2017-18 ILR Data checking tool: Student characteristics data summary technical document

Enquiries to ILRData@officeforstudents.org.uk

Publication date: 01 August 2018

Contents

Purpose

1. This document describes the algorithms used to generate the data needed to re-create the figures in the Student Characteristics workbook from ILR data.

2. The student characteristics data summary identifies fields taken from the ILR that are used in the construction of institutional performance metrics and their associated benchmarks. These are fields that classify a student's background or other characteristics for use in the benchmarking process for quality and institutional performance measurements, including the Teaching Excellence and Student Outcomes Framework (TEF).

3. This data summary is provided to help identify potential errors and reduce the numbers of unknown or unpopulated student characteristics in ILR data that will affect our uses of the data. Figures from 2016-17 ILR were calculated on the same basis as the latest year and are included in the workbook file to illustrate year on year changes in absolute numbers and proportions.

4. This document also contains a description of the algorithms used to create the populations for two of the metrics used on the Unistats website: Unistats entry population and Unistats classification of degree population (see paragraphs 34-40). Along with the other student characteristic information described below, you can use these population markers and the algorithms provided to preview a close approximation of these Unistats metrics for any relevant courses submitted. These markers have been included in this output to reduce the number of different outputs created for each provider making it easier to find useful information.

5. This document is aimed at readers with in-depth knowledge of the data. Readers are advised to have a copy of 'Specification of the Individualised Learner Record for 2017 to 2018' (available from the Education & Skills Funding Agency via https://www.gov.uk/government/publications/ilr-specification-validation-rules-and-appendices-2017-to-2018) to hand when using this document.

Student characteristics (SC) data summary workbook

6. The Student Characteristics workbook can be accessed from the OfS portal. The Excel workbook SC17_DCT_FXXXXXXXX.xlsx (where XXXXXXXX is the UK Provider Reference Number (UKPRN) for the provider) contains the following worksheets:

Table 1 Excel workbook 'SC17_DCT_FXXXXXXXX.xlsx'

Worksheet*	Title
Coversheet	Title page
Age	Student numbers in each age category
Degree class	Student numbers in each degree class
Disability	Student numbers in each disability category
Domicile	Student numbers in each domicile category
Ethnicity	Student numbers in each ethnicity category
Entry qualifications	Student numbers in each entry qualification category
Level of study	Student numbers at each level of study
Postcode	Student numbers with known and unknown postcodes
Subject	Student numbers in each subject category

* This worksheet reference corresponds to the spreadsheet tabs.

Using the individualised file

7. Full details of how to access this file are given on the OfS website (www.officeforstudents.org.uk/data-and-analysis/supplying-data/working-with-individualised-files/).

8. When working through this document it is necessary to use the individualised file, SC17_DCT_FXXXXXXXX_IND.csv, where XXXXXXXX is the UKPRN for the provider. This will show the allocation of students to cells within the tables. Full details of how to access this file are given on the OfS website (www.officeforstudents.org.uk/data-and-analysis/supplying-data/working-with-individualised-files/).

9. The individualised files each contain one record for each instance of HE level study in a subject area in the latest academic year. For example, a student who is studying for a first degree in biology in 2017-18 will have one record for that instance and will count as one full-person equivalent (FPE, a headcount measure). A student who is studying a joint course first degree with equal amounts of mathematics and physics in 2017-18 will have two records for that instance; one for each subject area, with each assigned an FPE value of a half (their single headcount for the instance being apportioned across the subject areas according to the proportion of the course that relates to each subject they are studying).

ILR fields used in the student characteristics tables

10. Only certain fields, detailed in Table 2, were used to generate the student characteristic figures. Fields taken from the ILR return or derived as part of the comparison tables are shown in capitals using the names given in Tables 2 and 3 respectively.

Table 2 Fields used in the student characteristics tables

Name	Description	Dataset
UKPRN†	UK provider reference number	ILR
LEARNREFNUMBER†	Learner reference number	ILR
AIMSEQNUMBER†	Learning aim data set sequence	ILR
ULN†	Unique learner number	ILR
NUMHUS†	Student instance identifier	ILR
LEARNAIMREF	Learning aim reference	ILR
DATEOFBIRTH	Date of birth	ILR
DOMICILE	Country of domicile	ILR
ETHNICITY	Ethnicity	ILR
LEARNFAM_DLA	Learner is in receipt of disabled students allowance	ILR
LEARNSTARTDATE	Learning start date	ILR
LEARNPLANENDDATE	Learning end date	ILR
LLDDHEALTHPROB	LLDD and health problem	ILR
MODESTUD	Mode of study	ILR
OUTCOME	Indicates whether the learner achieved the learning aim	ILR
OUTGRADE	The examination grade awarded to the learner for the learning aim	ILR

Name	Description	Dataset
PCFLDCS, PCSLDCS PCTLDCS	Proportion taught in LDCS_CO1-CO3 subject	ILR
POSTCODEPRIOR	Postcode	ILR
QUALENT3	Qualification on entry	ILR
STULOAD	Student instance FTE	ILR
ENG_LEVE	England FE/HE status	LARS
LDCS_CO1, LDCS_CO2, LDCS_CO3	Learning directory classification system codes	LARS
QUAL_TIT†	Learning aim title	LARS
QUAL_TYP	Learning aim type	LARS

† These fields are not used in the student characteristics calculations but are included in the individualised file to allow easy identification of students.

Derived fields used to recreate student characteristic tables

11. Here we give details of the derived fields in the individualised file. These fields are used to build the key dimensions of the Student Characteristics tables.

Table 3 Student characteristics derived fields

Derived field name	Description	Paragraph
HEFQAIM†	Recognised HE qualification aim	12
SCCRSELGTH	Expected duration of study	13
SCMODE	Mode of study	14
SCLEVEL_DETAIL	Level of study	15
SCLEVEL	Level of study, broad categories	16
SCPOP	Inclusion in population for student characteristics	17
SCYEAR_ST	Academic start year	18
SCAGE	Age	19
SCDEGCLASSPOP	Inclusion in population for degree classification	20
SCDEGCLASS	Degree classification	21
SCDISABLE	Disability	22
SCDSA	Disabled students allowance	23
SCDOM	Domicile	24
SCETHNIC	Ethnicity	25
SCENTQUALGRP	Highest qualification on entry	26
SCPOSTCODE	Postcode	27
SCLDCS	Learn direct codes	28
SCJACS	JACS subject codes	29
SCFPE	Full person equivalent	30
SCSBJ_CAH1	CAH1 subject code	31
SCSBJ_CAH2	CAH2 subject code	32
SCSBJ_CAH3	CAH3 subject code	33

† The algorithms for deriving these fields are given in ‘HEIFES17 comparison algorithms’ (see the ‘HEIFES17 comparison technical document’); the paragraph references refer to this document

HEFQAIM

12. This field classifies qualification aims into broad levels of study. Note: the calculation of HEFQAIM is based on the HEIFES17 comparison technical document for the relevant year. These can be found at https://www.officeforstudents.org.uk/data-and-analysis/data-checking-tools/2017-18-ilr-data-checking-tool/.

SCCRSELGTH

13. This field contains the number of years that the qualification aim is expected to last. If a student's planned end date falls within the first 14 days of a second year of study, the number of years is taken as one. Otherwise, it is rounded up to the nearest whole number. SCCRSELGTH is the difference in years between LEARNSTARTDATE and LEARNPLANENDDATE. For example, a course with LEARNSTARTDATE = 1 October 2014 and LEARNPLANENDDATE = 5 January 2016 will have SCCRSELGTH = 2.

SCMODE

14. This field allocates students to mode of study.

Value	Description	Definition
FT	Full-time	MODESTUD = 01, 02 or (MODESTUD = 99, BLANK and (SCCRSELGTH = 1 or (SCCRSELGTH ≤ 2 and HEFQAIM = HIGHER, FIRST, FOUDEG, DIPHE, HND) or (SCCRSELGTH ≤ 3 and HEFQAIM = HIGHER, FIRST) or (SCCRSELGTH ≤ 4 and HEFQAIM = ENHANCED)))
PT	Part-time	Otherwise

SCLEVEL_DETAIL

15. This field classifies qualification aims into specific levels of study.

Value	Description	Definition

HNC	HNC	HEFQAIM = HNC
HND	HND	HEFQAIM = HND
FOU	Foundation degree	HEFQAIM = FOUDEG
UGDIP	Undergraduate diploma	HEFQAIM = DIPHE, DTLLS, DET
FDBC	Foundation degree bridging course	HEFQAIM = FDBC
OUG	Other undergraduate	HEFQAIM = CERTED, UGOTHER
DEG	First degree	HEFQAIM = FIRST
INTM	Integrated masters	HEFQAIM = ENHANCED
PGCE	PGCE	HEFQAIM = PGCE
OPGT	Other postgraduate taught	HEFQAIM = PGDIP, PGOTHER
PGTM	Postgraduate taught masters	HEFQAIM = MASTER
PHD	PhD and MPhil	HEFQAIM = HIGHER
OTHER	Other HE level qualification	ENG_LEVE = H and not above

SCLEVEL

16. This field classifies qualification aims into broad levels of study.

Value	Description	Definition
OUG	Other undergraduate	SCLEVEL_DETAIL = OUG, FOU, HND, HNC, UGDIP, FDBC

DEG	First degree	SCLEVEL_DETAIL = DEG, INTM
PG	Postgraduate	SCLEVEL_DETAIL = PHD, PGTM, PGCE, OPGT
OTHER	Other	Otherwise

SCPOP

17. This field indicates whether a student is in the student characteristics population.

Value	Description	Definition
1	In the population	SCLEVEL ≠ OTHER
0	Not in the population	Otherwise

SCYEAR_ST

18. The academic year the student entered on to the course taken from LEARNSTARTDATE. If the month of LEARNSTARTDATE is between January and July, then SCYEAR_ST is the year prior to the calendar year of entry.

SCAGE

19. This field indicates the age category of the student at 30 September in the year they commence their studies.

Value	Description	Definition
UNKNOWN	Unknown	DATEOFBIRTH = BLANK or Year of DATEOFBIRTH = 9999 or DATEOFBIRTH > 30 September SCYEAR_ST - 10
U21	Under 21 on entry	DATEOFBIRTH > 30 September SCYEAR_ST - 21
21_30	Between 21 and 30 on entry	DATEOFBIRTH ≤ 30 September SCYEAR_ST - 21 and DATEOFBIRTH > 30 September SCYEAR_ST - 31
30+	Over 30 on entry	Otherwise

SCDEGCLASSPOP

20. This field indicates whether a student is a first degree qualifier.

Value	Description	Definition
1	In the population	SCLEVEL = DEG and OUTCOME = 1
0	Not in the population	Otherwise

SCDEGCLASS

21. This field indicates the degree classification awarded to first degree students.

Value	Description	Definition
FIRST	First class honours degree	SCDEGCLASSPOP = 1 and OUTGRADE = FI
2_1	Upper second class honours degree	SCDEGCLASSPOP = 1 and OUTGRADE = SU
DIST	Distinction	SCDEGCLASSPOP = 1 and OUTGRADE = DS, DS*
MER	Merit	SCDEGCLASSPOP = 1 and OUTGRADE = ME
PASS	Pass	SCDEGCLASSPOP = 1 and OUTGRADE = PA
OTH_HONOURS	Other classifications of honours degree	SCDEGCLASSPOP = 1 and OUTGRADE = SL, SE, TH, FO
UNCLASS	Unclassified awards after following an honours degree	SCDEGCLASSPOP = 1 and OUTGRADE ≠ BLANK and not above
UNKNOWN	Unknown classification	SCDEGCLASSPOP = 1 and OUTGRADE = BLANK and not above

SCDISABLE

22. This field indicates whether the student has a reported disability.

Value	Description	Definition
Y	The student has declared a disability	LLDDHEALTHPROB = 1
N	The student has declared no disability	LLDDHEALTHPROB = 2
UNKNOWN	No information provided by the learner	Otherwise

SCDSA

23. This field indicates whether the student is in receipt of disabled students allowance (DSA).

Value	Description	Definition
Y	The student is in receipt of DSA	LEARNFAM_DLA = 1
N	The student is not in receipt of DSA	Otherwise

SCDOM

24. This field indicates whether the student is domiciled in the UK, other EU countries or elsewhere.

Value	Description	Definition
E	England	DOMICILE = XF or (DOMICILE = XJ, XK, GB and (POSTCODEPRIOR is in England or (POSTCODEPRIOR = BLANK or POSTCODEPRIOR begins ZZ)))
OUK	UK, except England	DOMICILE = XH, XI, XG, XK, XJ, GB and not above
OEU	Other EU	DOMICILE = AI, AN, AQ, AT, AW, AX, BE, BG, BL, BM, BQ, CH, CW, CY, CZ, DE, DK, EE, ES, EU, FI, FK, FO, FR, GF, GI, GL, GP, GR, GS, HR, HU, IC, IE, IO, IS, IT, KY, LI, LT, LU, LV, MF, MQ, MS, MT, NC, NL, NO, PF, PL, PM, PN, PT, RE, RO, SE, SH, SI, SK, SX, TC, TF, VG, WF, XA, XD, XE, YT
UNKNOWN	Unknown domicile	DOMICILE = ZZ, BLANK
OTHER	Not EU	Otherwise

SCETHNIC

25. This field indicates the student's ethnicity.

Value	Description	Definition

A	Asian	ETHNICITY = 39, 40, 41, 42, 43
B	Black	ETHNICITY = 44, 45, 46
W	White	ETHNICITY = 31, 32, 33, 34
O	Other	ETHNICITY = 35, 36, 37, 38, 47, 98
UNKNOWN	Unknown ethnicity	Otherwise

SCENTQUALGRP

26. This field contains the broad grouping of the student's highest qualification on entry.

Value	Description	Definition
HEPG	HE: Postgraduate level	QUALENT3 = DUK, DZZ, D80, M41, M44, M71, M80, M90, MUK, MZZ, H71
HEFD	HE: First degree level	QUALENT3 = M2X, H11, HUK, HZZ, JUK
HEOUG	HE: Other undergraduate level	QUALENT3 = H80, J10, J20, J30, J48, J80, C20, C30, C44, C80, C90
BACC	Baccalaureate	QUALENT3 = P62, P63
LEV3	Other Level 3 qualifications	QUALENT3* = P (excluding P62, P63)
FOUND	Foundation course	QUALENT3 = J49
ACCESS	Access course	QUALENT3 = X00, X01
NONE	No formal qualifications	QUALENT3 = X02, X03, X05

OTHERS	Other qualifications (unknown level, or below level 3)	QUALENT3* = Q, R, X04
UNKNOWN	Unknown qualifications	Otherwise

* the first character of QUALENT3 is used

SCPOSTCODE

27. This field shows the postcode prior to enrolment.

Value	Description	Definition
UNKNOWN	Unknown postcode	POSTCODEPRIOR = BLANK or (SCDOM = E, OUK and POSTCODEPRIOR = ZZ999ZZ)
KNOWN	Known postcode	Otherwise

SCLDCS

28. This field shows the Learn Direct codes that have been assigned to the student's programme of study. This directly maps from LDCS_CO1, LDCS_CO2, and LDCS_CO3.

SCJACS

29. This field shows the full 4-digit JACS code that has been assigned to the student's programme of study. The Learn Direct codes used to identify subject areas of study for students returned to the ILR (LDCS_CO1, LDCS_CO2, and LDCS_CO3) have been mapped to full 4-digit JACS codes.

SCFPE

30. This field shows the nominal full person equivalence (FPE) associated with the SCJACS code. This is derived using PCFLDCS, PCSLDCS and PCTLDCS.

SCSBJ_CAH1

31. This field shows which of the Common Aggregation Hierarchy level 1 codes the LDCS code maps to. It is only shown in the individualised file.

SCSBJ_CAH2

32. This field shows which of the Common Aggregation Hierarchy level 2 codes the LDCS code maps to. Where we cannot map to a subject, we set SCSBJ_CAH2 = CAH23-01. A mapping document is available alongside this technical document.

Value	Description
CAH01-01	Medicine and dentistry
CAH02-01	Nursing
CAH02-02	Pharmacology, toxicology and pharmacy
CAH02-03	Subjects allied to medicine not otherwise specified
CAH03-01	Biosciences
CAH03-02	Sport and exercise sciences
CAH04-01	Psychology
CAH05-01	Veterinary sciences
CAH06-01	Agriculture, food and related studies
CAH07-01	Physics and astronomy
CAH07-02	Chemistry
CAH07-03	Physical, material and forensic sciences
CAH08-01	General and others in sciences
CAH09-01	Mathematical sciences
CAH10-01	Engineering
CAH10-02	Technology
CAH11-01	Computing
CAH12-01	Geographical and environmental studies

CAH13-01	Architecture, building and planning
CAH14-01	Humanities and liberal arts (non-specific)
CAH15-01	Sociology, social policy and anthropology
CAH15-02	Economics
CAH15-03	Politics
CAH15-04	Health and social care
CAH16-01	Law
CAH17-01	Business and management
CAH18-01	Communications and media
CAH19-01	English studies
CAH19-02	Celtic studies
CAH19-03	Languages, linguistics and classics
CAH20-01	History and archaeology
CAH20-02	Philosophy and religious studies
CAH21-01	Creative arts and design
CAH22-01	Education and teaching
CAH23-01	Combined and general studies
UNKNOWN	LDCS code missing or unknown, or proportion taught not specified when LDCS code is known, or the proportion taught does not sum to 100

SCSBJ_CAH3

33. This field shows which of the Common Aggregation Hierarchy level 3 codes the LDCS code maps to. It is only shown in the individualised file.

Unistats

Description of derived fields used in Unistats metrics

34. Here we give details of the derived fields in the individualised file*. These fields are used to build the key dimensions of the Unistats tables.

Table 3 Student characteristics derived fields

Derived field name	Description	Paragraph
UNISTATS_FYEAR	Student in first year of study	38
UNITSTATS_ENTPOP	Student is included in Unistats entry population	39
UNISTATS_CODPOP	Student is included in the Unistats class of degree population	40

* The individualised file, SC17_DCT_FXXXXXXXX_IND.csv, downloadable from the OfS portal (see www.officeforstudents.org.uk/data-and-analysis/supplying-data/working-with-individualised-files/).

35. Below you will find a description of the algorithms used to create the populations for two of the metrics used on the Unistats website: entry population (UNISTATS_ENTPOP) and classification of degree population (UNISTATS_CODPOP). Along with the other student characteristic information described above, you can use these population markers and the algorithms provided to preview a close approximation of these Unistats metrics for any relevant courses submitted. The entry population can be used with SCENTQUALGRP to recreate the mix of entry qualifications for a course and the classification of degree population can be used with SCDEGCLASS to determine the outcomes achieved by students on degree level courses. You can also use the SCSBJ_CAH1/2/3 fields to see which subject areas your courses could aggregate with on Unistats using the Common Aggregation Hierarchy subject groupings.

36. The population markers have been included in the individualised file named SC17_DCT_FXXXXXXXX_IND.csv (where XXXXXXXX is your provider's UKPRN) which can be found in your results package to allow you to examine how our algorithms would apply to the data you have submitted. Full details of how to access this file are given on the OfS website (www.officeforstudents.org.uk/data-and-analysis/supplying-data/working-with-individualised-files/).

37. The Unistats continuation summary, which can be found in your results package from its release in week commencing 3 September 2018, also contains information about how your data would be shown on Unistats. This output focusses on the continuation of entrants from your previously submitted 2016/17 ILR R14 data into the 2017/18 data submitted to the data checking tool. This is presented along with technical guidance similar to this document on how to understand and recreate the output from individualised data. Full details of how to access this file are given on the OfS website (www.officeforstudents.org.uk/data-and-analysis/supplying-data/working-with-individualised-files/).

UNISTATS_FYEAR

38. This field indicates whether the student is in the first year of their learning aim.

Value	Description	Definition
1	Student is in the first year of their learning aim	LEARNSTARTDATE > 31 July 2017 and LEARNSTARTDATE < 1 August 2018
0	Otherwise	Otherwise

UNISTATS_ENTPOP

39. This field indicates whether the student is included in the Unistats entry population

Value	Description	Definition
1	Student is in the entry table	SCLEVEL ≠ OTHER, PG and SCLEVEL_DETAIL ≠ HNC and UNISTATS_FYEAR = 1 and STULOAD ≠ BLANK
0	Otherwise	Otherwise

UNISTATS_CODPOP

40. This field indicates whether the student is included in the Unistats class of degree population

Value	Description	Definition
1	Student is in the class of degree table	SCDEGCLASS ≠ UNKNOWN, BLANK
0	Otherwise	Otherwise