



Department
for Education



Office for
National Statistics

Accessing LEO data via the ONS SRS Frequently Asked Questions (v1.3)

November 2023

Introduction

Summary

This publication provides additional non-statutory guidance from the Department for Education. It has been produced to answer Frequently Asked Questions regarding access to the Longitudinal Education Outcomes (LEO) standard extract via the ONS Secure Research Service. The main initial points of guidance can be accessed [here](#).

Who is this publication for?

This guidance is for researchers who wish to apply for access to variables within the LEO standard extract via the ONS SRS.

FAQ Contents

The FAQs are broken down into two sections below – questions about applying for access and using the service in the first section, then questions about the data that can be accessed through the service in the second section.

I have a question that has not been answered below

DfE and ONS will work to update this document frequently, adding further information and answers to questions that have been asked by researchers.

If you have a question that has not been answered below, please contact the SRS customer support team (srs.customer.support@ons.gov.uk) and we will work to answer it in the next version of the FAQ document.

Application & service FAQs

Who can apply to access data in the LEO via SRS Service?

Researchers who are fully [accredited under the Digital Economy Act \(2017\)](#) can apply for access. Each researcher accessing data within the service requires this accreditation.

Researchers who were ONS approved researchers under the SRSA can still apply for LEO and will be recognised as an accredited researcher under the DEA (2017).

What level of skill do I or my team require to be able to work with this data?

Research teams will need to possess strong Structured Query Language (SQL) skills to be able to work effectively with the extract.

Am I able to store LEO data as a csv or txt file(s) within the SRS so that I can open it using STATA software?

Regarding storing the files as CSVs – unfortunately, this is not possible due to the size of the data. However, it is possible to connect to SQL server from STATA, and to work with the data using STATA.

What does the application process look like?

A summary can be found on the DfE guidance page [here](#), the following text provides additional detail.

Once you have submitted your application on the RAS, the following steps will take place:

Stage 1:

1. ONS carries out a first check of your project to flag up any immediate and obvious issues and to ensure that it is ready to be processed.
 - This may result in your application being queried with you and a member of ONS SRS Project Applications team will work with you to update your application.
2. ONS carries out an in-depth feasibility check of your project. This includes elements such as ensuring that the purpose, method and data required is proportionate to your research aims.
 - This may result in your application being queried with you and a member of ONS SRS Project Applications team will work with you to update your application.
3. An ethical review will be carried out by the UKSA's Centre for Applied Data Ethics (CADE).

Stage 2:

4. When ONS checks are complete your application will be reviewed by the DfE, HMRC and DWP who will consent or reject the use of LEO for your project.

If all data owners approve the use of LEO for your project the application will proceed to the next stage.

Stage 3:

5. The project will be reviewed by the Research Accreditation Panel. Apart from full accreditation, projects can be accredited with minor (advice or feedback which does not delay access to a project) or major (to be addressed before data access is given) comments requiring a revision of the application.
6. If approved, RAP will notify the ONS SRS Project Applications team who will contact you to complete the above forms to sign and setup your project area with LEO data.

What do I need to submit to ONS to apply for access?

You will be required to complete the following to apply to access Longitudinal Education Outcomes (LEO) data in the ONS Secure Research Service (SRS):

- Main project application form (completed via [ONS RAS](#))
- Guidance on project applications and an exemplar application can be found [here](#)
- [NSDEC ethics self-assessment tool](#) (and/or equivalent institutional ethical approval)
Note: if you have received institutional ethics approval for your project, please attach this as part of your application, but you may also be required to fill in the NSDEC ethics self-assessment tool if your current ethics approval does not cover in detail access to secure data in a secure research environment
- LEO I2SE Variable Request Form – to select the variables required for your project. All researchers on the project must be fully accredited by the UKSA.

If a project is approved each researcher associated with your project is required to sign the following data access agreements:

- Office for National Statistics Secure Research Service Accredited Researcher Assurance Registration form
- Department for Education's Individual Declaration form.

How can I seek support with my application and/or ask questions about the service itself?

To start an application or apply to be an accredited researcher please visit the ONS [Researcher Accreditation Service \(RAS\)](#) and their FAQs page [here](#).

If you require support with your application and questions about the service, please contact srs.customer.support@ons.gov.uk.

The LEO research community is hosted on Knowledge Hub and allows researchers to access relevant information and to ask questions of each other regarding LEO analysis on the SRS. The community also helps DfE and ONS to make the LEO service as useful and

accessible as possible and to develop the LEO knowledge base and is a platform for DfE to post announcements about the LEO via SRS service.

How do I join the LEO Research Community?

To access the community you require a [Knowledge Hub \(KHUB\) account](#) and UKSA researcher [Accreditation](#) (see “Applying for Accredited Researcher status” section).

To gain access once you have a KHUB account and accreditation, please submit an access request to the community [here](#) and send an email to DfE’s LEO mailbox (leo.programme@education.gov.uk) with your UKSA accreditation number. Your request to join the community will then be approved.

Where can I find details of which variables are included in the LEO standard extract?

The LEO Standard Extract Variable Request Form contains information about all the data available to request as part of the LEO via ONS SRS service. This can be found in Related Downloads on the ONS website [here](#).

Can I add further data to my selection from the LEO standard extract?

Additional non-person level data can be specified in your application form or brought in later through ONS’ ingestion process (by submitting an ingest form available [here](#)). Where this is approved, ONS will retrieve the data you have provided it, check it according to the ingestion process, and then make it available with the research project environment.

Can I add organisation-level data to my project area?

An update to the second iteration of the LEO via ONS SRS service is that pseudonymised organisation-level data can now be linked into the LEO Standard Extract. This data must be pseudonymised before it can be linked. This process will be done by ONS before making it available for you in your LEO project area in the SRS. If you require any further information about this process please contact srs.customer.support@ons.gov.uk.

Is it possible to link bespoke data to LEO?

At this stage, we cannot definitely say it will be feasible however this is something we are working towards. Currently, linking LEO to other person-level data is not permitted.

Can I access education data on its own?

If you require access to only the NPD, LILR and/or HESA datasets and do not require any of the LEO data tables please contact data.sharing@education.gov.uk for information on how to access these datasets.

I already have a project I am working on using the ONS SRS. Can I apply to add LEO to this project?

Linking to other person-level data is not currently permitted in the LEO via ONS SRS service. If you wish to access LEO data, you must make a new application and therefore would have a new project area on the SRS.

When completing the Variable Request form, how much justification is required in the Data Tables tab? Should justification be provided for each variable, or will more general justification suffice?

When selecting the data tables on the Variable Request Form you will also need to ensure that you justify how the data table and variables contained within will help to answer your research questions. For each table, justification of at least a short paragraph should be provided.

Do I need to publish my research findings?

Yes, researchers are required to commit to publishing their findings as a condition of accessing the service.

While working on an approved LEO project on the SRS, do I need to notify the DfE prior to publishing research outcomes?

Yes, once ONS have cleared an output for publication, the researcher is responsible for providing copy and notification of all publications to DfE's data.sharing@education.gov.uk mailbox. This notification must be given at least 72 hours ahead of the planned publication time.

Can I access LEO from home?

There is no home access permitted for LEO projects in the SRS. Currently, accessing LEO via the SRS is restricted to ONS Safe Rooms or through an organisation which has an Assured Organisational Connectivity Agreement (AOC) with ONS, and which maintains a current certification.

How long does an application take?

Depending on a number of factors (see below), we expect that it will typically take approximately 4-8 weeks between an application being submitted and researchers with successful applications having access to their selection of data on the SRS.

The process may take longer than this and it is important that applicants are aware of the following factors, amongst others, which may affect the length of time required.

- Requirement for full accreditation of researchers
- Familiarity of researchers with the application documentation
- Strength of the ethical considerations in research proposals
- Responsiveness of applicants in relation to queries regarding the application
- Complexity of request

Researchers are advised to use the [project application example guidance](#) and an [exemplar of a research project application](#) to understand the information that is required for a research project application to be considered. The better the quality of the initial application, the faster ONS will be able to progress it to the Research Accreditation Panel (RAP).

Does the LEO via ONS SRS Service allow exploratory projects?

While exploratory projects are allowed in the LEO via ONS SRS service, please note that in order to meet Digital Economy Act 2017 requirements the project must result in the publication and dissemination of research outcomes. Exploratory projects are permitted, however close attention must be paid to this requirement and the application must be clear that this requirement is met.

How can I see what other researchers and organisations are doing with LEO data?

You can find a list of DfE publications which will be of interest and give you a sense of the focus in DfE in the community library held in the LEO Research Community. The list also includes research carried out under secondment to DfE (the LEO Secondee Pilot) and outside of DfE too, so you can see what other teams have been working on as well. Equally useful, there is a list of all of the LEO projects on the SRS available at the UKSA website [here](#) (search "LEO" in the search box).

How would I go about making a case for an extension to LEO, or possibly another linked data, which focused on linking student records to HMRC data?

Please share feedback and ideas in the LEO research community (details on how to join and access the community described above). You can also e-mail the LEO Programme mailbox at LEO.PROGRAMME@education.gov.uk .

Data FAQs

What data is included in the LEO Standard Extract?

The LEO standard extract connects individuals' education data with their employment, benefits, and earnings data to provide a de-identified longitudinal person level administrative research dataset. The LEO Standard Extract is a relational database, comprising a range of tables where access via the ONS SRS is restricted on a needs basis.

This is a relational database comprising four distinct sets of data:

- National Pupil Database data (School and early years data)
- Individual Learner Record data (College and apprenticeship data)
- Higher Education Statistics Agency data (Higher Education data)
- LEO data (earnings, benefits and tax data).

How can I develop my understanding of the data?

The best place to find information and guidance about LEO is to join the LEO Research Community. This can be found [here](#). Please note that you must be a UKSA accredited researcher to access the community.

Additional information about the datasets can be accessed via the following links:

- Information about the data held in the National Pupil Database can be found [here](#).
- HESA/JISC provide information and user guides for Higher Education data on their website [here](#).
- The LEO user guide can be provided on request to accredited researchers, and is also available on the [LEO Research Community](#).

What is the coverage of the LEO extract? Is data for Wales, Scotland and Northern Ireland included?

The cohorts contained within the LEO labour market outcomes data are dictated by the education data which has been sent for matching.

The coverage is:

- All people born since 1985 who have engaged with the schools education system in England at least once (individuals located within the NPD); and
- People who are older than this and have been in Higher Education (HE) at a UK institution at least once since 1994/95 (individuals located within the HESA data) or English Further Education (FE) institution since 2002/03 (individuals located within the ILR data). Please note that HE institutions in Northern Ireland are not included in this data sharing so students only attending these institutions will not be included.

Can I access the relevant metadata while working with the data on the SRS?

Yes, the metadata for the LEO, NPD & ILR data will be available in the research areas for approved projects.

There is currently no HESA metadata available in ONS SRS, however users can request for this to be ingested following ONS' usual data ingestion process.

What geographical data is included in the LEO standard extract?

The LEO geography table uses residential address information from DWP's Customer Information System (CIS) and covers the period 04/05 – 18/19 for all individuals included in the LEO lookup dataset. Census Output Areas (COAs) are linked to individuals via postcode and higher-level geographies are added from the National Statistics postcode look-up.

The LEO standard extract also includes education data sources: NPD, ILR and HESA. These different data sources include their own geographic variables relating to individuals and institutions. The coverage and scale vary as these are disparate admin datasets which have been collected together for use with the LEO outcomes data. We would recommend that users review the supporting information available for these data sources to understand the geographic data included.

The ONS RAS online application specifies that only Local Authority (LA) and Regions are available in LEO. However, the LEO Variable Request Form allows users to request variables containing information about the Lower Super Output Area (LSOA). Could you please advise us on whether we can select the LSOA variables in the Request form?

There is LSOA information available in some of the NPD tables and HESA. The LEO outcomes geography information does not cover LSOA, and only includes the variables in the Geography table. So, the result will be that LSOA is not available for everyone on the LEO spine.

Is employer a field in LEO?

While there is not an employer field in LEO, so employers aren't able to be identified, there is IDBR (inter-departmental business register) data. This will enable researchers to identify the industry a person works in.

With the LEO dataset do you include variables such as industry of work so that we could filter on work relevant projects?

The Inter-Departmental Business Register (IDBR) is included in the second iteration of the LEO Standard Extract, and this enables individuals to be located within businesses. It can identify those who work together and can enable analysis of links between education and training and business outcomes.

Will the SIC codes be available in LEO?

SIC codes are available in the second iteration of the LEO Standard Extract. The link from IDBR to individuals in LEO is made at the enterprise level only via PAYE employer reference.

We are unable to link to individuals at local unit level with the result that sic codes represent the primary industry of the enterprise as a whole. For each tax year we give the industry sector for which the person earned the most money within the year.

Is it possible to identify siblings in the LEO data?

No, it is not possible to identify siblings.

In the LEO variable definitions file there is a School unique reference number variable, which we would like to use to get further data on schools. Our question is whether this variable within LEO is anonymized or is the same unique reference number that one can use to find schools on the Ofsted site, for example? Would this kind of linking by ourselves be permissible under the current access and safety standards?

Ofsted data can be brought into use with LEO; however, this must be in line with Digital Economy Act regulations and schools cannot be identified. As part of the newest iteration of the LEO via SRS service, organisation-level data can be ingested into project areas. Accepted requests for data will include a process whereby ONS will pseudonymise organisation-level data, so pseudonymised organisation-level data can now be accessed alongside LEO data.

I was wondering if the LEO data available in the SRS includes Scottish LEO data? Or is it just England only? If not, are there any plans to include Scottish LEO data on the SRS?

The LEO data available in the ONS SRS covers English learners only. The Scottish government are working on their own version of LEO, however it is not yet confirmed when this will be available for public access.

Can you determine if a student is studying for a doctorate in the LEO data?

Yes – using the HESA variable COURSEAIM = D00.

Can you determine if a student is UKRI funded?

Yes – using the HESA variable RCSTDNT.

I am looking at the variable request sheets of the employment and earnings data, and wonder if there would be information/variables on location of a company a

learner works for? I was aware that there is a residential location but did not see a work location there.

There is no information about the location of a company an individual works for in the LEO data.

Can ILR data be linked to LEO data and, if so, what variable do I need to ensure I have in the ILR data to do this?

The LEO Standard Extract contains Longitudinal ILR data, and you can select which variables you wish to access. These can be linked to the LEO Data by using the Education Records Lookups, which are included as part of the standard extract. The link is achieved via a multi-part join, please see the LEO Data User Guide for full details.

We are focusing on and identifying the impact of having children on life chances in the first instance, are there fields including child tax credits and investigate both the impact of having a child on earnings and outcomes, and the impact of flexible working on earnings and outcomes?

The LEO dataset does not contain child benefit or any indicator that a person has had a child. There is a developing dataset which is being developed in DfE called the Pupil Parent Matched Dataset (PPMD) which will link parents' tax records to their children's education outcomes. This will enable much richer analysis of the links between the circumstances in which individuals grew up and their circumstances later in life, providing a detailed picture of social mobility both at national and local levels. However, we need to wait for a prolonged period (multiple years) before reliable analysis against LEO subjects could be undertaken due to the lead times/longitudinal nature of LEO and the present nature of PPMD.

Why are HESA counts for graduates so much larger than those found in the LEO data?

The NINO is not present on the HESA student record itself and has been matched on where possible by fuzzy matching with personal data from the Student Loans Company. This process increases the likelihood of finding a match with DWP's CIS (Customer Information System). Accordingly, groups less likely to take a student loan, for example international students who are not eligible for one, are likely to have lower match rates. More information about the methodology can be found [here](#).



Department
for Education

© Crown copyright 2023

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. To view this licence, visit nationalarchives.gov.uk/doc/open-government-licence/version/3.

Where we have identified any third-party copyright information you will need to obtain permission from the copyright holders concerned.

About this publication:

enquiries www.education.gov.uk/contactus

download www.gov.uk/government/publications



Follow us on Twitter:
[@educationgovuk](https://twitter.com/educationgovuk)



Like us on Facebook:
facebook.com/educationgovuk