2021-22 Data Futures Beta data checking tool

2021-22 Student numbers technical document and rebuild instructions

Contents

Summary

1. This document describes the algorithms applied to the Data Futures Beta (21056) record. These are used to determine a provider's higher education numbers as reported to the Designated Data Body (DDB). These are produced so providers can see the method and changes to the algorithms to create student numbers as a result of the new student data model. This is in preparation for 2022-23 where we intend to use student numbers derived from the new student data model for various regulatory purposes, including for setting registration fees, assessing applications for degree awarding powers and assessing applications for university title and university college title. These algorithms are illustrative and we may make changes where we believe they will improve our algorithms.

2. Throughout the document, fields taken or derived from the Data Futures Beta (21056) record are shown in capitals.

3. The student numbers outputs are formed of two files. These are:

   a. Student numbers workbook 'SN21_DCT_BETA_XXXXXXXX.xlsx'

   b. Student numbers engagement individualised file 'SN21_DCT_BETA_XXXXXXXX_ENG.csv'

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

4. The data in the student numbers data checking tool outputs are provided as a data quality tool, to aid providers in getting their data correct and to give providers an opportunity to feedback on any potential issues with new Office for Students (OfS) algorithms and must not be used for any other purpose.

5. This document is aimed at readers with in-depth knowledge of the data. Readers are advised to refer to a copy of the Student (Beta) Record – 21056 coding manual https://codingmanual.hesa.ac.uk/21056/home/.

6. We will document any changes or corrections to the algorithms in this document on the OfS website (www.officeforstudents.org.uk/data-and-analysis/data-checking-tool/2021-22-data-futures-beta-data-checking-tool/)

Changes

7. This is the first publication of these algorithms based on the Data Futures Beta (21056) record. Any changes after initial publication will be listed here.

Note on terminology

8. The algorithms below apply to the Data Futures Beta (21056) fields.

9. We use the generic term 'record' to refer to a student engagement.

10. The year '20Y1-Y2' below refers to the academic year starting on 1 August 20Y1 and ending on 31 July 20Y2.

Student numbers workbook

11. The student numbers workbook is the Excel workbook SN21_DCT_BETA _XXXXXXXX.xlsx which contains the following worksheets:

Table 1: Worksheets in the student numbers workbook

Worksheet*	Title
Coversheet	Number of student full-time equivalents (FTEs) by higher education study level

* This worksheet reference corresponds to the spreadsheet tabs.

Student numbers engagement individualised file

12. When working through this document it is necessary to use the engagement individualised file, SN21_DST_BETA_XXXXXXXX_ENG.csv The engagement individualised file contains the values of all fields used to generate student numbers.

13. Fields taken from Data Futures Beta (21056) or derived as part of the student numbers outputs are shown in capitals using the names given in Tables 2 and 3.

14. Specific instructions for rebuilding the student numbers workbook from the engagement individualised file are provided in this document.

15. The engagement individualised file contains one record per engagement in the latest academic year. There is one row per engagement (each unique combination of UKPRN, SID and NUMHUS).

20Y1-Y2 fields used to create the student number outputs

16. The fields used to generate the student numbers outputs are listed in Tables 2 and 3.

Table 2: Student-Data Futures fields used in the student number calculations

Entity	Name	Description	Column in engagement individualised file
Engagement	UKPRN	UK Provider Reference Number	A
Engagement	SID	Student identifier	B
Engagement	NUMHUS	Engagement number	C
Engagement	ENGSTARTDATE	Engagement start date	O
Engagement	ENGENDDATE	Engagement end date	N
Engagement	INCOMINGEXCHANGE	Identifies if the student is an incoming or visiting exchange student	P
Engagement derived field	Z_PRINONUK	Data Futures derived field identifies if student is studying primarily outside the UK	R
Engagement derived field	Z_STATUS_CYC	Data Futures derived field records the student status during the cycle	S
Engagement derived field	Z_STULOAD_CYC	Data Futures derived field represents the total student load to date within the cycle	T
Qualification*	QUALCAT	Records the qualification category	Q

*Qualification from the latest student course session

Description of derived fields

17. The derived fields used to generate the student numbers outputs are listed in Table 3.

Table 3: Description of derived fields for recreating tables

Derived field name	Description	Paragraph	Column in engagement individualised file
SNEXCL	Reason(s) for exclusion from student numbers population	22	E
SNEXCL1 – SNEXCL8	Student number population exclusions	18 - 21	J-M
SNLEVELNUM	Level of higher education record	24	F

Derived field name	Description	Paragraph	Column in engagement individualised file
SNREL_TEF	Undergraduate higher education provision relevant to the Teaching Excellence Framework (TEF)	26	I
SNTOTHEFTE	Total higher education FTE per student	25	G
SNTOTHEFTE_TEF	Total higher education undergraduate FTE per student relevant to TEF	27	H

Derived fields for calculating total higher education FTE

SNEXCL1

18. This field identifies records that start after the academic year of interest.

Value	Description	Definition
1	Excluded	ENGSTARTDATE > 31 July 20Y2
0	Included	Otherwise

SNEXCL2

19. This field identifies records that finish before the academic year of interest.

Value	Description	Definition
1	Excluded	ENGENDDATE < 01 August 20Y1 and ENGENDDATE ≠ BLANK
0	Included	Otherwise

SNEXCL4

20. This field identifies records in the Data Futures Beta (21056) data that have been inactive for the reporting year and recorded as dormant or on an intercalating year.

Value	Description	Definition
1	Excluded	Z_STATUS_CYC = 02,03
0	Included	Otherwise

SNEXCL8

21. This field identifies incoming exchange students and students who are primarily outside the UK for their engagement in the Data Futures Beta (21056) data.

Value	Description	Definition
1	Excluded	INCOMINGEXCHANGE = 01,02,03 or Z_PRINONUK = 1
0	Included	Otherwise

SNEXCL

22. This field indicates whether the record will be included in the calculation of student numbers SNEXCL captures exclusions using applicable values from the table below.

Value	Description	Definition
1	Record starts after academic year of interest	SNEXCL1 = 1
2	Record ends before academic year of interest	SNEXCL2 = 1
4	Student is inactive for the entire reporting year	SNEXCL4 = 1
8	Incoming exchange student	SNEXCL8 = 1

23. This field contains the exclusion reason(s) for the instance. It is calculated as $(1 \times SNEXCL1) + (2 \times SNEXCL2) + \ldots + (8 \times SNEXCL8)$.

SNLEVELNUM

24. This field indicates the level of study for higher education records, using the sector recognised standards as set out in condition of registration B5.[1]

Value	Description	Definition
8	Records at Level 8	QUALCAT = D0003, D0004, D0005, E0000, E0001, E0002, E0003, E0004, E0005, L0000
7	Records at Level 7	QUALCAT = L0001, L0002, L0003, M0002, M0003, M0004, M0006, M0007, M0008, M0009, M0010, M0011, M0012, M0013, M0015, M0016, M0017, M0018, M0020, M0021, M0022, M0023, M0024
6	Records at Level 6	QUALCAT= H0003, H0004, H0005, H0006, H0007, H0008, H0009, H0010, H0012 H0013, H0014, H0015, H0016, H0018, H0019, H0020, I0001
5	Records at Level 5	QUALCAT = I0002, I0004, I0005, I0006, I0007, I0008 I0009, I0010, I0012, I0013, J0000, J0001, J0002, J0003, J0004, J0005, J0006, J0007, J0010, J0011, J0012
4	Records at Level 4	Otherwise

---

[1] www.officeforstudents.org.uk/media/53821cbf-5779-4380-bf2a-aa8f5c53ecd4/sector-recognised-standards.pdf

SNTOTHEFTE

25. This field calculates the total higher education FTE per student including any capping.

Value	Definition
(Min(Z_STULOAD_CYC,150))/100	SNEXCL = 0 and Z_STULOAD_CYC ≠ BLANK
0	Otherwise

Additional fields required to calculate total undergraduate higher education FTE relevant to TEF

SNREL_TEF

26. Identifies undergraduate provision relevant to TEF.

Value	Description	Definition
1	Provision in-scope for TEF	SNEXCL = 0 and QUALCAT in (C0000, C0001, C0002, C0003, C0004, C0005, C0006, C0007, C0008, C0009, H0003, H0004, H0005, H0006, H0007, H0008, H0012, H0015, H0016, H0018, H0019, H0020, I0001, I0004, I0008, I0009, I0010, I0012, I0013, J0000, J0001, J0002, J0003, J0004, J0005, J0006, J0007, J0010, J0011, J0012, M0002)
0	Provision not in-scope for TEF	Otherwise

SNTOTHEFTE_TEF

27. The total undergraduate students higher education FTE relevant to TEF per student including any capping.

Value	Definition
SNTOTHEFTE	SNREL_TEF = 1
0	Otherwise

Rebuild instructions

28. The engagement individualised file, SN21_DCT _XXXXXXXX_ENG.csv, can be used to interrogate the data and our calculation of derived fields by filtering to certain groups of records. 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).

29. Specific instructions to rebuild the student number summary figures in tables 1 and 2 of the student numbers data summary workbook, SN21_DCT_BETA_XXXXXXXX.xlsx, are provided here.

30. A number of additional fields not described in the algorithms above have been added to the individualised file. These are listed in Table 4.

Table 4: Additional fields in the individualised file

Field name	Description
REBUILDLEVFLAG	A column provided to enable the figures in Table 1 to be rebuilt.

31. The following instructions assume the user has the engagement individualised file open in Microsoft Excel.

32. To apply filters to the column headings, highlight the headings. From the 'DATA' tab in the toolbar select 'Filter'.

33. You will need to use the SUBTOTAL function with option 9 to sum the values in the relevant columns once the filters have been applied. Using the SUBTOTAL function instead of the SUM function will ensure that you only include those values that have been filtered to.

Table 1 Split of higher education by SNLEVELNUM

34. To allow you to calculate the total HE FTE at each individual level (4, 5, 6, 7 or 8), the REBUILDLEVFLAG is provided. This field can be used to select one row of data per student per level of higher education study, to prevent double counting the FTE where an individual has multiple aims at the same level of study.

35. To rebuild the figures in table 1 of the workbook, filter to REBUILDLEVFLAG = 1. The subtotal for SNTOTHEFTE, will give the total FTE in Table 1. To rebuild the figures for each level of study, filter SNLEVELNUM to 4, 5, 6, 7 or 8 as required.

Table 2 Higher education provision relating to undergraduate levels of study relevant to TEF (SNREL_TEF)

36. To rebuild the figures in table 2 of the workbook, first clear any filters. Filter to REBUILDLEVFLAG = 1 and SNREL_TEF = 1. The subtotal for SNTOTHEFTE_TEF, will give the total FTE in Table 2.