

Next Steps

Linked Student Loans Company
administrative datasets

User Guide (2nd Edition, October 2021)

Contact

Data queries should be sent to: help@ukdataservice.ac.uk

Questions and feedback about this user guide should be sent to clsfeedback@ucl.ac.uk.

How to cite this guide

Rihal, S., Gomes, D., Henderson, M. (2021) *Next Steps: Linked Student Loans Company administrative datasets User Guide (2nd edition, October 2021)*. London: UCL Centre for Longitudinal Studies.

Acknowledgements

CLS would like to thank the Student Loans Company for the provision of the linked data.

You should also acknowledge CLS following the guidance from <https://cls.ucl.ac.uk/data-access-training/citing-our-data/>

This guide was first published in September 2021 by the UCL Centre for Longitudinal Studies (CLS).

The UCL Centre for Longitudinal Studies (CLS) is an Economic and Social Research Council (ESRC) Resource Centre based at the UCL Institute of Education (IOE), University College London. It manages four internationally renowned cohort studies: the 1958 National Child Development Study, the 1970 British Cohort Study, Next Steps, and the Millennium Cohort Study. For more information, visit www.cls.ucl.ac.uk.

This document is available in alternative formats. Please contact the Centre for Longitudinal Studies:

tel: +44 (0)20 7612 6875

email: clsfeedback@ucl.ac.uk

Contents

About the Next Steps Cohort Study	2
1. Introduction	3
2. Education data linkage	3
2.1 Consent to education data linkage.....	3
2.2 Linkage to Student Loans Company (SLC).....	4
3. Linked Research Data.....	6
3.1. Licensing and data access.....	6
3.2 Available datasets.....	6
3.3. Identifiers	7
3.4. Applicant data description	7
3.5. Payments data description.....	9
3.6. Repayments data description.....	10
3.7. Overseas data description	11
4. Standard Disclosure Control.....	12
4.1. UKDS requirements	12
4.2. SLC policies and procedures	12
Appendix A: HEI Tuition Fee Types.....	13

About the Next Steps Cohort Study

Next Steps is a longitudinal cohort study, following a nationally representative group of nearly 16,000 people born in England in 1989-90. The study began when cohort members were 14 years old. With sweeps every year for the first seven years, it has captured incredibly rich information about their educational trajectories during adolescence and early adulthood.

Today, Next Steps continues to chart this cohort's experiences with a wider disciplinary scope, providing unparalleled insight into the many different aspects of their lives and transitions into adulthood.

A vital source of evidence, Next Steps has had a major influence on national education policy and cast light on a wide range of important social issues, including the effects of zero hours contracts and bullying.

1. Introduction

This guide describes the data linkage of education administrative records from the Student Loans Company (SLC) to survey data for cohort members in the Next Steps cohort study. The main aim of this data linkage exercise is to enhance the research potential of the study, by combining administrative record with the rich information collected in the surveys.

2. Education data linkage

2.1 Consent to education data linkage

In sweep 8, Age 25 survey, data linkage was a major part of the study, with 10 different linkage consents collected covering education, health, economics, and criminal justice. One of these consents was linkage to Student Loans Company data.

Considerable effort was expended in developing an approach that would maximise consent, particularly in the context of a mixed-mode survey, these are noted below:

A data linkage leaflet was included in the advance mailing, sent at the start of each batch of fieldwork. It gave information on the purpose, types, value, and process of data linkage, and encouraged study members to contact the study team with any questions they might have.

During the interview, and following an introduction page, consents were recorded electronically directly into the survey instrument. All participants were then sent a confirmation of their consents (as part of their 'Thank you' letter) by post or email.

More detailed information on the fieldwork and consent collection can be found in the Next Steps Age 25 Technical report and Next Steps Age 25 User Guide. All documents can be found under 'documentation' at <https://cls.ucl.ac.uk/cls-studies/next-steps/next-steps-age-25-%20sweep/>

2.2 Linkage to Student Loans Company (SLC)

The SLC is a non-profit making government-owned organisation that administers loans and grants to students in colleges and universities in the UK. Further details about the SLC can be found at

<https://www.gov.uk/government/organisations/student-loans-company>.

Linked SLC records include data on higher education loans. SLC matched the participants in the Student Finance Database which holds details on the applicant, payments, and repayments data.

The linked data does not include data on Further Education (Education Maintenance Allowance and Advanced Learner Loan) and Post Graduate Loans.

In 2019 CLS made a request to SLC to link all Next Steps participants to their student loan records in the Student Finance Database. A total of 4501 participants agreed to SLC linkage out of 7,707 who took part in sweep 8 of the survey, corresponding to a consent rate of 58.4%

A total of 4,501 cohort members consented to the linkage of the Student Loans Company data. Of those who consented to the linkage, 3,622 had studied at university by age 25 and therefore were likely to have taken out a loan (although some may have taken out a loan and studied in a non-university setting). Of those who consented and studied at university, 2,401 Next Steps participants were successfully matched to the SLC data, corresponding to a linkage rate of 66.3%

Table 1 below shows the number of successful matches to SLC records following data linkage.

The matching was completed by the SLC in June 2021. The personal identifiers supplied by CLS for matching against the SLC were: CLS identifier, anonymised Pupil Matching Reference (PMR), National Insurance number (NINO), first name, middle name, surname, title, other surname, date of birth, sex, current address and postcodes, town, county, country, address date.

Table 1: Participation, consent, and linkage

	Age 25 Sweep
Number of participants in sweep	7,707
Number with valid consent	4,501
Consent rate	58.4%
Number with linked SLC data	2,401
Number studied at university (by Age 25)	3,622
Linkage rate	66.3%

The data was matched by SLC following the criteria below:

Match 1: NINO and sex and date of birth

Match 2: NINO and sex and surname

Match 3: Surname and first name and sex and date of birth

There was no fuzzy matching¹ - the applicant record was matched to one of 3 above.

¹ Fuzzy matching is an inexact string-matching technique which allows the matching of an individual identifier despite small differences between two sources. For example, a person may have their name recorded as 'Elisabeth' in one dataset but 'Elizabeth' in the other. With a series of matching criteria looking at first name, surname and date of birth, the fuzzy matching method achieves a link despite the slight difference in first name

3. Linked Research Data

3.1. Licensing and data access

All datasets are available from the UK Data Service (UKDS).

All users of the data need to be registered with the UKDS. Details of how to do this are available at <https://www.ukdataservice.ac.uk/get-data/how-to-access/registration>.

The linked SLC data have been processed by CLS and supplied to the UK Data Service (UKDS) under Secure Access Licence. Applicants wishing to access this data need to establish the necessary agreement with the UKDS and abide by the terms and conditions of the UKDS Secure Access licence.

For details on Disclosure Control, please refer to section 4 of this document.

3.2 Available datasets

The following data was provided by the SLC:

1. SLC applicant
2. SLC payments
3. SLC repayments
4. SLC overseas assessment

All datasets are hierarchical in structure, that is, at least one case exists per cohort member (CM). The available datasets and a broad description of what is covered by each dataset is described in Table 2 and sections below.

Table 2: List of available datasets

Name of the dataset	Content summary
NS_SLC_CM_Applicant_2007_2020	SLC data on cohort member's application for student finance between academic years 2007 and 2020
NS_SLC_Payments_2007_2021	SLC data on payment transactions made to cohort member between financial years 2007 and 2021.

Name of the dataset	Content summary
NS_SLC_Repayments_2009_2021	SLC data on cohort member's repayment transactions between financial years 2009 and 2021.
NS_SLC_Overseas_2012_2021	SLC data on overseas assessment for cohort member between 2007 and 2020

3.3. Identifiers

NSID is the anonymised unique cohort member identifier which can be used to merge this data and other deposited Next Steps datasets.

3.4. Applicant data description

- Records are distinct for the applicant and include academic year, and study mode (full time undergraduate, part time undergraduate and postgraduate).
- All applicants have applied for student finance, the application may or may not have been approved and paid
- 2,218 unique applicants

Table 3: Variables in the NS_SLC_CM_Applicant_2007_2020 dataset

Variable Name	Description
NSID	NSID – Anonymised cohort member identifier
APPLICATION_TYPE	Indicated the support type applied for, FT Full Time, PT, Part Time
ACADEMIC_YEAR	The Academic Year of the applicant application.
COUNTRY_OF_STUDY	The Country of the Award Authority under which the applicant is studying.
DOMICILE_OF_STUDENT	The body (country) responsible for paying the student support. England/Scotland/Wales/Northern Ireland/Europe values E/S/W with NA for Advanced Learner Loans. For EU domiciled students this will be the country of study.
STUDENT_COHORT_CATEGORY	Indicates the entry cohort rules under which student finance is awarded to this applicant for this study mode in this year.

Variable Name	Description
DEPENDENCY_STATUS	An indicator (“D/ependent” or “I/ndependent”) of the applicant’s dependency status. There is a list of circumstances, any one of which would class an applicant as Independent. An applicant classed as Independent would be means tested on their (and maybe a partner’s) income rather than their parents’.
INCOME_INDICATOR	For applicants with Household Residual Income recorded, an indicator of whether it was their income or their parents’ that was assessed. “A” for applicant’s (and/or partner’s, spouse’s, step-father/mother’s), “P” for parent’s (mother/father only). Null if no HRI recorded. The value will default to an “A” if no parental sponsor is found.
HOUSEHOLD_INCOME_BEFORE_DISREGARDS	The gross income of the household before any “disregard” is taken into account.
HRI_HOUSEHOLD_RESIDUAL_INCOME	This is the income used to determine means tested support. It is the income after “disregards” of those people on whom the applicant is considered to be dependent. They are called “sponsors” and are normally parents or guardians but could instead be partners / spouses. Not all applicants will supply income data. A proportion of applicants will choose to be non means tested (i.e., just apply for a fee loan and the non means tested element of the maintenance loan). In addition to this, if the applicant is a returning student, then they do not necessarily have to supply income. They may choose to self-certify their income has not changed in the last year and therefore income data will not be held against the current academic year application. Instead, the rules engine will refer to the previous year’s income data.
HEI_NAME	The name of institution at which the applicant applied to attend.
HEI_TUITION_FEE_TYPE	The type of HEI, for example Public (PU), Private (PI) see Appendix A for details.
COURSE_CODE	The code of the course which the applicant applied to study
COURSE_NAME	The name of the course which the applicant applied to study

Variable Name	Description
NHS_BURSARY_IND	Y/N. Indicates if the student is in receipt of an NHS bursary (which limits SLC support available).
WITHDRAWN_DATE	The date on which the application moved to the withdrawn work stage.
YEAR_OF_STUDY	The year of study the applicant has applied for.
LOCATION_RATE_TERM_1	Indicates the loan rate applied to the applicant assessment based on living location. Parental Home, Elsewhere, Elsewhere London, Abroad or Unknown
LOCATION_RATE_TERM_2	Indicates the loan rate applied to the applicant assessment based on living location. Parental Home, Elsewhere, Elsewhere London, Abroad or Unknown
LOCATION_RATE_TERM_3	Indicates the loan rate applied to the applicant assessment based on living location. Parental Home, Elsewhere, Elsewhere London, Abroad or Unknown
MT_ASSESS_AMOUNT	The Means Tested element of the Maintenance loan
NMT_ASSESS_AMOUNT	The Non-Means Tested element of the Maintenance Loan
TUITION_FEE_LOAN	Tuition Fee Loan Entitlement Amount

3.5. Payments data description

- Represents the payment transactions posted to the applicants' accounts.
- Distinct per product type/financial year.
- Payments incurred across all financial years and academic years to date are included up to the end of April of 2020.
- All payments for loan products, excluding non-repayable allowances and grants.
- The financial year is based on the transaction date of the payment. For some-one starting in September 2012, payments where the transaction date is March 2013 or earlier would be classed as financial year 2012(/13) and payments between April 2013 and March 2014 as financial year 2013(/14).
- Any reversal transactions where the payment has for any reason been reversed are included.

- 2,124 unique applicants

Table 4: Variables in the NS_SLC_CM_Payments_2007_2021 dataset

Variable Name	Description
NSID	NSID – Anonymised cohort member identifier
FY_END	The financial year, starting 1 st April, to which the payments relate. Format is YYYY to refer to the end of the financial year. For example, the 2019-20 financial year is written as 2020.
PRODUCT	The product paid. FL – Tuition Fee Loan, SS -Student Support Loan, PTFL – Part Time Tuition Fee Loan, PTML – Part Time Maintenance Loan
AMOUNT	The net value of payments posted to the applicant’s account in the financial year.

3.6. Repayments data description

- Repayments are shown against the financial year in which they were repaid – this is calculated from the transaction date of the repayment for all repayment types
- The values for Self-Assessment (SA) repayments reflect the financial year the liability was incurred (i.e., the income on which the repayment was based) and not the financial year based on transaction date. This means that repayment types are consistently reported.
- Any reversal transaction where the repayment has for any reason been reversed are included.
- 1,929 unique applicants.

Table 5: Variables in the NS_SLC_CM_Repayments_2009_2021 dataset

Variable Name	Description
NSID	NSID – Anonymised cohort member identifier
TY_END	The tax year in which the applicant made the repayment – has no bearing on when the repayment was posted to SLC accounts. The tax year officially begins on April 6th, but this report uses April 1st. In practice, this will amount to the same thing, since HMRC repayments and the application of interest are transacted on the last day of each month
PRODUCT	The product paid. FL – Tuition Fee Loan, SS -Student Support Loan, PTFL – Part Time Tuition Fee Loan, PTML – Part Time Maintenance Loan
REPAYMENT_TYPE	The type of repayment made by the applicant. PAYE, SA, Refund, Direct Voluntary UK, Direct Obligatory UK, Direct Voluntary Overseas, Direct Obligatory Overseas, RTL, Repayment of Overpayment or Undefined

3.7. Overseas data description

- Shows a history of overseas assessments for the applicant, where the applicant is in the process of repaying the loan and has had a successful overseas repayment assessment at some point
- The applicant may or may not be overseas currently.
- 117 unique applicants

Table 6: Variables in the NS_SLC_CM_Overseas_2012_2021 dataset

Variable Name	Description
NSID	NSID – Anonymised cohort member identifier
OVERSEAS_START_DATE	Date when applicant went overseas (or moved to country of residence), DD/MM/YYYY
OVERSEAS_END_DATE	Date when applicant returned from overseas (or moved out of country of residence) If this field is blank the customer is still
COUNTRY_OF_RESIDENCE	Overseas Country where matched on 3-digit ISO code

4. Standard Disclosure Control

4.1. UKDS requirements

Data included under **Next Steps: Linked Administrative Datasets - Student Loans Company Records (SLC), 2007 – 2021** is only available via the UKDS Secure Lab. Access to this data is controlled and the UK Data Service will always perform a certain level of disclosure control on the outputs generated by researchers, as outlined in their SDC Handbook which can be downloaded from: <https://securedatagroup.org/sdc-handbook/>

The two UK Data Service Secure Lab rules of thumb that will be applied to all outputs are:

- Threshold rule: No cells should contain less than 10 observations
- Dominance rule: No observation should dominate the data to a huge extent

4.2. SLC policies and procedures

The SLC adhere to the UK Statistics Authority Code of Practice for Statistics (<https://code.statisticsauthority.gov.uk>)

Appendix A: HEI Tuition Fee Types

Table 7: Values in the HE_TUITION_FEE_TYPE variable

Category	HE_TUITION_FEE_TYPE	Description
Public	PU	Authority funded Institution
Private	PV	Specifically Designated (with or without DAPs)
Approved	AP	Providers registered in the Approved section of the OfS register
Approved (fee cap) with an APP	AA	Providers registered in the Approved (fee cap) section of the OfS register
Approved (fee cap) without an APP	AW	Non-registered providers on behalf of the Approved (fee cap) or Approved providers (Franchisee)