

Written evidence submitted by defenddigitalme (DIG0021)

About defenddigitalme

defenddigitalme is a non-profit, non-partisan, data privacy and digital rights group. We aim to make all children's data safe, fair, and transparent across the education sector in England. Our work is funded by a grant from the Joseph Rowntree Reform Trust Ltd.

Executive summary

1. We address two questions set out in the enquiry terms of reference, *“the extent to which Government datasets are made available to private-sector and academic service developers,”* and *“How well the Government and its agencies deploy their datasets to maximise their value for money,”* specific to the applied use of national education data, and pupil level data from schools
2. Some of the current issues with national pupil data handling, were included our evidence¹ to the previous Committee's 2016 *'Big Data Dilemma' Report*².
3. One 2016 report conclusion drawn by the Committee at that time was that *“there is more to do to breakdown departmental data silos, to bring data together in order to further improve public services.”* Some of the more significant foreseeable problems as a result, at national and local level³, have become reality. These must be solved, before encouraging greater data use.
4. Data quality is being compromised in the rush to extract value from national education datasets, and the resulting flawed policy and decision making by individuals could have lasting detrimental effects on lives and institutions.
5. Government ignores the lack of public agreement process for secondary re-use of administrative data at its peril and with associated institutional reputational risk. The Lords AI Committee report⁴ published in March, recognised that there is an inherent risk how the public would react to Data Trusts, a proposed model to be established to facilitate more data sharing between organisations, because there is no social license, saying:
6. *“Under the current proposals, individuals who have their personal data contained within these trusts would have no means by which they could make their views heard, or shape the decisions of these trusts.”*

¹ Big Data Dilemma evidence submitted by defenddigitalme (December 2015) <http://data.parliament.uk/WrittenEvidence/CommitteeEvidence.svc/EvidenceDocument/Science%20and%20Technology/Big%20data%20dilemma/written/25380.html>

² Science and Technology Committee Report 2016 Big Data Dilemma <https://publications.parliament.uk/pa/cm201516/cmselect/cmsctech/468/468.pdf>

³ Data on thousands of children used to predict risk of gang exploitation (Guardian, September 2018) <https://www.theguardian.com/society/2018/sep/17/data-on-thousands-of-children-used-to-predict-risk-of-gang-exploitation>

⁴ The Lords AI Committee report, (March 2018) <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf>

7. In 2015, a total of 37,000 students responded to UCAS' Applicant Data Survey. 90% only agreed with sharing personal details outside of the admissions process only with active and informed consent⁵. This is ignored by government.
8. 69% of 1,004 parents in England of children age 5-18 polled, are unaware of third-party individual pupil level data access from the school census, according to a survey we commissioned in February 2018, carried out by Survation.⁶
9. We respectfully recommend that the committee should refer to public administrative datasets, rather than *government datasets* if the intention is to refer to the datasets gathered about individuals in the course of our interactions with the state. These are not data about government, nor are they owned by government. The government has a role often of joint Data Controller, shared with the organisation at the point of collection. The rights to use cannot be assumed, and remain in a partnership of responsibility between the individual and the public bodies. These are generally balanced in favour of the person the data are about to be informed, and where applicable, with rights to object to and restrict processing, to correct errors, to have safeguards in place on automated decision making with significant effect, and to expect data security. All use should preserve human dignity, universal rights to privacy, and come with expectation throughout the data-cycle lifetime of professional codes of conduct, and common law duty of confidentiality.
10. As set out in the Council of Europe Recommendation CM/Rec (2018)7⁷ of the Committee of Ministers, it is for member States to respect, protect and fulfil the rights of the child in the digital environment. Data used in childhood for profiling, and predictive analysis in particular, has the potential to have opaque lifetime effects. The full range of human rights enshrined in the United Nations Convention on the Rights of the Child (UNCRC), in the Convention for the Protection of Human Rights and Fundamental Freedoms (ETS No. 5), and their protocols, should be fully respected, protected and fulfilled. We call on the Committee to put this ahead of making datasets "*available to private-sector and academic service developers,*" and "*maximising their value for money*".

Recommendations

11. Understand the extent to which Government datasets are made available to private-sector and academic service developers, is often in stark contrast with the lack of access afforded to the person the data are about.
12. Consider the implications for lawful data sharing as the government is keen to expand the *extent to which Government datasets are made available to private-sector and academic service developers* these often involve uses far beyond the original intended purpose which was

⁵ 37,000 students respond to UCAS' Applicant Data Survey (2015)
<https://www.ucas.com/corporate/news-and-key-documents/news/37000-students-respond-ucas%E2%80%99-applicant-data-survey>

⁶ A survey of parents in England has found that only half of parents agree they have enough control of their child's digital footprint.
<https://defenddigitalme.com/2018/03/only-half-of-parents-think-they-have-enough-control-of-their-childs-digital-footprint-in-school/>

⁷ Recommendation CM/Rec (2018)7 of the Committee of Ministers to member States on Guidelines to respect, protect and fulfil the rights of the child in the digital environment (*Adopted by the Committee of Ministers on 4 July 2018 at the 1321st meeting of the Ministers' Deputies*)
https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=09000016808b79f7

communicated at the time of data collection, many years ago, from millions of people with whom the public body no longer have no contact, and has no satisfactory answer.⁸.

13. Encourage all government departments to fully meet their obligations to subject access under the Data Protection Act 2018 and GDPR. The right of access gives individuals the right to obtain a copy of their personal data and other supplementary information. It helps individuals to understand how and their data are processed, and check it is being done lawfully.
14. Ensure government turns the recommendation into reality from the 2014 Science and Technology Committee Report “Responsible Use of Data”⁹; *“the Government has a clear responsibility to explain to the public how personal data is being used,”*
15. Note that public engagement work carried out about public data uses has already been extensive, though it sometimes does not return the answers those who call for entirely new public engagement want to take into account. When the Administrative Data Research Network was set up in 2013, a new infrastructure for “deidentified” data linkage, extensive public dialogue¹⁰ was carried across across the UK. It concluded in very similar findings as was apparent at dozens of care.data engagement events in 2014-15.
 - There is not public support for
 - *“Creating large databases containing many variables/data from a large number of public sector sources,*
 - *Establishing greater permanency of datasets,*
 - *Allowing administrative data to be linked with business data, or*
 - *Linking of passively collected administrative data, in particular geo-location data”*
 - The other ‘red-line’ for some participants was allowing *“researchers for private companies to access data, either to deliver a public service or in order to make profit. Trust in private companies’ motivations were low.”*
16. Promote staff awareness across all levels of government, and third parties with access to these datasets of the Digital Economy Act 2017 Codes of Practice¹¹.
17. Promote accountability and institutional trustworthiness through obligatory data registers of all data processing of all public administrative datasets.
18. Promote public trust in institutional data processing by ending unlawful and unfair secondary re-use of personal confidential data in bulk in public administrative datasets for operational reasons, not for purposes for which it is collected. For example, use by the Home Office.¹²

⁸ 109014 Darren Jones MP To ask the Secretary of State for Education, what legal advice her Department has received on meeting data protection requirements for data held on the national pupil database that relates to former pupils with whom schools have no contact.

⁹ Science and Technology Committee Report 2014-15 <http://www.publications.parliament.uk/pa/cm201415/cmselect/cmsctech/245/245.pdf>

¹⁰ Administrative Data Research Network Public Dialogue (2013) <https://defenddigitalme.com/wp-content/uploads/2018/09/Dialogue-on-Data-Exploring-the-public%E2%80%99s-views-on-using-linked-administrative-data-for-research-purposes.pdf>

¹¹ Digital Economy Act 2017 part 5: Codes of Practice <https://www.gov.uk/government/publications/digital-economy-act-2017-part-5-codes-of-practice>

¹² MOU between the DfE and HO footnote comment page 5

<https://www.whatdotheyknow.com/request/495973/response/1229011/attach/7/DfE%20Draft%20MOU%20earliest%20version%20found%20Redacted.pdf>

19. Commission an authoritative list of data sharing gateways that currently exist across government at all levels. This would give clarity and confidence to departments and the public of the lawful basis of datasharing and processing at local and national levels of government, and where commercial third parties increasingly intersect with those public bodies in a changing landscape with new data protection laws in 2018. A 2014 data sharing report¹³ from the Law Commission pointed out that no authoritative list of data sharing gateways exists and its first recommendation included mapping these existing gateways.

Case study: National pupil data handling and changes in its access model

20. At local government level and at public body (school) level there is no transparency or independent oversight at all of the extent of children's administrative data commercial reuse after distribution to app developers and other third parties, due to lack of any go-to register or public visibility.

21. The Department for Education makes available data extracts to third parties from the National Pupil Database (NPD), School Workforce, Individualised Learner Record and Higher Education Statistics Agency.

22. The Department for Education does not make available extracts to the individuals the data are about through any working Subject Access process. A journalist can get more access to a child's record from the Department, than the child or parent themselves¹⁴.

23. The NPD alone contains 23 million children's personal confidential records at named level and are made available to private-sector and academic users through the Department for Education.

24. This degree of identifying, individual level data distribution is an outlier in other government departments, and across education in the UK regions.¹⁵

25. The identifying and sensitive data distribution model was partially stopped in May 2018¹⁶ and a new model began on September 25, 2018. Personal data have been given away between late 2012 and September 2018 to thousands of private-sector and academic users, without standard audit processes¹⁷.

26. Cost to the Department: Data extracts are provided free of charge¹⁸. The Department was unable to provide the total extraction costs for NPD third-party register requests since 2011 when we asked via FOI in 2016.¹⁹

¹³Law Commission: Data Sharing between Public Bodies A Scoping Report

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/330403/41831_HC_505_Law_Commission_351_Web.pdf

¹⁴ We found out 10 journalists at The Telegraph were given sensitive and identifying data in 2013 and since they requested pupil-level data no small number suppression was applied. <https://www.whatdotheyknow.com/request/293030/response/738135/attach/2/Annex.pdf>

¹⁵ Defendigitalme UK pupil datasets comparison 2018 (England, Northern Ireland, Scotland and Wales) https://defendigitalme.com/wp-content/uploads/2018/03/UK_pupil_data_comparison_27032018.pdf

¹⁶ Sharing of school pupils' data put on hold, BBC, May 15, 2018 <https://www.bbc.co.uk/news/technology-44109978>

¹⁷ PQ 106878 what audit processes her Department has implemented to ensure that third-party organisations which access the National Pupil Database store and process that data in compliance with their original agreements with the Government

<https://www.parliament.uk/business/publications/written-questions-answers-statements/written-question/Commons/2017-10-10/106878/>

¹⁸ PQ 61684 To ask whether her Department provides National Pupil Database data to named bodies and third parties free of charge.

27. Cost to the individual: Data extracts can include millions of identifying and sensitive records. The Department is unable to tell an individual to which third parties their personal data has been given away, when asked in a parliamentary question they replied, "The Department does not maintain records of the number of children included in historic data extracts".²⁰
28. As the CEO of the Open Data Institute, the ODI wrote in 2012 in submission to the public consultation on the changes to give away pupils' personal data, "the National Pupil Database is not open data" and "access to the National Pupil Database should be tightly restricted." But this advice was ignored.
29. Five years later, the Department has spent considerable time and effort to rectify this going forwards, using a new safe data access process with the ONS launched in September 2018.²¹ The potential harm done has not been assessed.
30. Dataset error rates should be published and transparent. For example, these data are longitudinal and processing involves probability based linking. There are issues where the Department blends two children's data together into one record but they're actually separate individuals and vice versa. While this is seen as a barrier to distributing the data for Subject Access, it is not for distributing for third-party use. Use of data with opaque error rates should end to encourage understanding of where conclusions drawn may be uncertain.
31. 69% of parents in England of children age 5-18 are unaware of third-party individual pupil level data access from the school census, according to a survey we commissioned in February 2018, carried out by Survation.²²
32. Pupil data are already being used for purposes far beyond children or parents expect when their data are given to a school for their education ,and without any meaningful explanation on third party use and there is no consent process for these re-uses, including linkage by academics and commercial users with Police National Computer data²³, for use in research with predictive tools²⁴ , for benchmarking²⁵, promoting private tutors, or access by journalists.²⁶
33. A number of explicit data sharing gateways for research and general ancillary powers pre-date the Digital Economy Act 2017. Local authorities, for example, may have relied on Section 2 of the Local Government Act 2000. But the encouragement to promote increased data sharing, has not

<https://www.parliament.uk/business/publications/written-questions-answers-statements/written-question/Commons/2017-01-25/61684/>

¹⁹ FOI Pupil data: security and costs 2017 https://www.whatdotheyknow.com/request/pupil_data_security_and_costs

²⁰ PQ 109065 To ask the Secretary of State for Education, how many children's personal confidential identifying data from the national pupil database was given to journalists working for the Telegraph in 2013 <https://www.parliament.uk/business/publications/written-questions-answers-statements/written-question/Commons/2017-10-23/109065/>

²¹ How to access Department for Education (DfE) data extracts <https://www.gov.uk/guidance/how-to-access-department-for-education-dfe-data-extracts>

²² A survey of parents in England has found that only half of parents agree they have enough control of their child's digital footprint. <https://defenddigitalme.com/2018/03/only-half-of-parents-think-they-have-enough-control-of-their-childs-digital-footprint-in-school/>

²³ Case study PNC data linkage with link to original DfE documents via FOI https://defenddigitalme.com/wp-content/uploads/2018/04/PNC_NPD.pdf

²⁴ Case study with link to original Institute of Criminology and DfE documents via FOI https://defenddigitalme.com/wp-content/uploads/2018/04/Cambs_Crimi_NPD.pdf

²⁵ Case study Mime consulting with original FOI links https://defenddigitalme.com/wp-content/uploads/2017/06/Mime_Consulting.pdf

²⁶ Original data for identifying pupil data by journalist <https://www.whatdotheyknow.com/request/293030/response/723407/attach/10/BBC%20Newsnight.pdf>

yet been matched by a similar enthusiasm for any duty of responsibility to transparent documentation and publication of registers by public authorities, or by accredited access facilities, about data sets they are linking, and providing to whom and for what purpose, in any format.

34. In January 2018 Local Authorities failed to carry out²⁷ any data protection impact assessment or fair processing of a new pupil dataset in the Alternative Provision education sector, labelling children with 'pregnancy', 'youth offender' and 'mental health' among other reasons for their transfer out of mainstream education, without telling families or asking for agreement or correction.

Case study: Higher Education, and Longitudinal Education Outcomes (LEO) data

35. In 2015, a total of 37,000 students responded to UCAS' Applicant Data Survey and 90% only agreed with sharing personal details outside of the admissions process only with active and informed consent²⁸. This is ignored by government.
36. The Small Business, Enterprise and Employment Act (2015), without public consultation or consent from individuals, authorised the permanent linking of the National Pupil Database with HMRC and DWP records, by DfE.
37. The result is the Longitudinal Education Outcomes (LEO) data, now on Unistats published as **experimental statistics**²⁹, but the experimental nature of the data and its substantial caveats³⁰ are left out of the published materials by the OfS, and others encouraging use by students and the public.³¹
38. It is highly unlikely these data are fit for use by Organisations which can apply to develop digital tools using graduate outcomes data to help prospective higher education students make decisions, in the SBRI: higher education open data, digital tools for prospective students competition.³²
39. David Morris, former Deputy Editor of Wonkhe wrote⁷ "LEO is phenomenally complicated and riddled with caveats", "as all financial investment disclaimers say, 'past performance is no indicator of future earnings'. Perhaps universities will have to add that to their marketing materials in the future."

²⁷ We made FOI requests to 156 local authorities with educational responsibility who have responsibility for the AP census. None had prepared or distributed any appropriate communication for families.

²⁸ 37,000 students respond to UCAS' Applicant Data Survey

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

²⁹ Experimental statistics on employment and earnings of higher education graduates using matched data from different government departments.

<https://www.gov.uk/government/statistics/graduate-outcomes-for-all-subjects-by-university>

³⁰ A beginner's guide to Longitudinal Education Outcomes (LEO) data

<https://wonkhe.com/blogs/a-beginners-guide-to-longitudinal-education-outcomes-leo-data/>

³¹ The Longitudinal Education Outcomes (LEO) data refers to a dataset of education records joined to tax (HMRC) and welfare (DWP) data. It shows whether some graduates were employed and how much they were paid.

³² SBRI: higher education open data, digital tools for prospective students

<https://apply-for-innovation-funding.service.gov.uk/competition/181/overview>

40. There is a real risk of selling students snake oil ‘knowledge’ and solutions with lifetime lasting effects on individuals’ and policy thinking.

Case study: Student Loans Company access to students’ data

41. Higher Education data protections have been removed in the Higher education and Research Act 2017, and the Technical and Further Education Act 2017, without necessary oversight infrastructure in place to prevent harm. The ‘barriers’ some see between data silos, are often ‘protections’ if seen from another’s point of view, and in place for good reason.
42. The Student Loans Company (SLC), HMRC, and private bodies including Pearson Education Ltd. have been granted potential access to all student data through the Office for Students, in data sharing agreements made without consultation, or public transparency, but in closed meetings³³ and changes passed into regulations through negative statutory instrument in July 2018 on which the House of Commons held an annulment debate³⁴. The Shadow Minister for Higher Education said, “*privacy impact assessment should have assessed human rights and that the issue will require considerable scrutiny.*” The regulations passed without scrutiny for impact.
43. The SLC now stands accused of spying on vulnerable students' social media, and support workers reportedly say estranged students lost funding despite no finding of guilt of fraud.³⁵ These uses of student data demand independent oversight, safeguards and redress process, all lacking today.
44. Freedom of Information requests in 2018 shows that the data sharing between the OfS and a number of third parties, is focused on fraud detection, and powers for the OfS who “might become enforcers of consumer law in future.”³⁶ The regulations also enable commercial reuse of data³⁷, with no safeguards.

Case study: Children’s pupil data misuse for secondary re-uses

45. Secondary re-use of school census data should be controlled, because pupil data are collected for a narrow, specified and trusted purpose to be sent to the DfE for the purposes of education, under s537a of the Education Act 1996.
46. Secondary bulk reuse of data for the purposes of the Home Office was recognised as unlawful in a 2015 working draft³⁸ of the data sharing document, which nevertheless went on to be an operational Memorandum of Understanding between the Departments. This same lawful basis for collection goes for all uses. These are similarly exceeded in some uses by *private-sector and academic service developers*, and must not be ignored.

³³ Notes obtained via FOI Student data rights: The Higher Education and Research Act 2017 (Cooperation and Information Sharing) Regulations 2018 No. 607 <https://www.whatdotheyknow.com/request/491880/response/1210408/attach/4/FOI%202018%200026457%20Jen%20Person%20Annex%20A.docx.pdf>

³⁴ Higher Education and Research Act 2017 (Cooperation and Information Sharing) Regulations 2018 July 2, 2018 [https://hansard.parliament.uk/Commons/2018-07-02/debates/b015265b-dbf3-48e3-b944-d72d99b5ec78/HigherEducationAndResearchAct2017\(CooperationAndInformationSharing\)Regulations2018](https://hansard.parliament.uk/Commons/2018-07-02/debates/b015265b-dbf3-48e3-b944-d72d99b5ec78/HigherEducationAndResearchAct2017(CooperationAndInformationSharing)Regulations2018)

³⁵ Student Loans Company ‘spied on vulnerable students’ social media’ August 2018, The Guardian <https://www.theguardian.com/education/2018/aug/02/student-loans-company-spied-on-vulnerable-students-social-media>

³⁶ FOI meeting minutes on data sharing powers of the OfS p6/34 <https://www.whatdotheyknow.com/request/491880/response/1210408/attach/4/FOI%202018%200026457%20Jen%20Person%20Annex%20A.docx.pdf>

³⁷ University students’ data to be shared with private companies
The Telegraph (June 2018) <https://www.telegraph.co.uk/education/2018/06/18/university-students-data-shared-private-companies/>

³⁸ Version 0.2 MOU between the DfE and the Home Office in late 2014 (end of page 5 of 16) <https://www.whatdotheyknow.com/request/495973/response/1229011/attach/7/DfE%20Draft%20MOU%20earliest%20version%20found%20Redacted.pdf>

Submitted September 2018