YOUTH ENDOWMENT FUND



Submitting data to the YEF archive

Guidance for evaluators

Version 1.1

Contents

Version history 3

Introduction 4

Recap on the archiving process 5

How data is submitted 7

Frequency of submission 10

Formatting data 11

Submitting your data 16

Step-by-step guide 20

Annex 1 – Formatting of mandatory variables 21

Annex 2 – Formatting of non-mandatory variables 23

Annex 3 – Variables not permitted in the ONS Secure Research Service 25

Annex 4 – Data dictionary checklist 26

1. Version history

Version	Date published	Main changes from previous version
1.0	December 2022	First published guidance
1.1	October 2026	Updated for clarity and to improve quality of submissions



2. Introduction

As part of our mission to understand what interventions and services best prevent children from becoming involved in violence, we are creating a data archive. The YEF archive will include all data collected on the young people who take part in our projects. We'll use this data to follow-up on their outcomes, to find out what works to reduce offending and improve other important outcomes.

This document sets out the process for archiving data once evaluations have finished. In particular, it covers the agreements that we will put in place to facilitate archiving, the processes for submitting data and what evaluators need to do in order to prepare their data for submission.

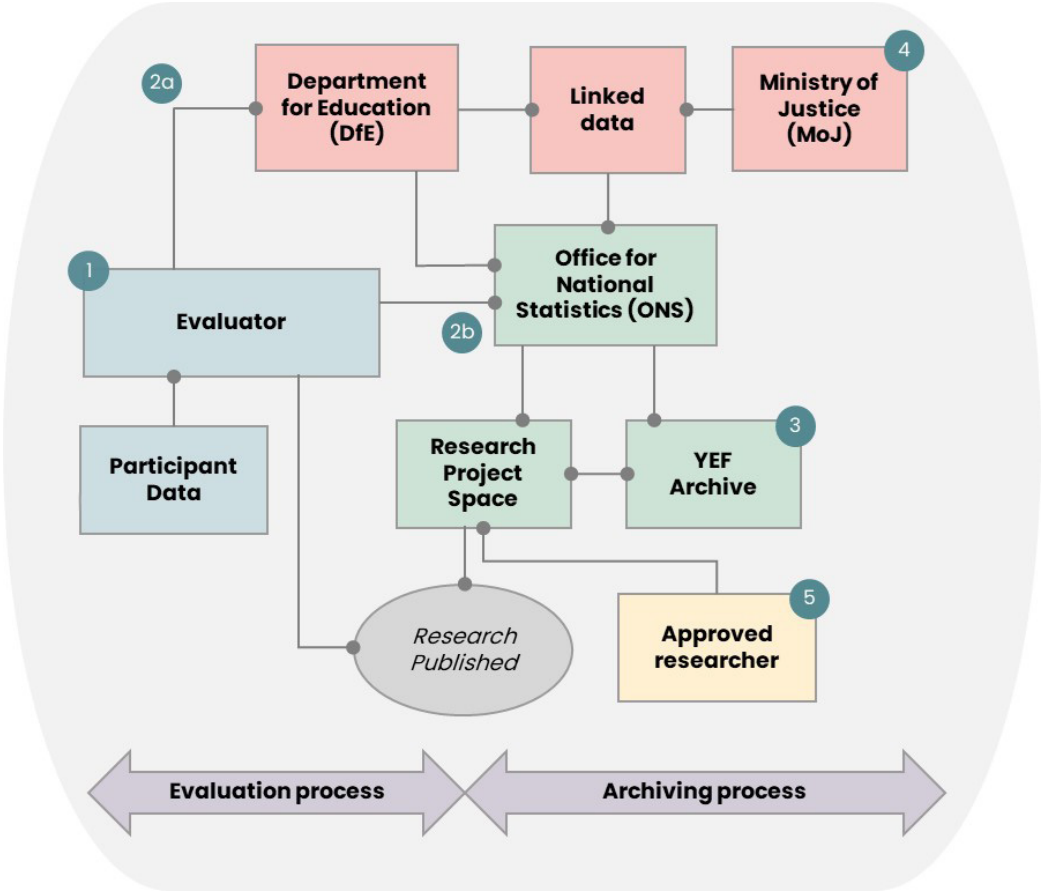
This guidance applies to you if **all** of the following are true:

Does this guidance apply to me?	
You are an evaluator on a YEF funded project	✓
You are collecting individual level data on the participants in your trial	✓
You are evaluating outcomes via a Pilot or Efficacy/Effectiveness Study	✓

3. Recap on the archiving process

Our technical guidance for projects and evaluators and our DPIA contain detailed information on how the YEF archive works. Below is a diagram that summarises the process.

Figure 1: Overview of the YEF archiving process



1. **Evaluation:** Personal data is collected from participants by the evaluator and analysed. Data may also come from parents, providers or practitioners. Evaluators analyse the data and publish findings in a deidentified form. During this step, evaluators are controllers of the data.

2. **Data submitted for archiving:** Two datasets are created and submitted separately to the DfE and ONS:

    a. One dataset contains just children's identifying data and a unique project-specific reference number for each child. This is submitted to the DfE and pseudonymised – a process where the personal identifying data is removed and replaced with DfE's pupil matching reference numbers (PMRs). The PMRs and the project-specific reference numbers are then submitted to the ONS for storage in the YEF archive.

b. The second dataset contains all the evaluation data and the project-specific reference numbers. This is submitted directly to the ONS, where it's also stored in the YEF archive.

The inclusion of the project-specific reference numbers in both the data submitted to DfE and the data submitted to ONS allows both datasets to be recombined once they're in the ONS secure environment. Once the data has been submitted to the DfE and the ONS, the YEF will become the controllers.

3. **Secure storage:** Both datasets are held securely in the YEF archive by the ONS, who manage the archive on our behalf. Nobody can access the evaluation data without approval from us, nor without being an ONS accredited researcher. If access to PMRs is required, approval must additionally be obtained from the DfE.

4. **Data-linking:** The DfE and MoJ routinely link together education and crime records. On request, this can be made available in the ONS Secure Research Service (SRS). Due to the inclusion of the PMRs in this data and the records of the children in the YEF archive, it's possible to match both datasets in the future, in order to follow-up on education and offending outcomes.

5. **Access to the YEF archive:** In the future, approved researchers may apply to access YEF data via the ONS SRS. This will be accessed via their own project space created in the ONS SRS by the ONS. They may apply to the DfE and MoJ to access the linked NPD-PNC data, and if successful it will be made available here to combine with the YEF evaluation data, using the PMRs. Researchers will only be able to access pseudonymised data. All results will be published in a deidentified form.

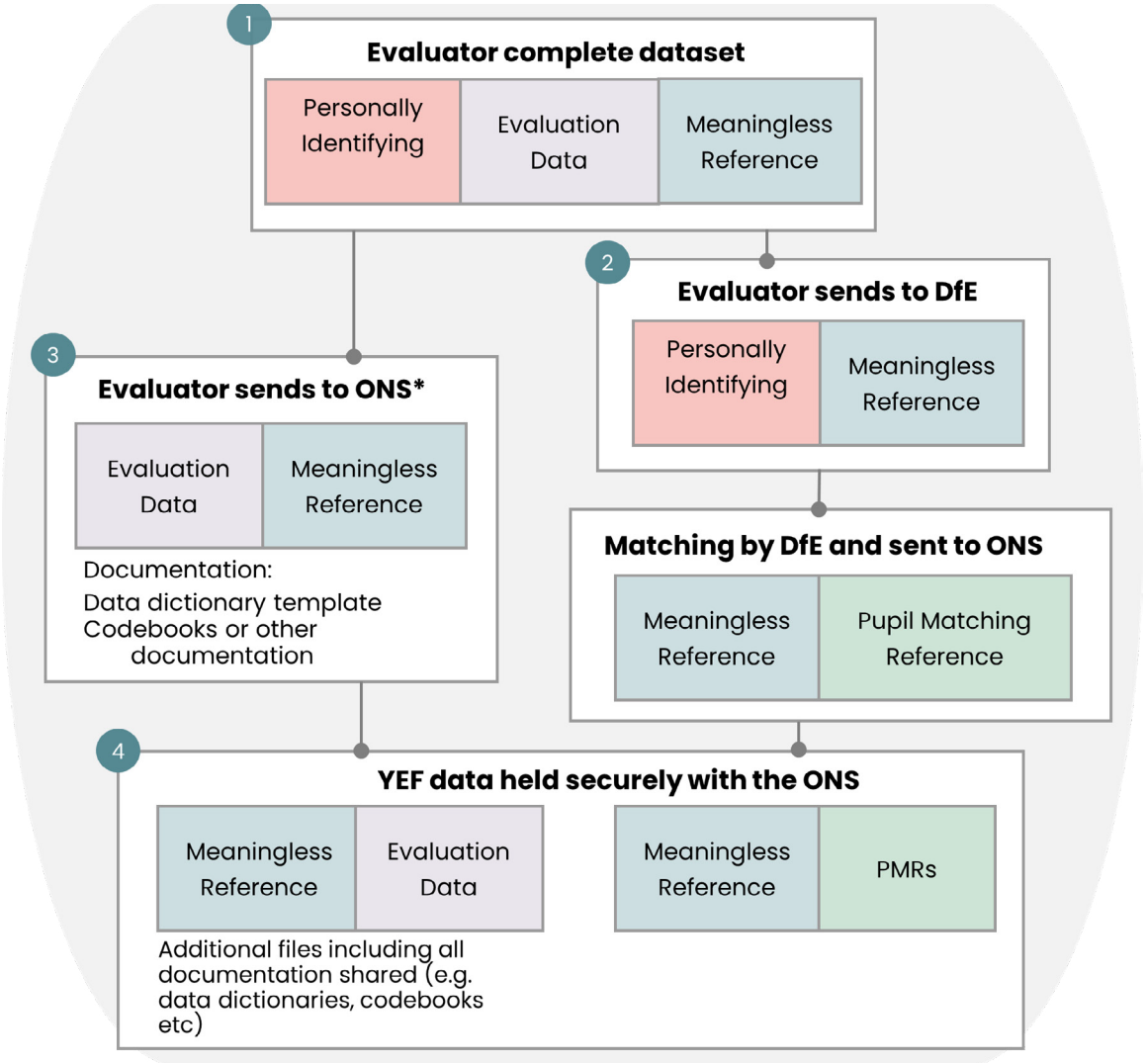
The above process applies to cases where evaluators do not need to access DfE data during the evaluation (step 1). Our technical guidance for projects and evaluators explains how the process works where evaluators do access DfE data during step 1.

The remainder of this guidance relates to step 2 – how data is submitted for archiving; how it's prepared, and the wider supporting documentation and agreements needed.

4. How data is submitted

Figure 2 below summarises how data will be submitted to the YEF archive.

Figure 2: Overview of the data submission process



*In some cases, evaluation data will already be in the SRS environment before it's submitted to the YEF archive and will have PMRs already attached. This might be because evaluators are looking at school outcomes from the National Pupil Database during the evaluation period. In these cases, evaluators will not need to submit data to DfE to be matched again to their PMRs.

The main steps are:

Step 1: YEF evaluators' complete dataset

At the end of YEF projects, evaluators will have prepared a single child level dataset containing three types of data:

- **Personally identifying data:** Data on the names, dates of birth, postcodes and unique pupil numbers (UPNs) when delivering in schools.
- **Evaluation data:** All the data collected as part of the evaluation including information on the background characteristics and home life of the participants, details of the intervention received (i.e. timing, dosage etc.), and data on key outcomes.
- **Meaningless reference numbers:** An identifier created by the evaluator that uniquely identifies each child in the dataset. The meaningless reference numbers will take a specific form, including their YEF Evaluation Reference Number (ERN) (provided to the evaluator by the YEF) and a random five-digit number generated by the evaluator that is unique to every child in the evaluation. See below for details on formatting data for submission.

When the evaluation is finished and evaluators are ready to submit their data for archiving, the YEF will enter into a **controller-to-controller agreement**. This is an amendment to the evaluator agreement between YEF and evaluators. It codifies the point at which YEF becomes the controller of the evaluation data and instructs evaluators to share data with the ONS and DfE, on the basis set out in this guidance. This agreement will also contain the '**evaluation reference number' (ERN)**, a unique number given to each evaluation so we can track projects consistently in the future.

Step 2: Evaluators submit identifiable data to DfE to be pseudonymised

Evaluators submit a dataset to the DfE containing only the personally identifying data (i.e. name, address etc.) and the list of meaningless reference numbers. None of the evaluation data will be sent to the DfE. The data will be shared using the secure Galaxkey data sharing platform (DfE's preferred secure file transfer system). See below for details on how to format data for submission.

Once received, the DfE will link the personally identifying data to the education records they hold, replacing names, dates of birth etc. with DfE's Pupil Matching Reference numbers (PMRs), contained in the National Pupil Database (NPD). The DfE will then submit to the ONS a list containing just the PMRs and meaningless reference numbers, for storage in the YEF archive dedicated project space. Evaluators will not be involved in this part of the process.

The DfE will act as processor on behalf of the YEF whilst they match children to the NPD. The DfE will only process YEF data for the purpose of pseudonymisation. Once this is completed and data transferred to the ONS, all YEF data will be removed from their systems.

Step 3: Evaluators send deidentified data to ONS

Evaluators submit to the ONS their evaluation data and the list of meaningless reference numbers on each of the children. No directly identifying data (e.g. names and addresses) can be included. All data will be submitted to the ONS using the secure MoveIT data sharing platform. The ONS will provide the login details for MoveIT.

Evaluators will also submit to the ONS a data dictionary, and any codebooks or a description of the analysis used in generating the results presented in the final evaluation report. For more details on preparing the data and accompanying documentation for submission to the ONS, see below.

Once the data has been submitted to both DfE and ONS, evaluators will either permanently destroy all the personal data they have collected or retain a copy they will use in future research[1].

Step 4: ONS stores YEF data

The ONS will store the data received from evaluators and the DfE in two ways:

- Each dataset from evaluators will be stored in its own unique project folder within the YEF archive, along with the accompanying data dictionaries and codebooks
- The ONS will also separately store a data table for each project that includes separate columns for PMRs and meaningless identifiers provided by the DfE.

1 If evaluators wish to retain any data once data has been submitted to the archive, they will be sole controllers of it. They will need to explain to participants at the start of the evaluation their intention to keep a copy and provide all the necessary information on how they will keep the data safe and how participants will be able to exercise their rights over the data evaluators retain.

5. Frequency of submission

To reduce the amount of work involved for the DfE and ONS, we've agreed evaluators will only be able to submit data in four months of the year, as shown below. This means if you miss a particular window, you'll have to wait three months until the next available month.

When can data be submitted?

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
✗	✗	✓	✗	✗	✓	✗	✗	✓	✗	✗	✓

We ask that evaluators store their data for at least three months after they have first submitted it. There may have been issues with the original upload, in particular when submitting data to the DfE. If the match rate is particularly low, we may ask you to resubmit in the next upload window.



6. Formatting data

6.1 Preparing the master dataset

Once evaluations have finished, evaluators will need to create a single child-level file, containing all the data that will go to the DfE and ONS. This means there will be a single row for every child in the dataset.

Where evaluators have used data from different sources (e.g. child and parent report surveys separately) they will need to ensure these are combined into one file, where all the data is collapsed at the child-level. Any identifiers relating to the person who completed the survey or supplied the data, other than the child, will need to be removed.

The file will need to be 'readable'. By this we mean:

- there are no empty rows or columns (such as in the top row or first column);
- the top row contains the variable labels or acronyms;
- the left-most columns contain the child's identifying information, including their unique reference number; and,
- all missing values should be consistently coded with "NA".

An illustration of how the data should be set out is below.

	A	B	C	D	E	F	G
1	YP_Ref	Var2	Var3	Var4	Var5	Var6	...
2	value	value	value	value	value	value	...
3	value	value	value	value	value	value	...
4	value	value	value	NA	value	value	...
5	value	value	value	value	value	value	...
6	value	NA	value	value	value	NA	...
7	value	value	value	value	value	value	...
8	...	...	...	...	...	...	...

As mentioned above, evaluators will need to create a meaningless reference number for each child in the dataset. This should be labelled 'YP_Ref'. This reference number should be constructed using the Evaluation Reference Number (ERN) and a random unique five-digit number appended with an underscore '_'. The table below gives an illustration of how these should be constructed.

Observation	Evaluation Reference Number (ERN)	Random Number	Meaningless Reference Number (YP_Ref)
Child 1	LG0001	00021	**LG0001_00021**
Child 2	LG0001	00301	**LG0001_00301**
Child 3	LG0001	00119	**LG0001_00119**

It is critically important that the random number uniquely identifies the child in the dataset – no two should have the same reference numbers. When the meaningless reference number is constructed, evaluators should check that there are no duplicates.

The Evaluation Reference Number (ERN) will be provided to evaluators by the YEF as they prepare the data for archiving and will be referenced in the controller-to-controller agreement signed between the YEF and the evaluator (see above). The ERN will allow the YEF to locate the relevant project data in our archive.

6.2 Preparing the data for the DfE[2,3]

Using the single child-level data file, evaluators will need to generate a separate dataset for DfE, saved in a comma-separated values '.csv' format. It will need to contain the children's YEF meaningless reference numbers (YP_Ref) and the additional information contained in the table below. Not all evaluations will have collected data for each of the mandatory and non-mandatory fields. **However, even where the data is missing, we ask that the submitted data file contains headings for all nine variables below.**

2 For evaluators that are already working in the SRS and have PMRs matched to their data, prior to submission to the YEF archive (i.e. in cases where the original evaluation is using other data matched in from the NPD), then this step is not required. You proceed directly to the step on submitting data to the ONS.

3 For projects delivered in Wales, this step does not have to be followed. Welsh children are not in the NPD so cannot be found by the DfE. We will however expect data from Welsh projects to be sent to the ONS. Whilst we can't track Welsh participants' future outcomes, we will still use their data in the future for reanalysis.

Variable Name	Mandatory	Description	Format
YP_Ref	Yes	Child's YEF evaluation reference number.	Text
Forename	Yes	Child's given forename.	Text
Surname	Yes	Child's given surname.	Text
Middlenames	No	Child's middle name(s).	Text - "Name1, Name2..."
DOB	Yes	Date of birth of the child.	Text - "DD-MM-YY"
Postcode_Child	No	Postcode of child's main residence.	Text - "AA9A 9AA"
UPN	No	Unique Pupil Number (UPN) of the child held by the school.	Text
URN	No	Unique Reference Number (URN) of the school that the child attends. This is a 6-digit reference number assigned to each school in England and can be found via the Get Information about Schools (GIAS) system.	Numeric "XXXXXX"
LAESTAB	No	Local Authority Establishment number. The combination of the school's local authority code (first 3 digits) and the establishment code (last 4 digits).	Numeric "XXXXXXX"

Below are a worked example and a blank template for evaluators to use.

Worked example



Blank example



6.3 Preparing the data for the ONS

Formatting data

Using the single child-level data file (as set out in 6.1 above), evaluators will need to generate a second dataset, this time for the ONS. This should be saved in the comma-separated values '.csv' format. This time the dataset will need to contain just the child's meaningless reference numbers and **all the other data collected and used as part of the evaluation.** This dataset **must not contain any identifying data** (i.e. names, addresses etc. of the children in the trial).

In addition to the meaningless reference number, there are broadly three categories of data which we'd expect evaluators to report on:

- **Background and characteristic information:** Information about where the child lives, their characteristics and past experiences, the make-up of their families and peer groups etc.;
- **Programme and delivery intervention:** this relates to the nature of the programme they were part of, the number and type of sessions they attended, as well as information about the trial they were part of (i.e. when was data collected, were they part of the intervention or control group etc.);
- **Outcome:** Data on the outcomes that were being evaluated.

No two evaluations will be identical. This means each evaluation dataset will contain a different set of variables. We're asking that all data used in generating the final evaluation report is included in the archive.

To help evaluators determine the structure of the data to include, see Annexes 1 and 2, which respectively provide a list of 'mandatory' and 'non-mandatory' fields. Annex 1 includes a list of variables we expect most evaluators will submit. Annex 2 includes a list of variables we expect some but not all evaluations will include. Neither list is exhaustive. They're intended to provide consistency in how data is to be formatted before submitting to the ONS.

Annex 3 contains a list of identifying variables that evaluators must omit from the ONS dataset. However, if school identifiers (i.e. URN, LAESTAB) are available in the data collected, evaluators should retain a pseudonymised version in the ONS dataset.

Other documentation

There are two other things evaluators will need to complete to accompany the data being sent to the ONS. Your evaluation manager will contact you about this as you prepare your data for submission.

1. **Data dictionary:**

   - This is a requirement of all data being sent to the ONS. It's a file that contains a description of all the variables included in the dataset being submitted. It also contains a high-level summary of what's in the dataset. We will send evaluators a template for this.

   - Annex 4 provides a checklist to support evaluators when preparing the dictionary. This checklist reflects feedback we've previously provided to evaluators, and will help minimise any revisions to the dictionary and dataset ahead of submitting to ONS.

2. **Codebook or description of analysis:**

   - As one of the goals in creating the archive is to allow for the reanalysis of the trials on which data is stored, it's important we also collect details of the analysis carried out originally by the evaluator, so others can understand how the results were generated. This can include the underlying codebooks used in generating the final outputs in the evaluation report (e.g. in .txt file format). If no syntax is available, evaluators should provide a description of the steps taken in the analysis.

   - These should include the codes used or steps taken to set up the data (cleaning, variable generation, etc.) and to conduct the analyses (primary, secondary, compliance, sensitivity, robustness, multiple imputation, etc.). Codes should be accompanied by brief annotations explaining their purpose or intended outcome.

Once these have been created, evaluators must submit them to the YEF to check, before they are submitted to the ONS. **Note, no data from the evaluation should be directly shared with the YEF at any point.**

7. Submitting your data

7.1 Submitting your data to the DfE

Notifying DfE of your intention to upload

Prior to submitting your data to the DfE, you will need to provide advance notice. This should be at least five days before you intend to send the data. Below is a template email for you to complete and send. Please ensure that your YEF evaluation manager is copied into this email (and subsequent correspondence emails with DfE and ONS regarding archiving the data).



TEMPLATE

To: data.sharing@education.gov.uk

Cc: data@youthendowmentfund.org.uk; *‹email address of your evaluation manager at YEF›*

Subject: *Advance notification of matching request to DfE for submission to YEF data archive (DR210811.02)*

Dear DfE,

My name is *‹insert name›* from *‹insert organisation›* and we are an evaluator on behalf of the Youth Endowment Fund (YEF) for the *‹insert name of the intervention being evaluated›* project. Our Evaluation Reference Number (ERN) is *‹insert ERN›* and this email is to provide advance notice of data we will be submitting to DfE for matching before being de-identified and provided to ONS for inclusion in the YEF archive (DR210811.02) within the ONS Secure Research Service (SRS).

We expect the data to be submitted to DfE via Galaxkey on *‹insert estimated date, please provide a minimum of 5 days' notice›*.

Sending your data

For those that haven't previously used Galaxkey, navigate to: https://manager.galaxkey.com/services/registerme. You'll be asked to enter your email address. Enter the text on the screen and click Sign Up. There's a free version, which is the default option.

To submit your data:

1. Sign in to your Galaxkey account: https://manager.galaxkey.com/Account/LogOn?ReturnUrl=%2f

2. On the left-hand menu, click on Compose Email

3. In the To: box, enter: data.sharing@education.gov.uk

4. In the Subject field, enter: *"OFFICIAL SENSITIVE - Matching request to DfE for submission to YEF data archive for evaluation <insert ERN>"*

5. In the 'Message' field (Message summary), please enclose a message which reads: *"Dear DfE, Following our previous email notifying you of our intention to submit data on behalf of the YEF for evaluation <insert ERN>, please find here the relevant file(s). The dataset has been prepared in accordance with the guidance and template prepared by the YEF".*

6. Click on the Add button to upload your data. Please note that the data must be encrypted (e.g. password set in Excel), or ideally compressed and encrypted using WinZip or 7Zip.

7. Complete any other details in the email as required, then click on Send Secure.

8. Please email the password used to encrypt the dataset separately to the data.sharing@education.gov.uk mailbox.

7.2 Submitting your data to the ONS

Documents to be checked by YEF before submitting

Before submitting anything to the ONS, you'll need to have your data dictionary and description of analysis checked by the YEF. Your evaluation manager will contact you about this as you prepare your data for submission. The YEF will check and confirm these documents have been completed correctly. **Please ensure you do not send the YEF any personal data relating to your trial participants.**

Notifying ONS of your intention to upload

While the YEF is checking your data dictionary and description of analysis, you should notify the ONS of your intention to submit data. Below is a template email for you to complete and send. You should email the ADR Data Curation team at the ONS (adrcuration@ons.gov.uk), and copy in the ONS Data Management Team (SRS.data.management@ons.gov.uk), along with your evaluation manager.

Note, if the YEF has already signed off on the data dictionary ahead of your advance notice to ONS, you can also submit your data dictionary for ONS' review with this email. In this case, this and the next email template can be combined.



TEMPLATE

To: adrcuration@ons.gov.uk

Cc: SRS.data.management@ons.gov.uk; *<email address of your evaluation manager at YEF>*

Subject: Advance notification of submission of data to the YEF archive

Dear ONS,

My name is *<insert name>* from *<insert organisation>* and we are an evaluator on behalf of the Youth Endowment Fund (YEF) for the *<insert name of the intervention being evaluated>* project. Our Evaluation Reference Number (ERN) is *<insert ERN>* and this email is to provide advance notice we will be submitting data on behalf of the YEF to the ONS, for inclusion in the YEF archive.

We expect to submit our data dictionary for initial checking on *<insert estimated date, please provide a minimum of 5 days' notice>*. Please find below a description of the dataset:

- The data has been collected as part of YEF's *<name of project>* trial. The project works to *<one to two sentence description on the nature of intervention and the main aims>*
- The study design is *<one to two sentences on the study design and the pre-post-test measures>*

Submitting your Data Dictionary

Once you have YEF sign-off, you should submit your data dictionary to the ONS, to the same people listed above. A template email is below.



TEMPLATE

To: adrcuration@ons.gov.uk

Cc: SRS.data.management@ons.gov.uk; *<email address of your evaluation manager at YEF>*

Subject: Re: Advance notification of submission of data to the YEF archive

Dear ONS,

Following my previous email on *<insert date ONS previously contacted>* regarding YEF project *<insert ERN>*, please find attached the data dictionary for you to check. I understand that it's previously been agreed between yourselves and the YEF that no Data Deposit Agreement (DDA) will be needed as there's already a DDA in place covering all data submitted to the YEF archive.

Please can you send me MoveIT account login details, so I may upload the data once you're content with the data dictionary? Additionally, please can you indicate who in the ONS data management team the link to the data should be shared with, once it's been uploaded to MoveIT?

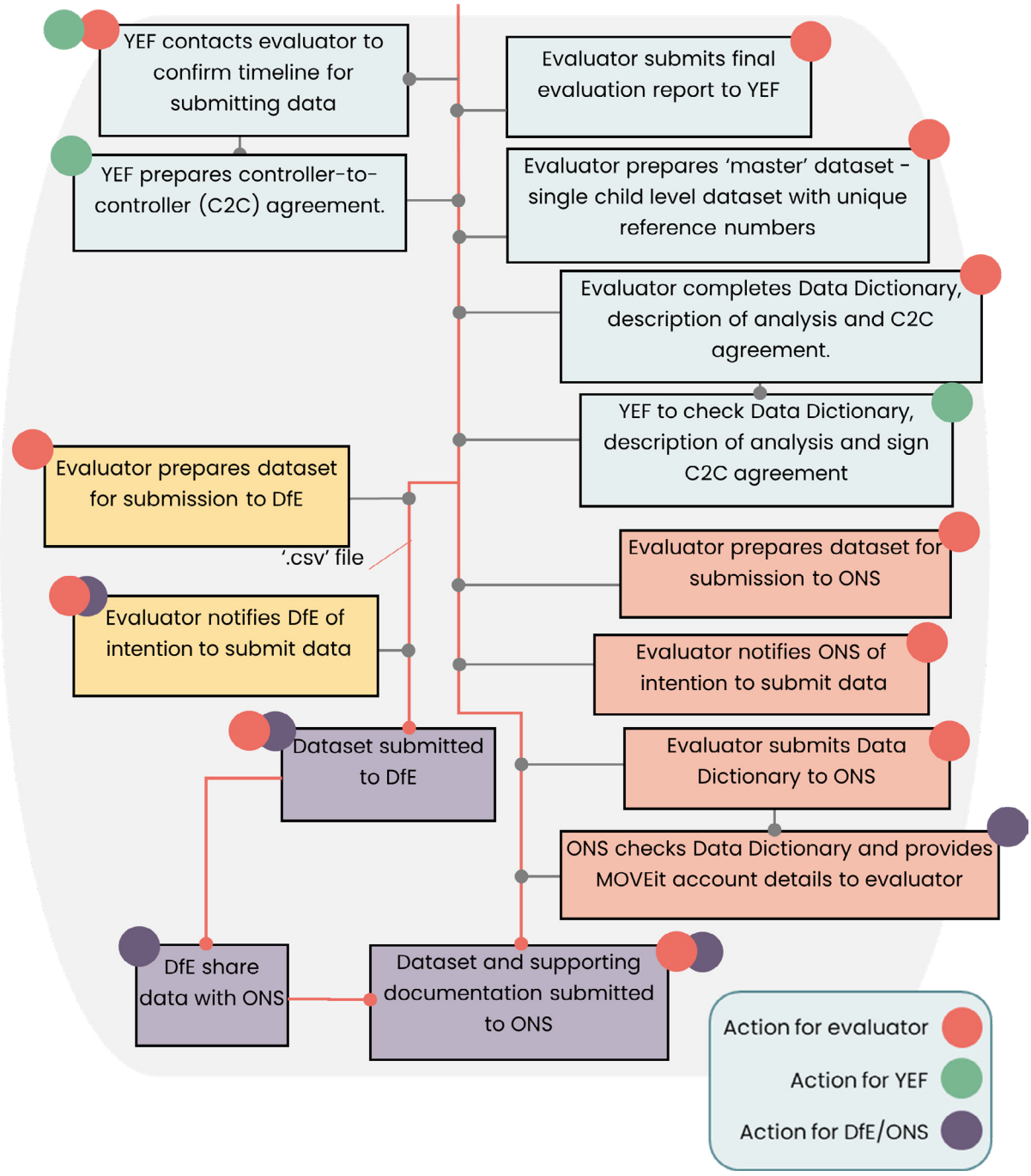
The ONS may follow-up with you if there are any outstanding issues for you to address. They will also provide you with login details for the MoveIT account via which you should submit your data as well as the email address of a named person at the ONS who will receive the data once it's been submitted.

Submitting your data

Once you've uploaded your data to MoveIT, there will be a button that says 'send files' which opens an email on the MoveIT webpage. Please send this to the nominated member of the ONS SRS Data Management team. An automatic email will be generated, sharing the data with them.

8. Step-by-step guide

The diagram below sets out the step-by-step guide for submitting data to the archive. The diagram summarises the steps outlined elsewhere in this document.



Annex 1 - Formatting of mandatory variables

Table of variables we expect the majority of our projects to collect and the recommended formatting.

For all outcome variables, please include both the **derived scores used in the analysis** and **all individual items** from which these scores are calculated. For example, see sdq_1_pre and sdq_1_post in the table below.

Variable Abbreviation	Variable Description	Variable Type	Variable Format
YP_Ref	Young person's unique evaluation reference number	Programme/ delivery	Text ('XXXXXX_XXXXX")
group	Group that the participant was allocated to	Programme/ delivery	Numeric (1 = treatment; 0 = control)
pre_test_dt	Date of the pre-test baseline data collection	Programme/ delivery	Text (DD/MM/YY)
post_test_dt	Date of the post-test data collection	Programme/ delivery	Text (DD/MM/YY)
start_dlvry_dt	Date of the start of delivery	Programme/ delivery	Text (DD/MM/YY)
end_dlvry_dt	Date of the last day of delivery	Programme/ delivery	Text (DD/MM/YY)
sdq_ver	Version of SDQ used i.e., self, parent or teacher-report	Outcome	Numeric (1 =self, 2 = teacher, 3 = parent)
sdq_tot_pre	Pre-test score on SDQ scales emotional problems, conduct problems, hyperactivity, and peer problems	Outcome	Numeric (0-40)
sdq_cndct_pre	Pre-test score on SDQ conduct problems scale	Outcome	Numeric (0-10)
sdq_hyper_pre	Pre-test score on SDQ hyperactivity scale	Outcome	Numeric (0-10)
sdq_peer_pre	Pre-test score on SDQ peer problems scale	Outcome	Numeric (0-10)

Variable Abbreviation	Variable Description	Variable Type	Variable Format
sdq_emot_pre	Pre-test score on SDQ emotional problems scale	Outcome	Numeric (0-10)
sdq_prosocial_pre	Pre-test score on SDQ prosocial scale	Outcome	Numeric (0-10)
sdq_1_pre	Pre-test score on SDQ item 1: "XXX XXX"	Outcome	Numeric (0 = not true, 1 = somewhat true, 2 = certainly true)
sdq_tot_post	Post-test score on SDQ scales emotional problems, conduct problems, hyperactivity, and peer problems	Outcome	Numeric (0-40)
sdq_cndct_post	Post-test score on SDQ conduct problems scale	Outcome	Numeric (0-10)
sdq_hyper_post	Post-test score on SDQ hyperactivity scale	Outcome	Numeric (0-10)
sdq_peer_post	Post-test score on SDQ peer problems scale	Outcome	Numeric (0-10)
sdq_emot_post	Post-test score on SDQ emotional problems scale	Outcome	Numeric (0-10)
sdq_prosocial_post	Post-test score on SDQ prosocial scale	Outcome	Numeric (0-10)
sdq_1_post	Post-test score on SDQ item 1: "XXX XXX"	Outcome	Numeric (0 = not true, 1 = somewhat true, 2 = certainly true)

Annex 2 – Formatting of non-mandatory variables

Table of variables that may be collected as part of evaluations, with recommended formatting.

For all outcome variables, please include both the **derived scores used in the analysis** and **all individual items** from which these scores are calculated.

Note that school URN and LAESTAB are only included in the dataset for submission to the DfE. Any school identifier should be pseudonymised in the dataset for submission to the ONS. For all outcome variables, please include both the derived scores used in the analysis and all individual items from which these scores are calculated.

Variable Abbreviation	Variable Description	Variable Type	Variable Format
cluster	Identifier if cluster randomisation used	Programme/ delivery	
pair	Identifier of pair or triplet if paired or matched randomisation used	Programme/ delivery	
urn	School's unique reference number	Programme/ delivery	Numeric (000000-999999)
laestab	School's Local Authority establishment number	Programme/ delivery	Numeric (7 digit)
sch_post	School's postcode	Programme/ delivery	Text - 'AA9A 9AA'
praction_type	Type of practitioner delivering the intervention	Programme/ delivery	Numeric (1 = teacher; 2 = therapist; 3 = social worker; 4 = other practitioner)
sessions	Number of sessions completed	Programme/ delivery	Numeric (0-999)
sessions_total	Number of sessions offered	Programme/delivery	Numeric (0-999)
sdq_impact_pre	SDQ pre-test impact score	Outcome	Numeric (0-10 parent and self-report; 0-6 teacher report)
sdq_impact_post	SDQ post-test impact score	Outcome	Numeric (0-10 parent and self-report; 0-6 teacher report)
pbfs_tot_pre	Total Problem Behaviour Frequency Scale pre-test score	Outcome	

Variable Abbreviation	Variable Description	Variable Type	Variable Format
pbfs_tot_post	Total Problem Behaviour Frequency Scale post-test score	Outcome	
srds_ver	Version of the Self-Report Delinquency Scale used	Outcome	
srds_variety_pre	SRDS variety of delinquency pre-test score	Outcome	Numeric (0-19)
srds_volume_pre	SRDS volume of delinquency pre-test score	Outcome	Numeric (0-11)
srds_variety_post	SRDS variety of delinquency post-test score	Outcome	Numeric (0-19)
srds_volume_post	SRDS volume of delinquency post-test score	Outcome	Numeric (0-11)
police_cont_6mnth	Any police contact in the past 6 months	Outcome	Numeric (1 = yes; 0 = no; 99 = unknown)
police_cont_12mnth	Any police contact in the past 12 months	Outcome	Numeric (1 = yes; 0 = no; 99 = unknown)
police_cont_post	Any police contact at follow-up	Outcome	Numeric (1 = yes; 0 = no; 99 = unknown)
type_police_contact	Type of police contact	Outcome	Text (1 = stop & search, 2 = ….)
type_last_off	Type of last offence	Outcome	Text (1 = offence type 1, 2= ….)
date_last_off	Date of last offence	Outcome	Text (DD/MM/YY)
gender	Gender	Background/ characteristic	Numeric (1 = male; 0 = female)
fsmeligible	Child recorded as eligible for Free School Meals	Background/ characteristic	Numeric (1 = true; 0 = false)
ethnicity	Child's ethnic identity	Background/ characteristic	Text: ABAN = Bangladeshi AIND = Indian ….
send	Special educational need and disability status	Background/ characteristic	Numeric: (1 = yes 0 = no 99 = unknown)
eal	English as an additional language	Background/ characteristic	Numeric: (1 = yes 0 = no 99 = unknown)

Annex 3 – Variables not permitted in the ONS Secure Research Service

Please see below the list of variables that must be excluded from the dataset for submission to the ONS. If evaluators have collected a school identifier, this should be pseudonymised in the dataset for submission to the ONS.

- Postcode
- Day of Birth (natural persons only)
- School/ Academy Names
- URN (School's unique reference number)
- UKPRN (UK Provider Reference Number)
- LAESTAB (School's Local Authority establishment number)
- Any free text variables

The direct identifiers below are also not permitted in the SRS:

Direct identifier	Description
Name	Person or company name
Street address	Person or company address
Email	Personal or company email address
NI number (NINO)	National Insurance Number
NHS number	NHS number
Phone number	Phone number
Grid	Grid location of the property
Eastings	One of the two parts of a grid
Northings	One of the two parts of a grid
UPRN	Unique Property Reference Number
CRN	Company Registration Number
UPN	Unique Pupil Number
Reference Case	Any reference to a person or organisation that can be traced using publicly available data
VAT registration	HMRC reference for companies using VAT
Passport ID	Personal passport number
Wow/ wowent/ wowref/ wowentref	IDBR business identifier

Annex 4 – Data dictionary checklist

Below is a compilation of feedback we've previously provided to evaluators when reviewing data dictionaries, and should help minimise any revisions to the dictionary and dataset for submission to the ONS.

The data dictionary template will be provided to evaluators by their YEF evaluation manager.

	Description	✓ / ✗
	Dataset details sheet	
1	The description of the dataset is clear and informative	
2	The number of variables matches the number of variables listed in the 'Data Dictionary' sheet	
3	It is clear how the 'Number of observations' was derived, and this is consistent with the number of CYP randomised in the final report • The data should be submitted for all CYP who were randomised (minus those who withdrew consent, if applicable), not only those who have complete data (i.e. filled out both baseline and endline surveys) • If 'Number of observations' is different from the number of CYP randomised as reported in final report, please describe this difference and provide a rationale in cell C37 "additional comments on changes"	
	Data dictionary sheet	
4	All variables from the evaluation have been included, including all variables used in the primary and secondary outcome analyses	
5	All relevant fields have been completed for each variable	
6	Variable names and definitions (and notes where provided) are intuitive so readers can understand what they mean	
7	The 'data type' column has been completed accurately, following definitions in the 'Reference' tab	
8	The 'values' column has been filled out and very clearly describes how variables have been coded. For variables listed in Annex 1 and Annex 2, the coding provided in the guidance should be followed. Examples: • group: 1=treatment, 0=control • gender: 1=male; 0=female • "The child exhibits behaviour X": Values range from 1=strongly disagree to 7=strongly agree	
9	Where ethnicity is provided, this has been labelled as sensitive (see ICO guidance)	
10	If there were assumptions used in the construction/ derivation of a variable, these have been described in the 'Additional Notes' column	

	Description	✓ / ✗
11	All mandatory variables from Annex 1 have been included, in the format required.	
12	For questionnaires, in addition to the derived subscale and total scores, individual item scores/ responses have been provided	
13	No identifying variables (names, dates of birth, addresses, postcodes, etc.) have been included in the data dictionary • The dataset for ONS should not contain any identifying data – only the dataset for DfE will have these • If school Unique Reference Number (URN) or LAESTAB is available, please pseudonymise this in the dataset for ONS (for future replication and reanalysis). Include the actual URN and LAESTAB in the dataset for DfE	
14	No free text variables have been included	



youthendowmentfund.org.uk

hello@youthendowmentfund.org.uk

@YouthEndowFund

Youth Endowment Fund

This document was last updated in **October 2026.**

We reserve the right to modify the guidance at any time, without prior notice.

The Youth Endowment Fund Charitable Trust

Registered Charity Number: 1185413