

Version 0.3.0

## SUMMARY OF DATA FUTURES MIGRATED DATASET

### Drafting note

This is a working document being produced for statistical purposes to enable analysts to understand the best way for data to be mapped from the legacy data model to the Data Futures data model. The migrated data process is to facilitate QA and to enable submission of accurate data to HESA by providers.

### Summary

Data Migration is the translation of legacy data submitted against the Student record (051) and Student Alternative record (054) collections into the Data Futures data model (056). This is being undertaken to allow quality assurance checks to be run against data submitted by providers for the first data futures live collection in 2022/23.

Migrated data will be loaded into the HESA data platform (HDP) to be used as part of the quality assurance process for the first year of Data Futures live collection. It will form the basis of historical continuity and credibility checks in the transition from legacy to the Data Futures model. Since data models differ at both entity and field level, it is important we migrate the data as accurately as possible to ensure robust assurance can be undertaken.

The goal of the Alpha data migration activity was to refine the migration specification in collaboration with Alpha providers to ensure it is fit for purpose. We will only migrate data that will be used either directly or to support transition year quality assurance. We will only use migrated data for quality assurance where the migration process is deemed robust enough to do this.

To assist participating providers with the refinement, we provided a migrated dataset to each Alpha participant along with a detailed migration specification to support understanding of this data for the transition year submission/QA process. The dataset contained migrated fields, Data Futures derived fields (prefix Z\_), original 'legacy' fields (prefix F\_) and 'legacy' derived fields (prefix F\_X) to allow for comparisons from old to new.

For Beta, the migrated data has been loaded into the HDP and entity based migrated data files with Data Futures derived fields (prefix Z\_) enhancements appended are available to providers.

The migration documents include details of the mappings between the legacy and Data Futures fields, highlighting where there is no direct mapping between the legacy data and the Data Futures model. Where there is no direct mapping, a change to the codes or the migration labels deviate from those listed in the coding manual, this is detailed within in document.

Please be aware that whilst field labels may be the same for legacy and Data Futures, the coverage may differ. The coding manual provides full lists of Data Futures valid entries and coverage of the data: <https://codingmanual.hesa.ac.uk/22056/>

The migrated dataset includes information drawn from the latest collection year and from the previous collection for students whose year of study appears to continue into the latest academic year (only applicable for non-standard year types). For the Alpha pilot, data was taken from the Student and Student Alternative records 2019/20 and in part from 2018/19. For Beta, data was taken from the Student and Student Alternative records 2020/21 and in part from 2019/20.

Where a student is present in both collection years, the Student and Engagement information (including sub-entities) have been drawn from the latest collection year. Where appropriate, the Student course session information has been drawn from the collection year itself.

Details of how the migration population has been created can be found in the Migration.MigrationPopulation document. A number of new concepts have been produced to use in the derivation of the population:

- **Z\_EXPECTEDENDDATE** – the expected end date of the course, based on the commencement date (COMDATE) and length of course (SLENGTH, UNITLGTH).
- **Z\_NONSTANDARD (Student record only)** – a course with a non-standard year type (TYPEYR = 2,3,4 or 5) or unknown year type and the month of commencement (month of COMDATE) is between January and July.
- **Z\_SHORTCOURSE** – a course which is expected to last for less than one year and the student has not already passed the expected end date.
- **Combined instance periods (Student Alternative record only)** – consecutive instance periods combined where the student identifier (HUSID), provider identifier (UKPRN), instance identifier (NUMHUS), course identifier (COURSEID), mode of Study (MODE), year of programme (YEARPRG) and year of student (YEARSTU) are the same. The resulting combined instance period can't be longer than 1 year + 4 days, so there may be multiple combined instance periods created for the same instance.
- **Duplicated instances (Student record only)** – Student course sessions created using data from the latest collection year which are “duplicated” to cover the entire academic year, the Student course session which starts in the previous academic year and ends in the latest one is the duplicated instance. Duplicated instances only apply to non-standard year types (Z\_NONSTANDARD = 1).

The migration population has been used as the basis for the mapping the various entities. No attempt has been made to migrate all entities or all fields in each migrated entity.

**Important note:** It is important to note that although placeholder Session year and Student course sessions have been created, they have been generated using the anniversary of the student's commencement date and may not be a true reflection. For this reason, the start and end dates associated with these entities will not be used as part of the quality assurance process.

The following documents provide detail of the migration and highlight which fields have been migrated and the detail of how it has been done:

- **06Migration.Student** – Student entity, including Disability sub-entity.
- **07Migration.Engagement** – Engagement entity, including Leaver, CollaborativeProvision, EntryProfile, EntryQualificationAward, EntryQualificationSubject, StudentAccreditationAim related entities.
- **08Migration.StudentCourseSession** – Student course session entity, including SessionYear, Course, Course.Qualification, Course.Qualification.QualificationSubject and Funding and monitoring related entities.
- **09Migration.SessionStatus** – SessionStatus entity.
- **10Migration.DerivedSubject** – A derived subject of study and full person equivalence entity.
- **11Migration.ModuleInstance** – Module instance entity, including Module, ModuleCostCentre, ModuleDeliveryRole related entities.
- **12Migration.Course.Accreditation** – Course accreditation entity.

- **13Migration.QualificationAwarded** – QualificationAwarded entity, including QualificationAwarded.Qualification, QualificationAwardAccreditation related entities.
- **14Migration.AwardingBodyRole** – Awarding body role entity.

Please refer to the individual migration documents for description of migrated datasets.

## Revision history

2021-10-07 0.1.0 Updated text following Alpha pilot. Additional text relating to course accreditation (not supplied as part of Alpha). Updated link to coding manual. Important note relating to Session years and Student course sessions added.

2022-01-31 0.1.1 Updated text to remove reference to colour coding of mapping types. References to SR and SAR changed to 051 and 054 respectively.

2022-02-23 0.2.0 National Identity migrated entity removed to reflect removal from the model. Description of migrated datasets removed from summary (available in individual documents) and link to coding manual updated to 21056.

2022-07-21 0.3.0 Updated to reflect the data available as part of Beta. List of entities covered in each document updated to reflect changes made to the migration documents.