Written evidence submitted by Jen Persson (BIG0085)

The National Pupil Database - Big Data Dilemma

defenddigitalme's campaign asks the Department for Education (DfE) to change their policies and to protect 8+ million[1] children's identifiable personal data in the National Pupil Database (NPD):

- Stop handing out identifiable personal data to commercial third parties and press
- Start telling pupils, their guardians and schools what DfE does with personal data
- Be transparent about policy and practice

We want to see legal and regulatory frameworks fit for our children's digital future and call for:

- secure handling of sensitive identifying pupil data
- statutory privacy impact assessments and public consultation
- the legislative review of DfE sharing of children's personal data
- the separation of consent for identifiable data required for school administration from secondary use commercial purposes

We are supported by a number of parents, pupils, legal, data privacy and data protection experts.

8 million children in "one of the richest education datasets in the world"

The National Pupil Database affects the identifiable personal data of over 8 million children and is "one of the richest education datasets in the world," [2] ca 600,000 new pupils are added every year.[3]

Any discussion of the opportunities of "Big Data" tends to focus on gains for organisations, summed up in the enquiry as 'business data worth £31bn by 2016'.[4] These benefits will only be delivered by "a highly skilled digital workforce"[5] that is missing due to an "educational gap."

That report recommended government takes steps to increase young people's understanding of the digital world and how data are used, but in reality it has not happened. Children who are our future workforce and will live with the privacy impacts of decisions made today in the regulation of "Big Data," often have no voice[6] or visibility of how their own data are used in "Big Data".

Nearly every child in the country has their personal data stored in this database, shared with third parties including the press, and our enquiries to date show that the public know nothing about it. It is illogical to train data-skilled adults in some areas but keep them intentionally ignorant in others.

The risks of "Big Data" for children should be given particular attention when considering how children's personal data are used; based on their vulnerability, lack of ability to consent, and reliance on adults making decisions on their behalf, which may affect their present and future.
The recommendation of the 2014 Report "Responsible Use of Data" [7] should now be enacted; "*the Government has a clear responsibility to explain to the public how personal data is being used*."

---

1 http://www.fft.org.uk/FFT/media/fft/Downloads/FFT-Story.pdf

2 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/472700/NPD_user_guide.pdf

3 http://www.ons.gov.uk/ons/rel/vsob1/birth-summary-tables--england-and-wales/2013/stb-births-in-england-and-wales-2013.html

4 http://www.parliament.uk/business/committees/committees-a-z/commons-select/science-and-technology-committee/news-parliament-2015/big-data-tor/

5 2013 Information Economy Strategy and Science and Technology Committee Report 2014 Responsible Use of Data

6 http://www.raeng.org.uk/publications/reports/privacy-and-prejudice-views

7 http://www.publications.parliament.uk/pa/cm201415/cmselect/cmsctech/245/245.pdf

Written evidence submitted by Jen Persson (BIG0085)

About the National Pupil Database

1. The national pupil database (NPD) contains the detailed information of approximately 8+ million pupils in schools and colleges in England, aged 2-19. The data include:

    - personal, identifying and sensitive [8] individual level information, such as full name and address, gender, ethnicity, first language, eligibility for free school meals, special educational needs (SEN), pupil absence and exclusions, indicators for looked after in care status and service family
    - test and exam results, prior attainment and progression at different key stages for pupils in the state sector, and in non-maintained special schools, sixth-form and further education colleges for children and young people
    - individual pupil level attainment data for pupils in non-maintained and independent schools who partake in the tests/exams [9]

2. The data available in the NPD is divided into 4 tiers and can be requested in a number of different combinations including identifying and sensitive data[10] in the top two, though all 4 are identifying:
    a. Tier 1: highly sensitive, identifying personal data (e.g. name, DOB, postcode, ethnicity, SEN, disability, service family, unique pupil number, looked-after status, exclusions)
    b. Tier 2: other personal, sensitive, identifying information (higher level SEN, FSM, language)

3. Data are released and sent to the applicants in their own setting. This distribution of raw data at users own sites should change for what may be identifying, sensitive, individual level data. We do not feel that this is a secure practice for sharing sensitive personal data fit for the 21st century. The data recipient should come to the data in a safe-setting[11], not have data sent to their 'home' setting.

4. Case Study releases from the NPD: identifiable and sensitive data given to press
    a) a television journalist in August 2014.[12]
    b) ten Telegraph journalists in February 2013[13]. This release was of an estimated 10m individuals' records. The items released did not include name, but included SEN and more sensitive data. [14]

    Further releases include commercial third parties such as data management consultancies, think tanks, charities and "one-man shows" as well as more Fleet Street papers.

The impact of Big Data policy change on privacy in practice

5. The public should know and understand what is done with their personal data, by who, and why. The Department for Education amended the release policy and legislation to permit wider sharing in 2012. Brief public consultation[15] ran in late 2012. Of the total 95 responses the government received only two were teachers or Headteachers, and eight were parents. It is perhaps therefore unsurprising that staff and parents do not recall any announcement of the change. However public polling shows that the public want to be asked before sharing data, especially of under 16s.[16]

---

8 http://www.bristol.ac.uk/media-library/sites/cmpo/migrated/documents/2011censususerguide.pdf

9 http://www.bristol.ac.uk/media-library/sites/cmpo/migrated/documents/ks2userguide2011.pdf

10 https://www.gov.uk/guidance/national-pupil-database-apply-for-a-data-extract

11 http://www.adrn.ac.uk/protecting-privacy/secure-environment

12 https://www.whatdotheyknow.com/request/293030/response/723407/attach/10/BBC%20Newsnight.pdf

13 https://www.whatdotheyknow.com/request/293030/response/723407/attach/3/Daily%20Telegraph.pdf

14 https://www.whatdotheyknow.com/request/293030/response/738135/attach/2/Annex.pdf

15 https://www.education.gov.uk/consultations/downloadableDocs/Prescribed%20Persons%20Consultation%20ResponseFinal.pdf

16 http://www.jrrt.org.uk/sites/jrrt.org.uk/files/user-uploads/IpsosJRRTprivacypollMay2014full.pdf

6. In our enquires with 40 schools to date none recall any communication since 2012 to inform them, or ask them to inform pupils that their personal data may be given to third parties. One replied simply, 'we've never heard of it', and another said that the most recent update about data privacy was sent from the local authority in 2010 - that preceded the legal and policy change in 2012/13.

7. No schools that responded, in a range of primary and secondary schools, give their pupils or parent/guardians any statement about the use of the NPD data by commercial third parties in their data protection policies and/or privacy notices and to date this appears to be because the DfE has failed to effectively communicate these onward sharing purposes to schools, staff or pupils.

8. The majority of policies viewed online to date include only one sentence from the suggested privacy notice wording from July 2015, *"The school is required to share some of the data with the Local Authority and with the DfE."* Some schools link to the DfE webpage, and that also fails to transparently mention the National Pupil Database or onward sharing to commercial third parties.

9. The DfE suggested privacy notice template [17] published in July 2015, is the only mechanism that the DfE actively provides to schools to communicate this to pupils and guardians. The template fails to transparently mention onward sharing with commercial third parties or press. In discussion DfE agreed it is unreasonable to expect pupils/guardians to 'jump through hoops' to find it.

Transparency of governance and oversight, security and destruction

10. Transparency of releases in the public domain was the goal of the NPD third party register[18] begun in 2012. However its trustworthiness is only as good as its accuracy. According to an application [19] made by The Cabinet Office (CO) to receive named, identifiable data in 2013, the CO had already received similar data of 16-18 year olds from the NPD for the purposes of "piloting an approach in the electoral register firstly in June 2011 and March 2013". Those two releases do not appear in the release register. It is also updated erratically, and only three times since 2012 in total.

11. Transparency of governance in how this "Big Data" store is managed and its release decision- making process would be increased if the register were published more frequently, audited for accuracy to add any missing requests, and if the Terms of Reference [20] were published regards the oversight and data release panel (DMAP) at the Department for Education - responsible for Tier 1 data approvals and Fast Track process - and also for all other data approval decisions. Additional independent panel oversight would be a further beneficial step towards increased transparency.

12. Data must be held securely, is Principle 7 of the Data Protection Act 1998.[21] This requires oversight. As of July 2015, no audits had been carried out of organisations or individuals that were data recipients before July 2015. An audit of all recipients is now advisable, as when another state body did this in the past for health data[22], it not only found poor practices that were changed, but restored public and professional trust in the process going forward. The cost of this in future could be reduced or avoided if raw data were not released to third parties in their own settings.

---

17 http://jenpersson.com/?attachment_id=4728

18 https://www.gov.uk/government/publications/national-pupil-database-requests-received

19 https://www.whatdotheyknow.com/request/293030/response/723407/attach/6/Cabinet%20Office%20060913.pdf page 8

20 https://www.whatdotheyknow.com/request/pupil_data_application_approvals

21 https://ico.org.uk/for-organisations/guide-to-data-protection/principle-7-security/

22 https://www.gov.uk/government/publications/review-of-data-releases-made-by-the-nhs-information-centre

13. Case study: on data retention and destruction. In response to a Freedom of Information request, the Department for Education confirmed that the recipient and department have failed to comply with the destruction requirement after data are used, stated in its data release process. The 2012 application form states that written confirmation is required from recipients to confirm destruction, when the agreed time period expires, but in this case study it had not been received almost two years later[23]. We await confirmation how many notices have never been received.[24]

Consent: the importance of clear privacy notices to understand what we sign up to

14. While consent may be one condition for processing, this does not mean that by consenting to use people trade all their privacy rights. The processing must still be fair and lawful, and pupils and guardians retain their rights under the DPA. UCAS was forced by the Information Commissioner's Office in early 2015 to change their blanket consent process which enabled and obliged commercial use. [25] UCAS since did a survey in which 37,000 students replied and 90% said they wanted to be asked before data sharing. These same students' data are likely stored in the NPD.[26]

15. Consent cannot be binary. At the current time, pupils and guardians must consent to the school sharing personal data with the Local Authority and Department for Education on all data protection and privacy notices reviewed date. However, the Department for Education does not request consent for any of the secondary uses of the data. Data that are submitted by guardians with the assumption it is for the purposes of administering a pupil's schooling. [27] Consent cannot fairly and reasonably be required for school administration, and also mean consent for use by press. DfE's privacy notice template [28] and usage statement are inadequate and misleading.

16. Signing the annual personal data print-out that schools offer - the current consent mechanism for processing the data and sharing with the DfE - should be able to separate the consent for pupil data sharing with schools for direct administration from commercial secondary purposes. Since this mechanism exists already, it would not be unduly burdensome to add text to an existing form.

17. Further concern may be felt if the public finds their data stored abroad. We have concerns that the Department for Education is considering an application to the National Pupil Database from the US.

Big data uses in public interest research and open access: assessing risks, measuring benefits

18. Bona fide public interest research enjoys public support. [29] Creating barriers to this which do not fix the problems they are said to be trying to solve should be avoided. Recent changes introduced in the NPD application process in November 2015 [30], include a requirement on researchers for the lowest level criminal record check, and comes at a cost. Due to the expiration of relevant offences, the check is illogical, creating a new barrier to public interest research that will not make data any more secure, and unlikely to be of any benefit to children whose data are released 'into the wild'.

---

[23] https://www.whatdotheyknow.com/request/293030/response/723407/attach/3/Daily%20Telegraph.pdf

[24] https://www.whatdotheyknow.com/request/pupil_data_sensitive_data_releas

[25] https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2015/04/ucas-application-form-to-be-changed-following-ico-concerns-over-marketing-practices/

[26] https://www.ucas.com/corporate/news-and-key-documents/news/37000-students-respond-ucas%E2%80%99-applicant-data-survey

[27] https://www.gov.uk/government/publications/data-protection-and-privacy-privacy-notices

[28] NPD data privacy template: https://www.gov.uk/government/publications/data-protection-and-privacy-privacy-notices

[29] Royal Statistical Society / Ipsos MORI work on the Data Trust Deficit and Recommendations for Policy Makers (incl. public concern on anonymisation and commercial use): http://www.statslife.org.uk/images/pdf/rss-data-trust-data-sharing-attitudes-research-note.pdf

[30] http://defenddigitalme.com/2015/11/access-to-school-pupil-personal-data-by-third-parties-is-changing/

19. These releases include identifiable and sensitive data. No Privacy Impact Assessment had been made, at the time of enquiry with the DfE (July 2015) and given the sensitive and identifiable releases of data, independent Privacy Impact Assessments must be made with urgency and published for the various databases, datasets and providers which store, use and release pupil data.

20. Big data must assess its proportionate risks, the bigger the dataset on an individual the bigger the number of pieces that can be used in 'jigsaw' identities together. Often rendering data anonymous is seen as unwanted.[31] Identifiers are not stripped, but used to link data in predictive big data models, risking discriminatory intervention. There can be confusion between correlation and causality[32].

21. Big Data gathering can also be burdensome so must deliver benefits that can be evaluated against its tangible and intangible costs. The benefits of NPD sharing is not measured today by DfE as is done in other organisations[33]. The founder of a schools' information provider, reportedly said in Nov 2015 [34], *"Teachers are having to enter more and more data, because the government wants more insight, so the problem is going to exacerbate. It's causing a lot of workload issues."*

22. Open data that are safe to be used by all, are already widely and safely published from the NPD. However a key problem with release of individual level data from this dataset is that by identifying one unique characteristic even in a whole class (such as an unusual GCSE combination) means you can read every other piece of data about an individual from the record. Data sharing needs privacy and risk assessment before various kinds of release are considered.

Broader big school data concerns: extending children's digital footprint

23. Children's data sharing concerns have also been raised by other organisations more widely. An international project looking at websites and apps[35] used by children raised concerns over the personal information collected about 41% of the 1,494 websites and apps considered, particularly around how much personal information was collected and how it was then shared with third parties. The Global Privacy Enforcement Network (GPEN) Privacy Sweep saw 29 data protection regulators around the world look at websites and apps targeted at, or popular among children.

    - Only 31% of sites/apps had effective controls limiting collection of children's personal information.
    - 67% of sites/apps examined collected children's personal information

24. When it comes to children we must ask whether it is a good idea to extend children's digital footprints to join location, personal identifiers and behaviours, between physical and digital assets, when they are not in a position to understand notions of consent. The risks this exposure to Big Data can also create in schools is highlighted in the recent breach at an educational toy provider.[36]

25. With reference to the Science and Technology - Second Special Report, Current and future uses of biometric data and technologies, September 2015 [37] we urge policy- and lawmakers to remember to

---

[31] http://nuffieldbioethics.org/wp-content/uploads/Biological_and_health_data_web.pdf Nuffield Council on Bioethics The collection, linking and
use of data in biomedical research and health care:ethical issues

[32] http://www.bbc.co.uk/news/technology-29824854

[33] https://adrn.ac.uk/research-projects/case-studies

[34] Data demands: Are management information systems a help or a hindrance? J. Stauffenberg Schools Week November 2015 http://schoolsweek.co.uk/coping-with-the-demands-of-data/?

[35] https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2015/09/questions-raised-over-children-s-websites-and-apps/

[36] Blog: http://www.troyhunt.com/2015/11/when-children-are-breached-inside.html

[37] http://www.publications.parliament.uk/pa/cm201516/cmselect/cmsctech/455/455.pdf

include children's data and their particular uses and needs when considering big biometric data. The use of eye scanning analysis software[38], RFID tagging[39] and fingerprinting[40] have become common in schools with little public consultation, transparency or privacy impact assessment.

Conclusions

26. Communication of what is done with individuals' data - by whom and for what purpose - through privacy policies and consent forms must be clear and complete to fulfil their obligation. Not only to meet organisations' legal requirements, but to help people grow to be trusting 'digital citizens.' The ethical duty to make sure it is understood by children who are open to exploitation in lack of understanding, must be met by commercial and public bodies, if they are to remain trustworthy with the public. Consent must mean we understand what we sign up to and be told if it changes.

27. Children and young people must be trained and empowered to understand how to navigate their digital rights as well as technology without commercial pressures, and at the same time not be expected to first need to learn where systems may exploit them in order to prevent it happening.

28. Children's personal data should not be seen as a commodity to exploit unless we are willing to consider their exploitation in the digital world as more acceptable than in the real one. As they grow up their lives will have less distinction between the two in commercial transactions, interactions with the state, and socially. We cannot continue to separate the exploitation of data, from the exploitation of the person, particularly when it comes to children's personal data.

29. Regards the National Pupil Database, its current levels of security, transparency, governance and oversight do not meet that required due to the sensitivity and volume of its "Big Data" held.

30. The Department for Education must change its current practices in handling of NPD to:
    a. Stop handing out identifiable personal data to commercial third parties and press
    b. Start an immediate audit of data sharing to date with commercial third parties
    c. Start independent oversight and increased governance of data handling and release
    d. Start statutory Privacy Impact Assessments and public consultation
    e. Start secure handling of sensitive identifying pupil data
    f. Separate consent for data used for school admin from secondary commercial uses
    g. Start telling pupils, their guardians and schools what DfE does with personal data
    h. Start a legislative review of DfE sharing of children's personal data
    i. Be transparent about policy and practice, with regular publication of releases and their benefits

31. Big data uses must assess privacy risks, and measure the benefits of data use, set against the burden of collection, storage and sharing. As regards 'big pupil data' specifically in the National Pupil Database, "collect it all", is neither a strategy nor solution. The DfE current policies and practices place the burden of risk of Big Data on our children, without measured benefit.[41]

We are happy to discuss any details or questions the committee has.

December 2015

---

[38] https://lccdigilit.our.dmu.ac.uk/2015/05/13/eye-gaze-for-assessment-project-update/

[39] http://rfidinschools.com/2014/12/20/final-update-on-west-cheshire-college-rfid-tagging-students/

[40] http://www.independent.co.uk/news/education/education-news/privacy-concerns-raised-as-more-than-one-million-pupils-are-fingerprinted-in-schools-9034897.html

[41] https://www.whatdotheyknow.com/request/pupil_data_application_approvals#outgoing-482241