

Specification for the migration of legacy data to the Data Futures data model.

MigrationPopulation

Version 0.1.1: Revision history at foot of document

Overview

This document provides the specification for the migration of legacy data student population identifiers into the Data Futures data model.

This is one of a series of specifications produced to translate or map legacy data submitted against the CXX051/54 collections into the Data Futures data model (056). The data migration activity will allow quality assurance checks to be run on the first Data Futures collection in 2022/23 for data submitted by providers.

Fields covered in this specification

Field name	Field description	Field version	Field length	Field type	Field level
Z_ANNCOM	Identifies the anniversary of the commencement date in the collection year.	0.0.0		Date	1
Z_COLLECTION_YEAR	Identifies the collection year from which the data has been drawn. For the Alpha pilot, the collection years are 2019/20 and 2018/19. For the Beta pilot, the collection years are 2020/21 and 2019/20.	0.1.0	7		1
Z_POPMIGRATION	Identifies which student instances and collection years have been migrated in the creation of Student course sessions.	0.1.0	1	BIT	2
Z_PERIODSTARTDATE	Identifies the start date of the migrated instance year.	0.1.0		Date	3
Z_PERIODENDDATE	Identifies the end date of the migrated instance year.	0.0.0		Date	3

Migration.Z_ANNCOM – the anniversary of the commencement date which falls within the latest collection year.

Migration.Z_COLLECTION_YEAR – the year from which the data has been drawn. The Student and Engagement migrated entities use the last submitted data. The Student course session migrated entity uses data from the year it was drawn from, as defined by Z_POPMIGRATION.

Migration.SCS.Z_POPMIGRATION – identifies which student instances and corresponding collection years are included in the migration population.

Migration.SessionYear.Z_PERIODSTARTDATE (Student record 051) – the anniversary of the commencement date representing the start of the instance year.

Migration.SessionYear.Z_PERIODSTARTDATE (Student Alternative record 054) – the earliest start date of the combined instance periods.

Migration.SessionYear.Z_PERIODENDDATE (051) – for courses with standard year type, the end of the academic year and for courses with non-standard year type, the period start date plus one year minus one day. The exception is for students on short courses which, in which case the expected end date is used.

Migration.SessionYear.Z_PERIODENDDATE (054) – the latest end date of the combined instance periods.

The terms **combined instance period** (054) and **duplicated instances** (051) are introduced in this specification.

Valid entries

Z_ANNCOM, Z_PERIODSTARTDATE, Z_PERIODENDDATE

A valid date.

Z_COLLECTION_YEAR

A string in the format 20XX/YY.

Z_POPMIGRATION

Code	Label
0	The record is NOT counted within the HE migration population
1	The record is counted within the HE migration population

Dependent fields

- Migration.SessionYear.SYSTARTDATE
- Migration.SessionYear.SYENDDATE
- Migration.StudentCourseSession.SCSSTARTDATE
- Migration.StudentCourseSession.SCSENDDATE

Depend upon fields

- Instance.HUSID
- Instance.UKPRN
- Instance.NUMHUS
- Instance.COMDATE
- Instance.COURSEID
- Instance.ENDDATE
- Instance.LOCSDY
- Instance.MCDATE
- Instance.MODE
- Instance.NOTACT
- Instance.REDUCEDI
- Instance.YEARSTU
- Instance.YEARPRG
- Instance.XLEV301
- Migration.Z_NONSTANDARD (051)
- Migration.Z_SHORTCOURSE [Z_PERIODENDDATE]
- Migration.Z_EXPECTEDENDDATE [Z_PERIODENDDATE]
- Migration.Z_POPMIGRATION [Z_PERIODSTARTDATE, Z_PERIODENDDATE]

Additional information

In preparation for Data Futures, the latest available data from the legacy collection is being migrated into the Data Futures model to allow certain comparisons to be made between legacy data and the incoming Data Futures return. This will help ensure the incoming data is credible and support data quality.

The migrated data includes student instances which would be returned in Data Futures if a provider was returning data for the cycle relating to the latest legacy collection. The term collection year will be used to identify the academic year from which the data has been drawn.

For the alpha pilot, data was drawn from the Student and Student Alternative records 2019/20 (Z_COLLECTION_YEAR = 2019/20) with some data taken from 2018/19 (Z_COLLECTION_YEAR = 2018/19). For the beta pilot, data will be drawn from the Student and Student Alternative records 2020/21 with some data taken from 2019/20. Where a student is present in both academic years, the student and instance characteristics (including entry profile) has been drawn from the latest collection year.

The migrated population includes all HE level instances returned in the latest legacy collection (Student and Student Alternative record) plus a subset of HE level instances from the previous year whose instance year of study overlaps the latest collection year.

For the Student record, the migrated Engagement population includes all student instances (based on the combination of HUSID-UKPRN-NUMHUS) returned in the latest collection year plus any additional instances returned in the previous collection year for which the course was non-standard (spans the standard academic year 1 August to 31 July) and where no end date has been recorded. Student instances with a standard year type in the previous collection year are not included in the migration population as their year of study will end before the latest collection year.

For the Student Alternative record, the migrated Engagement population includes all student instances (based on the combination of HUSID-UKPRN-NUMHUS) returned in the latest collection year plus any additional instances returned in the previous collection year with an instance period which ends in the following academic year and where no end date has been recorded.

The migrated Student course session population (Z_POPMIGRATION) identifies which student instances and collection years have been used to create the Student course sessions. The migrated instance year start and end dates are used to create the dates for the SessionYear and StudentCourseSession entities. The derived instance years are created for all types of students, including postgraduate research and fully flexible students for which the SessionYear entity is not required.

Student record

In the absence of start dates of each year of a programme in the legacy collection, the migrated SessionYear entity is based on anniversary of the commencement date which falls within the latest collection year (Z_ANNCOM).

Standard year type (Student record)

For standard year types, only those returned in the latest collection year are included in the migration and will have their instance year start date set to the anniversary of the commencement date which falls within the latest collection year.

The instance year end date will be one year on from the instance year start date, minus one day, unless the course has been identified as a short course, in which case the expected end date will be used.

Non-standard year type (Student record)

For non-standard year types (see Z_NONSTANDARD specification), the data returned in the latest collection year is used to identify if the information relates to the tail end of a year of study which started in the previous collection year or the start of a year of study starting in the latest collection year and continuing into the next or both.

For students who suspend studies before the anniversary of their commencement date in the latest collection year (Z_ANNCOM) and those who have been dormant, on sabbatical or who have been studying primarily overseas for the entire latest collection year, the data returned in the latest collection year will be used to create two student course sessions spanning the whole period. One of these instance years will start in the latest collection year and one will start in the previous collection year.

Students who end before the anniversary of their commencement date in the latest collection year (Z_ANNCOM) + 4 days will have only one instance year created which will start in the previous collection year.

For remaining student instances which are active at the end of the latest collection year, the instance year relating to that instance will start on the anniversary of their commencement date in the latest collection year (Z_ANNCOM). Where available, data from the previous collection year will be pulled forward to create the instance year starting in the previous collection year.

As with the standard year type, the instance year end date will be one year on from the instance year start date, minus one day, unless the course has been identified as a short course, in which case the expected end date will be used.

Student Alternative record

In the Student Alternative record, instance periods will be combined to create the student course sessions. Where two or more instance periods relate to the same instance and have the same course identifier, mode of study, year of programme and year of student, these instance periods have been combined (**combined instance period**) where they do not result in a combined period of longer than 1 year plus 14 days.

The combined instance periods have start date as the earliest of the instance period start dates and have end date as the latest of the instance period end dates. Where any of the instance periods in the combination have overlapping dates, the one returned in the latest collection is retained and the overlapping ones from the previous collection disregarded.

Combining instance periods in this way may result in combined instance periods being longer than 1 year plus 14 days (the maximum instance period length permitted). In this scenario, the combined instance period has been split to form multiple combined instance periods for the same instance. If one or more instance periods commence in the same month as the original commencement date, the combined instance period is split at this point. Otherwise, take each instance period in ascending instance period start date order and split such that each part is no longer than 1 year plus 14 days, but including the next instance period would form an instance period longer than 1 year plus 14 days. This means that the last split may leave a relatively small part.

For any combined instance periods with length between 1 year plus 5 days and 1 year plus 14 days, the start date is rolled forward so that the resulting combined period is exactly 1 year plus 4 days long. This is because instance years in Data Futures cannot be longer than 1 year plus 4 days, but instance period in legacy can be up to 1 year plus 14 days. This avoids splitting instance periods.

Notes







Each migrated instance year can be no longer than 1 year plus 4 days in length. If a student ends their course up to 4 days after the anniversary of their commencement date, an instance year of just over a year is permitted and the instance year end date is adjusted to be the end date of the student.

Instance years are only shorter than a year in length for students on courses which are expected to last for less than one year and the expected end date did not pass before the student ended.

Students for which there was no instance returned in the latest collection year, but there was an instance returned in the previous collection year are included in the migrated dataset provided that they are on a course that spans the standard academic year and they do not have an end date.

Technical Specification

Student record (051)

Key	
	Anniversary of commencement date in latest collection period (Z_ANNCOM)
	Continuation of session year
	Data from latest collection period (active)
	Data from latest collection period (dormant or suspended studies)
	Data pulled forward from previous collection period
	Data excluded from migration

	Previous reporting period	Latest reporting period
1	Standard year type Data excluded from migration	Active
2		Course starting
3	Continuation of session year	Course ending
4	Data excluded from migration	Course suspended
5	Data excluded from migration	Suspended / dormant
6		Dormant
7		Short course
8	Non-standard year type	Active
9	Data pulled forward from previous collection period	Active
10		Course starting
11	Data pulled forward from previous collection period	Course starting (previous period)
12		Course ending (latest period)
13	Data pulled forward from previous collection period	Course ending
14		Course suspending
15		Suspended / dormant
16		Course suspending
17		Suspended / dormant
18		Dormant
19		Dormant
20		Primarily overseas

The illustration above represents how student instances are migrated into instance years, based on the information returned in the latest collection.

Scenarios 1-6: One instance year is required for standard year types.

- The instance starts and ends within the collection year.

Scenarios 8 and 10: One instance year is required for students starting within the latest collection year and students ending their year of study which started in the previous collection.

- The instance ends within the collection year, and the end date is on or before the anniversary of the commencement date in the collection period plus 4 days. but not before commencement date itself.
- The instance starts within the collection year.

Scenarios 7, 9 and 11: Two instance years are required for non-standard year types. Where the student is active (excluding studying primarily overseas), data from the previous collection year represents the first instance year and data from the latest collection represents the second instance year.

- The instance relates to an active or writing up student.

Scenarios 12-15: For the remaining scenarios, the data returned in the latest collection represents the student's activity since the start of the collection period and therefore two instance years are created using the data from the latest collection.

- The instance relates to a mode change to dormant and the mode change took place before the anniversary of the commencement date in the collection period.
- The instance relates to sabbatical or dormant mode. When a student is returned with a dormant or sabbatical mode of study, it suggests the student has been in that mode for the entire reporting period.
- The instance relates to a dormant reduced return.
- The instance relates to a record for which the student has suspended their studies. As with legacy, students who are on non-standard year type who suspend studies are assumed to be inactive from the start of the reporting period and are excluded from the standard registration population.
- The instance relates to a record for which the student is studying primarily overseas. As with legacy, students who are on non-standard year type who are studying primarily outside the UK are excluded from the standard registration population.

Z_POPMIGRATION (051: Student course session population)

Identifies which student instances are included in the migrated Student course session population. Where a student instance is returned in both the latest and previous collection years, Instance_Latest denotes the information returned in the latest collection.

Note that for certain scenarios, instances returned in the latest collection year are duplicated to create two Student course sessions which span the entire year, rather than pulling forward data from the previous collection. Examples include instances which are dormant, or the study is primarily overseas in the latest collection year, this ensures we do not wrongly include some activity which took place in the previous cycle.

Migration. Z_COLL ECTION_ YEAR	Instance. XLEV301	Instance. ENDDATE	Instance. Z_NON STANDARD	Instance _Latest. ENDDATE	Instance _Latest. MCDATE	Instance _Latest. MODE	Instance _Latest. REDUCEDI	Instance _Latest. LOCSDY	Instance _Latest. NOTACT	Migration. SCS. Z_POP MIGRATION
Latest	3									0
	1, 2									1
Previous	3									0
	1, 2	Not NULL								0
		NULL	0							0
			1		Not NULL and ENDDATE<= Z_ANNCOM +4					0
					Otherwise	Not NULL and <ANNCOM				0
					Otherwise	51, 63, 64				0
						Otherwise	04			0
							Otherwise	S		0
								Otherwise	1,2	0
								Otherwise	1	

Z_PERIODSTARTDATE (051)

Migration. SCS.Z_POP MIGRATION	Instance. Z_NON STANDARD	Instance. ENDDATE	Instance. MCDATE	Instance. MODE	Instance. REDUCEDI	Instance. LOCSY	Instance. NOTACT	Migration. SessionYear. Z_PERIOD STARTDATE
0								NULL
1	0							Z_ANNCOM
	1	Not NULL and ENDDATE<= Z_ANNCOM+4						Z_ANNCOM - 1 year
		Otherwise	Not NULL and <Z_ANNCOM					
			Otherwise	51, 63, 64				
				Otherwise	04			
					Otherwise	S		
				Otherwise	1,2			
						Otherwise	Z_ANNCOM	

Create additional instance years as follows and where Instance.ENDDATE is NULL:

Migration. Z_COLLEC TION_YEAR	Instance. Z_NON STANDARD	Instance. MCDATE	Instance. MODE	Instance. REDUCEDI	Instance. LOCSY	Instance. NOTACT	Migration. SessionYear. Z_PERIOD STARTDATE
Latest	1	Not NULL And <Z_ANNCOM					Z_ANNCOM
		Otherwise	51, 63, 64				
			Otherwise	04			
				Otherwise	S		
					Otherwise	1,2	

Instances taken from the latest collection which have no end date (ENDDATE is NULL) and relate to a non-standard year type (Z_NONSTANDARD = 1) are referred to as **duplicated instances** where any of the following occur (they can be identified in the migrated data as having D appended to the end of their SCSESSIONID):

- TYPEYR = 5 (non-standard year ending)
- MCDATE not NULL and < Z_ANNCOM (mode change before the anniversary of commencement date)
- NOTACT = 1, 2 (suspended studies)
- MODE = 51, 63, 64 (sabbatical or dormant)
- LOCSDY = S (primarily overseas)

Note that for **duplicated instances** ending in the latest collection (TYPEYR = 5) and those suspending studies before the anniversary of the commencement date, there will be no corresponding Student course session starting in the latest collection as these students appear to be dormant for the remainder of the academic year.

Z_PERIODENDDATE (Student record)

Migration.Z_POPMIGRATION	Migration.Z_SHORTCOURSE	Migration.Z_NONSTANDARD	Migration.SessionYear.Z_PERIODENDDATE
0			NULL
1	1		Z_EXPECTEDENDDATE
	0	0	(Y2)-07-31
		1	Z_PERIODSTARTDATE + 1 Year – 1 day

Student Alternative record (054)

The instance year start and end dates are set to the earliest instance period start date (Z_EARLIESTSTART) and latest instance period end date (Z_LATESTEND) respectively of the combined instance periods. Where the combined instance period results in the length of the instance period being between 1 year plus 5 days and 1 year plus 14 days, Z_PERIODSTARTDATE is shifted forward by up to 10 days to ensure that the length of the combined period is exactly 1 year plus 4 days. Due to the way in which the instance years have been created, they will not overlap where there are multiple for a single each instance.

Z_PERIODSTARTDATE (054)

Migration.Z_POPMIGRATION	Migration.Z_LATESTEND	Migration.SessionYear.Z_PERIODSTARTDATE
0		NULL
1	> Z_EARLIESTSTART + 1 year + 4 days	Z_LATESTEND – 1 year – 4 days Z_EARLIESTSTART

Z_PERIODENDDATE (054)

Migration.Z_POPMIGRATION	Instance.Z_SHORTCOURSE	Migration.SessionYear.Z_PERIODENDDATE
0		NULL
1	1	Z_EXPECTEDENDDATE
	0	Z_LATESTEND

Revision history

2021-11-10 0.1.0 Updated text to add further clarification to highlight where the derivations are collection specific. Highlighted that Z_POPMIGRATION defines which instances and from which collection year the data is drawn from to create the migrated Student course sessions.

2022-02-14 0.1.1 Re-formatting of document, introductory text altered. Some references to Student Records and Student Alternative Record changed to 051 and 054 respectively.