidential identifying data" was shared. However, the response also contained an inaccurate statement, which we are satisfied was inadvertent, rather than a deliberate attempt to mislead. We have asked DfE to improve their communication about this and they are now working with external experts to address this concern.

DfE has told us that onward data sharing with third-parties only happens in a small minority of cases when data are shared. These are subject to the same data security assessment and access conditions as all other data shares. However, transparency about sharing data is important so that practices can be scrutinised using the Five Safes framework. We have asked DfE to ensure that information is published about all cases involving onward data sharing.

We will proactively monitor progress on these recommendations via our regular meetings with DfE. In addition, I have arranged to meet with DfE's Head of Profession for Statistics and Permanent Secretary, to discuss them. We will publish an update to this letter to report on progress in September 2018. As a result of our investigation, OSR, DfE and ICO will meet regularly – at this stage we envisage least three times a year - to discuss the NPD starting in April 2018.

Thank you for raising these important issues and for your patience while we conducted our detailed investigation.

Yours sincerely



Ed Humpherson

Director General for Regulation

---

<sup>4</sup> PQ 109065 response to Darren Jones MP from the Rt Hon Nick Gibb MP

<sup>5</sup> PQ 106644 response to Darren Jones MP from the Rt Hon Nick Gibb MP

## Annex

### Introduction

The concerns raised relate to:

- The accuracy of DfE's response to parliamentary questions (PQs) about data identifiability
- NPD data access protocols, including third-party onward sharing
- Privacy impact assessments and the Alternative Pupil Census
- Engagement with pupils, parents and schools
- Progress on previous OSR recommendations to DfE about the NPD
- Pupil census data validation
- Data Protection Act (DPA) compliance

### Parliamentary Questions (PQ)

You asked us to review the accuracy of the Schools Minister's PQ response (109065<sup>6</sup>) to Darren Jones MP's questions about data sharing. The PQs asked how many records provided to the Telegraph in 2013, and to the BBC in 2014, involved "personal confidential identifying data".

The data shared was de-personalised, meaning that identifiers such as name, date of birth and postcode had been removed. It did not include personally identifying data items. A short and accurate answer to these PQs could have been: None.

Instead, DfE provided a more detailed answer that attempted to describe the ways in which the data had been de-personalised using non-technical language. In doing so we believe they made an inaccurate statement, which you highlighted to us, that the data did not have the "potential to allow children to be identified". While de-personalised data cannot on its own be used to identify individuals, it carries a risk that this could happen if used in combination with other information. An organisation or individual doing this would be in contravention of the Data Protection Act (DPA), and the rules governing access to the NPD data. Given the nature of the de-personalised data that was shared, and the controls in place to assess its use, in this scenario the DfE deem that the risk of re-identification was minimal.

While the response was unhelpful in places, we are satisfied that the inaccurate language was inadvertent, rather than a deliberate attempt to mislead. To avoid this situation arising again, and to improve public understanding of this complex area, we are advising DfE to:

- Publish a "public guide to data identifiability and the NPD" drawing on the identifiability spectrum framework developed by the Wellcome Trust's Understanding Patient Data programme.<sup>7</sup>
- Incorporate the concept of the identifiability spectrum, and the suggested terms for different levels of data, into materials about NPD, and other relevant materials about data.

You also asked us to consider a PQ response (106644<sup>8</sup>) on the matter of onward data sharing with third parties. DfE told us that it estimates only a small minority of data shares involve a third party onward sharing agreement (which will be subject to the same data security assessment and access conditions as all other data shares). We welcome the steps DfE has taken to increase transparency about NPD use, culminating in the new public database of data shares released in December 2017. DfE does not feel that adding an extra field about third party onward sharing to this database that would be blank in almost all cases is helpful. However, the contents of the database are being reviewed as part of work to refresh information in line with GDPR and DfE is considering this issue.

---

<sup>6</sup> PQ 109065 response to Darren Jones MP from the Rt Hon Nick Gibb MP

<sup>7</sup> Understanding Patient Data guide to identifiability

<sup>8</sup> PQ 106644 response to Darren Jones MP from the Rt Hon Nick Gibb MP

We believe that transparency about all data sharing is important to enable scrutiny of practices using the Five Safes framework<sup>9</sup> and are recommending that DfE:

- Ensure that information about all onward data sharing is made public.

### **Alternative Provision (AP) Census**

You asked whether DfE should have conducted a privacy assessment prior to the AP census. We understand that this is not a legal requirement under the DPA. However, we think doing this on a voluntary basis could help build public trust. We therefore recommended that DfE:

- Consider conducting a Data Protection Impact Assessment of the NPD.<sup>3</sup>

DfE told us that the changes to the AP census relate to information already required (and held) by local authorities during the process of commissioning placements in alternative provision and do not require the collection of any additional information from the individuals themselves. We have also spoken to data users who highlighted the need for more data to be available about vulnerable children to better understand their needs and improve their outcomes.

In addition to this voluntary impact assessment, and the steps we are recommending in this letter such as greater ICO input, we welcome the broader steps DfE is currently taking to reform their approaches to data ethics, governance and access.

### **Engagement with pupils, parents and schools**

You queried the pupil census privacy notice wording, correspondence with schools, input from the Information Commissioner's Office (ICO), and potential amendments to align with the new General Data Protection Regulation (GDPR). DfE told us that privacy notices are reviewed on an annual basis and will therefore be reviewed again during spring / summer 2018. The most recent review (in advance of the 2017/18 academic year) focused on ensuring the wording was compliant with both the current DPA and GDPR. DfE proactively approached ICO for input on the autumn privacy notice, which they actioned in full, and we are pleased that they took this approach. DfE also sought input from representatives of schools and local authorities. However, direct testing was not carried out with children. The next review will take place once GDPR has come into force and so will remove aspects relating to the current DPA.

While the current notice is DPA / GDPR compliant we recommend that DfE:

- Conduct user-testing with both parents and children (with a range of abilities and ages) of the pupil census privacy notice and any associated information materials for parents and children, as part of all future reviews of these materials where substantive changes are planned.

Increasing public engagement activities could contribute valuable insights and perspectives about DfE's practices and ongoing proposals for NPD developments. We also recommend that DfE:

- Explore options for setting up parent and pupil panels, similar to the public panels that the Administrative Data Research Network has established in Scotland and Wales,<sup>10</sup> and the ALPHA young people's advisory group established by DECIPHer in Wales.<sup>11</sup>

### **Follow-up on previous commitments made by DfE**

You raised DfE's progress on some of the recommendations arising from issues you raised with us in 2016<sup>12</sup> and 2017.<sup>13</sup> These were: an invitation for you to join DfE's Data Access Management

---

<sup>9</sup> See information here about the Five Safes - and the section at end of annex

<sup>10</sup> Administrative Data Research Network public panels

<sup>11</sup> ALPHA young people advisory group

<sup>12</sup> Letter from Ed Humpherson to Jen Persson April 2016

Panel (DMAP); a suggestion to write to Head Teachers before the autumn census; a commitment to publish DfE's review of data access policy by April 2017. We accept that there is some learning for us here in terms of monitoring actions following our suggestions. We now have regular scheduled meetings with the DfE and these discussions include progress updates on recommendations arising from casework.

DfE's data governance arrangements were updated after the original suggestion for you to join DMAP. A new strategic Data Sharing and Publication Board sitting above DMAP has been introduced and DMAP now meets weekly to review applications. We understand that you were instead invited to join the NPD user group (which you have done), but the NPD user group is not currently meeting regularly while DfE undertakes detailed research with users as part of the discovery phase of its modernising access project.

Correspondence with head teachers is very limited to minimize burden on schools. We understand that DfE decided not to send a letter to head teachers prior to the 2017/18 autumn census, as we had suggested, because other information was being sent to schools concerning the school census to highlight key information. DfE felt that an additional letter was not needed, but we were not informed of this at the time.

The announcement in April 2017 of the general election to be held in June meant that departmental publications ceased. We followed this up with DfE in June, following the election, and understood at that time that a statement from DfE was due imminently. Work on the data access review is still underway, and is covering three areas:

- 1) Increasing the transparency of routine data sharing  
In December DfE published a new database containing all external DfE data shares. We are very pleased to see this information in the public domain.
- 2) Harmonising DfE's governance processes  
Progress on this has been slower than initially expected as the scope of the work was expanded. DfE is expecting to have a single approvals panel for all routine and ad hoc third-party data shares from March 2018. We know that users will welcome this development.
- 3) Modernising access channels  
The first phase of this work looked at options for a privacy controlling API, we understand that DfE have shared a copy of the report with you. The work is now in its discovery phase, exploring other options that DfE think could be more suitable to meet the need to ensure safe settings for data access. This phase is expected to be completed at the end of February 2018, after which DfE expects to begin developing prototypes of solutions, aligned with the five safes framework.

Our regular meetings with DfE include progress updates on this wide package of reform work. DfE could be more proactive about reporting progress on the various stages of this work, so we are recommending DfE:

- Publish a statement outlining progress on all the modernising data access work currently being conducted, with details of the expected next steps and timeframes for key milestones. Aligning this work with the Five Safes<sup>14</sup> framework would be useful. The statement should be prominent, easily accessible, and its availability publicised widely.

### **Data safeguarding and Data Protection Act compliance**

Some of your questions require us to make judgements about compliance with the Data Protection Act (DPA). These include: whether pupil census data are processed lawfully and fairly; children's rights to access individual level data on themselves; and whether the NPD meets the requirements

---

<sup>13</sup> Letter from Ed Humpherson to Jen Persson February 2017

<sup>14</sup> See information here about the Five Safes - and the section at end of annex

for s33 exemption from the DPA. In the context of the new *Code of Practice for Statistics*<sup>15</sup> we are interested in the extent to which the department looks after people's information securely and manages data in ways that are consistent with relevant legislation and serve the public good. However, it is the ICO who is competent to assess whether specific data collection processes are DPA compliant. We cannot, therefore, provide judgements about these specific issues. However, we understand that DfE is currently reviewing its procedures in line with the adoption of the GDPR provisions in May 2018; this review includes subject access requests. Greater clarity about how data are safeguarded, privacy is protected and processes comply with data protection legislation would all help DfE demonstrate its compliance with the *Code of Practice's* data governance principles, and, ultimately, its trustworthiness. We are recommending that DfE:

- Align all relevant NPD outputs (e.g. NPD: data flows and data sharing<sup>16</sup>; NPD user guide<sup>17</sup>) with the Five Safes framework, and include details of any plans to develop DfE's practices in any of the five areas. Update these at regular intervals to underline the fact that safeguarding data privacy is a continually evolving process.
- Provide details outlining how the National Pupil Census and NPD will be GDPR compliant, highlighting any aspects that ICO has been consulted on or provided assistance with.
- Proactively involves the ICO in all major decisions covered by their remit for data processing and privacy safeguarding related to the pupil census and NPD.

ICO and OSR have complementary roles in this area and we would benefit from working together more closely. We have agreed to arrange joint meetings between us, ICO and DfE at least three times a year to discuss the NPD. The first of these meetings is happening in April 2018.

### **Pupil census data validation**

You asked for our view on the data validation carried out of the pupil census. In 2017 we assessed statistics based on data gathered in the pupil census and asked DfE to improve its documentation of the data validation and quality assurance processes.<sup>18</sup> These have now been judged to meet the high standards required of National Statistics and we highlighted these aspects in our recent decision to confirm this designation.<sup>19</sup> However, we expect all statistics producers to continually monitor their quality assurance processes, as underlined by the Assured Quality principle in the new Code of Practice, which includes the following practice:

*Systematic and periodic reviews on the strengths and limitations in the data and methods should be undertaken. Statistics producers should be open in addressing the issues identified and be transparent about their decisions on whether to act.*

### **The Five Safes framework**

The Five Safes framework is a set of principles that can be used by organisations to help safeguard data and privacy. Each of the principles sets out processes that can be used to minimise the risk that data will be used incorrectly. Depending on the context, data owners may choose to give greater weight to different principles but, in all cases, the controls in place across the five should ensure that confidentiality is maintained. The five principles, and the aspects they address are:

- Safe data: what steps have been taken to remove items that could identify individuals?
- Safe people: do the users have the necessary technical skills to use the data, do they understand the importance of data confidentiality and have they completed all necessary training?

---

<sup>15</sup> Code of Practice for Statistics

<sup>16</sup> NPD data flow and sharing document

<sup>17</sup> NPD User Guide

<sup>18</sup> Assessment of statistics for England on schools pupils and their characteristics and on absence and exclusions

<sup>19</sup> Confirmation of National Statistics designation

- Safe projects: is it an appropriate use of the data, ethical and clearly for the benefit of the public?
- Safe settings: where will the data be used and what steps are in place to ensure the data are kept safe?
- Safe outputs: what processes are in place to ensure that outputs produced from the data cannot be used to identify individuals?