

Behind the Baseline Test

We have raised concerns about the [government guidance](#) 'Changes to assessments in primary schools' on the Reception Baseline Assessment (RBA), published on May 14, 2019. In the section on Using the Data the Department for Education suggest that they will:

- *"collect the data from the assessments to create school-level progress measures for primary schools, showing the progress schools make with their pupils from reception to the end of year 6.*
- *use the data at the end of year 6 to measure pupils' progress from reception to the end of key stage 2 (KS2)."*

It also says that, *"The RBA will not be used to track individual pupils or as a performance measure for early years providers."* [our emphasis] We will hold them to that.

But our concern started with the gap between that statement, and what was published in [the Assessment Framework](#) published in February, that said:

- *"Raw scores will be recorded in the national pupil database and used to create a cohort level progress measure for schools at the end of key stage 2."* [our emphasis]

If the test results will be added to the [National Pupil Database \(NPD\)](#), then who has told schools, families and children exactly how their data will be used once collected by all of its end users and all of its purposes?

Reporting

Schools will be provided with a series of narrative statements to describe how each pupil performed on the different content domains presented in the assessment.

At the end of the assessment, the reception baseline will record a single raw score out of 45 for each pupil. This score will not be made available to schools. **Raw scores will be recorded in the national pupil database** and used to create a cohort level progress measure for schools at the end of key stage 2.

What is missing?

Will these be added to the National Pupil Database, but *not used*? Seems unlikely. No other data is. By contrast for example, the DfE says, that nationality data are not used, but [they were not put into the National Pupil Database at all](#).

The [privacy notice issued online by the NFER](#) seems very careful in its wording, but may be inadequate to explain the lawful requirements for such notices with multiple hard-to-follow links to click down through.

This demands a certain level of curiosity and the thought that you might need to look in depth, take the time and have the ability to go online, to dig down into information not on paper handouts.

If it will be used like other data in the National Pupil Database, then perhaps the notice should include:

- *'we can share these data with the following state and commercial bodies', and list each of those bodies. Purposes must be explicit.*
- *'data are kept indefinitely, and may be used even beyond your death'.*
- *'DfE will link this with other data throughout your education, and others may link it with a wide variety of other data sets about you, including commercially bought records from data brokers. It could be used for risk profiling by a wide variety of commercial and public bodies including to create profiles and flight paths by companies that teachers will act on, or by the Home Office, Police, in Court Cases.'*

The NFER has determined how the test will be run and which data are necessary to collect for it. The NFER will pass the data over to the Department for Education and also keep it for some time itself, until 2021.

So while the notice may say the DfE is the data controller of the data, we believe the NFER is a joint-data-controller. It is for a Data Controller to ensure its legal obligation of fair processing (ie telling people about what data are collected, why, who will use and about each of their applicable data rights). The NFER and guidance claims the basis for processing is public task, but neither have set out any balancing test or Right to Object.

The processing of each data item, must be *necessary*. There would be ways to reasonably perform the data processing tasks in a less intrusive way, such as through sampling, or pseudonymising data before national collection. There is certainly no necessity *for the child*, for many of the NPD reuses. This lawful basis, is therefore open to challenge.

[It is a legal challenge we are taking on.](#) And we would welcome support.

The screenshot shows a crowdfunding page on the CrowdJustice platform. The campaign title is '#LabelsLastaLifetime: DfE must respect children's data privacy rights' by defenddigitalme. The page features a central image of a book cover titled '#LabelsLastaLifetime' with the subtitle 'My family and me' and 'Drawn by Exclusion Resistor, "Sexual Gremlin" Age 5'. To the right of the image, it displays '25 days to go' and a progress bar showing '£3,866' pledged, with a goal of 'pledged of £16,000 stretch target from 22 pledges'. A 'Pledge now' button is visible at the bottom right of the campaign area.

A little bit about the National Pupil Database

How the NFER describes uses from the National Pupil Database might well be misleading by what it omits, saying, "It provides invaluable evidence on educational performance to inform independent research, as well as studies commissioned by the Department." What about the other ca. 60% of uses?

The National Pupil Database is, 'the richest education dataset in the world'.

It is longitudinal, meaning it stores a child's lifetime data from start to finish in their education: Early Years Foundation Stage age 2-4, Phonics at age 5, SATS at age 6 and 11, and every later exam are all sent for each child on a named basis, to the [National Pupil Database](#).

Since neither the NFER or the DfE has explained this adequately for any pilot and trial data yet to date, it is our opinion that these have not been processed lawfully and fairly. Parents cannot understand the potential implications of this new data collection from that privacy notice.

What we do know, is how the identifying personal confidential data of over 25 million children, and adults under 36, have been used from the NPD since 2012. [Around 60% in our 2012-16 analysis](#) was for commercial and other reuses, not academic research, and certainly not commissioned by the DfE. Uses included journalists from the Telegraph who got almost everything but the name, BBC Newsnight (Tier 1 data, the most sensitive), The Times, and the FT. They were allowed to through a 2012 change of law under Mr Gove, [without much input from parents, and no opt out](#).

These uses are not foreseeable at the point of collection, when data are collected without a child's knowledge, and for the purposes of a school.

Although thousands of data *users* know [how to apply to access](#) the database, millions of families are left completely in the dark. 69% of 1,004 parents asked in [our commissioned survey last year](#), said they had not heard of the National Pupil Database. And there is no meaningful communication [where children's data go to from there](#).

Assessing the risk of identification within data		Assessing the sensitivity of data	
DfE uses six levels of identification risk to describe data		DfE uses five different categories of sensitivity to describe data	
1	Instant identifier - These are items of personal level data that instantly identifies an individual from within the dataset. Typically, this is full names, full addresses but could also be email addresses, phone numbers or IP addresses.	A	Public commitment this data item will never be used outside of DfE - There are a few data items which the DfE have made a public commitment will only be used for internal DfE purposes. This category is used to ensure that those commitments are embedded into all data governance processes.
2	Meaningful identifier - These are identifiers that are assigned to people. An NHS Number or National Insurance Number are the examples most people are familiar with. In education, pupils have identifiers too (for example, Unique Pupil Number, Unique Learner Number, National Candidate Number). We call these meaningful identifiers because they directly identify the individual, are often known by the individual and can easily be used to link other educational data. Potentially, a user could use the meaningful identifier within our data to join it to other data they already hold and thus increase the chances of identification.	B	Contains highly sensitive data about interactions with Children's services. We collect data about the interactions some children have with children's services, such as Fostering or being Looked After or adopted. We consider this as highly sensitive.
3	Meaningless identifier - A lot of research is interested in how individual pupils progress over time. To achieve this whilst safeguarding the individual's identity, DfE makes use of identifiers that have no meaning outside of our data. These are less risky than meaningful identifiers as they can't be used to join DfE data to non-DfE data.	C	Sensitive data not captured as a special category under GDPR - The law defines areas of personal data that are particularly sensitive for individuals as 'special categories' (see category D below). Within education, we believe that there are items that citizens would treat as equally sensitive, but are not covered in GDPR legislation as they are very specific to education. (e.g. Free School Meal eligibility). We use this category to ensure such items are thought about in the same way as GDPR special category data during our decision making processes, even if legally there are differences.
4	Non-identifier that has a high re-identification risk associated with it - Within our personal level data, there are data variables that do not fall into Level One, Two or Three but can still be joined together to identify individuals. Even if the names, addresses, meaningful reference numbers have all been taken out of the data we know there is still a risk that certain items could result in an individual being identified. This is what we class as 're-identification risk'. Quantifying re-identification risk is not an exact science. We have consulted experts in the field and have found that certain combinations are more risky than others. For example, number of siblings, the school a child attends, postcode of home address increases that risk. DfE identify these combinations within the data requested and then questions whether they are essential to the user's project purpose or research.	D	Sensitive data captured as a special category under GDPR - GDPR special categories are clearly set out in law. Most relevant in the context of education data are gender, ethnicity, disability, elements of Special Educational Need (SEN) that have a health context.
5	Non-identifiers with lower identification risk - This is the level of identification risk we give to data variables that do not meet any of the above.	E	Other - Data that does not fit into any of the four categories above. This would be data items referring to exam results, or the school someone attended.
6	Anonymised Data - We use this term to describe counts of individuals that are at a level where re-identification is extremely remote. This can also refer to individual level data that has undertaken extremely high levels of stress testing to prove that re-identification risk is extremely low.		

The Data Protection Impact Assessment (DPIA)

Since the *'raw scores will be included in the National Pupil Database'*, we can only assume that they will be used exactly as other NPD data are, or may be in future, which means it may also be distributed to the same third parties who get pupil data from the rest of the National Pupil Database. By default this means additional processing of the data to link data with other records, but this is not explained in the privacy notice.

There is **nothing** in place to stop it being given away as the rest of our children's records are, and not anonymised, but as raw data from the NPD.

The DfE hasn't confirmed or denied it plans its distribution.

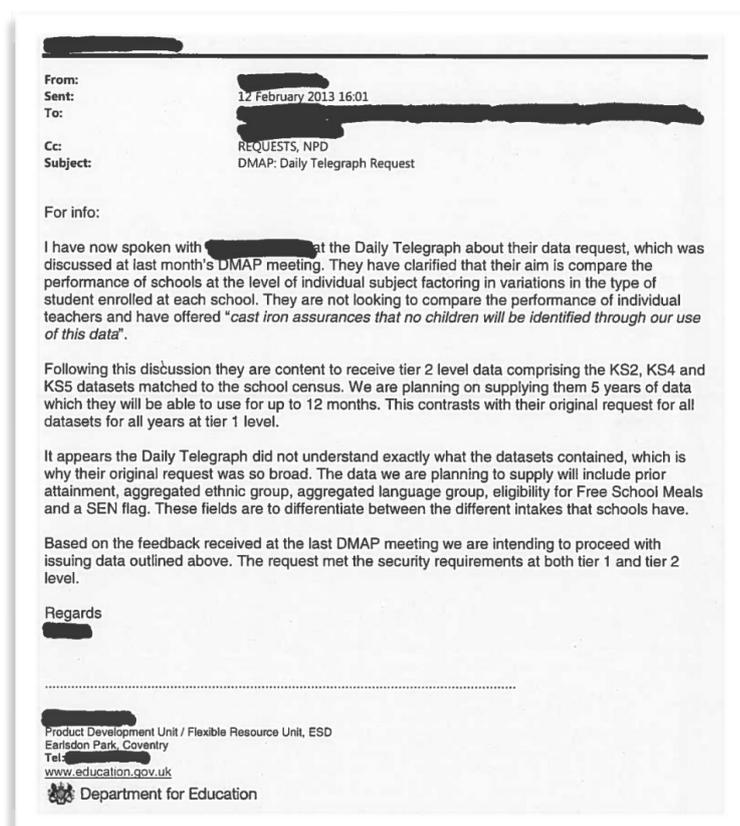
So in trying to find out, we requested [the Data Protection Impact Assessment for the Baseline Test](#) via FOI.

It was written in May 2018 before the trials of the reception baseline assessment in autumn 2018, and the Standards and Testing Authority says in the document, that they will be updating it ahead of the large-scale, national, voluntary pilot, starting in September 2019.

We are worried it is using weasel words to hide that these data will be available widely to third parties, as the rest of the National Pupil Database is today, or may be in future.

"There is no current intention of using this data for any other purpose within the DfE, and it will not be published anywhere," they say.

Not publishing the identifying pupil data, was something the ten Telegraph journalists had to agree to, before being given access to millions of pupils' records in 2013. But a user 'not publishing' the data, is not the same as DfE not releasing it, to third parties.



The DPIA also says that, "Data will be stored by the DfE in the National Pupil Database for use 7 years later when the pupils reach the end of Key Stage 2 in order to calculate the progress measure. NFER will not store the data."

need to do some quality checks to ensure that correct data is matched to pupils when it is stored in the National Pupil Database.

Storage limitation Data will be stored by the DfE in the National Pupil Database for use 7 years later when the pupils reach the end of Key Stage 2 in order to calculate the progress measure. NFER will not store the data.

Integrity and confidentiality See Annex A for the GDPR agreement reached between NFER and the DfE for the reception baseline assessment. As part of their contract, NFER are committed to

4:

But what it does not say is as important. It will be stored in the National Pupil Database but makes no mention of what that means in terms of linkage, secondary re-use, and safeguards on scope creep of future uses of the data, or how it will comply with indefinite retention requirements under GDPR, such as pseudonymisation, since it will be kept forever.

It also conflicts with NFER's own statements, and privacy notice, claiming that the NFER will not store the data, when [the privacy notice](#) says the NFER will retain the data, until Jan. 2021.

Which is true?

Summary

This risk assessment fails to measure any risks.

They conclude, "*This project does not present any significant change from the data schools are accustomed to giving for the Key Stage 1 tests – it only brings an assessment earlier to reception rather than the end of Year 2. No data is collected as part of the reception baseline assessment that should give rise to significant impact upon individuals.*"

No impact? Its use could create a very early (and potentially mis-)perception of a child's ability and failures. It will be, if added to the NPD (since the NPD is a linked set of ca. 25 databases), potentially linked with a whole range of other data increasing its own and the other data's sensitivity.

Did you know that today some Local Authorities may be using all your child's education records to rate you as a family whether you pose a risk of sexual abuse, or domestic violence? Or that National Pupil Database records about individuals have been linked to Met Police records? Do you know who has a copy of your own or your child's national record, where it has gone, or how accurate it is? [You have [the right to find out for free if you want to](#). We can help with questions.]

In 2002 politicians assured the public that collecting names for the National Pupil Database posed [no risk to privacy](#). Ten years later, the Home Office first trialled using NPD data, and then started using it monthly in 2015, to match their records [for immigration enforcement](#). Children have been removed from school as a

result, and misuse of school census data collected for the purposes of education, damages both professional and public trust.

No risks from adding new data to the NPD? Hell yes, it poses risks.

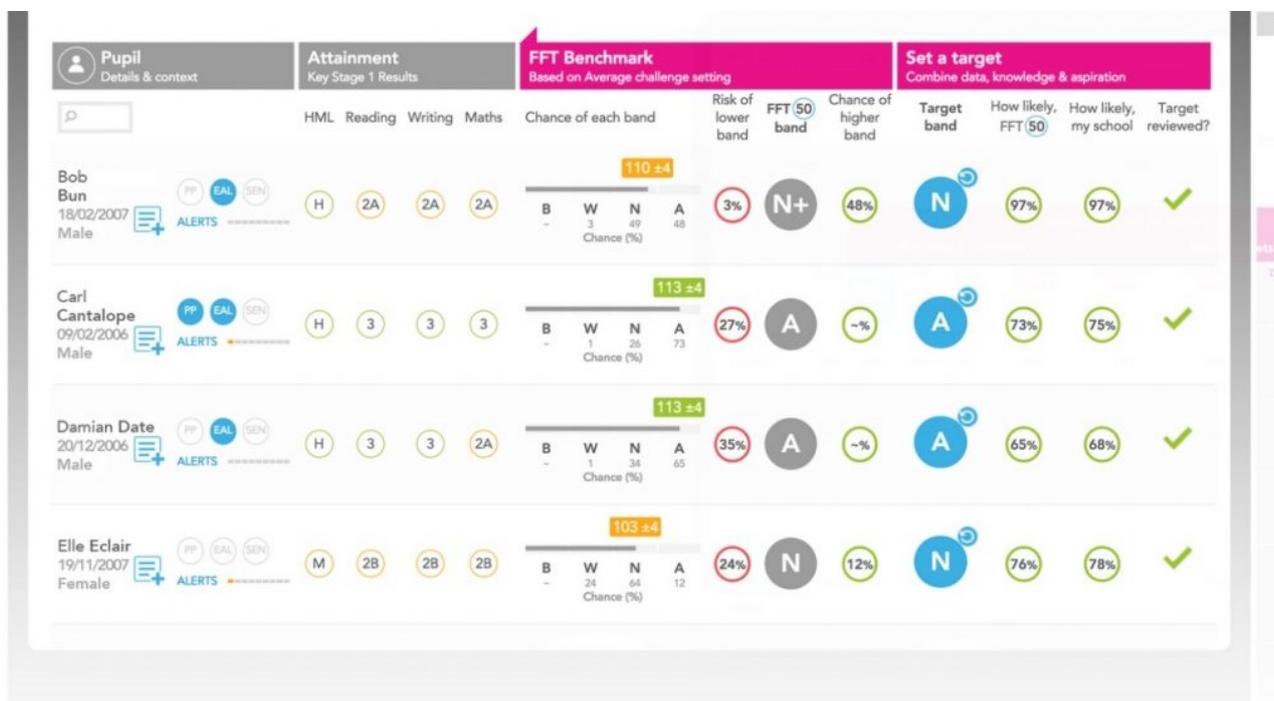
While the DfE is making significant changes to its handling processes at the time of writing, it currently still gives away raw, identifying and sensitive data about millions of individual children, to a wide range of third parties. As far as we know, there is no timetable to end this distribution.

In 2018, the Department for Education made important and expected progress on its data management improvements, with [support of the ONS](#).

DfE reduced the risks of *distributing* data by cutting down how often they give it away raw, and instead they started to grant access to the data for researchers to come to use the data in safe settings, rather than give it away.

In 2019, it seems to have slipped back to below the expected standard.

They are still [sending out raw data to far too many users](#). Including millions of records to processors scoring children, such as Fischer Family Trust, who then resell those processed data back to schools.



In fact, the DfE is in the middle of passing [a new law](#) to enable every child's full pupil records, not only to be given away to the Office for Students, but through them, potentially to Pearson Ltd, to the Student Loans Company, to the Tax office, and HMRC, among a named list of 13 organisations. [[The House of Lords](#) will debate those new Regulations on June 5th.] Forever.

No risk? [That Baseline Test Data Protection Impact Assessment](#) is either uninformed about the facts how pupil data are used, or it is deliberately misleading. Neither is enough.

There are conflicting statements on data retention being made.

There are statements which could be misleading due to telling only part of the truth.

Families and children need to know how their pupil records are used by companies, charities, the press, think tanks, and academic researchers. And how they [might be used for the rest of the child's lifetime](#).

We must be told if personal data from Baseline Tests will be given to businesses like other data in the National Pupil Database are -- because what parents are not being told right now, smacks of cover up.

Our concerns are compounded by the fact this is not the only expansion going on.

An equally inadequate risk assessment of [the Multiplication Tables Check](#) leaves out completely, that it will collect any personal data at all. But reasons for *not taking* the test, such as, '*just arrived EAL*' will be included alongside name and more.

Remember that the Home Office still gets monthly handovers of information from children's records entrusted by schools to the DfE, and misuses it for [immigration enforcement](#) and further purposes of [the Hostile Environment](#). How might 'just arrived' be a useful indicator, after [collective action](#) beat [the collection of nationality data](#). Think about it DfE. No way.

- The data risk assessment standards of these new tests are weak.
- There is still no sign of the long overdue *Data Protection Impact Assessment* of the National Pupil Database.
- The internal DfE group that decides on data expansions '[the Star Chamber Scrutiny Board](#)' publishes an 'annual report' in name only, devoid of meaningful content (2016-17 didn't mention the disaster of the school census nationality data expansion.)
- The [internal DfE group that decides on data releases](#) to third parties publishes no minutes or reports at all.
- Transparency, safeguards and oversight of current and proposed data handling are inadequate for all data processing, and are poor for the MTC and Baseline.

The government has added new tests, new items to the school census, new data collections and snuck through database changes on its use, year on year since 1996. The [collective impact](#) on our children's and family life is enormous, and much of it is hidden, or remains yet to be seen in future.

Schools and parents have rejected Baseline on principle. It is high time to reject the Department for Education's ongoing flawed data practices too.

Can I stop my child's personal data being used?

NFER handles personal data in accordance with the rights given to individuals under data protection legislation.

Under certain circumstances, you have the right:

- to ask us for access to information about you that we hold (Subject Access Request – SAR)
- to have your personal data rectified, if it is inaccurate or incomplete
- to request the deletion or removal of personal data where there is no compelling reason for its continued processing
- to restrict our processing of your personal data (i.e. permitting its storage but no further processing)
- to object to direct marketing (including profiling) and processing for the purposes of scientific/historical research and statistics
- not to be subject to decisions based purely on automated processing where it produces a legal or similarly significant effect on you

How long will personal data be retained?

We will only keep your child's personal data for as long as we need it for the purpose of this project, after which point it will be securely destroyed.

All pupil data relating to the pilot phase of reception baseline will be retained by NFER until January 2021.

The DfE will hold the data in the [National Pupil Database](#) (NPD) where it will be retained according to the DfE's data policy. The NPD is owned and managed by the DfE and contains information about pupils in schools in England. It provides invaluable evidence on educational performance to inform independent research, as well as studies commissioned by the Department. More information on how this data is shared can be found [here](#).

Please note that, under Article 5 (1)(e) GDPR 2018 and in compliance with the relevant conditions, we can lawfully keep personal data processed purely for research purposes indefinitely.

Who will the personal data be shared with?

NFER may need to make personal data available to other organisations. This includes Amazon Web Services as they are hosting our online system, BeP.

Data will also be shared with STA to analyse the effectiveness of the reception baseline assessment.

Where we need to share your child's personal data with others, we will ensure this complies with data protection legislation.

More background

- All state-funded primary schools with a reception cohort (year group) will need to carry out the reception baseline assessment (RBA).
- The RBA is scheduled for introduction in September 2020.
- Over 9,600 schools are participating in the reception baseline assessment (RBA) pilot in autumn 2019.
- Schools will no longer have to carry out key stage 1 (KS1) assessments from September 2022, following the scheduled statutory introduction of the RBA.

Selected key points

- The lawful basis for the Baseline Test will be made statutory for March 2020 through amending the [Childcare Act 2006](#). There was no change of law so far, to enable the new data collection.
- The reception baseline assessment will collect information on all reception age pupils in England.
 - full name
 - date of birth
 - date of starting school
 - gender
 - Unique Pupil Number (13 digit code across all child's education)
 - results from the assessment / time and day it was taken
- It admits that timing may be problematic. A child starts school in ca. the first week of September. The first school census which creates the core National Pupil Database record for some children, is not until later, in the first week of October. The Baseline will therefore be collected *before* the child's core record may exist in the National Pupil Database. That means there may be nothing to link it *with* for pupils that only start school at four or five without previous state education. Pre-schoolers, will already have a core record with a unique pupil number, and their Early Years Foundation Stage scores. As the DfE says, "*As the assessment occurs before the census data is collected by the DfE, then the DfE will need to do some quality checks to ensure that correct data is matched to pupils when it is stored in the National Pupil Database.*" [*Those data items are set out in this guide, [pages 127-133](#).*]
- The statement on the Early Years Foundation State Profile seem to suggest it has quality issues, because, "*the existing assessment in reception – the early year foundation stage profile (EYFSP) – does not provide for a sufficiently reliable baseline in its current form. This is because it is an observational assessment made by teachers over the whole of the reception year, whereas we require the baseline to be a task-based ‘snapshot’ in time. In addition, whilst the EYFSP assesses early literacy and numeracy, its scope is expansive and *does not correlate sufficiently with outcomes at the end of key stage 2*. Therefore, the only way to achieve the policy intent and realise the benefits of that policy is to create a new assessment.*" One wonders why EYFS data are collected at national level at all.
- By contrast, the NFER and DfE have not stated how the new Baseline **will** correlate with the outcomes at the end of key stage 2. **Nothing we can see, sets out how the new progress measure will be a measure of progress at all.** Ostensibly its purpose. [*Despite saying the test is needed for*

[accountability measures on the whole of page 8.](#)]

- It's interesting to note that the document suggests that while the National Foundation for Educational Research will develop and deliver the reception baseline assessment until 2022, that will change, because, "*a longer-term, subsequent solution will then be adopted, with the current intention being to bring design and delivery of the assessment in-house to DfE.*"

References

- [Baseline: Data Protection Impact Assessment and Consultation](#)
- The Higher Education and Research Act 2017 (Further Implementation etc.) Regulations 2019 (000/2019) [[links](#)]
- The National Pupil Database (NPD)
 - Wikipedia [[link](#)]
 - Our NPD Backgrounder [[link](#)]
 - Full briefing [[download .pdf 3.2 mB](#)] updated November 20, 2017

Related posts

- [SATs are not just like going for a health check, Minister.](#) April 28, 2019
- [SATs and scores that Last a Lifetime](#), March 2, 2019
- [The Multiplication Tables Check \(MTC\)](#) May 10, 2019
- [Multiplication tables check admin guidance missing vital information](#) April 3, 2019
- DfE on the MTC check [[link](#)]

More on The National Pupil Database

The [National Pupil Database](#) grows by around 700,000 records each year, and having begun in 1996 it currently links together the personal confidential data from over 23 million people. Its core is created from the collection of children's personal details in the [school census](#) three times a year. Which data items, around 400 eventually on any one child, are listed on [pages 127-133 of the DfE guidance](#).

Parents and pupils are not told that after children's national tests, personal confidential data are shared with a wide range of third-party organisations, and in over 1,000 releases of data to third parties since April 2012.

'External organisation data shares' includes data distribution to:

- DfE external organisations supported by memorandum of understandings, data sharing agreements and service level agreements
- data shares with Home Office
- with police
- and through court orders

[The DfE published National Pupil Database Third Party Requests](#) list only the currently active users, -- those who have asked for data and the request had not yet been completed, -- or to whom identifying and sensitive data were given and their agreement has not yet expired.

Previous versions of the data showing releases to a wider range of third parties including newspaper and TV journalists, can be found archived:

- [September 2017 to April 2018](#)
- [January 2017 to May 2017](#)
- [April 2012 to June 2016](#)