KIS Extract and data population

Technical Design document

Purpose

This document covers the software developed to provide the facility to extract the required data for the KIS data set return, allow user review and amendment and the generation of the XML format file for upload to HESA.

Background

The KIS collection (C12061) is a new data feed required by HESA, starting from the 12/13 academic year. See HESA - Higher Education Statistics Agency
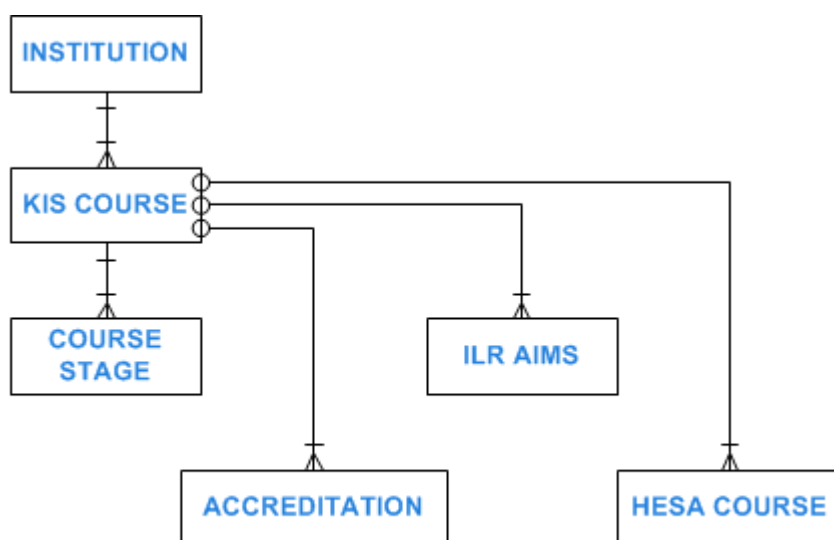
The data model from HESA is as follows.

Diagram last saved 2011-08-17

The model consists primarily of a parent Institution record containing data relative to the submitting institution (The University of Sussex in our case), a number of Course records detailing the courses provided that prospective students can apply for (The old Sussex terminology was Programmes) at UG level only (no PG level courses are included in the returns) and details of the teaching and assessments for each year of the course. Further details show any external accreditations available as a part of the course and linkages to the same course in previous HESA reporting sets. For further education institutions (not Sussex) ILR aims also are supplied, but as they are not relevant to Sussex, these will not be described further.

Functionality

The functionality for the extract is provided by

- New database tables KIS_INSTITUTIONS, KIS_COURSES, KIS_COURSE_STAGES, KIS_ACCREDITATIONS and KIS_HESA_COURSES based on the HESA data model above.

- Screen SSTU0289 to allow the user to control the data extract, review/update the extracted data and produce the XML file for upload.
- Package HESA_C12061_RETURN_PACK that has the required extract and XML production code.
- Views KIS_INSTITUTIONS_VIEW, KIS_COURSES_VIEW, KIS_COURSE_STAGES_VIEW, KIS_ACCREDITATIONS_VIEW and KIS_HESA_COURSES_VIEW to support the XML file production.

For the initial release, a number of data management scripts and other changes were required, but it is not anticipated that these will need to be altered in future years.

- Change to Screen SSTU0080 to use one of the unused spec value codes (SPEC_VALUE9) on HES_ASSESSMENT_METHODS to hold and maintain the KIS assessment classification.
- Data population script KSTU0001 to set the initial values for the KIS assessment classifications.
- Script LSTU0500 to add additional columns to SUSX_PROG_PROPOSALS and its Archive table.
- Change to SSTU0033 to allow for the maintenance of the KIS data columns on the Course records (Programmes in old terminology).
- Data population script KSTU0002 to set the initial values for KIS course types.
- Changes to package SUSX_PROG_PACK to roll forward the KIS information when courses are rolled forward.
- FOR 2012 ONLY, there is an additional script estimates.sql to override the calculated teaching and assessment figures with supplied estimates from the schools.

Data Population

Call HESA_C12061_RETURN_PACK. pr_populate_data with parameters

pa_kis_reporting_year = reporting year eg: 2012

pa_kis_last_session = last academic session (for the teaching/assessment details) eg: 11/12

KIS_INSTITUTIONS

The data for this entity is the institution identifier and student accommodation information. As this is not held in the main database, this is populated as fixed values within the script based on business supplied information. The UKPRN and RECID will not change as these are the Institution identifier and data set identifier, but the number of beds and accommodation cost range is likely to be different next year. The URL for the accommodation information is less likely to alter, but this should be checked.

KIS_COURSES

Basic selection is to select all courses from SUSX_PROG_PROPOSALS where the academic year is the supplied reporting year , the classification is UG and the status is CONFIRMED. The qaf and programme type are also tested as only those courses directly run by Sussex for UG students are needed. The checks on the hesa_reportable flag and KISTYPE field are to allow user suppression of Courses by managing this data against the Course. Setting (or clearing) the hesa_reportable flag

would suppress the Course from all HESA reporting, just removing the KIStype would only exclude it from the KIS data collection.

If the Course has a UCAS code then all fee related information comes from a lookup done by HESA against the UCAS database, and therefore information is blanked out prior to writing the data record.

Within the KIS_COURSE record are a series of URLS to allow prospective students to access aspects of the course information such as course details, employment prospects, support available, teaching and assessment details. Generating these URLs is complicated by the way that a different department code is used for the website prospectus compared with the actual department running the course, and the structure is subject to change, so this needs to be reconfirmed in future years.

There are three different types of Courses that may be reported, signified by the KISTYPE, and the data reported differs for each. KISTYPE = 1 is in essence single degree courses. By default (and certainly for the 2012 reporting) all Sussex courses fall into this category. KISTYPE = 2 is in essence placeholder for a true joint degree course. The data for these is minimal apart from the reference via the RELATEDKIS fields pointing to the single degree courses for lookup. The URLs are likely to be different in content, so these are required. KISTYPE = 3 are subject level courses, for which there is no single degree option required. They describe the subject element of a joint course to allow the KISTYPE = 2 to be the placeholder. See the HESA information for further details.

KIS_COURSE_STAGES

For any KIS_COURSE record, there should be a KIS_COURSE_STAGES record for every year of the course (FT only), so a 3 year degree programme would have 3 records. A one year foundation course would have one etc. The stages are determined by the FT entries in SUSX_PROG_PERIODS, but NOTE the requirement is a number between 1 and 9, so if there is a period 0 (foundation), the period numbers are all incremented by 1.

For each stage, the requirement is to provide the breakdown of the Course learning split between Scheduled (ie: Taught), Independent and Placement (no differentiation between YAB and commercial placement) and also the split between the different methods of assessment (coursework/written (exams)/practical). Each of these two breakdowns must add up to 100 (or be 0) across the three elements contained within them, and therefore rounding needs to be considered.

When calculating the splits, where a Course consists of a number of module options (student choice is a factor) then the requirement is to base the information on the most common/popular set of options and therefore in the selection of this data the modules within the course (SUSX_PROG_COURSES) are ordered by the number of students associated with the module descending. A further wrinkle is that they want the splits based on total module number of credits being a maximum of 120 credits, and therefore the calculation selection stops once 120 credits have been read, and the final module read may be pro-rated.

The package includes the code for working out these splits, but for 2012 Sussex is providing estimated figures due to some course elements (particularly practical and placement) not being well defined in current courses. This is the reason that for 2012, the separate scrip estimates.sql is run. For future years, just do not run the separate script that populates the estimates.

NOTE the rounding effect of calculating percentages means that the figures may not always add up to 100 (eg: even split between coursework/written and practical would be 33/33/33 = 99. A priority is therefore built in to round up one of the elements to ensure that the total is 100.

KIS_ACCREDITATIONS

This data is based on SUSX_PROG_ACCREDITATIONS linking in the list of accrediting bodies supplied by HESA and incorporated into the system earlier.

KIS_HESA_COURSES

Except for new courses, there is a requirement to link the courses to the data for the same course reported to HESA in previous years (primarily to allow linkage of student satisfaction and actual employment data). The requirement is to provide the course identifiers used in the previous two HESA returns. Due to the timing for 2012 as it is actually based on 2011 module data, so these are 2010 and 2009, in future years this will not be the case, so there is a line to remove from the procedure artificially stepping back a year.

 As the prior returns also had elements of recursion in them, this means that the course may have relevant data just for 2009, just for 2010 or is referenced in both. This is determined by matching the prior prog_seq values for the same course (programme) code to the STU_HESA_COURSE_ARC data.

Gary Wilkinson

Aug 2012