

**Office for
Students**



2022-23 Student data checking tool

**2022-23 Student characteristics technical
document and rebuild instructions**

Enquiries to hesastudentdata@officeforstudents.org.uk

Publication date 23 May 2023

Contents

Summary	2
Student characteristics data summary workbook	3
Student characteristics individualised file	3
Student (22056) return fields used to create the student characteristics tables	5
Description of derived fields	8
Fields used to describe the nature of study undertaken	9
Fields used to derive populations of students	15
Fields used to describe student characteristics	20
Rebuild instructions	23
Rebuild instructions for student numbers summary table	23
Rebuild instructions for student characteristics, study location and teaching provider summary tables	23

Summary

1. The 2022-23 student characteristics data summary shows the number of students by different types of characteristics, including personal characteristics and characteristics of their course. The algorithms used to generate the data summary are indicative of those the Office for Students (OfS) intends to use to categorise students in the Student (22056) record submitted to the designated data body (DDB).
2. The OfS expects to use characteristics derived from the Student (22056) record to inform an understanding of provider performance on student outcomes and experiences for a variety of regulatory purposes. These include:
 - a. Regulating access and participation through registration condition A1.¹
 - b. Regulating student outcomes through registration condition B3, and for risk-based monitoring of quality and standards more generally.²
 - c. Assessments through the Teaching Excellence Framework (TEF).
3. The student characteristics data output is formed of two files. These are:
 - a. The student characteristics data summary workbook (SC22_DCT_XXXXXXXX.xlsx).
 - b. The student characteristics individualised file (SC22_DCT_XXXXXXXX_IND.csv).

where XXXXXXXX is the UK Provider Reference Number (UKPRN) for the provider.

4. Details of how to access and use individualised files are given on the Office for Students (OfS) website at www.officeforstudents.org.uk/data-and-analysis/supplying-data/working-with-individualised-files/. Further information specific to this data summary is provided in this document.
5. This document is aimed at readers with in-depth knowledge of the data. Readers are advised to refer to the Student (22056) return available at <https://codingmanual.hesa.ac.uk/22056/home/>.
6. This document lists the fields used to build the data summary tables, which are either taken directly from the Student (22056) return, derived by the DDB from the Student (22056) return,

¹ The OfS registration conditions are described in the Regulatory framework for higher education in England, and its amendments, at [Securing student success: Regulatory framework for higher education in England - Office for Students](#).

² As set out in the revised ongoing conditions of registration B1, B2, B4 and B5, which came into effect from 1 May 2022, and the revised initial and ongoing condition of registration B3, which came into effect from 3 October 2022.

or derived from them by the OfS. Fields with the prefix 'IP' or 'SC' are derived by the OfS using the algorithms provided in this document.³

7. Fields with the prefix 'IP' use algorithms which the OfS intends to apply to individualised student data from the Student (22056) record to generate student outcome and experience measures for OfS regulatory purposes. Each of these fields corresponds to an equivalent field of the same name in our technical algorithms document published in spring 2023.⁴ This document describes the algorithms applied to the DDB's legacy data collections, the Student record (CXX051) and Student Alternative record (CXX054). The approach we have taken in converting existing algorithms from the legacy data model to the Student (22056) data model is described at <https://www.officeforstudents.org.uk/data-and-analysis/data-checking-tool/2022-23-student-data-checking-tool/>.
8. All fields with the prefix 'SC' are derived specifically for this data summary and are written in this document. These fields are necessary because it has not been possible to derive the equivalent 'IP' field for this data summary due to the need to link a provider's student data returns to other data sources. We expect to publish the 'IP' versions of these algorithms, and all other algorithms which the OfS intends to use to construct student outcome and experience measures, in summer 2023.

Student characteristics data summary workbook

9. The student characteristics data summary workbook, (SC22_DCT_XXXXXXXX.xlsx), contains the following worksheets:

Table 1: Worksheets in the student characteristics workbook

Worksheet*	Description
Coversheet	Title page
Student numbers	Number of students by cohort, mode and level of study
Student characteristics	Number of students by characteristic
Study location	Number of students by regional location of study
Teaching provider	Number of students by teaching provider

* This worksheet reference corresponds to the spreadsheet tabs.

Student characteristics individualised file

10. The individualised student data files (SC22_DCT_XXXXXXXX_IND.csv) that we have supplied to each provider are intended to maximise the transparency of our approaches. The individualised file contains the values of all the fields used to generate the data summary tables.

³ The DDB publish the specification for the Student (22056) return and the fields they derive from the Student (22056) return at HESA, [Student Record HESA Coding Manual](#).

⁴ See 'Technical algorithms for student outcome and experience measures' available at [Description and definition of student outcome and experience measures - Office for Students](#).

11. Fields taken from the Student (22056) return or derived by the DDB from the Student (22056) return are shown in capitals using the names given in Table 2. Fields derived as part of the data summary tables are prefixed with 'IP' or 'SC' and are listed in Table 3.
12. Specific instructions for rebuilding the student characteristics data summary workbook from the individualised file are provided in this document.
13. The fields used from the Student (22056) return are associated with an engagement, identified by its engagement number (NUMHUS); a student may have more than one engagement. We will use the generic term 'student' in this technical document and the student characteristics data summary workbook to refer to one student engagement.
14. The individualised file contains one row per student engagement per subject studied. Subject is defined using level 3 of the Common Aggregation Hierarchy (CAH3) as given by the derived field IPSBJ_CAH3.⁵ The individualised file is provided at CAH3 level for consistency with individualised files supplied to providers alongside published updates to our student outcome and experience measures.
15. To enable users to rebuild student headcounts from a file that is one row of data per CAH3 subject, we provide a column called 'SUBWT' on the individualised files that users can sum over. This column apportions the student headcount across the different rows of data that relate to them, according to the proportion of their engagement associated with the CAH3 subject for each row. In the rebuild instructions provided in this document, this sum over the column 'SUBWT' step is always included.

⁵ See [Common Aggregation Hierarchy \(CAH\)](#).

Student (22056) return fields used to create the student characteristics tables

16. The Student (22056) return fields used to generate the data summary tables are listed in Table 2 below.

Table 2: List of fields used from the Student (22056) return or derived by the DDB from the Student (22056) return

Entity	Name	Description
Engagement	UKPRN	UK provider reference number
Engagement	SID	Student identifier
Engagement	NUMHUS	Engagement number
Engagement	ENGSTARTDATE	Engagement start date
Engagement	ENGENDDATE	Engagement end date
Engagement	INCOMINGEXCHANGE	Identifies if the student is an incoming or visiting exchange student
Student	ETHNIC	Ethnicity
Student	BIRTHDTE	Date of birth
Student	SEXID	Sex identifier
Course.Qualification*	QUALCAT	Qualification of course aim
Course*	PREREQUISITE	Prerequisite qualifications for the course
Course initiative*	COURSEINITID	Course initiative identifier
Course initiative*	COURSEINITVALIDFROM	Course initiative valid from date
Course initiative*	COURSEINITVALIDTO	Course initiative valid to date
Collaborative provision	COLPROVTYPEID	Collaborative provision type
Curriculum accreditation*	CURACCID	Curriculum accreditation identifier
Curriculum accreditation*	CURACCIDVALIDFROM	Accreditation valid from date

Entity	Name	Description
Curriculum accreditation*	CURACCVALIDTO	Accreditation valid to date
Reference period student load	RPSTULOAD	Full-time equivalence in the reference period
Student course session	SCSSTARTDATE	Student course session start date
Student course session	SCSENDDATE	Student course session end date
Student initiative	STUINITID	Student initiative identifier
Study location*	PRINONUK	Identifies if the student is mainly studying outside the UK throughout their engagement
Study location	VENUEID	Venue identifier
Study location	STUDYPROPORTION	Proportion of activity at study location
Venue	VENUEUKPRN	UKPRN of provider delivering learning at the venue
Engagement derived field	Z_ACT_CYC	DDB derived field, categorises whether the student was active during the cycle
Engagement derived field	Z_DISABILITYGRP1	DDB derived field, categorises the student's declared disability categories
Engagement derived field	Z_DISABILITYMRK	DDB derived field, categorises whether a student has declared a disability from the latest student disability entity submitted
Engagement derived field	Z_MODEGRP2	DDB derived field, categorises the student's mode of study, based on the latest student course session
Engagement derived field	Z_PERMADDGRP4	DDB derived field, categorises the student's permanent address information from the latest entry profile submitted
Engagement derived field	Z_QLEVELGRP1_CYC	DDB derived field, categorises the student's highest qualification awarded during the cycle
Engagement derived field	Z_STATUSEND	DDB derived field, records the status at the end of the last student course session, or the end of the reference period if it hasn't finished
Engagement derived subject	Z_SUBJHECOSSCS	DDB derived field, records the latest HECoS subject information based on the latest student course session

Entity	Name	Description
Engagement derived subject	Z_SUBJFPE	DDB derived field, records the full person equivalence associated with Z_SUBJHECOSSCS
Student course session derived field	Z_INACTFROMSCS	DDB derived field, records the start date of a student's period of inactivity
Student course session derived field	Z_INACTTOSCS	DDB derived field, records the end date of a student's period of inactivity

* The values of these fields are taken from the latest student course session associated with the student's engagement.

Description of derived fields

17. The derived fields used to generate the data summary are listed in Table 3 below.

18. Where the definition of a derived field remains unchanged from that published in our spring 2023 technical algorithms document, the algorithm has not been repeated in this document and we instead direct the reader to the previously published algorithm. In these cases Table 3 contains the paragraph reference in the spring 2023 technical algorithms document.⁶

Table 3: List of derived fields

Derived field name	Description	Paragraph
IPCOMDATE	Engagement start date	19
IPACTENDDATE	Engagement end date	20
IPMODE	Mode of study	21
IPAPPRENTICE	Apprenticeship status	22
IPOFSQAIM	Higher education category by qualification	23
IPLEVELNUM	Numeric level of study	24
IPLEVEL	Level of study	Technical algorithms paragraph 48
IPLEVELBROAD	Level of study (broad grouping)	Technical algorithms paragraph 50
IPDENT	Indicates students on a dentistry programme of study	26
IPHECOS	Higher education classification of Subjects (HECoS) code	27
IPSBJ_CAH2	CAH level 2 subject code	29
IPSBJ_CAH3	CAH level 3 subject code	30
IPCAH3FPE	Full person equivalence per CAH3 subject	31
SUBWT	Subject weighting	33
IPLOCPOSTCODE	Study location postcode	35
SCREGION	Study region	37
IPIKPRNTC	Majority teaching provider	38
SCOFSHE	Higher education student status	40
SCHECAT	Higher education category	41

⁶ See 'Technical algorithms for student outcome and experience measures' available at [Description and definition of student outcome and experience measures - Office for Students](#).

Derived field name	Description	Paragraph
IPACTANN	Identifies activity after the student's anniversary	42
IPAYDUP	Identifies duplicate activity within the academic year	43
SCCONTEXTPOP	Indicates students included in the contextual population that informs the summaries	44
IPBIRTHDATE	Birth date	47
IPSEX	Sex	48
IPDISABLETYPE	Disability type	49
IPDISABLE	Disability status	50
IPETHNICDETAIL	Detailed ethnic group	51
IPETHNIC	Broad ethnic group	Technical algorithms paragraph 131
IPDOM	Domicile	52
IPSTARTAGE	Age in the year of commencement of study	Technical algorithms paragraph 125
IPSTARTAGEBAND	Age category in the year of commencement of study	Technical algorithms paragraph 126

Fields used to describe the nature of study undertaken

IPCOMDATE

19. This field shows the start date of a student's engagement. IPCOMDATE is equal to ENGSTARTDATE.

IPACTENDDATE

20. This field shows the end date of a student's engagement. IPACTENDDATE is equal to ENGENDDATE.

IPMODE

21. This field allocates students to a mode of study.

Value	Description	Definition
WUPFT	Writing up (previously full-time)	Z_MODEGRP2 in (01, 02) and Z_STATUSEND = 04 and Z_ACT_CYC = 1

Value	Description	Definition
WUPPT	Writing up (previously part-time)	Z_MODEGRP2 = 03 and Z_STATUSEND = 04 and Z_ACT_CYC = 1 and not above
APPR	Apprenticeship	IPAPPRENTICE = 1 and Z_MODEGRP2 in (01, 02, 03) and Z_ACT_CYC = 1 and not above
FT	Full-time	Z_MODEGRP2 in (01, 02) and Z_ACT_CYC = 1 and not above
PT	Part-time	Z_MODEGRP2 = 03 and Z_ACT_CYC = 1 and not above
OTH	Other	Otherwise

IPAPPRENTICE

22. This field indicates whether the student is studying on an apprenticeship at any level.

Value	Description	Definition
1	The student is studying on an apprenticeship at any level	At least one value of student initiative, STUINITID, in (004, 020) or In the latest student course session, at least one value of course initiative, COURSEINITID, in (004, 020) where (COURSEINITVALIDFROM < SCSENDDATE or SCSENDDATE = BLANK) and (COURSEINITVALIDTO >= SCSSTARTDATE or COURSEINITVALIDTO = BLANK)

Value	Description	Definition
0	The student is not studying on an apprenticeship	Otherwise

IPOFSQAIM

23. This field allocates qualification aims to categories of higher education. The values of fields in the Student (22056) return are taken from the latest student course session associated with the student's engagement.

Value	Description	Definition
PHD	PhD and MPhil	QUALCAT in (D0003, L0000)
OTHL7_Q_R	Other Level 7 research-based qualification	QUALCAT in (L0001)
MASTER	Masters'	QUALCAT in (M0003, M0004, M0006, M0007)
PGCE	PGCE and other postgraduate initial teacher training (ITT)	QUALCAT in (H0013, M0016)
DTLLS_PG	Postgraduate diploma in teaching in the lifelong learning sector	QUALCAT in (M0020)
PGCERT	Postgraduate certificate	QUALCAT in (M0012)
PGDIP	Postgraduate diploma	QUALCAT in (M0009)
PROCONGRAD	Professional, conversion and other graduate entry programmes	QUALCAT in (H0009, H0010, H0014, I0002, I0005, I0006, I0007) or (QUALCAT in (H0016, I0010) and PREREQUISITE = 02)
ENHANCED	Enhanced first degree (or integrated masters)	QUALCAT in (H0004, M0002)
MEDVETDENT	Pre-registration first degree with honours leading towards obtaining eligibility to register to practise with the General Medical Council, General Dentistry Council (as a dentist) or the Royal College of Veterinary Surgeons	QUALCAT in (H0003, H0005, I0001) and ((At least one value of CURACCID in (05901, 12001, 05803) where (CURACCVALIDFROM < SCSENDDATE or SCSENDDATE = BLANK) and (CURACCVALIDTO >= SCSSTARTDATE or CURACCVALIDTO = BLANK)) or

Value	Description	Definition
		(IPDENT = 1 and at least one value of CURACCID = 05802 where (CURACCVALIDFROM < SCSENDATE or SCSENDDATE = BLANK) and (CURACCVALIDTO >= SCSSTARTDATE or CURACCVALIDTO = BLANK)))
FIRST	First degree	QUALCAT in (H0003, H0005 I0001) and not above
CTLLS	Certificate in teaching in the lifelong learning sector	QUALCAT = C0006
DET	Diploma in Education and Training	QUALCAT = I0008
DIPHE	Diploma of Higher Education (DipHE)	QUALCAT = J0002
DTLLS	Diploma in teaching in the lifelong learning sector	QUALCAT in (H0015, I0009)
FOUDEG	Foundation degree	QUALCAT in (J0000, J0001)
HIGHCERT	Higher certificate	QUALCAT = C0000
HNC	Higher national certificate	QUALCAT = C0001
HND	Higher national diploma	QUALCAT = J0003
PTLLS	Preparing to teach in the lifelong learning sector	QUALCAT = C0005
OTHL[X]_Q	Other Level X qualification, where X is the level indicated by IPLEVELNUM	QUALCAT in (C0002, C0003, C0004, C0007, C0009, E0000, E0001, E0002, E0004, H0006, H0007, H0008, H0012, H0019, I0004, J0004, J0005, J0006, J0007, J0010, J0012, M0008, M0010, M0011, M0013, M0015, M0017, M0018, M0021, M0023) or (QUALCAT in (H0016, I0010) and PREREQUISITE ≠ 02)
OTHL[X]_U	Other Level X unit, where X is the level indicated by IPLEVELNUM	QUALCAT in (C0008, D0004, D0005, E0003, E0005, H0018, H0020, I0012, I0013, J0011, L0002, L0003, M0022, M0024)

IPLEVELNUM

24. This field allocates qualification aims to a numeric level of study according to the sector-recognised standards relating to the OfS's ongoing condition of registration B5 and initial condition B8, available at www.officeforstudents.org.uk/publications/securing-student-success-regulatory-framework-for-higher-education-in-england/. This also aligns with FHEQ and NVQ levels.
25. The value of QUALCAT from the latest student course session associated with the student's engagement is used.

Value	Description	Definition
8	Doctoral degree	QUALCAT in (D0003, D0004, E0000, E0001, E0002, E0003, E0004, L0000)
7	Masters' degree, postgraduate diplomas, postgraduate certificates	QUALCAT in (L0001, L0002, L0003, D0005, M0002, M0003, M0004, M0006, M0007, M0008, M0009, M0010, M0011, M0012, M0013, M0015, M0016, M0017, M0018, M0020, M0021, M0022, M0023, M0024, E0005)
6	Bachelors' degrees, graduate certificates and diplomas	QUALCAT in (H0003, H0004, H0005, H0006, H0007, H0008, H0009, H0010, H0012, H0013, H0014, H0015, H0016, H0018, H0019, H0020, I0001)
5	Foundation degrees, diplomas of higher education and other higher diplomas	QUALCAT in (I0002, I0004, I0005, I0006, I0007, I0008, I0009, I0010, I0012, I0013, J0000, J0001, J0002, J0003, J0004, J0005, J0006, J0007, J0010, J0011, J0012)
4	Certificates of higher education	QUALCAT in (C0000, C0001, C0002, C0003, C0004, C0005, C0006, C0007, C0008, C0009)
BLANK	Not applicable	Otherwise

IPDENT

26. This field indicates whether a student has at least one instance of a 'Dentistry', 'Pre-clinical dentistry' or 'Clinical dentistry' programme of study.

Value	Description	Definition
1	The student has at least one instance of a 'Dentistry', 'Pre-clinical dentistry' or 'Clinical dentistry' programme of study	At least one value of IPHECOS in (100266, 100268, 100275))
0	The student does not have an instance of a 'Dentistry', 'Pre-clinical dentistry' or 'Clinical dentistry' programme of study	Otherwise

IPHECOS

27. This field shows the full 6-digit Higher Education Classification of Subjects (HECoS) code that has been assigned to the student's programme of study.

28. IPHECOS is equal to Z_SUBJHECOSSCS, which records the latest HECoS subject information for the engagement based on the latest student course session.

IPSBJ_CAH2

29. This field shows which of the Common Aggregation Hierarchy level 2 (CAH2) codes the IPHECOS code maps to. Where we cannot map to a subject, we set IPSBJ_CAH2 = CAH23-01. The mapping of HECOS codes to Common Aggregation Hierarchy codes can be found on the HESA website.⁷

IPSBJ_CAH3

30. This field shows which of the Common Aggregation Hierarchy level 3 (CAH3) codes the IPHECOS code maps to. Where we cannot map to a subject, we set IPSBJ_CAH3 = CAH23-01-01. The mapping of HECOS codes to Common Aggregation Hierarchy codes can be found on the HESA website.⁸

IPCAH3FPE

31. This field shows the nominal full person equivalence (FPE) associated with the IPSBJ_CAH3 subject code. The concept of FPE student numbers is defined in full on the HESA website.⁹

32. IPCAH3FPE is calculated using Z_SUBJFPE, which records the latest apportioned FPE associated with each HECoS subject (IPHECOS). Z_SUBJFPE is aggregated to give the FPE associated with each CAH3 subject (IPSBJ_CAH3).

SUBWT

33. This field shows the subject weighting associated with the IPSBJ_CAH3 subject and can be summed over to facilitate rebuilding student headcounts from individualised files.

34. SUBWT is calculated as IPCAH3FPE divided by 100.

IPLOCPOSTCODE

35. This field shows the student's location of study postcode.

36. For providers in England:

- a. We sum the STUDYPROPORTION of each venue (VENUEID) for which VENUEUKPRN matches the majority teaching provider identified in IPUKPRNTC. Where there is more than one student course session associated with the engagement in the academic year, we use RPSTULOAD to weight the summed STUDYPROPORTION across the student course sessions.
- b. Then IPLOCPOSTCODE is set as the postcode of the venue (POSTCODE) with the largest summed STUDYPROPORTION in the academic year. Where there is more than one

⁷ See https://www.hesa.ac.uk/files/HECoS_CAH_Version_1.3.4_final.xlsx.

⁸ See https://www.hesa.ac.uk/files/HECoS_CAH_Version_1.3.4_final.xlsx.

⁹ See HESA, [Definitions: Students](#).

student course session associated with the engagement in the academic year, this is the sum of the weighted values described above.

- c. In the event of a tie between venues with the same summed STUDYPROPORTION, IPLOCPOSTCODE is set to Unknown.

SCREGION

37. This field assigns the location of study postcode (IPLOCPOSTCODE) to broader regions of study defined by the International Territorial Levels, level 2 (ITL 2). This mapping is based on National Statistics Postcode Lookup (NSPL) data from May 2022.¹⁰

IPIKPRNTC

38. This field shows the UKPRN of the provider where the student is taught for the majority of this academic year. It is calculated using the method described below.

39. For providers in England:

- a. For each VENUEUKPRN associated with a student course session, we sum the STUDYPROPORTION across all associated VENUEIDs. Where there is more than one student course session associated with the engagement in the academic year, we use RPSTULOAD to weight the summed STUDYPROPORTION across the student course sessions.
- b. Then IPIKPRNTC is set as the VENUEUKPRN with the greatest summed STUDYPROPORTION in the academic year. Where there is more than one student course session associated with the engagement in the academic year, this is the sum of the weighted values described above.
- c. In the event of a tie between the registering provider and another teaching provider, IPIKPRNTC is set as the UKPRN of the registering provider. If there is a tie between teaching providers who are not the registering provider, IPIKPRNTC is set to Unknown.

Fields used to derive populations of students

SCOFSHE

40. This field determines whether a student could be counted as a higher education (HE) student for any OfS purpose. It is designed to align the coverage of different student records. The following are excluded:

- incoming exchange students
- students that left within two weeks without any award
- students that are not on a higher education aim.

¹⁰ See Office for National Statistics, [National Statistics Postcode Lookup \(May 2022\) | Open Geography Portal](#).

Value	Description	Definition
1	Student is counted as a higher education student	IPLEVELBROAD ≠ NA and INCOMINGEXCHANGE = BLANK and (IPACTENDDATE = BLANK or IPACTENDDATE – IPCOMDATE > 14 or Z_QLEVELGRP1_CYC in (01, 02)))
0	Student is not counted as a higher education student as they are excluded by one of the clauses in paragraph 40	Otherwise

SCHECAT

41. This field categories students into key subsets of the higher education population for the purpose of understanding student lifecycle indicators.

Value	Description	Definition
1	Student is registered at a UK provider but is mainly studying abroad	SCOFSHE = 1 and (COLPROVTYPEID = 02 or PRINONUK = 01)
2	Student is mainly studying in the UK and is aiming for credit or modular provision rather than a qualification	SCOFSHE = 1 and IPLEVEL in (UGCREDIT, PGCREDIT) and not above
3	Student is mainly studying in the UK and is aiming for a qualification but is dormant or sabbatical	SCOFSHE = 1 and IPMODE = OTH and not above
4	Student is mainly studying in the UK and is writing up on a qualification aim	SCOFSHE = 1 and IPMODE in (WUPFT, WUPPT)

Value	Description	Definition
		and not above
5	Student is mainly studying in the UK and is actively studying on a qualification aim	SCOFSHE = 1 and not above
0	Student is not counted as a higher education student as they are excluded by one of the clauses in paragraph 40	Otherwise

IPACTANN

42. This field determines whether the student was actively studying at any point in the academic year beyond the anniversary of the day 14 days after their starting date.

Value	Description	Definition
1	The student was actively studying during the academic year beyond the anniversary of the day 14 days after their starting date	<p>IPCOMDATE < 17 July 20YY+1 and for any student course session in the academic year:</p> <p>SCSENDDATE > anniversary of IPCOMDATE + 14 and</p> <p>((Z_INACTFROMSCS > anniversary of IPCOMDATE + 14 and Z_INACTFROMSCS > SCSSTARTDATE) or (Z_INACTTOSCS < 31 July 20YY+1 and Z_INACTTOSCS < SCSENDDATE))</p>
0	The student was not actively studying during the academic year beyond the anniversary of the day 14 days after their starting date	Otherwise

Note: We consider the anniversary of the day 14 days after IPCOMDATE such that it lies within the current academic year.

IPAYDUP

43. This field determines whether the student record is used in calculations of student headcounts where we count each student's year of programme of study once. It ensures that similar activity is counted in a similar way irrespective of when it occurs. It primarily de-duplicates activity for students on non-standard academic years so that each student record is counted once and only once for each year of programme of study.

Value	Description	Definition
1	The student has been recorded with a starting date beyond the current academic year	IPCOMDATE > 31 July 20YY+1
1	The student left within 14 days of their starting date without an award, or they had no activity in the academic year more than 14 days after the anniversary of their start date.	IPACTENDDATE < 1 August 20YY or IPACTANN = 0 or (IPACTENDDATE ≤ IPCOMDATE + 14 and OFSHE ≠ 1)
0	The student record is used in calculations of student headcounts where we count each student's year of programme of study once	Otherwise

SCCONTEXTPOP

44. This field indicates whether a student is included in the population used to create the workbook summary tables.
45. A student may fall into multiple populations, for example a student on a one year course could count towards the entrant and qualifier population. To account for this, the population is constructed as follows:
- For the all student population, values 1, 2, 3 and 4 are used
 - For the entrant population, values 1 and 2 are used
 - For the qualifier population, values 1,3 and 5 are used.
46. SCCONTEXTPOP is calculated once per student at mode and broad level. This means that the following deduplication is applied:
- A student, as identified by SID, is only counted once per IPMODE and IPLEVELBROAD.
 - If the student appears multiple times at a single combination of IPMODE and IPLEVELBROAD, the record with the highest level (according to IPLEVELNUM) is prioritised.
 - If there are multiple records at the highest level, the record that would be counted towards the entrant, qualifier or all student contextual populations is prioritised.
 - If there are still multiple records, the record is chosen consistently by considering identifiers UKPRN, SID, and NUMHUS alphabetically.

Value	Description	Definition
1	The student is counted in the all student, entrant and qualifier contextual populations	SCHECAT in (1, 2, 5) and

Value	Description	Definition
		<p>IPAYDUP = 0 and</p> <p>Z_QLEVELGRP1_CYC in (01, 02) and</p> <p>IPCOMDATE ≥ 17 July 20YY and</p> <p>IPCOMDATE < 17 July 20YY+1</p> <p>and not above</p>
2	The student is counted in the all student and entrant contextual populations but not the qualifier contextual population	<p>SCHECAT in (1, 2, 5) and</p> <p>IPAYDUP = 0 and</p> <p>IPCOMDATE ≥ 17 July 20YY and</p> <p>IPCOMDATE < 17 July 20YY+1</p> <p>and not above</p>
3	The student is counted in the all student and qualifier contextual populations but not the entrant contextual population	<p>SCHECAT in (1, 2, 5) and</p> <p>IPAYDUP = 0 and</p> <p>Z_QLEVELGRP1_CYC in (01, 02)</p> <p>and not above</p>
4	The student is counted in the all student contextual population, but not in the as a qualifier or entrant contextual populations	<p>SCHECAT in (1, 2, 5) and</p> <p>IPAYDUP = 0</p> <p>and not above</p>
5	The student is counted in the qualifier contextual population but not the entrant or all student contextual populations	<p>SCHECAT in (1, 2, 5) and</p>

Value	Description	Definition
		Z_QLEVELGRP1_CYC in (01, 02) and not above
0	The student is not counted in the contextual population	Not above or Student is not counted towards the contextual population after deduplication (see paragraph 46)

Note: The DDB data model requires that PGR students transferring to a new provider as part of a collaborative provision arrangement be treated as entrants to a new engagement at the new provider. The engagement start date, IPCOMDATE, must be returned as the date that reporting responsibility transferred to the new provider. Reporting under the HESASTU and HESASAR data models did not include this requirement for PGR students in these arrangements.¹¹

Fields used to describe student characteristics

IPBIRTHDATE

47. This field shows the date of birth of the student. IPBIRTHDATE is equal to BIRTHDTE.

IPSEX

48. This field indicates the sex of the student.

Value	Description	Definition
2	Female	SEXID = 10
1	Male	SEXID = 11
9	Other sex	SEXID = 12
0	Unknown	SEXID in (96, 99)

IPDISABLETYPE

49. This field indicates the type of disability the student has reported.

Value	Description	Definition
COG	The student has cognitive or learning difficulties	Z_DISABILITYGRP1 = 05
MH	The student has a mental health condition	Z_DISABILITYGRP1 = 07

¹¹ See HESA, [Further Guidance on PGR Collaborative Supervision Arrangements](#).

Value	Description	Definition
MULTI	The student has multiple or other impairments	Z_DISABILITYGRP1 in (04, 10, 11)
NONE	The student has no disability reported or an unknown disability type	Z_DISABILITYGRP1 in (01, Z9)
PHY	The student has a sensory, medical or physical impairment	Z_DISABILITYGRP1 in (02, 03, 06, 08)
SOC	The student has a social or communication impairment	Z_DISABILITYGRP1 = 09

IPDISABLE

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

Value	Description	Definition
Y	Disability reported	Z_DISABILITYMRK = 1
N	No disability reported	Otherwise

IPETHNICDETAIL

51. This field indicates the student's ethnicity, split into 17 groups.

Value	Description	Definition
A_01	Asian – Bangladeshi or Bangladeshi British	ETHNIC = 100
A_02	Asian – Chinese or Chinese British	ETHNIC = 101
A_03	Asian – Indian or Indian British	ETHNIC = 103
A_04	Asian – Pakistani or Pakistani British	ETHNIC = 104
A_05	Any other Asian background	ETHNIC in (102, 119)
B_01	Black – African or African British	ETHNIC = 120
B_02	Black – Caribbean or Caribbean British	ETHNIC = 121
B_03	Any other Black Background	ETHNIC = 139
M_01	Mixed or multiple ethnic groups – White or White British and Asian or Asian British	ETHNIC = 140
M_02	Mixed or multiple ethnic groups – White or White British and Black African or Black African British	ETHNIC = 141
M_03	Mixed or multiple ethnic groups – White or White British and Black Caribbean or Caribbean British	ETHNIC = 142

Value	Description	Definition
M_04	Any other Mixed or Multiple ethnic background	ETHNIC = 159
O_01	Other ethnic group	ETHNIC in (180, 899)
O_02	Gypsy, Roma, Traveller, Irish Traveller, Showman or Showwoman	ETHNIC in (163, 164, 165, 168, 170)
W_04	White	ETHNIC in (160, 161, 162, 166, 167, 169, 179)
U	Refused, Unknown, Prefer not to say or not collected	Otherwise

IPDOM

52. This field indicates whether the student's domicile is a country in the UK, an EU country or elsewhere.

Value	Description	Definition
E	England	Z_PERMADDGRP4 = 01
N	Northern Ireland	Z_PERMADDGRP4 = 02
S	Scotland	Z_PERMADDGRP4 = 03
W	Wales	Z_PERMADDGRP4 = 04
EU	European Union	Z_PERMADDGRP4 = 06
OTHER	Other international	Z_PERMADDGRP4 = 05, 07
UNKNOWN	Unknown domicile	Otherwise

Rebuild instructions

53. The individualised file, 'SC22_DCT_XXXXXXXXX_IND.csv', can be used to interrogate the data and our calculation of derived fields by filtering to certain groups of records.

54. Specific instructions to rebuild the student characteristics data summary workbook, 'SC22_DCT_XXXXXXXXX.xlsx', are provided here. Full details of how to access and use an individualised file are given on the OfS website (www.officeforstudents.org.uk/data-and-analysis/supplying-data/working-with-individualised-files).

Rebuild instructions for student numbers summary table

55. Table 4 describes the steps involved in rebuilding the counts and percentages of students in the student numbers data summary table.

Table 4: Rebuild instructions for the student numbers summary table

Step	Description		Rebuild instructions
1	Select the cohort	All students	SCCONTEXTPOP = 1, 2, 3, 4
		Entrants	SCCONTEXTPOP = 1, 2
		Qualifiers	SCCONTEXTPOP = 1, 3, 5
2	Select the level of study	Undergraduate	IPLEVELBROAD = UG
		Postgraduate	IPLEVELBROAD in (PGT, PGR)
3	Select the mode of study	Full-time	IPMODE = FT
		Part-time	IPMODE = PT
		Apprenticeship	IPMODE = APPR
4	Calculate the count of students		Sum of SUBWT
5	Calculation the percentage of students	Calculate the total count of students in the same cohort and level	Sum of SUBWT after applying steps 1 and 2
		Calculate the percentage of students	(Count of students ÷ total count of students) x 100

Rebuild instructions for student characteristics, study location and teaching provider summary tables

56. Table 5 describes the rebuild steps for the student characteristics, study location and teaching provider summary tables.

Table 5: Rebuild instructions for summary tables other than student numbers

Step	Description		Rebuild instructions
1	Select the population of the table	All students	SCCONTEXTPOP = 1, 2, 3, 4
2	Select the level of study	Undergraduate	IPLEVELBROAD = UG
		Postgraduate	IPLEVELBROAD in (PGT, PGR)
3	Select the students in this row of the table		Refer to the appropriate section for this summary table below
4	Calculate the count of students		Sum of SUBWT
5	Calculation the percentage of students	Calculate the total count of students in the same cohort and level	Sum of SUBWT after applying steps 1 and 2.
		Calculate the percentage of students	(Count of students ÷ total count of students) x 100

Student characteristics

57. On the 'Student characteristics' sheet, each summary table shows the count and percentage of students categorised by a given characteristic. Table 5 describes how to rebuild these numbers from the individualised file. In step 3 of Table 5, each row of the summary table can be rebuilt from the individualised file by applying the corresponding filters given in the tables below.

Age

Row value	Rebuild algorithm
Under 21 years	IPSTARTAGEBAND = U21
21 to 25 years	IPSTARTAGEBAND = 21_25
26 to 30 years	IPSTARTAGEBAND = 26_30
31 to 40 years	IPSTARTAGEBAND = 31_40
41 to 50 years	IPSTARTAGEBAND = 41_50
51 years and over	IPSTARTAGEBAND = 51 +
Unknown	IPSTARTAGEBAND = U

Sex

Row value	Rebuild algorithm
Female	IPSEX = 2
Male	IPSEX = 1
Other sex	IPSEX = 9
Unknown	IPSEX = 0

Disability

Row value	Rebuild algorithm
Disability reported	IPDISABLE = Y
No disability reported	IPDISABLE = N

Disability type

Row value	Rebuild algorithm
Cognitive or learning difficulties	IPDISABLETYPE = COG
Mental health conditions	IPDISABLETYPE = MH
Multiple or other impairments	IPDISABLETYPE = MULTI
Sensory, medical or physical impairments	IPDISABLETYPE = PHY
Social or communication impairments	IPDISABLETYPE = SOC
No disability reported or unknown disability type	IPDISABLETYPE = NONE

Ethnicity

Row value	Rebuild algorithm
Asian	IPDOM in (E, S, W, N) and IPETHNIC = A
Black	IPDOM in (E, S, W, N) and IPETHNIC = B
Mixed	IPDOM in (E, S, W, N) and IPETHNIC = M
Other	IPDOM in (E, S, W, N) and IPETHNIC = O
White	IPDOM in (E, S, W, N) and IPETHNIC = W
Unknown or not applicable	IPDOM not in (E, S, W, N) or IPETHNIC = U

Domicile

Row value	Rebuild algorithm
UK	IPDOM in (E, S, W, N)
EU	IPDOM = EU
Other international	IPDOM = OTHER
Unknown	IPDOM = UNKNOWN

Subject of study (CAH2)

Row value	Rebuild algorithm
Medicine and dentistry (CAH01-01)	IPSBJ_CAH2 = CAH01-01
Pharmacology, toxicology and pharmacy (CAH02-02)	IPSBJ_CAH2 = CAH02_02
Nursing and midwifery (CAH02-04)	IPSBJ_CAH2 = CAH02_04
Medical sciences (CAH02-05)	IPSBJ_CAH2 = CAH02_05
Allied health (CAH02-06)	IPSBJ_CAH2 = CAH02_06
Biosciences (CAH03-01)	IPSBJ_CAH2 = CAH03_01
Sport and exercise sciences (CAH03-02)	IPSBJ_CAH2 = CAH03_02
Psychology (CAH04-01)	IPSBJ_CAH2 = CAH04_01
Veterinary sciences (CAH05-01)	IPSBJ_CAH2 = CAH05_01
Agriculture, food and related studies (CAH06-01)	IPSBJ_CAH2 = CAH06_01
Physics and astronomy (CAH07-01)	IPSBJ_CAH2 = CAH07_01
Chemistry (CAH07-02)	IPSBJ_CAH2 = CAH07_02
General, applied and forensic sciences (CAH07-04)	IPSBJ_CAH2 = CAH07_04
Mathematical sciences (CAH09-01)	IPSBJ_CAH2 = CAH09_01
Engineering (CAH10-01)	IPSBJ_CAH2 = CAH10_01
Materials and technology (CAH10-03)	IPSBJ_CAH2 = CAH10_03
Computing (CAH11-01)	IPSBJ_CAH2 = CAH11_01
Architecture, building and planning (CAH13-01)	IPSBJ_CAH2 = CAH13_01
Sociology, social policy and anthropology (CAH15-01)	IPSBJ_CAH2 = CAH15_01
Economics (CAH15-02)	IPSBJ_CAH2 = CAH15_02
Politics (CAH15-03)	IPSBJ_CAH2 = CAH15_03
Health and social care (CAH15-04)	IPSBJ_CAH2 = CAH15_04
Law (CAH16-01)	IPSBJ_CAH2 = CAH16_01
Business and management (CAH17-01)	IPSBJ_CAH2 = CAH17_01
English studies (CAH19-01)	IPSBJ_CAH2 = CAH19_01

Row value	Rebuild algorithm
Celtic studies (CAH19-02)	IPSBJ_CAH2 = CAH19_02
Languages and area studies (CAH19-04)	IPSBJ_CAH2 = CAH19_04
History and archaeology (CAH20-01)	IPSBJ_CAH2 = CAH20_01
Philosophy and religious studies (CAH20-02)	IPSBJ_CAH2 = CAH20_02
Education and teaching (CAH22-01)	IPSBJ_CAH2 = CAH22_01
Combined and general studies (CAH23-01)	IPSBJ_CAH2 = CAH23_01
Media, journalism and communications (CAH24-01)	IPSBJ_CAH2 = CAH24_01
Creative arts and design (CAH25-01)	IPSBJ_CAH2 = CAH25_01
Performing arts (CAH25-02)	IPSBJ_CAH2 = CAH25_02
Geography, earth and environmental studies (CAH26-01)	IPSBJ_CAH2 = CAH26_01

Study location

58. The study location summary table shows the count and percentage of students for each regional location of study defined by the derived field SCREGION (see paragraph 37). Table 5 describes how to rebuild these numbers from the individualised file. In step 3 of Table 5, each row of the table can be rebuilt by filtering the individualised file to the corresponding value of SCREGION.

Teaching provider

59. The teaching provider summary table shows the count of students by majority teaching provider as defined by the derived field IPUKPRNTC (see paragraph 38). Table 5 describes how to rebuild these numbers from the individualised file. In step 3 of Table 5, each row of the table can be rebuilt by filtering the individualised file to IPUKPRNTC equal to the UKPRN of the teaching provider.



© The Office for Students copyright 2023

This publication is available under the Open Government Licence 3.0 except where it indicates that the copyright for images or text is owned elsewhere.

www.nationalarchives.gov.uk/doc/open-government-licence/version/3/