National Data Strategy consultation submission from defenddigitalme

Contents

About defenddigitalme

defenddigitalme is a call to action to protect children's rights to privacy. We are teachers and parents who campaign for safe, fair and transparent data processing in education, in England, and beyond. We advocate for children's data and digital rights, in response to concerns about increasingly invasive uses of children's personal information. Funded by the Joseph Rowntree Reform Trust.
For more information please see defenddigitalme.org

1. Scope of our response

Our response is bounded to the area of our own remit of work and aligned with the data scope of the Department for Education including its redistribution and re-use by other third parties:

1. **Data types:** 'administrative data', personal data in the state education sector, processed about children and their families from the course of a child's state education age 0-25[1] (i.e. statutory data collections and with limited scope to those datasets processed by the Department for Education, also from children's social care such as the national Children in Need, and Looked After Children datasets)
2. **Geographic scope:** a focus on England, but some limited reach into UK wide policy and practice where explicitly mentioned, since education and therefore the data collected are devolved matters. (see the Annex of UK national pupil databases UK comparison chart)
3. **Data scope:** administrative data processed since 1996, to date.

Our response addresses a number of the topic areas of the questions but does not follow only the construct of the consultation, but we answer some selected structured questions in section six.

2. Finding a way forward on national education data and a wider public-facing strategy

After the ICO compulsory audit of the Department for Education, the question now is what will the Department and government learn and how and when will it address the 139 recommendations for improvement, with over 60% classified as urgent or high priority. How will the public know and profession have trust and confidence restored in department data handling?

> *"Internal cultural barriers and attitudes are preventing the DfE from implementing an effective system of information governance, which properly considers the rights and freedoms of data subjects against their own requirements for processing personal data to ensure data is processed in line with the principles of the GDPR."*

Nesta proposed in its 2019 report, Educ-AI-tion Rebooted?[2] that the Government should publicly declare an ambition to create a system of responsible education data sharing by 2030. We suggest that 2030 is too far away and the governance framework must be built with urgency.

Is the government intentional about delivering "*a forward-looking strategy that takes into account public opinion and delivers real change*"? If so it must operate within the rule of law. It must operate within a people-based framework, that is to say a human rights framework that goes beyond the GDPR, and other data protection frameworks and law that cover the processing of data that apply equally to the UK, including the Convention 108+ (for which new guidelines in data protection for educational settings were adopted in November 2020[3]) This cannot be ignored if the government is to deliver on its aims for the UK's domestic strategy and success in export markets.

---

[1] How to apply for data extracts from the National Pupil Database (NPD), School Workforce, Individualised Learner Record and Higher Education Statistics Agency. https://www.gov.uk/guidance/how-to-access-department-for-education-dfe-data-extracts
[2] AI Educ-AI-tion rebooted? Exploring the future of artificial intelligence in schools and colleges (2019) Nesta | https://www.nesta.org.uk/report/education-rebooted/
[3] The Committee of Convention 108 adopted Guidelines on Children's Data Protection in an Education Setting in November 2020 https://www.coe.int/en/web/data-protection/-/protect-children-s-personal-data-in-education-setting-

What is striking about this strategy is that it could be the same had it been published a year ago, pre-COVID 19 or five years ago, when an EU Referendum was not yet decided. An effective national data strategy cannot sit in a vacuum of the real world or other policy and public experience of government. The timing of the next steps for this strategy will be in parallel with an unprecedented recent time of poverty, instability, and uncertainty in the population at large. The implications for health and wellbeing, earnings, education, family life and the capacity for people to engage with these plans in any meaningful way, in particular those increasingly in debt to the State, do not feature in the strategy as influencing factors, and this needs to be accounted for.

The enthusiasm of the data for growth message, is often not what people see in their every day. 'Take back control' may have left the building, but the sentiment remains in the lived experience of people whose lives have been disrupted by digital first policies that are challenging without sufficient digital skills. Who cannot support homework even in pre-COVID times because school did not have devices to lend and there was no money at home to buy one or the data needed to be online. A world in which people may feel surveilled by their smart devices[4], worn down by incessant marketing emails, or more easily coerced in a cashless society.[5] People are fed up that buying a piece of music is no longer yours to lend to a friend, get harassed by debt collectors for an overdue parking ticket[6], and are worn down by a barrage of captchas and failed log-ins and automated telephone redirections, that add barriers to everyday living, rather than 'digital' making life easier. It's long been the case that many people just want to speak to 'a real human being'.[7]

The public experience of government will also affect how the public responds to its next stage of asks. A long list of bad government engagement over recent months, is a poor foundation for a new data regime that expects a trusted relationship with the public, and you don't fix that by talking about "the benefits" but by addressing concerns and delivering on public expectations.

Government must recognise
- failure over the last ten years, to make its own departmental data processing lawful
- introductions of toxic policy that relied on data processing in secret, and
- local government and public sector procurement practices of products that rely on large scale data processing without any form of public registration or transparency obligations, have enabled dangerous data processing to go without oversight.

Government must build the necessary infrastructure to enable understanding and control that people want when it comes to personal data.

1. Enhance and build the physical infrastructure:
   i. access to personal reports on what data is held and how it is used.
   ii. management controls and communications over reuse [opt-in to meet the necessary permissions of legitimate interests or consent as lawful basis for

---

[4] Which? (2020) Are Alexa and Google Assistant spying on us?
https://www.which.co.uk/news/2020/01/are-alexa-and-google-assistant-spying-on-us/
[5] Women's Aid (2015) Unequal trapped and controlled: Women's experiences of financial abuse and the potential implications for Universal Credit https://www.womensaid.org.uk/financial-abuse-report/
[6] Hull, R. (2019) ThisIsMoney | Councils passed almost 1.1m unpaid parking PCN cases to bailiffs (without regulation) in 2018-19
https://www.thisismoney.co.uk/money/cars/article-7456331/Councils-increasingly-turn-bailiffs-pursue-unpaid-parking-fines.html
[7] Which? Conversation (2012) Banks, I just want to speak to a human being
https://conversation.which.co.uk/money/bank-automated-call-helplines-telephone-speak-human-ivr/

> further data processing, conditions for sensitive data processing, or at very least opt-out to respect objections].

      iii.   secure state systems requires state investment

2. Enable the necessary transparency tools and create demonstrable accountability through registers of algorithms[8] and data sharing with oversight functions and routes for redress.
3. Empower staff with the necessary human skills at all levels in the public sector on the laws around data that does not just consist of a workshop on GDPR — but has go-to expertise on data related privacy law, communications law, equality, disability, discrimination laws
4. Empower the public with the controls that they want to have so their rights are respected
5. End harmful use of predictive algorithms in children's social care, across linked datasets in child justice, and in exam awarding
6. Examine toxic policy choices (i.e. Home Office data sharing) that drives bad data collection

Any fundamental lessons learned during the ICO compulsory audit of the Department for Education should be published and applied across government to get the basics right:

> *"There is no formal proactive oversight of any function of information governance, including data protection, records management, risk management, data sharing and information security within the DfE which along with a lack of formal documentation means the DfE cannot demonstrate accountability to the GDPR. Although the Data Directorate has been assigned overall responsibility for compliance, actual operational responsibility is fragmented throughout all groups, directorates, divisions and teams which implement policy services and projects involving personal data. Limited reporting lines, monitoring activity and reporting means there is no central oversight of data processing activities. As a result there are no controls in place to provide assurance that all personal data processing activities are carried out in line with legislative requirements."*

3. Summary of the need for change

In November 2014, the Science and Technology Parliamentary Committee published a report, *The Responsible Use of Data[9]*, and said, **"the Government has a clear responsibility to explain to the public how personal data is being used"**.

In 2016 The Government Internal Audit Agency rated the DfE assurance as 'limited' when it came to data issues. **Improvements were needed over vetting and validation of applications to access the National Pupil Database, information retention procedures, and data handling guidance**.

In 2020 the Information Commissioner's Office published the executive summary[10] of the compulsory audit of the DfE. It found that *still*, **data protection was not being prioritised and this had severely impacted on the DfE's ability to comply with UK's data protection laws.**

---

[8] Johnson, K (2020) Amsterdam and Helsinki launch algorithm registries to bring transparency to public deployments of AI https://venturebeat.com/2020/09/28/amsterdam-and-helsinki-launch-algorithm-registries-to-bring-transparency-to-public-deployments-of-ai/ and New Zealand Graham-McLay, C. (The Guardian) 2020 | New Zealand claims world first in setting standards for government use of algorithms https://www.theguardian.com/world/2020/jul/28/new-zealand-claims-world-first-in-setting-standards-for-government-use-of-algorithms
[9] Science and Technology Committee | Responsible Use of Data (2014) https://publications.parliament.uk/pa/cm201415/cmselect/cmsctech/245/24502.htm
[10] Statement on the outcome of the ICO's compulsory audit of the Department for Education (October 2020) https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2020/10/statement-on-the-outcome-of-the-ico-s-compulsory-audit-of-the-department-for-education/

Four whole years after the GIAA report, the ICO found "*no clear picture of what data is held by the DfE and as a result there is no Record of Processing Activity (ROPA) which is a direct breach of Article 30 of the GDPR. Without a ROPA it is difficult for the DfE to fulfil their other obligations such as privacy information, retention and security arrangements.*" There is "*limited oversight and consistency around how data is shared externally (including from the National Pupil Database)*"[11] and lack of appropriate procedures for the "*creation, storage and retention of records which include formal documentation of what information is added to, or not added to, the NPD from any given information collection, weeding of records and retention and disposal.*"

Six full years after the the Science and Technology Parliamentary Committee report, the ICO among its 139 recommendations, with over 60% classified as urgent or high priority, found that the Department "*are not fulfilling the first principle of the GDPR, outlined in Article 5(1)(a), that data shall be processed lawfully, fairly and in a transparent manner,*" and that, "*The DfE are not providing sufficient privacy information to data subjects as required by Articles 12, 13 and 14 of the GDPR.*"
i.e. **The Department fails to explain to the public how personal data is being used.**

In summary, since 2010, the government has made a successive series of changes in national data policy and practice always seeking to 'reduce the barriers'. It has changed various laws with impact on data handling, including the Digital Economy Act 2017 (as the strategy references in 6.2.1), The Higher Education and Research Act 2017, The Technical and Further Education Act 2017 and the Education Act 1996 has had various secondary legislation added to expand data collections through statutory instruments[12], and introduced a new Data Protection Act.

Despite all this it fails to meet the fundamental necessary principles of data processing: explain to the public how personal data is being used and process in line with data protection laws. The biggest barrier to good government data strategy, is the government itself.

The new strategy framing states that it aims to take public opinion into account and deliver "*real change.*" In order to do that, the government must change its own policy and practice at all levels, and not seek to retrospectively, as the ICO audit found at the DfE, "*find a legal gateway to 'fit' [use] rather than … provide assurance and accountability that the sharing is lawful in line with statutory requirements".* [13]

This strategy rather suggests that, like the DfE, where there is— "*no clear picture of what data is held by the DfE*" …. and that, "*without a ROPA it is difficult for the DfE to fulfil their other obligations such as privacy information, retention and security arrangements*— government does not yet know what data it holds, why, or recognise the governance necessary for processes, infrastructure and oversight.

It is exciting to see opportunity for change in national data practice, and we welcome the chance to contribute, in the hope of seeing the needed UK infrastructure and capability 'built better'.

---

[11] In May 2019, the National Pupil Database contained over 21 million individual pupil records on a named basis, and it grows every year by an annual cohort of around 700,000 FOI request: NPD Data Protection Impact Assessment: Public Summary 1.0 (May 2019) https://www.whatdotheyknow.com/request/533041/response/1375464/attach/html/3/DfE%20NPD%20DPIA%20Public%20Summary%20v1.0%20May%202019%20002.docx.html
[12] See the State of Data 2020 report fig 3 The legislation and data items expanded in the national pupil database https://defenddigitalme.org/state-of-data/#h.3hk1473dtulz
[13] Final paragraph of the executive summary from the ICO audit of the DfE (October 7, 2020)

4. Framing the strategy[14] and looking forward

The strategy framing states that, "*we want to ensure that we produce a forward-looking strategy that takes into account **public opinion** and delivers **real change**.*" In order to look forward and to deliver change or improvement, one must understand the starting point and status quo.

4.1 Public opinion

In the last six years public opinion has been both consistently stable and consistently ignored in the national and local government stewardship of public administrative datasets.

When asked, the majority of people both who are willing and less willing to have data about them reused, generally want similar things. Information and understanding, control via opt in to the uses of the data, safeguards through restricted distribution, and protections for redress and against misuse strengthened in legislation.

The common law duty of confidentiality appears currently unfashionable but perhaps is the simplest principle to consider in terms of public expectations: is information I give in confidence, kept in confidence? The 2009 Government document[15]: Information Sharing: Further Guidance on Legal Issues includes reference to the duty, while more recent government guidance does not.

This expectation is reflected in law by Article 8 of the European Convention on Human Rights incorporated into UK law by the Human Rights Act 1998, recognising the right to respect for private and family life as the default, which requires a high bar of **necessary** and **proportionate** grounds to justify interference by the State:

> *Article 8.1: **Everyone has the right to respect for his private and family life, his home and his correspondence.***

> *Article 8.2: **There shall be no interference by a public authority with exercise of this right except such as is in accordance with the law and is necessary in a democratic society** in the interests of national security, public safety or the economic well-being of the country, for the prevention of crime or disorder, protection of health and morals or for the protection of rights and freedoms of others.*

Public engagement work has been extensive by various bodies.

2013: Extensive work was carried out across the UK in workshops and surveys over several months in 2013 before the Administrative Data Research Network began allowing accredited researchers to reuse 'deidentified' personal data in safe settings, obtained from public administrative datasets. The red lines in the report[16] remain valid today, even more so for identifying data and in relation to:

---

[14] Policy paper National Data Strategy (September 2020)
https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy
[15] Information Sharing: Further guidance on legal issues (2009) ISBN: 978-1-84775-276-5
http://webarchive.nationalarchives.gov.uk/20130401151715/https://www.education.gov.uk/publications/eOrderingDownload/Info-Sharing_legal-issues.pdf
[16] ADRN/ ESRC (2013) Public dialogues on using administrative data | The project was directed by a steering group including researchers from Loughborough and Edinburgh universities, members of ESRC's Methods and Infrastructure committee, and representatives from the Department of Business, Innovation and Skills (BIS), the Wellcome Trust, and Sciencewise.
https://esrc.ukri.org/public-engagement/public-dialogues/public-dialogues-on-using-administrative-data/

- Creating large databases containing many variables/data from a large number of public sector sources
- Allowing administrative data to be linked with business data
- Linking of passively collected administrative data, in particular geo-location data

"All of the above were seen as having potential privacy implications or allowing the possibility of reidentification of individuals within datasets. 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."

2014: Ipsos MORI polls[17] and work by Wellcome[18] and the care.data summaries[19] after public engagement work collapsed, shared many similar findings.

2018: Our State of Data 2018 poll commissioned by defenddigitalme through Survation. 69% of parents of children age 5-18 in England asked, had not heard of the National Pupil Database or that the data are distributed to third parties by the Department for Education. Over 50% did not feel they had enough control over their child's digital footprint in educational settings.

2018-2020: Doteveryone's public attitudes research work[20] *" found that people care deeply about the use of their personal information - 95% say it's important to know their data is secure, 94% say it's important to know how their data is used. And they would like more control over it — 91% say it's important to be able to choose how much data they share with companies, but half (51%) can't currently find out that information."*

In 2019: The ICO Annual Track Survey found, *"The biggest change since 2018 is on 'sharing my personal information to third parties without my permission'. The public is significantly more likely than in 2018 to feel if this happened it would cause a negative impact on the trust and confidence they have with the organisation (81% in 2019 compared with 77% in 2018).*

4.2 Real change

The aim of the National Data Strategy (NDS) is *"to drive the collective vision that will support the UK to build a world-leading data economy,"* and *"The NDS will also provide coherence and impetus to the wide range of data-led work across government, while creating a shared understanding across the economy of how data is used."*

A shared understanding across the economy is largely irrelevant. Business will use data as it needs. A shared understanding across society is necessary, but there is no strategy how to do this set out in the policy proposals. We suggest the conflation of too much in a high level approach is a barrier to delivering both strategy and identifying the tactics required to deliver it. It cannot be all things for all

---

[17] The Joseph Rowntree Reform Trust commissioned lpsos MORI to complete a poll of the general public across Great Britain on privacy and personal data. (2014) https://www.ipsos.com/ipsos-mori/en-uk/privacy-and-personal-data and research for the Royal Statistical Society carried out by Ipsos MORI (2014) The data trust deficit with lessons for policymakers https://www.ipsos.com/ipsos-mori/en-uk/new-research-finds-data-trust-deficit-lessons-policymakers
[18] Wellcome (2017) The One-Way Mirror: Public attitudes to commercial access to health data https://wellcome.figshare.com/articles/journal%20contribution/The_One-Way_Mirror_Public_attitudes_to_commercial_access_to_health_data/5616448/1
[19] Care.data summary presentation (download) https://jenpersson.com/wp-content/uploads/2020/10/care-data-presentation.pdf
[20] DotEveryone (2020) Introducing People, Power and Technology: The 2020 Edition https://www.doteveryone.org.uk/2020/02/introducing-people-power-and-technology-the-2020-edition/

purposes and for all users, in the same way that a food distribution strategy post Brexit would fail if it did not consider its logistical needs for consumables with a short shelf life and long life separately, or understand that food collected for feeding cattle cannot be safely used to feed people and falls under different legal obligations.

The longitudinal datasets held by the government need to be processed with the understanding that data have different properties, different permissions, different purposes and different limitations that come from the intrinsic nature of the data itself — not artificial barriers to access.

4.3 Understanding 'the data'

Data is or data are? The policy says that data means different things to different people. While true, there should also be some things that are understood about data as facts, while others are opinion.

The language of the term 'data' has changed over recent use, and it may help to consider why and to its implications when thinking about data and its dynamic properties. In 2012 the Washington Post, the Guardian[21] and others, addressed the shift from plural to singular as a contentious issue. "*In Latin, data is the plural of datum and, historically and in specialized scientific fields , it is also treated as a plural in English, taking a plural verb, as in the data were collected and classified . In modern non-scientific use, however , despite the complaints of traditionalists, it is often not treated as a plural. Instead, it is treated as a mass noun, similar to a word like information, which cannot normally have a plural and which takes a singular verb. Sentences such as data was (as well as data were) collected over a number of years are now widely accepted in standard English.*" The concept of data may have shifted to be discussed more of a singular object, than a multitude of information, but intrinsic characteristics have not changed.

The concept of personal data is defined in data protection law, and we use the terms here as defined relating to the processing of personal data set out in the Data Protection Act 2018.[22]

> *(2)"Personal data" means any information relating to an identified or identifiable living individual (subject to subsection (14)(c)).*
>
> *(3)"Identifiable living individual" means a living individual who can be identified, directly or indirectly, in particular by reference to— (a) an identifier such as a name, an identification number, location data or an online identifier, or (b) one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of the individual.*

It is important therefore to note this early on, and mention where the NDS policy paper (September 2020)[23] premise of what "the data" is, is wrongly framed because it affects the whole approach. Fundamentally the strategy fails because it tries to treat all 'data' as the same 'thing'.

---

[21] Rogers, S. (2012) The Guardian | Data are or data is? https://www.theguardian.com/news/datablog/2010/jul/16/data-plural-singular
[22] The UK Data protection Act 2018 (Part 1, section 3) Terms relating to the processing of personal data
https://www.legislation.gov.uk/ukpga/2018/12/section/3/enacted
[23] Policy paper National Data Strategy published 9 September 2020
https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy

"Data is a non-depletable resource in theory, but its use is limited by barriers to its access…"
A successful National Data Strategy cannot be about "the data". There is no such thing as "the data" as an amalgamated single static ***"non-depletable"*** resource, insofar as it cannot be treated consistently for all people, for all users, for all purposes, for all time, in the same way. One must stop thinking of public administrative data as abstract information, or seeing personal data as a commodity when it comes to national policy.

To the State and commercial actors administrative data is business intelligence. What resources are used, where, by whom and who costs the business / The Treasury how much? Any change in National Data Strategy about public administrative data is about the rewiring of state power in the delivery of those services. It is most clearly demonstrated in the recent abuse of purposes in national education data by other government departments including the Home Office.

To individuals, administrative data is the knowledge about the workings of the interactions between the public and the State, about us as people and communities. The strategy should recognise that not every person whose data it will process is either a citizen or a consumer as described in the NDS. In fact, some of the most marginalised groups of children, in education, and who may be treated at most disadvantage, are those who have an insecure status and are undocumented[24] and have barriers to access public services. There are now estimated to be 215,000 undocumented children in the UK. Many more are now at risk of becoming undocumented as a result of Brexit. (GLA research report, 2020)

Administrative data is the story of selected life events, and when it comes to education that is admission to school, every statutory attainment test result, termly school censuses transfers in and out of new settings and interactions with children's social care. Data is therefore not static, not a non-depletable resource, but is often a 'living' set of characteristics that change according to people's life events (address change, school move). Public administrative datasets *do* deplete over time in quality and validity, and two of the seven data principles[25] come into play as a result:

1. Data Accuracy

The Department for Education amasses data as an ever growing set of snapshots in time, collected from educational settings through seven school censuses. These data are not all provided by parents, but can be provided by or inferred by school staff.

A record of an opinion is not necessarily inadequate or irrelevant personal data just because the individual disagrees with it or thinks it has not taken account of information they think is important. However, in order to be adequate, records should make clear that it is opinion rather than fact. If an opinion is likely to be controversial or very sensitive, or if it will have a significant impact when used or

---

[24] Following changes in Government policies and cuts since 2010, more individuals have risked becoming undocumented with fewer prospects to secure their status. Research was commissioned through the Mayor of London's Citizenship and Integration Initiative and undertaken by the Institute for Community Research and Development at the University of Wolverhampton.https://www.london.gov.uk/what-we-do/communities/migrants-and-refugees/londons-children-and-young-people-who-are-not-british-citizens
[25] ICO | The GDPR sets out seven key principles https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/

disclosed, it is even more important to state the circumstances or the evidence it is based on.
(ICO, data minimisation principle)
Data is dynamic and can deteriorate over its life cycle.

The nature of the data can change and affect its **validity and value** and therefore brings with it certain obligations. Data can become inaccurate if it is changeable data, led by the data subject, such as home address that has not been updated in school records and is no longer current, or changed by a system requirement i.e the term "free school meal" has had different meaning associated with it over time, including the proxy definition of the financial earnings threshold required behind the conditionality calculation in order for the state to choose whether or not to support a child in need; or if / when the 'universal infant free school meal' was conflated into the same label. Over time, data with the same description can have different meanings.

Data **relationships also change,** and in a particular way in education for more than most datasets, but also worth noting for other public administrative datasets. And this changes the nature of the obligations of the state to the data subject. This process of changing data and changing duties needs to be reflected in any strategy— *the question to ask is how much of the longitudinal datasets that the government controls today, has exceeded their validity for keeping in identifying format.*

Time changes the obligations of the controller to the rights holder *as a child,* and when the data subject becomes a competent child and/or adult and has their own autonomy as a rights holder *distinct from their paren*t. Data processed on the basis of consent given by a parent, continues to require a lawful basis for processing that means permissions given on the data by the parent of a competent offspring will no longer be valid as the child reaches adulthood.

Few national data processing activities rely on consent for data processing, however it may apply to the collection of special category or sensitive data in the school census such as ethnicity, religion, sexual orientation, and such data retained beyond eighteen may no longer have a lawful basis for processing where used in projects that result in interventions, beyond scientific and historic research purposes.

> *"In educational settings children are disempowered in their relationship with a public authority and are also recognised as vulnerable due to their lack of understanding and evolving capacities, and state of being in the process of development into adulthood. From the static point of view the child is a person who has not yet attained physical and psychological maturity. From a dynamic point of view, the child is in the process of developing to become an adult. **Children are also active rights holders, and agents who require not only protection but also provision of information, training, and guidance.***

> *"Data processing on the basis of consent, may be invalid where a power imbalance exists, notably between a public authority and an individual, which impairs the freely given nature of the consent. This imbalance is even more significant where the data subject is a child. Another basis is therefore more likely to be valid for routine processing activities and such processing should be based in law."*

"The legal guardian's powers to exercise rights on behalf of a child as a data subject expire when the competent child reaches the age of maturity as laid down in law. The data subject (the child) should be informed of any ongoing data processing about them, to which the legal guardian gave consent, so as to be able to exercise the rights of the data subject, as an adult." (Council of Europe Guidelines (2020) on Convention 108, on data protection in educational settings).[26]

Data collection systems can be far out of step with parental expectations, or not involve them when it comes to school data submissions, and accuracy suffers as a result.

4.3.1 Case study: A diplomatic incident caused by data inaccuracy

The data input fields used by thousands of schools across England to record school census information in 2016, allowed administrators to ascribe a child's ethnicity. Italian parents complaints to the Embassy in London resulted in an apology from the DfE about country of birth and language forms sent out by UK schools asking parents whether their child was "Italian", "Italian-Sicilian" or "Italian-Neapolitan". The Ambassador pointed out after complaints, that Italy has been a unified country since 1861.[27]

Some data are retained indefinitely, others are written over in school data systems. It is not clear how Department for Education datasets deal with this. It is suspected that the data are kept historically, so that if a staff member submitted *Italian-Sicilian* in the 2001 census, and parents were asked instead in 2002 and provided *Italian*, that the national pupil dataset now shows that what was an error to be corrected, was a change of language over time. There is missing context.

There is also the infrastructure missing that enables data subjects to action their rights. And we are sure the government does not want the standard route for redress to be at the ambassador level. The obligation on data controllers and processors towards data accuracy creates a right in the data subject to access the information, and to be able to correct it. This is not respected by the government today in its stewardship of administrative data, as the Department for Education unwieldy Subject Access Request process demonstrates.[28] Despite assurances made in 2002 by a previous government, that the pupil (or their parents or guardians) will be able to request a copy of their own record in order to confirm its accuracy; in 2020 the process is poorly communicated, badly fulfilled and the DfE is unable to tell you to which third parties it has sent your child's data. All of which are obligations under the data accuracy principle necessary in a subject access request.

---

[26] Council of Europe Convention 108+ | Guidelines on Children's Data Protection in an Education setting (adopted in November 2020) https://rm.coe.int/t-pd-2019-6bisrev5-eng-guidelines-education-setting-plenary-clean-2790/1680a07f2b

[27] BBC (2016) UK school pupil ethnicity question angers Italy https://www.bbc.co.uk/news/blogs-news-from-elsewhere-37631062

[28] If you or your child has been state educated in England since 2002, you have the right to ask the Department for Education (DfE) for all the information they hold in your named records. You can make a "Subject Access request." It's a free process to help you protect your rights under data protection law. As well as giving you a copy of the data they hold, the Department should also be able to tell you who they have given it to https://defenddigitalme.org/my-records-my-rights/#MyRecordsMyRights   Stephen Timms said in January 22, 2002 https://publications.parliament.uk/pa/cm200102/cmhansrd/vo020128/text/20128w29.htm#20128w29.html_sbhd0 "the pupil (or their parents or guardians) will be able to request a copy of their own record in order to confirm its accuracy"

2. Purpose limitation

Data minimisation and purpose limitation matter in order to both keep people informed at the point of collection (GDPR recital 61) what data will be used for, and prevent misuse for incompatible purposes over time.

Intervention purposes: Data are collected in an educational setting for the purposes of delivering a child's education. They are not provided for repurposing for the benefit of commercial companies and families do not expect this reuse. The Department not only fails to meet public expectations, and their obligations on lawful processing, but the ICO also found:

"The Commercial department do not have appropriate controls in place to protect personal data being processed on behalf of the DfE by data processors." (ICO audit of the DfE , 2020)

Reasonable expectations are not met for direct interventions in the re-use of pupil data by:

- Police[29]
- DWP[30]
- Home Office

To individuals and communities, this is information about them, used about them, to influence decisions and affect their lives. This is vital to appreciate, if the strategy aim is to deliver 'change' because without the willingness to address people in the strategy; to acknowledge people as the data subjects and rights' holders, see people as data stewards with responsibility and positions that can be held to account, and identify people as 'go to' points of contact for routes of redress across the whole changing data life-cycle, and life of an individual, this strategy will fail.

4.3.2 Case study: Purpose limitation of names in the national pupil database

Prior to 2007, the Schools Census dataset was a collection of statistical information known as the Pupil Level Annual Schools Census (PLASC). Comprehensive PLASC data was first collected in 2002, and included individual pupil names for the first time.

MPs in the House of Commons were assured on the changes to the "Central Pupil Database" in 2002 by then Minister of State for Education and Skills, Stephen Timms, that, *"The Department has no interest in the identity of individual pupils as such, and will be using the database solely for statistical purposes, with only technical staff directly engaged in the data collation process having access to pupil names."[31]*

---

[29] In November 2019, personal data from all 2,136 pupils who attended an unspecified Merseyside school during a four year period was provided to police from the National Pupil Database. https://www.whatdotheyknow.com/request/pupil_data_and_workforce_data_ho
[30] In April 2018 DfE permitted the use of the National Pupil Database for a DWP welfare support fraud investigation of 185 children. https://www.whatdotheyknow.com/request/pupil_data_and_workforce_data_ho#incoming-1630439
[31] Timms, Stephen (January 22, 2002). Hansard, Central Pupil Database debate https://publications.parliament.uk/pa/cm200102/cmhansrd/vo020128/text/20128w29.htm#20128w29.html_sbhd0

Thirteen years later, a Freedom of Information request[32] revealed that Education officials had an agreement in place since July 2015 to share the personal details including names which are used to match records, of up to 1,500 school children a month with the Home Office, for a range of Border Force purposes under section 24 or 24A of the Immigration Act 1971, or section 35 of the Asylum and Immigration (Treatment of Claimants) Act 2004.

Neither the Home Office or the Department for Education today take responsibility or provide any public transparency for the outcomes on the affected children and families, of the Home Office reuse of administrative data collected for the purposes of education, but now used punitively. Asked in a Parliamentary question in September 2020[33], (a) how many and (b) what type of (i) individual and (ii) family intervention her Department made based on the 1545 individual pupil records received from the Department for Education between July 2015 and July 2020, the Home Office replied, "*The specific information you have requested is not readily available and could only be obtained at disproportionate cost.*"

As long as such lack of care, transparency and trustworthy behaviour is the norm, the government should not expect any improvement in public or professional trust in the education sector or families, over the use of national pupil data by third p[arties across government. This internal behaviour must change if the strategy is to be successful.

Without addressing policy misuse in the processing of public administrative datasets of personal data, including across other government departments, then public trust will fail.

While the strategy sets out a broad aim for *"wide range of data-led work across government"* it does not set any boundaries on what this may be. Data cannot simply be thought of as 'collect once, use many times.' as might be suggested from reading *"Non-standardisation and a lack of coordination on data mean that data collected by one organisation cannot easily be used by another,"* in 2.4. Purposes of use and controllers (including joint controllers) of data, must be assessed, defined and communicated to the public, for fair and lawful processing at the time of data collection, and re-communicated later where purposes may not be compatible with a view to asking for consent. Otherwise the data may need to be processed into a state that it is no longer personal data, or for the data to be destroyed where it may no longer continue to be lawfully processed.

The Department for Education in England has been aware for years that they share too much data with third party users, calling it an ' excessive' amount of data in the underlying datasets.'[34]

---

[32] FOI Ref: 2016-0042333 confirmed Home Office access to previously collected school census pupil data includes name, home and school address. Purposes of all cases within the Home Office requests "are a) dependant(s) of a parent/guardian who is suspected of an offence under section 24 or 24A of the Immigration Act 1971, or section 35 of the Asylum and Immigration (Treatment of Claimants) Act 2004 has been, or is being committed (or b) the child in question is an unaccompanied minor.)" "Each Home Office request includes details of the individuals concerned as held to enable us to search for them in the NPD."
https://www.whatdotheyknow.com/request/sharing_national_pupil_database?nocache=incoming-878444#incoming-878444
[33] https://questions-statements.parliament.uk/written-questions/detail/2020-09-21/92745
[34] DfE data dissemination discovery report (2018) "Users are required to download the entire dataset, then remove and manipulate extraneous data reducing it to a specific subset. Many expressed a desire to be able to customise the data they downloaded."
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/721729/HiveIT_-_DfE_disseminatio n_discovery.pdf (p.29)

The data minimisation principle, requires personal data shall be: adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed. You should not have more personal data than you need to achieve your purpose. Nor should the data include irrelevant details. "If you are holding more data than is actually necessary for your purpose, this is likely to be unlawful (as most of the lawful bases have a necessity element) as well as a breach of the data minimisation principle. Individuals will also have the right to erasure."[35]

Scope creep and repurposing, unlawfully when for incompatible purposes, must be prevented in public administrative datasets.

Despite such poor practice evidenced by the ICO audit, the DfE continues further and deeply sensitive data expansions. The Children in Need Census[36], also controlled by the DfE as part of its remit for Children's Social Care, ends up in the National Pupil Database, albeit with some of what is thought the most deeply sensitive data held back but every item of data in such a database should be considered sensitive. Linkage in our opinion, into the NPD removes the intended protections of collecting such data *without names* by adding the newly collected data to *named* records in the National Pupil Database. This is not only illogical but potentially dangerous, since the database is open to Subject Access Requests and without context it is a risk that data could be released that had intentionally been withheld from parents. There appears to be no overarching oversight or independent holistic view of all data held by the DfE about a single child and the national data strategy does not include any thinking on what infrastructure should be in place in Departments for those at risk, children, and their most sensitive data processing.

Such data policy and practice needs careful consideration across government, and should be addressed as part of any overarching strategy on the basis of a wide cooperation with education practitioners, academia, industry, with organisations representing teachers, families at risk, and civil society.

4.3.4 Case study: Dangerous blind spots in Children's Social Care and predictive algorithms

Children's data collected in Children's Social Care at local government level needs urgent action to be made safe and must not wait for a child to come to harm as the result of an algorithm that a staff member over relies on due to the trust we place into computer generated data, or that children are missed or harmed through over intervention, i.e. removal from a family.

The What Works Centre commissioned The Alan Turing Institute and the Rees Centre, University of Oxford to undertake the review of machine learning and ethics in children's social care in 2019-20. At the report launch, the Ethics Review of Machine Learning in Children's Social Care[37], recommended that such models were at the time of research, not fit for purpose.

---

[35] The ICO | Principles of domestic and international data protection law include data minimisation and privacy by design https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/data-minimisation/
[36] Children in Need Census FOI https://www.whatdotheyknow.com/request/pupil_data_children_in_need_cens#incoming-1639044
[37] Leslie, D. et al (2020) Ethics of machine learning in children's social care https://www.turing.ac.uk/research/publications/ethics-machine-learning-childrens-social-care and What Works for Children's Social Care https://whatworks-csc.org.uk/research-report/ethics-review-of-machine-learning-in-childrens-social-care/

"These issues related to the safe and ethical functioning of a predictive ML model are magnified in high impact and safety-critical domains such as CSC, for system errors, unreliable performance, and lurking biases may have life and death consequences."

More recently, writing about further predictive model functionality testing, Michael Sanders of the What Works Centre for Children's Social Care found, *"Across 32 models, none meet the threshold we set in advance for success, with most of them falling far short of it. Models that attempt to predict the future — i.e. those that are actually useful in practice — do even worse — meaning that more families could see unnecessary intervention in their lives, and more opportunities for support could be missed. The models don't perform any worse for specific groups — defined by race, age, or disability — but this is a cold comfort when the models don't perform well anyway. It seems that increasing the sample size may help but the population changes quickly enough that in waiting for more data the previous data becomes obsolete and local authorities have different enough contexts that combining data is unlikely to help"*

He recommended that, *"At the moment, the case has not been made. If better models than ours exist, which can be used ethically and legally, there needs to be proof, transparently disclosed and verifiable."* *"Now is a good time to stop. With the global coronavirus pandemic, everything has been changed, all our data scrambled to the point of uselessness in any case."*[38]

4.4 Understanding data degradation means lawful obligations are dynamic not static

It is vital to understand what the strategy means by data 'use'. We therefore consider 'use' any form of processing across the data life cycle and that cycle has dynamic obligations:

"Processing", in relation to information, means an operation or set of operations which is performed on information, or on sets of information, such as—

(a) collection, recording, organisation, structuring or storage,
(b) adaptation or alteration,
(c) retrieval, consultation or use,
(d) disclosure by transmission, dissemination or otherwise making available,
(e) alignment or combination, or
(f) restriction, erasure or destruction,

For the strategy to succeed, it must include the factor of time. The concepts of data degradation and the requirements for government and other data controllers and processors to act in different ways **at different times** need to be understood. In order to do so, rather than thinking of data as a thing, but thinking of data as a dynamic by-product of public sector service processes, you can see the data as moving flows of information across the different tasks carried out in order to complete an activity (ie the school admissions process, with the tasks of registering a child by parents,

---

[38] Sanders, M. (2020) What Works for Children's Social Care | Machine Learning; Now is a time to stop and think
https://whatworks-csc.org.uk/blog/machine-learning-now-is-a-time-to-stop-and-think/

confirming a place by the school, informing the parents) and all overlaid with the roles of who is involved and over time:

> a) within which at certain times data is used* until its retention point is reached and processing concludes with changing the data to be anonymised or destruction
> and
> b) the role based rights' management framework defines who has which right and what responsibility when

Admin data is not free, but a product of labour from the public sector funded workforce involved in its creation, collection, cleaning, submission. Retention, maintenance, and support of ongoing relationships with the data subject need maintained (ie contact details) and / or as a by-product from the labour and time of the children who the data are about, from their statutory test data for example. (See State of Data 2020 Report, section 2.2 Statutory national data collections).[39]

Data across all of its use, is governed by a rights' based framework in law. Those rights are assigned to rights holders, people, as data subjects. (Noting that not all data subjects will be citizens such as or consumers). Data protection legislation requires accountability and that duty also rests with people, in terms of regulatory and legally responsible owners. It cannot function as a data strategy for public admin data unless it is designed in terms of expressing what data is, that it is changeable, and that over time it has different permissions, changing obligations, and that over time the accuracy, quality and validity of the knowledge within the data may deplete, and this is part of the reason that retention periods are limited for identifying data, and require data to be changed in its characteristics over time, such as obligations for aggregation and anonymisation.

The other moving part to understand, is that the environment in which the data are used is dynamic, for example in its threat models and changing state of security requirements. This imposes **changing obligations on the data controller or processors (i.e. government at all levels).**

Article 32 of the GDPR, which provides more specifics on the security of states*: 'Taking into account **the state of the art,** the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk'.*

Pseudonymisation and encryption are two examples of measures that may be appropriate and the government cannot on the one hand propose that the UK have 'world leading' data practices, and at the same time, the Home Secretary[40] calls for the end of e2e encryption. If the UK government is serious about its aspiration 'to become a global leader in data' then it must meet world leading security expectations and standards in technology. It is simply incongruous and makes government

---

[39] State of Data 2020 Report, s2.2 Statutory national data collections https://defenddigitalme.org/state-of-data/#h.m0h3lm6vel2l
[40] International statement: End-to-end encryption and public safety | Statement from Five Eyes nations plus India and Japan (2020) signatories incl Rt Hon Priti Patel MP, United Kingdom Secretary of State for the Home Department https://www.gov.uk/government/publications/international-statement-end-to-endencryption-and-public-safety

not only appear not joined up, but technologically incompetent to call for weaking end to end encryption for data in transit, or the security of data at rest.

4.5 Aims for wider data linkage and distribution

Central government chose in 2012 to change legislation to distribute identifying pupil data to third parties[41] after 2011 discussions on 'opening up' anonymised data from education.

The distribution of national pupil data since has not been anonymous, but identifying and highly sensitive. The government / Department for Education did not inform the 15 million individuals in the database at the time who had already left the school system, that their data would be repurposed for new uses, by new users. It continues to fail the people the data have been added to the database since the 2012 change of law, and who are still in school.

This fails the first principle of data protection law, fair and lawful processing.

Uses agreed for 'research purposes' goes far beyond what is permitted in data protection research exemptions. Commercial actors receive identifying national pupil data (around 40% of all distribution, in ca 1,600 releases of millions of records in each release since 2012) as well as journalists, think thanks, charities and bona fide academic researchers. (2.5.4 State of Data Report)

Increasingly, distribution has not only been for academic research, but commercial 'research' and development of products, and more recently not only commercial purposes, but punitive interventions. **These purposes are not met by research exemptions, and are outside the legal basis that the Department had at the time of the data collection. That is a fundamental failure that must not be repeated across other government datasets.**

4.5.1 Case study: When 'research 'are not research purposes in data protection law terms

The definition of scientific research purposes has substantial ramifications for the range of data processing activities a controller may undertake and in this context, for local and national government processing and processing by third parties. The term 'scientific research' is not defined in the GDPR. Recital 159 states "(…) For the purposes of this Regulation, the processing of personal data for scientific research purposes should be interpreted in a broad manner. (…)", however the EDPB considers the notion may not be stretched beyond its common meaning and understands that 'scientific research' in this context means a research project set up in accordance with relevant sector related methodological and ethical standards, in conformity with good practice. Recital 33 does not disapply the obligations with regard to the requirement of specific consent. (EDPB Guidelines, 2020).[42] DfE practice does not conform to these standards at the time of writing.

Research purposes under data protection law to enjoy any exceptional permissions, but "research" cannot mean "operational background work" before interventions, or any purposes that the

---

[41] Section 114 of the Education Act 2005, and section 537A of the Education Act 1996, together with the 2009 Prescribed Persons Act, updated in 2013, specifically allows the release of individual children's data to third parties which in practice has permitted data to get given to journalists, commercial businesses, think tanks, and charities.
[42] EDPB Guidelines 05/2020 on consent under Regulation 2016/679
https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202005_consent_en.pdf

controller has not determined the means and purpose of processing for, or lawful and compatible purposes.

Research purposes and statistical purpose implies that the result of processing for statistical purposes is not personal data, but aggregate data, and that this result or the personal data are not used in support of measures or decisions regarding any particular natural person. (GDPR, Recital 162) But distribution of pupil data for research is identifying.

In 2016 ten years' worth of national pupil data across various years for each data set (2000-2015) were linked with police data, including children's highly sensitive personal data, looked after, exclusions, SEND and ethnicity and language. These identifying and sensitive items, or identifying data items were matched at individual pupil level with census data for the Early Years (age 2-5) KS2, KS4 and KS5 datasets, Alternative Provision 07/08-14/15 (plus T1 and T2) and ten years of Children Looked After from 2005 until 2015 and Children in Pupil Referral Units 2009-2013.

This was done at individual pupil level, and without suppression of small numbers which was considered important only for externally 'published' outputs.

Even truly academic 'research uses' of data should not be considered 'neutral', when the conclusions may be used for interventions and may cause harm. A research project in 2013 through the University of Cambridge Institute of Criminology, the Greater London Authority and the Education Endowment Foundation (EEF) set up a randomised control trial to assess the effectiveness of an intervention for children at risk for fixed term exclusion from school in October 2013 with roughly 800 children in Years 9 and 10 in 40 selected schools. The NPD data requested was used to create a model to predict exclusion based solely on administrative data for London schools and schoolchildren. The trial was an independent evaluation of a 12-week-long intervention, Engage in Education-London (EiE-L), delivered by Catch22.

> *"Anecdotal evidence from the EiE-L core workers indicated that in some instances schools informed students that they were enrolled on the intervention because they were the "worst kids"; this may not only hinder any engagement in intervention but also jeopardise the teachers' relationships with the students and thus contributed to negative effects."* [43]

The researchers asked for a lot of data. "As it's not clear what the best predictors would be, we (necessarily) require as much information as possible on the pupils, short of knowing their names and addresses, hence the breadth of our request." They said in their application, "We would also require as much information as possible about the schools in terms of school level measures of ethnicity, gender, academic performance, type of school (e.g. secondary, academy, specialist disability school) and so on. Once we have this model, we will then repeat the exercise with data from the 40 schools in the study. This will be a separate NPD application and/or may be collected directly from the participating schools." When we asked for a copy of the original request via FOI, the Department for Education was unable to find and provide this data governance information.[44]

---

[43] Obsuth et al (2016) London Education and Inclusion Project (LEIP): Results from a Cluster-Randomized Controlled Trial of an Intervention to Reduce School Exclusion and Antisocial Behavior https://link.springer.com/article/10.1007/s10964-016-0468-4
[44] FOI request for the Institute of Criminology application for national pupil data (2013)
https://www.whatdotheyknow.com/request/pupil_data_licensing_agreements_2#incoming-1079063

The programme (strategy point 6.2), "which is itself one of the biggest data-linking exercises in government", brings together administrative data on over one million individuals and around three hundred and fifty thousand families, from local government and four central government departments. MHCLG may work with the Office for National Statistics (ONS), acting as a trusted third party, to match the data together from the local authorities and government departments to create an individual and family level dataset for analysis, but again, data are used for data analytics at scale (ie beyond ONS such as the Bristol ThinkFamily database) and identifying interventions without the knowledge or consent of families involved, including those who receive no intervention.

We suggest that the "individual and family level dataset" is likely currently unlawful where the data linkage creates new insights about the person, i.e. new personal data, and therefore needs new fair processing. Any predictive algorithms and risk scoring based on profiles needs special attention and basis for processing set out in law and with safeguards. Profiling should not routinely apply to a child and everyone is entitled to an explanation of the processing (recital 71, GDPR and para 77, Convention 108 Explanatory Note).[45] We have not seen these obligations met in practice.

The evaluation report suggests the total number of individuals matched have been:

- 114,916 to the Police National Computer;
- 376,235 to the National Pupil Database;
- 767,275 to the Work and Pensions Longitudinal Study and/or the Singl
- Housing Benefit Extract (317,033 adults; 450,242 children).

Evidence from Stephen Crossley (2018)[46] points out that the construction of troubled families as a social problem has led to, *"dirty data — information that is potentially discrediting and damaging — has been uncovered in a number of key areas, which casts doubt on almost all of the government claims in relation to the TFP. "*

The fact that the national evaluation of impact[47] of the Troubled Families Programme found that in the two years after joining the programme, the proportion of children on the programme persistently absent from school fluctuated and there was no clear trend, should itself be a sign that such an indicator used as a risk predictor is unlikely of significance, and should not be used.

The proportion of children either classified as a Child in Need or with a Child Protection Plan decreased in the two years after joining the programme. The proportion of Looked After Children increased in the same period. This information is given again as if significant, yet there is no contextual assessment on the change in use of s47 child protection law or discussion of academic

---

[45] Data subjects should be entitled to know the reasoning underlying the processing of data, including the consequences of such a reasoning, which led to any resulting conclusions, in particular in cases involving the use of algorithms for automated decision making including profiling. For instance in the case of credit scoring, they should be entitled to know the logic underpinning the processing of their data and resulting in a "yes" or "no" decision, and not simply information on the decision itself. Having an understanding of these elements contributes to the effective exercise of other essential safeguards such as the right to object and the right to complain to a competent authority.
[46] Crossley, S. (2018) Troublemakers: The construction of 'troubled families' as a social problem, Policy Press, University of Bristol
[47] National evaluation of the Troubled Families Programme 2015 to 2020: findings (March 2019)
https://www.gov.uk/government/publications/national-evaluation-of-the-troubled-families-programme-2015-to-2020-findings

research that suggests concerns that the re-authorization of Section 47 work through its endorsement of practices skewed towards investigative child protection and a broader range of agencies drawn into child welfare's more forensic practices, *"may exacerbate the social exclusion of poorer working class communities and women, ignoring the socio-economic determinants of child `abuse'."* (Broadhurst, 2007)[48]

The national evaluation again found bad data, but it is seemingly given little weight in finding the conclusions that the policy is effective either way. *"Employment status is based on P45 returns, provided by data from Her Majesty's Revenue and Customs (HMRC). There are **significant data quality issues** that made it difficult to draw robust conclusions about the Troubled Families Programme using this data."*

We are also told that Troubled Family becomes the dataset for further analytics and purposes, for example in children's social care predictive algorithms, and we are concerned that such bias and bad data are reinforcing patterns, such as a family in poverty, rather than identifying need.

> *"Predictive analytics is not new to the public sector. There are many examples of risk algorithms that have been in use, without controversy, for some years, such as the Youth Justice Board's approach to determining the risk of re-offending and the 'DASH' risk model used by police and social workers to understand the risk of domestic violence escalating."*[49]

New awareness of such processes and their failures should encourage new risk assessment and fresh review of practices. When a company CEO says that their measure of product 'failure' is whether or not it saved the Local Authority money, the powers-at-be should ask themselves whether the impact on children's lives is being adequately assessed for the effects of these machine-led data based decision-making with lifelong consequences, before even 'wider data sharing' is encouraged.

---

5. Better data

The strategy does not concretely define what it means by "better use of data" or "data use that is more secure", "more innovative" or "more widely recognised as a force for good."

But for individuals in recent years, data about children in the education sector has been about none of these concepts. Instead it has been seen as being used against them, as a burden, and subject to boycott and breach. These would all be remedied by either fixing public trust in good data policy, following the rule of law, and using standard good data governance frameworks.

---

[48] Broadhurst K, Mason C, Grover C. (2007) Critical Social Policy: Sure Start and the `re-authorization' of Section 47 child protection practices;27(4):443-461. doi:10.1177/0261018307081807 https://journals.sagepub.com/doi/10.1177/0261018307081807
[49] Xantura blog (December 2020)
https://xantura.com/despite-the-headlines-new-predictive-analytics-have-the-power-to-improve-public-sector-decision-making-for-everyones-benefit/

Burdensome data (2010): Standardised testing that generates national 'accountability' data is seen without local value and a cause of unnecessary stress to children. It resulted in a boycott of thousands of Key Stage tests in 2010.[50]

Boycott after misuse through intentional government policy (2016+): in pursuit of the Hostile Environment, data from the school census is still handed over monthly to the Home Office from the National Pupil Database. This discovery in 2016 resulted in a data boycott in the school census of the new expansion, to collect nationality and country of birth[51] in 2016-17[52] to an extent that made the data both objectionable, and invalid. Collection ended in 2018.

Data breach (2020): as a result of sloppy information governance and security process management by the Department[53] data from the Learner Records' Service in 2020 was accessed by gambling companies, onboarding new clients.

Data harms (2020): The public perception of the unjust and unfair adoption of automated data driven decisions reached a new low in the August of 2020 as Ofqual had ignored concerns and/or recommendations from the Education Select Committee[54], The Royal Statistical Society, us and many others on the decisions it made over the processing of data to deliver exam grades for exams that had not been sat in the summer of 2020. The Prime Minister went on to publicly attribute harm to a 'mutant algorithm' and the Chief Regulator of Ofqual resigned her position.

There is very low public understanding of how the exams awarding process and system of comparable outcomes, uses administrative data about children collected over time, not only from the exam 'on the day' but even statutory testing from age 10-11. Parents are told these tests do not affect outcomes and should not be prepared for, so are not perceived as a measure of a child's personal attainment, but sold as a measure of the *school* system, in the national state education accountability model. But we believe that they do have a significant effect on young people because SATs scores are used to profile state educated children with progress measures, predetermine a child's educational path, and predict GCSE grades, as well as being used in the results standardisation. An Ofqual 2017 video demonstrates how it is done.[55]

While we had proposed ahead of the exams, that explainability was important (under Article 22 of the GDPR) not only under these exceptional circumstances, but every year, and that a report should be produced to enable each candidate to see and understand the data used in their grade

---

[50] The Times (2010) Thousands of Key Stage tests cancelled after teachers boycott exams
https://www.thetimes.co.uk/article/thousands-of-key-stage-tests-cancelled-after-teachers-boycott-exams-qhlrvtqrb7t
[51] Timeline of the school census expansion 2016 https://defenddigitalme.org/timeline-school-census/

[52] Staton, B. (2017) Sky News | School census boycott over child deportation fear
https://news.sky.com/story/school-census-boycott-over-child-deportation-fear-11067557
[53] Trandall, S. (2020) Civil Service World | DfE data protection 'tightened significantly' after massive breach of learner records
https://www.civilserviceworld.com/news/article/dfe-data-protection-tightened-significantly-after-massive-breach-of-learner-records
[54] Getting the grades they've earned: COVID-19: the cancellation of exams and 'calculated' grades: Response to the Committee's First Report (Education Select Committee) https://publications.parliament.uk/pa/cm5801/cmselect/cmeduc/812/81202.htm
[55] Using key stage 2 results data to predict GCSE results (Ofqual) 2017 https://www.youtube.com/watch?v=BDhS-Gx9-o8

calculation, such information was not made available in detail on results day, and the publication of code since[56], does not meet the explainability requirement, to each student, either.

We await to find out if and how the modelling was lawfully applied by Ofqual and exam boards.[57]

It is not within the remit of the Office for Statistics Regulation to regulate operational decisions made by government bodies about outcomes for individuals, nor to form a judgement on decisions about government policy. So the OSR Review will not address implications of the models for the results of individual students or evaluate policy decisions on the most appropriate way to award exam grades in the absence of exams. But the government must learn from this experience – not only that students will take to the streets if they perceive systemic injustice and that data has been used to hold back their lives, but to make sure that these statistical models are trustworthy and serve the public good.

> As Ed Humpherson, Director General, Office for Statistics Regulation wrote in October 2020, *"From the perspective of statistics serving the public good, the exams story looks worrying. Not only has there been widespread criticism of the statistical models because of the perceived unfairness in the results they created; but there is a risk that future deployment of statistical techniques may be held back by the chilling effect of the poor publicity. That would be a real setback: it would limit statistical innovation and mean that the public sector cannot use new approaches to providing services to the public."*[58]

5.1 Case study: Proposal on automated decision-making explainer for exam grades

Exam boards should produce 'better data' about how data is used. A template explainer for students could be viewed securely online, and made available for schools to download and print for students on demand, to set out in a simple max two sided A4 document how a grade is arrived at. Much of the explainer would be the same per subject, per board. And in addition, it would have a personalised section for each student. It would include information, data and charts, of any information that influenced the students' grade including aggregated/ national data, i.e. data such as the ranges using KS2 scores / grade boundaries, comparable outcomes calculations, and how any local CAG assessed data affected decision making. Students should be able to securely access the data online on 'exams day'.

On demand, each centre could access the explainers, per pupil, as a merged report and print to give to each individual student, with the additional section of their awarded grades and the data / calculations involved.

---

[56] Ofqual publishes computer code developed to support GCSE, AS and A level awarding of calculated grades in summer 2020. (December 2020) https://www.gov.uk/government/news/code-used-to-calculate-qualification-grades-in-summer-2020
[57] FOI request to the ICO did not disclose the discussions between the regulators on the algorithms, but confirmed an ongoing review.
[58] Ed Humpherson, Director General, Office for Statistics Regulation (October 2020) The exams algorithm story is about more than just exams https://osr.statisticsauthority.gov.uk/the-exams-algorithm-story-is-about-more-than-just-exams/

6. Summary answers to the structured survey questions

1. **Question: Taken as a whole, do the missions and pillars of the National Data Strategy focus on the right priorities.**

   The strategy pillars and mission are fundamentally flawed in that they try to express what government wants almost exclusively as economic outcomes from data ***use***, in effect in terms of 'we've got all this data, now how do we use it" whether generating income 'pro-growth' or in government spending efficiencies, and 'combating fraud' as per the key strand of the Digital Economy Act 2017. It does not express its aims and tactics or how to achieve them in terms of human need, roles and responsibilities which are fundamental as to understand *why* there is public administrative data at all and its life cycle, *where* in what parts of government, *who* must be involved in any strategy and activities involving the datasets, and *how* it is governed. Thinking about data use in terms of "the high watermark set by the pandemic"[59] fails to acknowledge that certain permissions on certain data use are *time-limited*, under the exceptional conditions of a pandemic that needs exceptional data governance and those exceptions expire under 'normal' circumstances. Uses that ignored consent or other lawful requirements in order to expedite processing, would no longer be valid and should and could not be lawfully 'sustained'.

   The 'five concrete and significant opportunities for data to positively transform the UK' while including 2.4 services transformation, again sees this from the perspective of centralised government with repeated focus on cost effectiveness and not from the point of view of the person at local level who engages with a public service.

   The core pillars of the strategy while mentioning law and the infrastructure for data use and access, among 'a number of interconnected issues [that] currently prevent the best use of data in the UK' fail to include what is needed to deliver, "*real change,*" that is the mechanisms to meet the lawful requirements to explain to the public how their data are used, and what will be done to assess the data currently held in order to know what is held, with what permissions, and therefore whether it can be used lawfully *at all* for the purposes under discussion and new in scope for the strategy.

2. **We are interested in examples of how data was or should have been used to deliver public benefits during the coronavirus (COVID-19) crisis, beyond its use directly in health and social care. Please give any examples that you can, including what, if anything, central government could do to build or develop them further.**

6.1 Case study A: Device allocation to schools under COVID-19

In 2020, in order to support a rapid response to remote learning, and the lack of previous funding for devices across the country to enable learning to continue online at home, the

---

[59] Defenddigitalme submission to the Education Select Committee related to the impact of Covid-19 on the digital environment in education | https://committees.parliament.uk/writtenevidence/8444/html/

DfE stepped in, and after court action was proposed in some London boroughs.[60] Schools were told however, the exact number and type of devices available would be confirmed at the time of ordering based on stock availability and the extent of coronavirus restrictions. Schools reported delays and not receiving equipment, and in October 2020 the Department for Education told schools a new approach, would be introduced after half term, will ensure "allocations are more effectively targeted to the children, schools and areas of the country that have greatest need". The new system meant schools would only be able to claim around 20 per cent of what they had been allocated. 'Headteacher Chris Dyson said the decision, communicated to heads on Friday evening, was "clueless". His allocation was cut from 61 to just 13.' (Schools Week, October 23)[61]

In response to our FOI, still pending a full response[62], the device allocation is the number of laptops and tablets DfE estimate are needed based on:

- the number of children in years 3 to 11
- free school meals data (*not taking into account the estimated exponential growth in need in the pandemic due to changes of employment income, or children with NRPF[63])
- an estimate of the number of devices a school already has

The data used in the calculation remain opaque at the time of writing, despite a request made to the DfE for the data to be made open. Assumptions seem weak, and using Teacher Tapp (an advertising and marketing driven app,[64] whose audience is primarily Twitter users in the teaching sector) as 'policy data' has not been independently assessed in terms of quality and bias. The DfE despite its vast data collection, does not necessarily have accurate data and rather than asking for and trusting schools requests, the DfE has made a series of assumptions. While we await concrete data, the explanations behind the calculations suggest little more than informed guesswork:

"Survey data suggests that there are fewer children without access to a device than there are FSM children so simply using FSM does not give an accurate picture of device access requirements. The

---

[60] Good Law Project (2020) Every child has the right to a suitable education: that is what the Education Act says. But the Government's plan to continue education online during the coronavirus lockdown risks leaving a million children behind. https://goodlawproject.org/case/children-will-be-left-behind/

[61]Dickens, J. Schools Week (October 2020) 100k more laptops for schools – but access is slashed https://schoolsweek.co.uk/100k-more-laptops-for-schools-but-access-is-slashed/

[62] FOI request to the DfE Covid-19 device provision: calculation data https://www.whatdotheyknow.com/request/covid_19_device_provision_calcul

[63] All children, regardless of their immigration status, can receive state school education whilst they are of compulsory school age. All children in reception, year 1 and year 2 at state schools in England automatically get free school meals, regardless of their immigration status. For older children at state schools, normally eligibility for free school meals is linked to the parent being in receipt of certain welfare benefits, all of which are public funds (Section 115 of the Immigration and Asylum Act 1999 and para 6 of immigration rules). http://guidance.nrpfnetwork.org.uk/reader/practice-guidance-families/introduction/#14-good-practice-points A 2020 briefing from the Migration Observatory at the University of Oxford shows that more than a million children under 18 in the UK (1,082,000) do not have British or Irish citizenship. More than one in six of these children (175,634) live in a family expected to have no recourse to public funds https://migrationobservatory.ox.ac.uk/press/a-million-children-in-the-uk-dont-have-british-or-irish-citizenship-and-175000-live-in-families-expected-to-have-no-recourse-to-public-funds/ "Fernández-Reino: We estimate that 175,000 children in the UK live in families whose immigration conditions mean that they are likely to have no recourse to public funds. Having NFPF could increase the financial hardship of children in families whose adult members had lost their jobs or had large income losses as a result of the COVID-19 crisis."

[64] Every day at 3.30pm, Teacher Tapp pings its members a short poll which they complete on their mobile phones in minutes. *"This produces a bank of rich, up-to-the-minute data on the daily lives of teaching staff which can prove invaluable to businesses who are researching and marketing products. There are three options for businesses and researchers: Commissioned surveys & questions, back-catalogue data & education insight reports, and targeted advertising."* https://teachertapp.co.uk/business/

Department uses this FSM data <u>alongside assumptions that private device needs will be met to some extent by the device endowments of schools and colleges</u>. To construct this estimate, we have used the BESA ICT Survey 2019 data on the average number of laptops and tablets in primary and secondary schools <u>and assumed that these are distributed between teachers and students.</u> This survey identifies that primary schools have an average of 27.2 Tablets and 49.1 Laptops, and Secondary schools have an average of 87.3 Tablets and 194.6 Laptops. We have also used The Teacher Tapp survey (March 2020) which shows the proportion of teachers provided with a device by their school. The survey found that 45% of primary school teachers and 29% of secondary school teachers were provided a laptop by the school.<u> We use this data to assume that the first devices in schools will go to teachers and remove devices equivalent to this percentage</u> of teachers in an average school from the average number of devices in primary and secondary schools, on the basis that some of these devices will go to teachers. <u>This produces an assumption</u> based on this data that 24.5% of pupils in primary and 27.1% of pupils in secondary would receive a device from their school."

6.2 Case study B: NAO report[65] Investigation into the free school meals voucher scheme

"On 31 March 2020, the Department announced a national free school meals voucher scheme, as a temporary substitute for the normal arrangements. The vouchers were worth £15 per week for each eligible child.

"The Department awarded a contract to administer the voucher scheme to a private contractor, Edenred (UK Group) Ltd (Edenred). Schools had a choice whether to use the scheme or to make local arrangements. Schools participating in the scheme ordered electronic codes (eCodes) online that could be converted into shopping vouchers for families to use at nominated supermarkets. The vouchers were mainly in the form of electronic gift cards, although paper vouchers could be printed." (NAO audit, December 2020)

The National Audit Office found data related failures during COVID that led to harm for families. Parents' email data accuracy meant vouchers were not received in timely ways or at all. On accountability there was no data produced by the Department on Edenred success or failure, no standard Service Level Agreement terms to meet, and no accountability for spend afterwards. The audit found no evidence that the Department knew how many children the vouchers had helped, and this was not as the audit suggested the DfE had framed it, due to data protection requirements, but in failure to think how and what data was needed --a simple count of pupil uptake, not named data would have sufficed at organisational level. Further, there was no audit requirement locally on the ordering process and voucher provision.

Public data about service provision should be collected and used to deliver accountability metrics about services not the use of data as a commodity, but data as a control measure.

---

[65] NAO Investigation into the free school meals voucher scheme (December 2020)
https://www.nao.org.uk/report/investigation-into-the-free-school-meals-voucher-scheme/

The Welsh Government decided to support remote learning provision, that "schools will no longer rely on consent and instead will provide these additional services as part of their public task" a change the government "has been considering making [...] for some time, but has brought forward the decision." But Google Inc. cannot rely on the public task and we argue that Google does more processing with children's personal data collected in the use of Google for Education product suite, than is necessary and the minimum required for the purposes of the school.

Across the UK, core national education infrastructure that process data for the routine functioning of the education system, must be put on the national risk register. Dependence on products such as Google for Education, MS Office 365, and cashless payment systems, all need to have a further duty to transparency reporting obligations. We are currently operating in the dark where remote learning is and is not supportable, and about the implications of dependence on these systems for the delivery of key school functions and children's learning. Rapid response by the Department for Education in England to support schools without any remote learning platform was exclusively supportive of the 'Tech Giants' Google and Microsoft, and indirectly supported their market foothold. Yes, "many educational settings lack the infrastructure" but that should never mean encouraging ownership and delivery by exclusively closed commercial partners. The current route risks the UK losing control of the state education curriculum, staff training and (e)quality, its delivery, risk management, data, and cost controls.

The controls on companies' access to children's data, is what controls the knowledge companies get about the UK delivery of education and the state education sector. That business intelligence is produced today by the public sector teachers and children who spend time administering and working in the digital systems. So while many companies offer their systems for free or at low cost to schools, schools have intangible costs in staff workload and support time, and donate those labour costs to companies for free. Our children and public sector staff are creating a resource that for-profit companies gain from for free. Is this either sensible or sustainable without a set of assurances and safeguards?

At national level there is no independent oversight of how any data infrastructure at local and regional school level is managed. The delivery of education fails to appear on the national risk register despite its brittleness and the problems in its vulnerability caused by remote learning demands in response to the shock of the COVID-19 pandemic. Who owns and has responsibility for the infrastructure? How much is dependent on Silicon Valley and what is known about future ownership, stability, security and costs? Where is it inadequate? Does it meet needs from the child's educational perspective? What plans exist if a company that provides its products plus teacher training for free today, nationwide, starts charging tomorrow? All questions that should be transparently in the public domain and yet any agreements between the State and the company tend to be kept entirely closed, rather than protecting only narrowly significant competitive

---

[66] Department for Education press release (April 2020) Schools to benefit from education partnership with tech giants (Google and Microsoft start up support) https://www.gov.uk/government/news/schools-to-benefit-from-education-partnership-with-tech-giants via The Key: Digital education platform hub

commercial information, "held under an obligation of confidence."[67] This fails to serve the public interest.

3. **If applicable, please provide any comments about the potential impact of the proposals outlined in this consultation on individuals with a protected characteristic under the Equality Act 2010.**

6.4 Case study: Protected characteristics in Higher Education applicant data

Government changed the law on Higher Education in 2017 and got it wrong[68] to start the collection of protected characteristics which was entirely foreseeable at the time. Now third parties pass around named equality monitoring records like religion, sexual orientation, and disability and it is stored forever on named national pupil records. The Department for Education (DfE) now holds sexual orientation data on almost 3.2 million, and religious belief data on 3.7 million people[69] in the National Pupil Database, without the individuals' knowledge. The defenddigitalme view is that these identifying data should be destroyed.

The DfE is an accountable data controller for these data even if sourced from HESA for example. For many worthy reasons, many organisations advocate for the collection of religion or sexual orientation among equality monitoring data. But any risk of loss, theft, leak or misuse could be devastating to individuals and communities. Risk of damage or distress is high. There can be no justifiable necessity for distribution of named data in this way that is proportionate to the risks to fundamental rights and freedoms.

We have asked the Universities and Colleges Admissions Service, with regards to the processing of any student or Higher Education staff personal data, to provide the most recent Data Protection Impact Assessment(s) regards processing or data that UCAS control, and that may contain "protected characteristics" of age, disability, gender reassignment, race, religion or belief, pregnancy and maternity, sexual orientation and, in certain circumstances, marriage and civil partnership.

The lack of accountability and passing the buck of responsibility, some to the point of not even assessing risk or clear ownership for such risk, is poor.

Protected characteristics in Higher Education (in NPD) controller responses

- UCAS FOI https://www.whatdotheyknow.com/request/student_and_staff_data_dpia_5
- OFS (pending reply / Internal Review as answer was incomplete) https://www.whatdotheyknow.com/request/student_and_staff_data_dpia_3
- Wales https://www.whatdotheyknow.com/request/student_and_staff_data_dpia

---

[67] WhatDoTheyKnow FOI request | **UK Crown Commercial Service (CCS): google cloud MoU** https://www.whatdotheyknow.com/request/705140/response/1689640/attach/html/3/FOI2020%2017808%20REPLY.pdf.html
[68] Higher Education and Research Act 2017 (s79) https://www.legislation.gov.uk/ukpga/2017/29/section/79/enacted The Higher Education and Research Act 2017 (Further Implementation etc.) Regulations 2019 (000/2019)
[69] The Department for Education (DfE) holds sexual orientation data on 3.2 million people, and religious belief data on 3.7 million. https://defenddigitalme.org/2019/07/statement-on-student-religion-or-belief-and-sexual-orientation-data-in-the-national-pupil-database/ The records go back to 2012/13, so include both current students and those who have finished university.

- Scottish Higher Education and Funding Council (not held)
  https://www.whatdotheyknow.com/request/student_and_staff_data_dpia_4
- HESA (declined)
  https://www.whatdotheyknow.com/request/student_and_staff_data_dpia_2

HESA says that, "*The information we collect includes equality and diversity information about students so that the funding and regulatory bodies, and HE providers themselves, can meet their public sector duty to ensure equality of opportunity and prevent discrimination,*" but makes no risk assessment of the data it holds and has therefore we can only assume not assessed the potential for collecting only aggregated data rather than individual identifying data, to distribute to meet that duty.

The proposals outlined in this consultation on individuals with a protected characteristic under the Equality Act 2010 should ensure that only aggregated statistical data processing is permitted beyond the point of collection, for the purposes of equality monitoring and revisit data legislation, such as the Higher Education Act 2017 to put in such safeguards.

---

4.   **We welcome any comments about the potential impact of the proposals outlined in this consultation on the UK across all areas, and any steps the government should take to ensure that they take account of regional inequalities and support the whole of the UK.**

6.5 Case study: Inconsistency of who is in the data, who can access it, and its support

Public admin data is collected from the school census. School census requirements are mandated by national governments and vary across regions. The data that each school holds about each child may therefore not be of a school's choice, but a statutory obligation. This brings with it additional burdens for the school to collect, clean, maintain, retain and process that data and the relationship between the data controller (the Department for Education) and data subject (the child and/or parent). Education data are fundamentally unequal and unrepresentative of 'education' per se in the UK. The national administrative datasets only tell part of the picture about how education is delivered and measured. Private and independent schools may not submit data to the Department.

Section 537A of the Education Act 1996 enables the Secretary of State to require state funded schools to provide certain information, including UPNs, to local authorities and the department including: key stage assessment information and pupil level census returns. Section 408 of the same Act enables the Secretary of State to require the transfer from school to school of pupil records, including UPNs, using CTF. Schools are legally entitled to transfer such data and do not need to seek pupils' or parents' *consent* to do so. National guidance on the use of the UPN was updated in December 2017 and 2019. **Both state that the UPN must lapse when pupils leave state funded schooling, at the age of sixteen or older. What this should mean in practice for providers is unclear because by age sixteen it has been widely distributed, and linked with other data including the unique learner numbers which take over identification post-16.** Previous Guidance suggested that the

UPN[70] is a 'blind number', not an automatic adjunct to a pupil's name and only transferred to those with a genuine right and requirement for its receipt. It should certainly not become a general identifier across children's datasets across government.

In fact the EU Commission High Level Expert Group in AI, proposed in 2019, that children should be ensured a free unmonitored space of development, and upon moving into adulthood should be provided with a "clean slate" of data storage by default with retention beyond compulsory education in administrative datasets, on a necessary and proportionate legal basis, as an exception not the rule.[71]

There is a legislative discrepancy in parental rights[72] to pupil data access in the educational record across different school types, beyond data protection law, due to inconsistent legislation. It is not standardised for children across all settings and inconsistent between Local Authority and academies and other models of education. **The government should undertake an assessment of all data legislation that affects the UK public administrative datasets, and ensure it is fit for purpose, and ensure new standards of transparency reporting, algorithmic registers, and Codes of Practice to promote awareness of expected high standards, where and how they should be effectively and consistently applied.**

Furthermore, there is a postcode lottery of how well school information management systems suppliers (all known to and in regular contact with the DfE) offer any standard reporting mechanisms to meet data protection law and this creates inconsistency in how Subject Access Rights are fulfilled by a wide variety of settings, their workload and cost. Schools can struggle to meet SARs due to the way in which information is managed, and some offer limited system ability to generate legally required documents. In a quid pro quo for MIS providers to access the public sector, providers should be required to demonstrate a high minimum standard requirement to support schools' SAR needs.

8.   **What could central government do beyond existing schemes to tackle the particular barriers that small and medium sized enterprises (SMEs) face in using data effectively?**

6.6 Case study: Longitudinal Education Outcomes Data opened up to SMEs

The effective use of data relies on data quality and accuracy. Government must improve the contextual understanding of data it can access and gives to others, in order to create effective and accurate policy. The worst of structural discriminations in society and education are reinforced by algorithms and reinforced in bad products by design based on bad data. Government could and should change the quality standards it sets itself.

---

[70] Department for Education UPN guidance March 2018 http://defenddigitalme.com/wp-content/uploads/2018/04/UPN_Guide.pdf
[71] AI High-Level Expert Group, Policy and Investment Recommendations for Trustworthy Artificial Intelligence (April 2019) (accessed July 1, 2019) https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence permanent copy https://defenddigitalme.com/wp-content/uploads/2019/07/AIHLEGPolicyandInvestmentRecommendationspdf.pdf
[72] The Education (Pupil Information) (England) Regulations 2005 do not apply to non-maintained schools (e.g. academies, free schools and independent schools). https://www.legislation.gov.uk/uksi/2005/1437/contents/made

In 2018, then Universities Minister Sam Gyimah launched a £125,000 competition for tech companies and coders to develop new digital tools to help students pick the university course that is best for them. "The aim of the competition is to spur private sector innovation", by providing the seed funding to cover the costs of the initial research and development. The increasing reliance on data to inform decisions and policy means that the lack of awareness of who or what is not represented in the data means misinformed decision making is exacerbated.

Aside from all other caveats this dataset is old, historic data and should not be used for forward thinking, future employment or economic planning or predictions on what is changing and not yet known.

> *"Longitudinal Education Outcomes (LEO) data enables us to know how much UK graduates of different courses at different universities are earning now, either one, three or five years since graduating. It does this by linking up tax, benefits, and student loans data. But it's just as important to state at the outset what LEO is not. LEO does not help us identify the universities with the best or most effective teaching, nor is it a measure of the 'value added' by a university degree. LEO is not a performance indicator. It isn't even a predictor of how much students at any university or on any course will earn in the future: it's historical data, and future labour market outcomes are arguably too variable to be predicted based on past outcomes." (Morris, 2017)[73]*

The data that are riddled with caveats and based very much on an ideological position rather than accurate information, was transformed from a set of mixed data explained with caveats and flaws by statisticians, described the Office for Students as "experimental and still being evaluated and are not yet fully developed,"[74] to a fact-basis for policy making not only by the government, but for third parties to develop products from, based on a fiction. Not even all subject courses have similar content.

Then Secretary of State for Education was reported in the press as saying, some degrees were "letting down thousands of students and costing taxpayer millions". The *Longitudinal Education Outcomes (LEO) dataset* is the data in education most akin to a pumpkin turned into Cinderella's coach, but in Parliament Liz Truss, as then Minister to the Department for Trade, pointed out not its weaknesses in Parliament but claimed it could predict graduate incomes, saying the data "*shows people how much they can expect to earn by studying particular courses.*"[75] It does not, but popular political description has promoted a belief in the beauty and power of the data, based on misinformation. It will at some point wear off, but may be dangerous before then. This makes bad policy and bad products that support bad decision making by potential new students and by policy makers. Central government should not mis-sell data whether to MPs, the public, or to small and medium sized enterprises (SMEs) and encourage data misuse.

[73] Morris, D. (2017) WONKHE | A beginner's guide to Longitudinal Education Outcomes (LEO) data
https://wonkhe.com/blogs/a-beginners-guide-to-longitudinal-education-outcomes-leo-data/
[74] Office for Students (February 2020) Graduate earnings data on Discover Uni from the Longitudinal Education Outcomes (LEO) data
https://www.officeforstudents.org.uk/data-and-analysis/graduate-earnings-data-on-discover-uni/
[75] Liz Truss MP, Secretary of State for International Trade | Global Britain | Volume 670: January 30, 2020
https://hansard.parliament.uk/Commons/2018-12-18/debates/DEB0C8A8-5EDA-4C73-9DCE-0516A4B17C47/ONSDecisionsStudentLoans#contribution-8717AC22-F1D6-4F10-B6A8-93B4D9OC474F

10. **How can the UK's data protection framework remain fit for purpose in an increasingly digital and data driven age?**

There is little wrong with the core frameworks -- the UK Data Protection Act 2018 and its GDPR foundations, or the Convention 108+, or privacy and communications laws, which are also about data and all apply to the UK equally. Fundamental rights and freedoms are not only about data but may be protected or enabled through its legislation.

What needs fixed by removal in the Data Protection Act, is the government's ideological clause: The immigration exemption, in Schedule 2 Part 1 Paragraph 4 of the Data Protection Act 2018. And what has been omitted, Article 80(2) should be added, support for rights of redress, especially for children. In the whole round, enforcement action is still limited compared with other countries.

Given the UK police propensity for misusing the identity data of dead children[76] without permission and harming families' dignity and welfare, we suggest that there could also be an additional protection introduced for children's identity data under Recital 27, *"Member States may provide for rules regarding the processing of personal data of deceased persons."*

11. **To what extent do you agree with the following statement: the functions for the Centre for Data Ethics and Innovation (CDEI) should be Artificial Intelligence (AI) monitoring, partnership working and piloting and testing potential interventions in the tech landscape? How would a change to statutory status support the CDEI to deliver its remit?**

We remain to be convinced that Artificial Intelligence merits such unique attention that could not be given through existing organisations. Artificial Intelligence (AI) monitoring from a data compliance perspective remains within the remit of the ICO. The CDEI operations are somewhat opaque and its board meeting agenda and minutes are published only erratically.[77] Does it have a thematic roadmap that can be published over a fixed time period? Outputs from minutes over soon to be three years, seem more weighted to its own administration rather than delivering value-add outputs to sectors that could benefit from; curated expertise, or guidance. Has there been a regular 'State of The Nation' report?

We are yet to see any work from CDEI on education, and we have not been invited to or seen any public engagement work. Therefore it is not possible for us to comment if a statutory footing is of merit and what its role would be.

AI in education is more problematic than in other settings, as we show in some of the case studies in our recent report, the State of Data 2020[78], because of the imbalance of power in an educational setting. AI companies' terms and conditions may set out that they process on the basis of consent. But children cannot freely consent to the use of such services due to capacity and in particular where the power imbalance is such that it cannot be refused, or easily

---

[76] Evans, R. The Guardian (2020) Met faces legal action over spies' use of dead children's identities https://www.theguardian.com/uk-news/2020/dec/07/met-police-legal-action-spies-use-dead-childrens-identities
[77] Agenda and minutes from CDEI's board meetings https://www.gov.uk/government/publications/board-meeting-agendas-and-minutes
[78] 3.10.1 Artificial intelligence and product development | Case studies | https://defenddigitalme.org/state-of-data/

withdrawn and it is rare to see schools offer children a right to object or alternative method of learning. "*Public authorities, employers and other organisations in a position of power may find it more difficult to show valid freely given consent.*" (ICO)[79]

61% of parents polled by Nesta for the report Educ-AI-tion rebooted? anticipate that AI will be fairly or very important to the classroom in the near future (there was no indicator of what type or why). However, many are fairly or very concerned about consequences of determinism (77%), accountability (77%) and privacy and security (73%). According to Nesta, they are concerned how this may affect their children including through discrimination and social equity.[80]

It is notable when the CDEI went on to report on algorithmic bias in the aftermath of the Ofqual / summer exams crisis in 2020, its first footnote set out the conflict of interest between the Chair of the regulator and the same Chair of the ALB.[81] This is not an optimal relationship for the Chair of a body supposed to offer oversight and neutrality when it comes to national data processing, to have such a significant national conflict.

12.   **We have identified five broad areas of work as part of our mission for enabling better use of data across government: Quality, availability and access, Standards and assurance, Capability, leadership and culture, Accountability and productivity, Ethics and public trust. We want to hear your views on any actions you think will have the biggest impact for transforming government's use of data.**

> **Capability:** Without teacher training in statistics and understanding bias and data discrimination, teaching staff are likely to inadvertently perpetuate any historical bias in the data they have to interpret. Given the significance of carrying out assessment it is a big gap in teacher training, as Dr Becky Allen told the Education Select Committee Enquiry on primary assessment in 2017[82], that "*we do not have a system of training for teachers that makes them in any way experts in assessment*". Some schools had resorted to buying commercial options of varying quality, as described by the Association of Teachers and Lecturers concerned about several dubious "solutions" commercially available to schools which do not offer value for money or a high-quality assessment framework. It was proposed in 2017 that the risks "*of schools purchasing low-quality assessment systems from commercial providers*" are to be mitigated by high quality advice and guidance, rather than change of policy and practice. That recommendation from the enquiry into Primary

---

[79] Information Commissioner (ICO) | The right to object to the use of your data [processing under the public task] https://ico.org.uk/your-data-matters/the-right-to-object-to-the-use-of-your-data/

[80] To obtain the perspective of parents on AI and education Nesta commissioned YouGov to undertake a survey of 1225 GB parents with children aged 18 and under https://media.nesta.org.uk/documents/Future_of_AI_and_education_v5_WEB.pdf Educ-AI-tion Rebooted? Exploring the future of artificial intelligence in schools and colleges

[81] Review into bias in algorithmic decision-making (November 2020): "Footnote 1: *Note that Roger Taylor, the chair of the CDEI Board, is also the chair of Ofqual, the English exams regulator. Following the controversy around August 2020 exam results, Roger has stepped away from involvement in any changes made to the final version of the review. CDEI has not had any direct role in assessing Ofqual's approach, at the time of writing we understand a number of regulators are looking into the issues raised in detail.*" https://www.gov.uk/government/publications/cdei-publishes-review-into-bias-in-algorithmic-decision-making/main-report-cdei-review -into-bias-in-algorithmic-decision-making

[82] Education Select Committee Enquiry on primary assessment in 2017 https://publications.parliament.uk/pa/cm201617/cmselect/cmeduc/682/682.pdf

Assessment has not been realised and must be with strengthened data standards requirements.

Leadership: A national oversight body is needed on a statutory footing to oversee data governance in education to address the lack and compliance and accountability found by the ICO in its audit of the Department for Education, and issues in the broad use of edTech. A National Guardian for education and digital rights, would provide a bridge between government, companies, educational settings and families, to provide standards, oversight and accountability which is not all about data protection, but ethics as well. Capacity and capability across the sector would further benefit from a cascading network of knowledge with multi-way communication, along the principles of the NHS Caldicott Guardian model.

Culture: The Department for Education Data Management Review Group 2016 report with the aim of reducing teacher workload had a finding we can strongly support four years later. *"Government, school leaders, and teachers, rather than starting with what is possible in collecting data, should challenge themselves on what data will be useful and for what purpose, and then collect the minimum amount of data required to help them evaluate how they are doing. Decisions about the identification, collection and management of data should be grounded in educational principles. In this way schools can have greater freedom to balance professional autonomy and agency against the demands of the accountability system."*[83]

Culture around compliant and qualified data use is not all about the money, and if the strategy aims to bring about culture, *"in the same way that the government has worked to instill a culture that emphasises value for money,"* then its approach will fail to engage the public sector staff in ways you want. In education, where teachers already pay for children's breakfasts, classroom paper and glue sticks, it was badly received when the Academies Minister Lord Agnew insisted he could find more back office money savings, "like a pig out hunting for truffles when it comes to finding waste", and insisted more can be done and bet any headteacher "a bottle of champagne and a letter of commendation" that he can identify more potential savings in their schools.[84]

13. **The Data Standards Authority is working with a range of public sector and external organisations to create a pipeline of data standards and standard practices that should be adopted. We welcome your views on standards that should be prioritised, building on the standards which have already been recommended.**

Standards should also consider the quality and availability of data that is available about the data, in the public domain. Freedom of Information laws should apply to all non-state actors, companies and arms length government bodies, providing education and children's services to the state sector including third country processors. This is often not the case today, where such laws are limited to public service providers, and even then when it comes to sensitive subjects such as children's data processing in the Prevent programme from education, educational

---

[83] Department for Education Data Management Review Group (2016) Reducing teacher workload: Data Management Review Group report https://www.gov.uk/government/publications/reducing-teacher-workload-data-management-review-group-report
[84] Whittaker, F. (2018) Schools Week | Minister bets heads a bottle of champagne he can find savings in their schools https://schoolsweek.co.uk/minister-bets-heads-a-bottle-of-champagne-he-can-find-savings-in-their-schools/

bodies and local authorities consistently refuse access to statistical information baked on security exemptions.

14. **What responsibilities and requirements should be placed on virtual or physical data infrastructure service providers to provide data security, continuity and resilience of service supply?**

"Education settings can be involved in processing children's data, on a large-scale over long periods of time. Applying appropriate security measures to this data, and its processing environments both at rest and in transit, is vital to ensure children's data is protected to the highest standards. As the Convention 108 for data protection sets out, security measures should take into account the current state of the art of data-security methods and techniques in the field of data processing. Their cost should be commensurate with the seriousness and probability of the potential risks. The protective measures applied to personal data should be based on a risk assessment following industry standards and best practice and using established technical guidance (such as the ISO 27000 series and others as appropriate).

> *"Risk assessment prior to processing must also assess whether data is protected against unauthorised access, modification and removal/destruction. … Due diligence must be done to establish the third party's ability to protect personal data appropriately, including confidentiality, integrity and availability.*

> *"Any measures put in place should be regularly tested, as set out in Article 7 of the Convention 108+ and take into account the changing data-security methods and techniques and risks and be kept under regular review and updated where necessary." (Extracts from Council of Europe Guidelines (2020) on Convention 108, on data protection in educational settings, section 7)*[85]

15. **Demand for external data storage and processing services is growing. In order to maintain high standards of security and resilience for the infrastructure on which data use relies, what should be the respective roles of government and data service providers, their supply chain and their clients?**

Council of Europe Guidelines on the impact of business on children set out that lawmaking and procurement at all levels of government must respect the UN General comment No. 16 (2013) on State obligations regarding the impact of the business sector on children's rights[86] and this should apply to educational settings in procuring all providers. A Code of Practice is required for schools to establish consistent standards.

> 2.8.1 *"a State should not engage in, support or condone abuses of children's rights when it has a business role itself or conducts business with private enterprises.""States must take steps to ensure that public procurement contracts are awarded to bidders that are committed to respecting children's rights. State agencies and institutions, including security forces, should not*

---

[85] Council of Europe Convention 108+ | Guidelines on Children's Data Protection in an Education setting (adopted in November 2020) https://rm.coe.int/t-pd-2019-6bisrev5-eng-guidelines-education-setting-plenary-clean-2790/1680a07f2b
[86] Committee on the Rights of the Child General comment No. 16 (2013) on State obligations regarding the impact of the business sector on children's rights https://www.unicef.org/csr/css/CRC_General_Comment_ENGLISH_26112013.pdf

collaborate with or condone the infringement of the rights of the child by third parties. States should not invest public finances and other resources in business activities that violate children's rights."

Staff in the public sector assume commercial competency. Researchers at LSE in 2019 have documented how children care about their privacy online, that they want to be able to decide what information is shared and with whom, and further that, "teachers are unclear what happens to children's' data and there is common misunderstanding of how much data leaves a school:", "The only time it does [to the government] is when we do the Year 11 data [...] Because obviously they'll do the tracking of different groups. (teacher, London)" and when it comes to using educational platforms, "I would've thought the fact that it's a school-based software, this has all been properly regulated." (teacher, London)"

Most state schools are poorly equipped to address the requirements under the Data Protection Act 2018 (and GDPR Article 2518) to minimise its data collections and ensure proper policy, technical and security measures to address excessive data collection and enforce retention (including at national levels or for children leaving compulsory school), limit unique identifiers, and ensure anonymisation. They do not have specialist staff in data protection or security. This creates a power imbalance for companies especially in international locations, and who rely on school staff lack of due diligence and capacity in the procurement process and associated data processing.

The respective role of government is mentioned elsewhere in the need for funding and delivering expertise and training across the education sector in terms of basic teacher training and continuous professional development (CPD).

17. **To what extent do you agree with the following statement: The government should play a greater role in ensuring that data use does not negatively contribute to carbon usage?**

Rachel Coldicutt, formerly of DotEveryone, and founder of Careful Industries, summarised in a 2019 speech[87], "*The carbon cost of collecting and storing data no one can use is already a moral issue. So before you add another field, let alone make a new service, can you be sure it will make enough of a difference to legitimise its impact on the planet? And can you be sure it will serve the public – all of the public, in a way that aligns to your organisation's mission and values?*

"Mary Meeker predicts that by 2020, only 16% of the world's data will be structured enough to be useful. Or to put it another way, 84% of the data collected about people and things is sitting unused. Using up servers, eating electricity. To quote a recent blog post by AI Now Institute:

"Researchers Lotfi Belkhir and Ahmed Elmeligi estimate that the tech sector will contribute 3.0—3.6% of global greenhouse emissions by 2020, more than double what the sector produced in 2007 (Belkhir and Elmeligi, 2018). The estimated 2020 global footprint is comparable to that of the aviation industry, and larger than that of Japan, which is the fifth biggest polluter in the world. Data centers will make up 45% of this footprint (up from 33% in 2010) and network infrastructure 24%.

---

[87] Coldicutt, R. (2019) Just enough Internet: Why public service Internet should be a model of restraint
https://www.doteveryone.org.uk/2019/10/just-enough-internet/

The government has an enormous role in ensuring that data use does not negatively contribute to carbon usage insofar as it is an enormous controller of data, and without audit, it is unlikely to know how much of the public administrative data is *sitting unused.* Not only has the government obligations to society to do all it can to protect our living environment, but obligations in law on data minimisation and data retention, that at least in DfE controlled datasets, are not met today. Enforcement action may be the only route to remedy if the government takes no action of its own accord.

18. **/ 19. How can the UK improve on current international transfer mechanisms, while ensuring that the personal data of UK citizens is appropriately safeguarded? What are your views on future UK data adequacy arrangements (e.g. which countries are priorities) and how can the UK work with stakeholders to ensure the best possible outcome for the UK?**

The requirements are clear in recital 104 of the GDPR; follow the rule of law that is not met today. The assessment will take into account how a particular third country respects the rule of law, access to justice as well as international human rights norms and standards and its general and sectoral law, including legislation concerning public security. Furthermore, all international and domestic transparency must be meaningful and demonstrate accountability. Third country transfers are opaque to children and their legal guardians. Such decisions are self determined and schools cannot independently assess them in contractual clauses or otherwise. Whether binding corporate rules pursuant to Article 47 are met, can be specialist knowledge beyond most schools internal expertise. To facilitate scrutiny by external bodies such as Supervisory Authorities and civil society, Public Authorities should publish an open register of:

- processors / sub processors they engage processing children's data
- a register of any commercially obtained sources of personal data collected for processing, or linkage with data provided by individuals in the course of their public sector interactions, and update it on a regular 14 basis. (i.e. Data brokers, third-party companies, social media)
- Data Protection Impact Assessments, Retention schedules, and GDPR s36(4) Assessments with periodic fixed review to address changes
- A duty on the controller to log all recipients of personal data; and an obligation to present the file to the data subject and where they do not have capacity, their legal guardians

The relationships between the data subject controller and processor should be maintained and be up to date due as long as necessary to meet the school obligations on fair processing, accuracy and to facilitate Subject Access requests, while processing may be abroad. The best position would be to 'bring data home' as Hamburg recommended to its education sector on use ofMicrosoft 10, Google and other cloud based services.

Defenddigitalme
December 2020

Annex: Comparison table of UK national pupil databases[88]

Identifying pupil data in national school census

	England June 2018	Northern Ireland October 2016	Scotland March 2017	Wales March 2017
Volume of data	21,230,000	327,122	1,265,501	1,034,907
Year begun	1998	2006	2007	2003
Pupil level	yes	yes	yes	yes
Named	yes	yes	no	yes
Sensitive data	yes, incl. indicators for children of parents in services, ethnicity, nationality, and looked after children. Not religion.	yes, incl. religion, ethnicity and indicators for children of parents in services	yes incl. ethnicity , language, nationality, in care, and indicators for children of parents in services	yes incl. ethnicity , language, nationality, in care, and indicators for children of parents in services
Subject Access permitted to your own record	no	TBC	TBC	yes
Third party access is restricted to those with accredited training or access only with oversight in safe setting.	no. "Applicants must complete license agreement, terms and conditions."	TBC	no	No. "Applicants must complete a Data Access Agreement (DAA) and a Security Aspects Letter (SAL)"
Distributed to third parties	yes	yes.	yes — in most cases, researchers make use of the eDRIS facilities which have been security accredited by the Scottish Government. (Trusted third party model)	yes
Identifying data distributed to commercial third parties for re-use	yes, including journalists, data analytics companies, charities, for turning into print and online school comparison tools, think tanks.	no	no	yes but limited to FFT for analytics and resale to schools and local authorities, see 'Third party distribution' below.
Small number suppression	"There is no suppression applied to data extracts from the NPD before release. Instead, Requesters are required to sign up to strict terms & conditions covering the confidentiality and handling of data, security arrangements, retention and use of the data. These include that no individual will be identified in published data. The Daily Telegraph requested pupil-level data and so suppression was not applicable." [source]	Typically the Department suppresses information of fewer than five pupils for any data that is deemed as potentially identifiable. Any figures greater than five can also be "counter-suppressed" to avoid working out a suppressed figure by subtracting it from a total.	Our general policy is that numbers under five are suppressed however other numbers may also be suppressed where it is felt necessary to do so to reduce risk of identification of individuals. Where applicants require access to disclosive data for research, different processes apply which restrict the use and sharing of the data.	Individual level data is fully anonymised in most cases. Disclosure control is applied to all aggregate data, most commonly by not publishing any actual counts less than 5 or any counts that can be derived to be less than 5 from other data in the dataset to be released.
Audit of recipients	no	no	We have not conducted any audits. We work with partners to encourage the use of recognised safe havens, and this is reflected in our application processes.	no
Typical release length	between 6 months to 3 years depending on the aims of the project, is often extended	not known	not known	Most cases when one calendar year has passed
Indefinite retention	yes	yes	yes	yes
Data collection sources	School Census is every third Thursday in each term (three times a year) and added to around 20 other data sources to create the National Pupil Database	Information is collected by the Department on the Friday of the first full week in October every year	School/Pupil Census from 2007 onwards, internally known as "SCH". Attendance, Absence and Exclusions from 2006 onwards, internally known as "AAE". Achievement of Curriculum for Excellence from 2016 onwards, internally known as "TJ" (Teacher Judgements)	School/Pupil Census from 2007
Public transparency of third party data distribution in any online register that children and parents can view	yes, since 2013. Includes publication of volume of police and Home Office use for immigration enforcement, since Dec.2017.	no	no	no
Third party distribution	https://www.gov.uk/government/publications/dfe-external-data-shares	Unknown. "Analytical Services Unit within the Department of Education (NI) does not keep a record of every single request for data, although it is known that no data has been supplied to the Home Office or Police in this time period." [source: FOI WDTK]	http://defenddigitalme.com/wp-content/uploads/2018/03/Scotland_datasharing.pdf	http://defenddigitalme.com/wp-content/uploads/2018/03/Wales_Datasharing.pdf
General school census information	https://en.wikipedia.org/wiki/England_school_census	http://defenddigitalme.com/wp-content/uploads/2018/03/Schools-Census-Information.pdf	http://defenddigitalme.com/wp-content/uploads/2018/03/Scotland_pupil_data.pdf	http://defenddigitalme.com/wp-content/uploads/2018/03/171204-technical-completion-notes-for-local-authorities-and-schools-plasc-en.pdf
FOI reference or other sources of fact		https://www.whatdotheyknow.com/request/pupil_data_n_ireland_the_nationa_2	https://www.whatdotheyknow.com/request/pupil_data_scotland_national_pup_2	

---

[88] Comparison table of UK national pupil databases available to download as .pdf via
https://defenddigitalme.org/wp-content/uploads/2018/03/UK_pupil_data_comparison-1.pdf