

Archiving evaluation data analysed using the ONS Secure Research Service



November 2020

1. Updates

November 2020:

- Minor edits to 3.7 and data linkage chart to make it clearer that the DfE deletes personal data submitted by the evaluators and doesn't send it to the SRS.

October 2020:

This is a revised version of the document dated 1st May 2020 and includes the following updates:

- The process for sharing data with the DfE for NPD matching¹ has been revised to reflect recent process updates. Evaluators will now send only a limited number of variables to the DfE (pupil identifiers) and they will upload the rest of the evaluation data directly to their SRS project space.
- Appendix 1 (EEF-DfE-ONS data journey) has been updated slightly to reflect this change.
- New chart added summarising key data linkage processes.
- New section added on archiving projects that do not use the NPD.
- Data specification for archiving in the SRS included in Appendix 3.
- Minor clarifications added throughout.

2. Background

The Education Endowment Foundation (EEF) funds high-potential and promising interventions and, in parallel, independent evaluations of their impact. Upon project completion, relevant datasets are submitted to the EEF data archive, managed by [FFT Education](#),² for quality assurance, longitudinal analysis and analytical developments.

Most EEF-funded evaluations require linking data collected during the project to pupil-level variables from the [National Pupil Database](#) (NPD), held by the Department for Education (DfE). In the past, the DfE provided NPD data extracts to researchers for evaluation purposes. At the end of an EEF-funded project, the evaluation team would submit the NPD-linked

¹ This document uses data 'matching' and 'linking' interchangeably.

² Links active at the time of publication.

evaluation data directly to FFT. FFT would check data protection safeguards, validate the data and add it to the EEF archive.

The process for granting access to NPD data to external research teams has recently changed. This can now only be accessed through the [Secure Research Service](#) (SRS) offered by the Office for National Statistics (ONS). It has also been agreed that the EEF data archive will be transferred to the ONS, where it will continue to be managed by FFT within the SRS. In addition to facilitating easy transfer of NPD-linked datasets from the project analysis space to the archive space within the SRS, this will significantly increase the security of the data, therefore better safeguarding the rights of the individual data subjects.

DATA PROTECTION

In order to provide maximum value to taxpayers, school leaders, policy makers and the wider education and research communities, we archive impact evaluation data with a view to tracking the long-term effect of the interventions we fund. In addition, we are committed to supporting open science and methodological developments by facilitating lawful access to our data archive for secondary analysis. These purposes, and the data sharing, linking and matching required for them, must be made clear to participants at recruitment stage. Clear opportunities to withdraw from such data processing in case of objections must also be provided.

External evaluation teams are data controllers during the evaluation, retaining all relevant responsibilities until the project has been completed, the data has been archived with EEF's archive manager, having passed quality and data protection checks, and the evaluator has deleted the data from their servers at the end of their retention period. The EEF becomes data controller for the data once it has been archived and has passed the required quality and data protection checks.

To ensure data submitted to the archive is processed in [accountable ways](#), we ask that evaluators apply strict data protection safeguards from the very start of the evaluation. This includes ensuring that evaluation participants (data subjects) and/ or their parents, carers or legal guardians are sent information letters and/ or privacy notices that are fully compliant with the [General Data Protection Regulation](#) and [Data Protection Act 2018](#). At the end of the project, these documents must be included in the evaluation report.

When confirming that the data is ready for archiving (see Appendix 2, *EEF project completion and archiving form*), the evaluator must also confirm data protection safeguarding during the evaluation.

It is therefore essential that the data protection section of the form in Appendix 2 is considered when producing the project recruitment documents.

Please note only pupil-level impact evaluation data will be archived from now on. This includes primary outcomes, secondary outcomes and any other data used in statistical analysis (e.g., compliance, missingness, sensitivity, moderation, mediation analyses). Unless a specific case is made otherwise, we will not archive non-pupil data, data from pilots and scale-up evaluations, or implementation and process evaluation data that is not used in statistical analysis for impact evaluation. Evaluators are encouraged to archive these with their own institutions or public repositories for potential future use.

The updated process for linking, analysing and archiving data from EEF evaluations is described further down and represented schematically in Appendix 1.

Please note this overview is provisional and may change slightly as the new process is tested and implemented.

3. Linking evaluation data to NPD variables

1. EEF approves project and selects external evaluator through competitive bidding.
2. Evaluator collects data from schools.
3. Evaluator submits NPD application to DfE, listing all of the following in the application form:
 - a. all the variables needed from the NPD (these must always include the Pupil Matching Reference, PMR, without which longitudinal analysis would not be possible – see below), plus
 - b. pupil identifiers collected by the evaluator from schools that will be submitted to the DfE for matching, along with the evaluator's own meaningless pupil identifiers (without which it is not possible to complete further matching in the SRS).
 - c. all the evaluation variables the evaluator intends to bring directly into the SRS for matching (e.g., bespoke test results, survey data, any publicly available data).
4. DfE approves NPD application and informs the ONS; ONS creates SRS project space.
5. Evaluator signs data sharing agreement with the DfE and submits evidence to the ONS of right to share data (e.g., privacy notice, controller-processor agreement).
6. Upon DfE instruction, evaluator sends pupil identifiers securely to the DfE for matching. These will include the evaluator's own meaningless identifiers and pupil personal data collected from schools, such as pupil name, surname, date of birth, home postcode, UPN, school identifier (e.g., URN, LAESTAB). The DfE uses this personal data to identify the pupils in order to extract the NPD variables requested.
7. DfE extracts the NPD variables, matches them to the PMR and the evaluator's own meaningless identifier and transfers the matched dataset to the SRS. The data sent to

the SRS by the DfE does not include the personal identifiers submitted by the evaluator, which the DfE discards after matching.

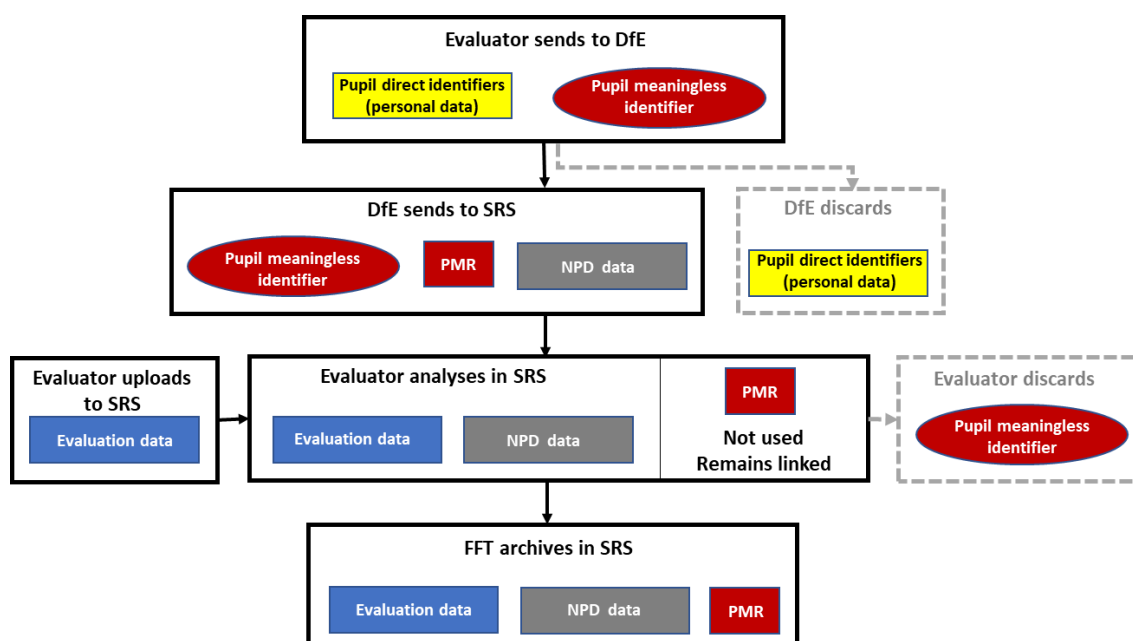
PUPIL MATCHING REFERENCE (PMR)

The PMR (NPD variable name: PupilMatchingRefAnonymous) is a meaningless identifier that will enable potential further linking by the DfE. This variable is meaningless outside the DfE and only the DfE, as data owner, can use it to link pupil data to further variables in the future.

With the exception of the PMR and the evaluator's own meaningless pupil identifier, no data can be transferred to the SRS that can be used to identify pupils. The evaluator's pupil identifier must be meaningless outside the evaluation team and only allow the evaluation team to link pupil data. Once the evaluator has matched the linked NPD dataset to the additional evaluation data they have brought into the SRS, the evaluator discards the meaningless pupil identifiers. The resulting dataset cannot, on its own, be used to identify any individual pupil in the SRS.

In order to estimate the long-term impact of our interventions, it is essential that all datasets being prepared for archiving in the SRS contain the DfE PMR, as this will be the only variable that can facilitate further linking or matching after project completion.

Key data sharing and matching are summarised in the chart below.



4. Analysing NPD-linked evaluation data

8. [ONS-accredited](#) evaluator accesses the linked data in the SRS, either remotely through [Assured Organisational Connectivity](#) or in [ONS safe rooms](#). (The report will be prepared for publication following the typical review and revision stages.)
9. Evaluator uploads further evaluation data needed for analysis to the SRS. The evaluator links the additional data to the NPD-linked dataset using their own meaningless pupil identifiers within the SRS. The final linked dataset must include the PMR and all evaluation data needed for analysis in the SRS, which will eventually be added to the EEF archive within the SRS by FFT (archive manager). No separate archive submission will therefore be necessary at the end of the project.
10. Evaluator analyses the data and submits outputs to the SRS team for clearance.
11. SRS team clear pre-publication and final publication outputs.
12. Evaluator discards own meaningless pupil identifiers.
13. Evaluator saves the final evaluation report in the SRS project space for archiving. (This is necessary to enable the archive manager to check the archived data against the report within the SRS.)
14. Evaluator submits final evaluation report to the EEF and prepares the data for archiving within the SRS.

5. Archiving evaluation data

15. Evaluator prepares data files and analysis syntax for archiving using the FFT SRS data specification provided (see Appendix 3).
16. Evaluator signals analysis completion and readiness for archiving by emailing the EEF, FFT and ONS, using the form and contact details provided in Appendix 2.
17. SRS team transfer the dataset from the project space to the FFT space within the SRS.
18. FFT checks the dataset and adds it to the archive.
19. Archive is refreshed with new NPD data regularly for quality assurance, longitudinal and developmental work completed by the overarching evaluator based at Durham University.

6. Accessing archived EEF data

Once the process above has been implemented and finalised, the archive will be open to the research community for analysis within the SRS. It is likely this will follow the typical process for accessing NPD data by applying to the DfE, with additional EEF involvement at approval stage. This approach is currently being piloted and an update will be provided in due course.

7. Archiving for projects that do not use NPD data

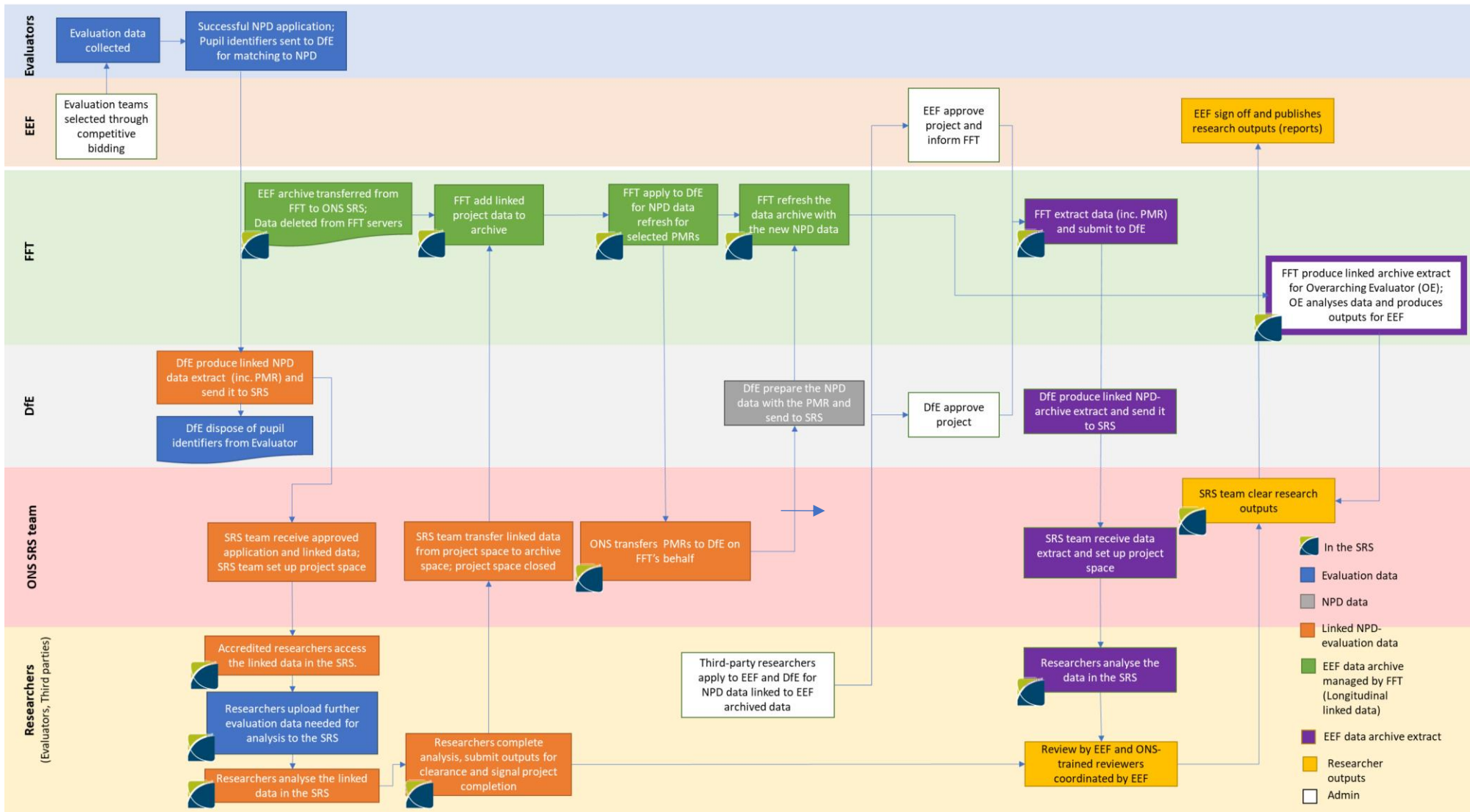
Any impact evaluations that do not use NPD data at all will continue to be archived via FFT. Along with the evaluation data, evaluators will need to submit personal identifiers to FFT, such as pupil name, surname, date of birth, home postcode, UPN, school identifier (e.g., URN, LAESTAB). At regular intervals, FFT will use these identifiers to request PMRs from the DfE, who will transfer them directly to the SRS. FFT will upload the remaining evaluation data to the SRS and link the PMRs. The PMR-linked evaluation data will then be added to the archive within the SRS.

It is important that evaluators are aware of this requirement from the start of the project, so data sharing can be adequately described in privacy notices.

This process will be tested and confirmed in due course.

Appendix 1

EEF-DfE-ONS data journey



Appendix 2

EEF project completion and archiving form

Please fill in this form on finalising your SRS analysis and email it to all of the following (Cc):

- the EEF evaluation manager responsible for your project
- FFT Education: laura.james@fft.org.uk; andrew.bibby@fft.org.uk
- ONS: research.support@ons.gov.uk

Project details	
DR number (from the NPD application)	
Project title	
Organisation (evaluator)	
Principal investigator (name and email address)	
Main contact for SRS analysis (name and email address)	
NPD licence expiry date	
SRS project space expiry date	
Date final evaluation report (will be) submitted to EEF	
Is the final evaluation report in the SRS project space?	YES/ NO
Files prepared for archiving	
Have you prepared your full dataset for archiving following the FFT SRS data specification?	YES/ NO
Does the dataset include the Pupil Matching Reference (PMR)?	YES/ NO
Have you prepared your full analysis syntax for archiving following the FFT SRS data specification?	YES/ NO
Names of data files to be archived	(Please list all relevant file names.)
Names of syntax/ do-files to be archived	(Please list all relevant file names.)
Names of any other files to be archived (if applicable)	(Please list all relevant file names.)

Data protection			
Your lawful basis for processing personal data			
Your lawful basis for processing special categories of personal data			
Were the data subjects and/ or their parents, carers or legal guardians given the opportunity to withdraw from data processing?		YES/ NO	
Is the withdrawal form, or other evidence for the above, included in the evaluation report?		YES/ NO (If NO, please specify where the evidence can be found.)	
Have you removed all the data for which withdrawal forms were received, before preparing your dataset for archiving?		YES/ NO	
Were the data subjects and/ or their parents, carers or legal guardians told that the data would be anonymised during or after the project?		YES/ NO (We do not expect this, but sometimes it may be deemed necessary. Please state if so.)	
Please confirm whether the data subjects and/ or their parents, carers or legal guardians were informed of the following data processing purposes, and whether evidence of this is included in the final evaluation report (e.g., information letter, privacy notice). If evidence is not included, please state where it can be found.			
Data will be:	Data subjects or parents informed	Evidence included in report	Notes
collected for evaluation	YES/ NO	YES/ NO	
linked to NPD for main evaluation	YES/ NO	YES/ NO	
shared with specific parties during main evaluation	YES/ NO	YES/ NO	
archived	YES/ NO	YES/ NO	
potentially shared with other parties after archiving	YES/ NO	YES/ NO	
potentially relinked to NPD after archiving	YES/ NO	YES/ NO	
potentially linked to other datasets after archiving	YES/ NO	YES/ NO	
Other information			
Is there any other information that would be useful for us to know about the dataset?			

Declaration

I confirm that the information above is accurate and the evaluation data is ready to be archived.

Name and project role:

Signature:

Email address:

Date:

Appendix 3

Data specification for archiving in the SRS

Updates

The specification has been updated to reflect the processing of evaluation data within the ONS SRS. It is anticipated that project data will have been imported into the SRS to be linked with the NPD for analysis and the names of pupils within the evaluation will have been replaced by the Pupil Matching Reference (PMR). As a result, pupil identifying fields Pupil_UPN, Pupil_Forenames, Pupil_Surname, and Pupil_DOB) are no longer required for archiving. Pupil_PMR is now mandatory.

The intention of the specification is to support the submission and potential reuse of all relevant items.

Background

Data collection for the EEF archive is comprised of three objects:

1. A form to record information about the project that needs to be stored with the data (see Appendix 2).
2. A data table (see below) to record information about individual pupils enrolled in the project, the treatment they receive and the outcomes of testing. The intention is to have one row per pupil and to standardise core data items to enable meta-analysis.
3. A metadata table (see below) whose purpose is to describe the data items in the data table. This table should be populated with relevant coding frames and descriptions to ensure the collection of a rich and meaningful dataset.

The three objects are designed to collect all data items that are relevant to the study. Each field in the data table should have corresponding entries in the metadata table listing the coding frames and descriptions, including missing value codes. In the data table specification there are a few fields flagged as mandatory to provide a sense of the minimum components expected from the project. However, all data items which are not temporary or placeholder variables should be carefully considered for submission regardless of the mandatory flag.

All variables included in the main analytical model should be submitted to enable secondary analyses.

The data submission must be in one of the following formats:

- SPSS .sav file (metadata table should be recorded in the Variable View); or
- STATA .dta file (metadata should be recorded for each variable); or
- Excel workbook (.xlsx or .xls, two worksheets – one for each table); or
- Two tab-delimited text files (.dat with field names in the header row).

We recommend that evaluators submit either a SPSS or STATA file, since these packages have functionality to record metadata without the need for a separate table.

Data table specification

Data table		
Field name	Mandatory	Description
School_Stratum_ID	Y	The stratum (block) of the school in your sample design (if stratification was used). Must be defined in metadata.
Pairing_ID	N	Identifier of pair or triplet if paired or matched randomisation is used
School_ID	Y	URN (6 digit Unique Reference Number) or LAESTAB (7 digit Local Authority and Establishment number). Can be retrieved from https://get-information-schools.service.gov.uk/
Class_ID	Y	Class identifier used for possible grouping or clustering for analysis (mandatory if a three-level hierarchical model was used)
Teacher_ID	N	Teacher identifier used for possible grouping or clustering for analysis.
Pupil_PMR	Y	NPD (National Pupil Database) Pupil Matching Reference.
Pupil_Month_Of_Birth	Y	Pupil's month of birth. Codes must be described in Metadata table.
Pupil_Year_Of_Birth	Y	Pupil's year of birth
Pupil_Gender	N	Pupil's gender. Used for matching. Codes must be described in Metadata table.
Pupil_Census_Series	N	Series of the census from which pupil background data has been taken, e.g. SPR12 means Spring 2011/12. NPD data items are suffixed with this code.
Pupil_Ethnicity	N	Pupil's ethnic code (EthnicGroupMinor) as recorded in school census
Pupil_FSM	N	FSM eligibility (FSMEligible) at the beginning of the intervention
Pupil_FSM6	N	Whether the pupil has ever been recorded as eligible for free school meals (EverFSM_6) in any termly Census in the six years prior to intervention
Pupil_SEN_Provision	N	Provision types under the SEN Code of Practice as recorded in School Census (SENProvision)
Pupil_EAL	N	Pupil whose first language is other than English as recorded in School Census (LanguageGroupMajor)
Treatment_Allocation	Y	Treatment group for the pupil. That is, the <i>intended</i> treatment. Must be described in Metadata table.
Treatment_Dose	Y	Treatment <i>actually</i> received by the pupil – must be described in Metadata table.
Treatment_Start_Date	N	Date of first intervention or treatment.
Treatment_End_Date	N	Date of last of intervention or treatment.
PreTest_Description	N	Description of baseline test undertaken prior to intervention.
PreTest_Date	N	The date the pupil took the baseline test.
PreTest_PupilYeargroup	N	The pupil's National Curriculum year group on the day of baseline test.

PreTest_Outcome_1	Y	The value of the primary test outcome. If this is a code rather than a numeric value then the code frame should be included in the metadata table.
PreTest_Outcome_2	N	The value of the second test outcome
PreTest_Outcome_3	N	The value of the third test outcome.
PreTest_Outcome_4	N	The value of the fourth test outcome.
PreTest_Outcome_5	N	The value of the fifth test outcome.
PostTest_Description	Y	Description of the test undertaken after the intervention.
PostTest_Date	Y	The date the pupil took the test.
PostTest_PupilYeargroup	N	The pupil's National Curriculum year group on the day of test.
PostTest_Outcome_1	Y	The value of the primary test outcome. If this is a code rather than a numeric value then the code frame should be included in the metadata table
PostTest_Outcome_2	N	The value of the second test outcome.
PostTest_Outcome_3	N	The value of the third test outcome.
PostTest_Outcome_4	N	The value of the fourth test outcome.
PostTest_Outcome_5	N	The value of the fifth test outcome.
In_Primary_Analysis	Y	Whether the record is included in the primary analysis
<Project specific item value 1>	N	Project specific field – must be referenced in Metadata table.
< Project specific item value 2>	N	Project specific field – must be referenced in Metadata table.
< Project specific item value 3>	N	Project specific field– must be referenced in Metadata table.
...
< Project specific item value n>	N	Project specific field– must be referenced in Metadata table.

Metadata table specification

Metadata table			
Field name	Mandatory	Code frame	Description
Item_Name	Y	N	The corresponding field name in the data table.
Item_Description	Y	N	A label or description of the field name in the data table.
Item_Code	N	N	The code used in the value/cell of the field in the data table.
Item_Code_Label	N	N	The label or description for the coded value.

Please see example metadata table below for code frames.

Example Metadata table including coding frames			
Item_Name	Item_Description	Item_Code	Item_Code_Label
PreTest_Outcome_1	Primary pre-test outcome.		
PreTest_Outcome_2	Second pre-test outcome.		
PreTest_Outcome_3	Third pre-test outcome.		
PreTest_Outcome_4	Fourth pre-test outcome.		
PreTest_Outcome_5	Fifth pre-test outcome.		
PostTest_Outcome_1	Primary post-test outcome.		
PostTest_Outcome_2	Second post-test outcome.		
PostTest_Outcome_3	Third post-test outcome.		
PostTest_Outcome_4	Fourth post-test outcome.		
PostTest_Outcome_5	Fifth post-test outcome.		
PreTest_Pupil_Yeargroup	National curriculum year group on the day of baseline test.	N1	Nursery first year
PreTest_Pupil_Yeargroup	National curriculum year group on the day of baseline test.	N2	Nursery second year
PreTest_Pupil_Yeargroup	National curriculum year group on the day of baseline test.	R	Reception
PreTest_Pupil_Yeargroup	National curriculum year group on the day of baseline test.	1	Year 1
PreTest_Pupil_Yeargroup	National curriculum year group on the day of baseline test.	2	Year 2
PreTest_Pupil_Yeargroup	National curriculum year group on the day of baseline test.	3	Year 3
PreTest_Pupil_Yeargroup	National curriculum year group on the day of baseline test.	4	Year 4
PreTest_Pupil_Yeargroup	National curriculum year group on the day of baseline test.	5	Year 5
PreTest_Pupil_Yeargroup	National curriculum year group on the day of baseline test.	6	Year 6
PreTest_Pupil_Yeargroup	National curriculum year group on the day of baseline test.	7	Year 7
PreTest_Pupil_Yeargroup	National curriculum year group on the day of baseline test.	8	Year 8
PreTest_Pupil_Yeargroup	National curriculum year group on the day of baseline test.	9	Year 9
PreTest_Pupil_Yeargroup	National curriculum year group on the day of baseline test.	10	Year 10
PreTest_Pupil_Yeargroup	National curriculum year group on the day of baseline test.	11	Year 11
PreTest_Pupil_Yeargroup	National curriculum year group on the day of baseline test.	12	Year 12
PreTest_Pupil_Yeargroup	National curriculum year group on the day of baseline test.	13	Year 13
PreTest_Pupil_Yeargroup	National curriculum year group on the day of baseline test.	14	Year 14
PreTest_Pupil_Yeargroup	National curriculum year group on the day of baseline test.	99	Missing/ not known
PostTest_Pupil_Yeargroup	National curriculum year group on the day of baseline test.	N1	Nursery first year
PostTest_Pupil_Yeargroup	National curriculum year group on the day of the test.	N2	Nursery second year
PostTest_Pupil_Yeargroup	National curriculum year group on the day of the test.	R	Reception
PostTest_Pupil_Yeargroup	National curriculum year group on the day of the test.	1	Year 1

PostTest_Pupil_Yeargroup	National curriculum year group on the day of the test.	2	Year 2
PostTest_Pupil_Yeargroup	National curriculum year group on the day of the test.	3	Year 3
PostTest_Pupil_Yeargroup	National curriculum year group on the day of the test.	4	Year 4
PostTest_Pupil_Yeargroup	National curriculum year group on the day of the test.	5	Year 5
PostTest_Pupil_Yeargroup	National curriculum year group on the day of the test.	6	Year 6
PostTest_Pupil_Yeargroup	National curriculum year group on the day of the test.	7	Year 7
PostTest_Pupil_Yeargroup	National curriculum year group on the day of the test.	8	Year 8
PostTest_Pupil_Yeargroup	National curriculum year group on the day of the test.	9	Year 9
PostTest_Pupil_Yeargroup	National curriculum year group on the day of the test.	10	Year 10
PostTest_Pupil_Yeargroup	National curriculum year group on the day of the test.	11	Year 11
PostTest_Pupil_Yeargroup	National curriculum year group on the day of the test.	12	Year 12
PostTest_Pupil_Yeargroup	National curriculum year group on the day of the test.	13	Year 13
PostTest_Pupil_Yeargroup	National curriculum year group on the day of the test.	14	Year 14
PostTest_Pupil_Yeargroup	National curriculum year group on the day of the test.	99	Missing/ not known
Pupil_Gender	Pupil gender	1	Male
Pupil_Gender	Pupil gender	2	Female
Pupil_Gender	Pupil gender	U	Unknown
Pupil_Ethnicity	Pupil ethnicity	1	White - British
Pupil_Ethnicity	Pupil ethnicity	2	White - Irish
Pupil_Ethnicity	Pupil ethnicity	3	Traveller of Irish Heritage
Pupil_Ethnicity	Pupil ethnicity	4	Any Other White Background
Pupil_Ethnicity	Pupil ethnicity	5	Gypsy / Roma
Pupil_Ethnicity	Pupil ethnicity	6	White and Black Caribbean
Pupil_Ethnicity	Pupil ethnicity	7	White and Black African
Pupil_Ethnicity	Pupil ethnicity	8	White and Asian
Pupil_Ethnicity	Pupil ethnicity	9	Any Other Mixed Background
Pupil_Ethnicity	Pupil ethnicity	10	Indian
Pupil_Ethnicity	Pupil ethnicity	11	Pakistani
Pupil_Ethnicity	Pupil ethnicity	12	Bangladeshi
Pupil_Ethnicity	Pupil ethnicity	13	Any Other Asian Background
Pupil_Ethnicity	Pupil ethnicity	14	Black Caribbean
Pupil_Ethnicity	Pupil ethnicity	15	Black - African
Pupil_Ethnicity	Pupil ethnicity	16	Any Other Black Background
Pupil_Ethnicity	Pupil ethnicity	17	Chinese
Pupil_Ethnicity	Pupil ethnicity	18	Any Other Ethnic Group
Pupil_Ethnicity	Pupil ethnicity	19	Refused
Pupil_Ethnicity	Pupil ethnicity	20	Information Not Yet Obtained

Pupil_SEN_Provision	Pupil SEN provision	0	No Special Educational Need
Pupil_SEN_Provision	Pupil SEN provision	1	School Action or Early Years Action
Pupil_SEN_Provision	Pupil SEN provision	2	School Action Plus or Early Years Action Plus
Pupil_SEN_Provision	Pupil SEN provision	3	Statement
Pupil_SEN_Provision	Pupil SEN provision	4	SEN support
Pupil SEN provision	Pupil SEN provision	5	Education, health and care plan
Treatment_Allocation	Treatment Group	1	In intended treatment group
Treatment_Allocation	Treatment Group	0	In control group
Pupil_FSM	FSM eligibility at the beginning of the intervention	1	Eligible
Pupil_FSM	FSM eligibility at the beginning of the intervention	0	Not eligible
Pupil_FSM6	FSM eligibility in the last six years at the beginning of the intervention	1	Eligible
Pupil_FSM6	FSM eligibility in the last six years at the beginning of the intervention	0	Not eligible
Pupil_EAL	Pupil whose first language is other than English	1	Yes
Pupil_EAL	Pupil whose first language is other than English	0	No
In_Primary_Analysis	Record is included in the primary analysis	1	Yes
In_Primary_Analysis	Record is included in the primary analysis	0	No

Notes

In Primary Analysis

This flag denotes whether the record was used as part of the main or primary analysis. It is intended to support replication of the original analysis.

NPD data items

As part of an NPD data request you will to obtain the pupil identifier (PMR), background information about the pupil and possibly baseline test results. Although the data collection includes items such as ethnicity there is no need to record this information unless it is relevant to secondary analysis. With appropriate permissions (when the schools and parents have been informed in advance, and forms for withdrawal from data processing have not been received) we will be able to link to the NPD for standardised data items to enable meta-analysis. The purpose of formally listing these fields is to encourage standardization of commonly collected items that should ease any subsequent analysis.

Pupil identifiers

The specification requests that the pupil identifier field PMR (Pupil Matching Reference) is submitted. Provided we collect PMRs we will be able to build a longitudinal dataset that incorporates changes in the school census variables collected.

Pupil data from the School Census

It is anticipated that evaluators will have obtained PMRs and School Census data from the DfE. Ideally, census data should be the most recent available at the time of the intervention. However, this may not always be the case in practice. We ask that you submit a limited amount of School Census data as received from DfE and include the series reference in your submission. In NPD, school census data items are suffixed with the series, e.g. SPR12, which denotes the academic year and term in which it was collected. If you have obtained census data for more than one term, please supply the series used in your analysis.

It is not necessary to define School Census data items in the metadata table if you have used the coding supplied in the NPD extracts. However, you should define them if you have recoded the variables. For instance, you do not need to define Pupil_Ethnicity if you have retained the four character codes used in school census (e.g. WBRI) but should supply metadata if you have recoded them (e.g. recoded WBRI as 1).

Treatment data

The treatment allocation variable should be used to denote all intended treatment groups, including combinations of multiple treatments. Control groups should also be identified.

The treatment dose should be used to give an indication of the amount of the intended treatment each pupil actually received.

Fidelity

It is anticipated that the assessment of fidelity will vary between projects and project specific fields should be included in the submission. Information that is available and meaningful for secondary analysis should be recorded.

Pre-test data

Include here details of any pre-tests used. In the case where pupils have taken multiple tests and you have used an average in calculating effect sizes, report the average as pretest_outcome_value1 and the component tests in _value2 to _value5 slots.

Post-test data

Include here the main post-test variable used to evaluate the effectiveness of the intervention. You may also wish to include other post-test variables, e.g. other variables you have tested for spillover effects. With the dataset structured with one row per pupil then additional outcomes at a higher level should be recorded against each associated record, for example, a teacher or class level outcome.

Optional data

In the optional data, include any other variables that you believe have a material effect upon your evaluation of the intervention. Include any other variables you have used in determining the effect

size of an intervention, for instance additional contextual data about the pupil (or the school) or process data about the way the treatment was given.

More than one year's data

When a project is conducted over more than one year then we would expect this to be flagged in the dataset. The time aspect should be recorded in project-specific fields as appropriate. For example, if there are two cohorts receiving treatment over two consecutive years, then simply adding a year field should be sufficient. If the intervention for a single cohort runs over one year and includes a notable interim test, then a single record should still be provided for each pupil. In such cases the project specific fields should be used to record the details of the interim results.

Blocks

The specification only lists one blocking field: School_Stratum_ID. Where relevant, additional project specific fields should be added to record more complex designs. The intention of the School_ID, Class_ID and Teacher_ID fields is primarily to record genuine attributes of the pupil to help identify potential spillover effects.

Pairing ID

If paired or matched randomisation is used, then this field should identify which schools or pupils are associated across the treatment and control groups.