Department for Education

National Pupil Database Data Protection Impact Assessment: Public Summary

Version 1.0

May 2019

1. The purpose and aims of the National Pupil Database (NPD).	
The NPD holds information including characteristics, attainment and social service interactions of pupils educated in England. The first version of the NPD was produced in 2002, and has grown to be DfE's primary data resource about pupils and one of the richest data resources about education in the world. NPD datasets are used externally for research purposes where legislation, security and public good criteria are met which enables rigorous independent evaluation and policy scrutiny to occur. DfE is committed to these principles, and to modernising processes, to ensure we can continue to deliver on them in future. The evidence and data obtained via the processing in NPD provide the Department, education providers, Parliament and the wider public with a clear picture of how the education system is working in order to better target (and evaluate) policy interventions to help meet the Department's strategic objectives and ensure all children are kept safe from harm and receive the best possible education.	
2. Details of the role of any data processor, suppliers or contractors if used.	A contracted data processor is used to collate, and quality assure, elements of source data (mainly attainment data) alongside application of defined matching algorithms used for the purpose of allocating a Pupil ID / Pupil Matching Reference to all data loaded into NPD. This matching is necessary to enable the NPD to effectively function as longitudinal research source. The Office for National Statistics (ONS), under a data sharing agreement with DfE, make DfE data available to external researchers within their Secure Research Service (SRS) for requests approved by the DfE Data Sharing

3. Details of any data transfer to or from any organisation, and or data sharing.	Data inputs into the NPD come from a variety of sources including statutory data collections, data feeds from awarding bodies, reference tables, etc. The data comprises of a combination of information which the data provider will already be required to hold for the purpose of delivering their day to day business alongside information collected, via the data provider, directly from the data subject or their parent / representative. The collected data is transferred from the relevant data collection system to a contracted data processor to fulfil their contractual obligations (see question 2 above). The data is transferred via the Department's secure file transport system (referred to as Egress) applying 256-bit AES encryption to the data and using Secure Sockets Layer (SSL) to secure the internet link. The processed data is returned by the contracted data processor to DfE for the purpose of populating the NPD via the Maytech secure transfer system which provides end-to-end data encryption and has UK pan-government accreditation for the transfer of official sensitive data. Maytech, as with Egress, is compliant with the requirements of the GDPR. Data is only shared where it is lawful, secure and ethical to do so. All sharing of individual data from the NPD is robustly governed by the DfE Data Sharing Approval Panel. Below is a link to organisations with which NPD data has been shared: https://www.gov.uk/government/publications/dfe-external-data-shares Data shared via the ONS SRS is transferred from DfE to ONS via Egress (referred to above). Upon receipt ONS move the data into a cloud based secure environment and create user profiles for each researcher that has been sanctioned to access the data by DSAP. The researchers then have the period of the license agreement to arrange access within the ONS data labs (either within 1 of the 5 physical labs across the UK or via remote access where organisations have met ONS specified security standards and have been approved for connectively to remotely access

4. Is this a change to an existing or previous DfE initiative?	No. The processing of data within the NPD is not a change to DfE policy although strengthened governance arrangements and facilitating external access to DfE data via the ONS SRS, reducing reliance on physical transfers of data, were introduced in 2018.
5. Which of DfE's functions does this initiative support?	All. Namely: ▪ teaching and learning for children in the early years and in primary schools ▪ teaching and learning for young people in secondary schools ▪ teaching, learning and training for young people and adults in apprenticeships, traineeships and further education ▪ teaching and learning for young people and adults in higher education ▪ supporting professionals who work with children, young people and adult learners ▪ helping disadvantaged children and young people to achieve more ▪ making sure that local services protect and support children In addition, the NPD provide the Department, education providers, Parliament and the wider public with supporting research and monitoring of educational policy.

About the scale and nature of processing

6. How many individuals will be having their personal data	Approximately 21m individuals (as at October 2018).

7. What categories of personal data will be processed?	▪ instant identifiers (e.g. name, address) ▪ meaningful identifiers (e.g. unique pupil numbers, unique learner numbers, candidate numbers) ▪ meaningless identifiers (e.g. pupil matching reference numbers) ▪ characteristics (e.g. ethnicity, FSM eligibility) ▪ special educational needs (e.g. SEN provision, type of need) ▪ placement / enrolment details (e.g. establishment, start date, attendance pattern, reason for placement) ▪ absence and exclusions ▪ attainment (e.g. EYFSP, KS1, KS2, KS4, KS5) ▪ children's services interactions (children in need, looked after children) ▪ destinations (YPMAD) Full details of all data available in the NPD are
8. What is the lawful basis for processing the personal data?	**GDPR Article 6(1)(e) Public task**; the processing is necessary to perform a task in the public interest or for DfE's official functions, and the task or function has a clear basis in law.
9. What personal data of a sensitive or highly personal nature will be processed? (special category)	Different categories of special category data are collected for different cohorts of individuals depending on the nature of their interaction with an educational establishment or local authority. For example: ▪ Ethnicity ▪ Reason for placement in AP – children in alternative provision only ▪ Special educational needs (type or category of need) ▪ Learner Learning Difficulty or Disability – FE and HE only ▪ religion or belief – HE only (collected by HESA) ▪ sexual orientation – HE only (collected by HESA) ▪ disabled student allowance – HE only (collected by HESA)
10. What is the lawful basis for processing special category personal data?	**GDPR Article 9(2)(j);** processing is necessary for archiving in the public interest, scientific, or historical research or statistical purposes.
11. Is the personal data of vulnerable individuals processed?	Yes, vulnerable individuals who are currently, or have previously, engaged in the English education or children's social care system.

12. Is new, innovative or unusual technology used to process personal data, including cookies or similar technologies in collecting or other processing of data?	No
13. Does processing personal data involve automated decision-	No
14. Does processing personal data involve profiling individuals?	No

Data Protection Summary
The minimum amount of personal data necessary to achieve the objectives is
The lawful basis for processing personal data and Special Category data, has been identified and recorded.
The processes are in place to extract all the personal data relating to a single individual on request to respond to a Subject Access Request.
The processes are in place to address all individual rights of the data subject.
There is a data retention period for this dataset. Data is collected, and processed, for research and statistical purposes in the public interest to promote the education and well-being of children in England. Data is retained for an indefinite period whilst ongoing reviews determine the data remains necessary for the purposes for
The processing of personal data by other organisations has been documented in a data sharing agreement, contract or similar document to ensure all data protection considerations have been addressed by all parties.
Internal security measure are in place to ensure only those people who have a need to access the personal data can do so. Personal data is processed securely using appropriate technical and organisational measures in line with the “security
This DPIA will be reviewed in six months (30th November 2019) and then annually.

Data Protection Risk Summary

These risks summarise the top risks identified at the time of publication of this DPIA summary. (Publication 31st May 2019)

This DPIA and risk summary will be reviewed by 30th November 2019.

Risk	Impact	Mitigations
1. The department is prevented from collecting personal information as we are not as transparent as we could be.	Individuals may be reluctant to share their personal information with the department as they do not understand what their information may be used for. This may impact on the department's ability to develop and implement new initiatives to achieve its objectives.	Annual review of all privacy notices. Gap analysis of existing privacy notices Privacy notice requirements built into DPIA process. Data Governance improvements implements improved controls and support transparency.
2. The scope of use of personal information is not restricted due to insufficient governance controls.	Personal information may be used for a purpose the individual is unaware of. Local Authorities, schools and others are unaware of this additional use and are vulnerable as a result. Individuals and organisations may be reluctant to share their personal	Improved Data Governance within the department including the DSAP, enhanced DPIA process and governance escalation to the Data Protection Board and the DPO. Data mapping will identify all data sets held and identify the lawful basis for processing enabling
3. Personal information will be held for longer than is necessary.	Personal information will not be deleted in line with retention schedules and policies, in potential breach of data protection legislation. The data subject will be unaware that we are still processing their personal information.	Improved Data Governance including DSAP and the Retention Review Board will implement better controls. The DPIA process requires data retention policies and schedules to be implemented.
4. The department may inadvertently breach the data	The department may inadvertently breach the data protection principles as a whole due to work processes that are still maturing.	Planned data protection awareness training for all staff on their responsibilities. Planned data governance

Risk	Impact	Mitigations
5. There is a risk that the data within [name of system] may be targeted and extracted by external organisations who wish to exploit it for financial or security gains.	A data protection breach may occur with consequences for the department and data subjects.	The system is protected with cyber security controls which are appropriate for the OFFICIAL data as outlined by the Cabinet Office and the National Cyber Security Centre The system is regularly pen tested by certified testers to ensure that any weaknesses in the system are identified quickly and rectified The network and servers which host the system are regularly patched The system has separate roles