September 2008/30

Core funding/operations

Assurance and funding

This report is for information and guidance

This document is a guide to HEFCE's web facility which generates several types of analysis from Higher Education Statistics Agency (HESA) student data. Use of the web facility will help higher education institutions to return accurate data to HESA, and to identify errors and forecasting discrepancies in HEFCE funding returns.

2007-08 statistics derived from HESA data

Guide to HEFCE web facility

HIGHER EDUCATION FUNDING COUNCIL FOR ENGLAND

Contents

2007-08 statistics derived from HESA data

Guide to HEFCE web facility

To	Heads of HEFCE-funded higher education institutions Heads of universities in Northern Ireland
Of interest to those responsible for	Student data, Audit, Finance
Reference	2008/**30**
Publication date	September 2008
For enquiries regarding:	For all enquiries (except on the use of HESA data to inform the 2009-10 widening participation allocations), contact: Lydia Booker tel 0117 931 7338 e-mail hesa_heses_stats@hefce.ac.uk For enquiries on the use of HESA data to inform the 2009-10 widening participation allocations, contact: Christine Daniel tel 0117 931 7373 e-mail hesa_heses_stats@hefce.ac.uk

Executive summary

Purpose

1. This document is a guide to HEFCE's web facility which is provided to institutions so that they can verify and correct, where appropriate, their 2007-08 Higher Education Statistics Agency (HESA) student data prior to submission. It generates the following outputs:

a. Higher Education Students Early Statistics Survey 2007-08 (HESES07) re-creation.
b. HESES07 re-creation based on cost centre sector norms.
c. Derived statistics that may be used to inform the 2009-10 widening participation (WP) allocations.
d. Derived fields that may be used to inform the 2009-10 partial completion weighting.
e. Re-creation of the aggregate return to monitor 2007-08 co-funded employer engagement student numbers (CFEE07) – a new output this year.
f. Research Activity Survey 2007 (RAS07) re-creation.
g. Research Degree Qualification Rates (RDQRs).
h. 2007-08 HEFCE-fundable student full-time equivalences (FTEs) for the Transparent Approach to Costing (Teaching), TRAC(T).
i. Derived statistics that are likely to inform HESES08 audits.
j. Lifelong Learning Network (LLN) student summaries.
k. Higher Education Students Early Statistics Survey 2008-09 (HESES08) non-completion toolkit.

Key points

2. The use of the web facility is strongly encouraged because it will help institutions to:

- return accurate data to HESA
- reduce the likelihood of selection for the derived statistics exercise for 2007-08 (selected institutions are typically subject to considerable additional work and the potential of funding adjustments)
- identify discrepancies between forecasts in HESES07 and the outturn position for 2007-08
- identify errors in HEFCE funding returns
- verify that the derived fields that may inform funding allocations are accurate
- verify that RDQRs are suitable for publication
- verify that 2007-08 HEFCE-fundable student FTEs for TRAC(T) are suitable to inform the periodic review of price groups
- verify that the derived statistics are suitable to inform HESES audits
- verify that the student summaries for LLNs are suitable to inform the evaluation of the initiative
- make informed estimations of non-completion rates when compiling HESES08.

3. This document (including the technical appendices which can be downloaded separately) provides:

- guidance for using the web facility
- details of the algorithms used to generate the derived statistics
- guidance on troubleshooting the differences between the HEFCE funding returns and the re-creations
- guidance on re-building the web facility outputs using the individualised data files supplied
- details of problems of fit with algorithms
- guidance for submitting overrides to our algorithms, and information on when this is appropriate
- guidance on submitting historical data error (HDE) files to correct RDQRs
- guidance on the preparation and submission of RDQR action plans
- guidance for using the HESES08 non-completion toolkit.

Action required

4. Use of this web facility is optional, but we strongly encourage institutions to use it as part of their data quality processes. Past reviews have confirmed that it is an essential element of most higher education institutions' HESA data quality processes. These reviews have also revealed that institutions that do not use the facility are more likely to be selected for the monitoring exercise described in paragraphs 7 – 13 below.

5. This will be the last opportunity for institutions to verify the accuracy of the data underlying the RDQRs for full-time students starting in 2001-02 and part-time and mode-switch students starting in 1999-2000 before they are asked to sign off these rates as suitable for publication. This will also be the last opportunity to ensure that 2007-08 HESA data, where these will inform future RDQRs, are accurate.

6. The deadline for receipt of RDQR action plans (to enable us to authorise the submission of 2006-07 HESA amendments and HDE files) is **6 October 2008**. The deadline for signing off 2006-07 HESA amendments and HDE files for the purpose of correcting RDQRs is **13 October 2008**.

Relationship with 'Statistics derived from HESA data'

7. We use the annual 'Statistics derived from HESA data for monitoring and allocation of funding' exercise to monitor institutions' HESES, RAS and CFEE returns using HESA student data. This exercise occurs after we have received a final copy of all institutions' data from HESA, typically in the December following the web facility launch.

8. Our funding allocations are informed by the data provided by institutions. If we find, either through reconciliations with HESA data, or any data audit, that erroneous data have resulted in institutions receiving incorrect funding allocations (including for WP and other targeted

allocations), then we will adjust their funding accordingly (subject to the appeals process and the availability of our funds).

9. Any funding adjustments arising from the reconciliation of a re-creation of HESES07 from HESA 2007-08 student data (the HESES07 re-creation) with HESES07 or from the comparison of cost centre assignments with the sector norms for subjects (the HESES07 re-creation based on cost centre sector norms), are likely to affect the funding previously announced for 2007-08 and all subsequent years, and WP funding and other targeted allocations for 2008-09.

10. Any funding adjustments arising from the reconciliation of a re-creation of CFEE07 from HESA 2007-08 student data (the CFEE07 re-creation) with CFEE07 are likely to affect the funding previously announced for 2007-08, and WP funding and other targeted allocations for 2008-09.

11. Any funding adjustments arising from the reconciliation of a re-creation of RAS07 from HESA 2007-08 student data (the RAS07 re-creation) with RAS07 are likely to affect the funding previously announced for 2008-09.

12. In recent years the vast majority of institutions selected for the monitoring exercise have had their funding adjusted because we have identified that erroneous data were returned to HEFCE and that this had resulted in their receiving incorrect funding allocations.

13. Institutions selected to make a response to the monitoring exercise must typically undertake a substantial amount of work to reconcile their data, which may take several months to complete. While the web facility is provided to complement the monitoring exercise it does not replace it.

New outputs

14. This year's facility will produce derived statistics that may inform the 2009-10 partial completion weighting. Annex F provides further information on the methods we will use, while Appendix 8 contains technical information regarding the algorithms used to produce these data.

15. We will collect the CFEE07 return in August 2008 to inform funding for co-funded employer engagement. In the meantime institutions will be able to use the web facility to produce a CFEE07 re-creation based on 2007-08 HESA student data. The CFEE07 re-creation is a new output that has not been produced by past web facilities. It provides institutions with an early opportunity to verify that 2007-08 HESA student data are accurate for their co-funded employer engagement students. On receipt of 2007-08 HESA data later this year we will use the CFEE07 re-creation to monitor the CFEE07 return. Annex G provides further information on the methods we will use, while Appendix 9 contains technical information regarding the algorithms used to produce these data.

16. The web facility will also generate derived statistics likely to inform HESES08 audits. Annex N provides further information on the methods we will use, while Appendix 18 contains technical information on the algorithms used to produce these data.

Changes to the 2007-08 HESA student return

17. Following the redevelopment of the HESA student record for 2007-08 we have made substantial changes to our derived statistics algorithms. We may need to further refine these algorithms after the release of the web facility to improve the accuracy of the outputs. However, we expect any changes to be minor.

18. This year we are releasing this web facility in two phases. This is to:

a. Provide institutions with the earliest opportunity to utilise the outputs as part of their data quality processes.
b. Collect feedback on any necessary refinements to our algorithms before formal release of this publication.

19. As each phase is released we will e-mail HESES and/or RAS contacts to let them know it is available to use. Below are the dates when the two phases are due to go live:

Phase 1 – HESES07 re-creation: 21 July 2008
Phase 2 – all other outputs, including WP and RDQR, RAS07 re-creation, cost centre sector norms and TRAC(T): 5 September 2008

Note that we also provided details of the RDQR deadlines (see paragraph 6) when we notified HESES and RAS contacts of the release of Phase 2.

20. We welcome comments on the suitability of the algorithms; these should be sent to Lydia Booker (e-mail hesa_heses_stats@hefce.ac.uk).

Introduction

21. This document provides guidance on using HEFCE's web facility and its outputs. The primary purpose of the web facility is to help higher education institutions to return accurate HESA data. It provides institutions with an opportunity to identify, and therefore rectify, any errors in data that affect the outputs generated by the web facility, before these data are submitted to HESA.

22. Use of this facility prior to submission of the HESA return, is strongly encouraged by both HESA and HEFCE, because both organisations regard it as an essential element of all institutions' data quality processes. The HEFCE web facility for 2007-08 statistics derived from HESA data is a separate utility to the HESA data collection system.

23. We believe that introduction of the web facility has contributed to an improvement in HESA student data returns. As a result of this improved data quality we have been able to significantly reduce the size of the selection thresholds for the 'statistics derived from HESA data for monitoring and allocation of funding' exercise. We have found that institutions that use the web facility are less likely to be selected for the statistics derived from HESA data exercise.

24. Use of the web facility can also help identify errors and discrepancies between the forecasts made in HESES07 and the outturn position for 2007-08 as recorded in the 2007-08 return. Where discrepancies occur, we expect institutions to take full account of the outputs from the web facility when preparing future HESES returns. We encourage institutions to analyse the web facility outputs as part of their planning and audit processes.

25. Changes to the web facility since '2006-07 statistics derived from HESA data: Guide to HEFCE web facility' (HEFCE 2007/15) are described in Annex A.

26. The web facility generates 11 outputs. These are:

- a HESES07 re-creation
- a HESES07 re-creation based on cost centre sector norms for subjects
- derived statistics that may inform 2009-10 WP allocations
- derived statistics that may inform the 2009-10 partial completion weighting
- a CFEE07 re-creation
- a RAS07 re-creation
- RDQRs
- 2007-08 HEFCE-fundable student FTEs for TRAC(T)
- derived statistics that are likely to inform HESES08 audits
- LLN student summaries
- HESES08 non-completion toolkit.

Using the web facility

27. The web facility allows institutions to upload their HESA data across the internet to a server at HEFCE. Instructions on how to use it are given in Annex B.

28. Typically institutions submit their data to the web facility and retrieve the resultant derived statistics several times before all identified errors are corrected. Therefore, we will not restrict institutions' use of the web facility. However, users should be aware that response times of the facility may be slower at times when there is high demand.

29. Institutions should include adequate time within their timetable to allow them to make full use of the web facility without jeopardising HESA quality arrangements or timetables.

30. The data submitted when using the web facility will not be viewed by HEFCE unless explicit permission has been provided by the institution. We will monitor usage in order to offer assistance where we see it has not been used by an institution. Please see paragraph 55 below for further information regarding data confidentiality.

Re-creations of HEFCE funding returns

31. The algorithms used to generate the re-creations of HEFCE funding returns (that is, HESES07, RAS07 and CFEE07) are intended to be the same as those that will be used for the 'Statistics derived from HESA data for the monitoring and allocation of funding' exercise. We may, however, make changes where we believe these will improve the algorithms.

32. We strongly advise that institutions use this opportunity to identify the cause of all discrepancies between the re-creations and their HEFCE funding returns, so that where errors in HESA data are a cause, these can be corrected before submission to HESA. The removal of such errors in HESA data reduces the likelihood of selection for the 2007-08 statistics derived from HESA data exercise.

<u>HESES07 re-creation</u>

33. The web facility generates a re-creation of HESES07 from HESA 2007-08 student data. This includes the calculation of formulaic adjustments to teaching grant and WP allocations for 2008-09. This output is coupled with a copy of the original HESES07 outputs for comparison and reconciliation. The HESES07 re-creation is generated using the methods described in Annex C and the algorithms given in Appendix 1.

<u>HESES07 re-creation based on cost centre sector norms</u>

34. The HESES07 re-creation is generated using cost centre data returned by individual institutions to determine price group assignments. In addition to this re-creation, the web facility generates a re-creation of HESES07 that uses cost centre sector norms for subjects to determine price group assignments, rather than the cost centres returned by the institution. The cost centre

sector norms are the most commonly returned cost centre for each subject area on the 2006-07 HESA return. Further details on how we generated the cost centre sector norm mapping and other information about this re-creation are provided in Annex D, and the algorithms are given in Appendix 4.

35. The 'HESES07 re-creation based on cost centre sector norms' includes a re-calculated grant adjustment report that is produced by applying the same formulae that were used to calculate the grant adjustment report for HESES07.

CFEE07 re-creation

36. The web facility generates a CFEE07 re-creation from 2007-08 HESA student data. Where appropriate this includes an indication of the funding associated with the recruited full-time equivalences (FTEs) and the difference in funding between the CFEE07 re-creation and the original CFEE allocation (where these data are available).

37. The CFEE07 return will not be collected until August 2008; however (until these data are available) the web facility provides an early opportunity to verify that the 2007-08 HESA student data are correct for these students.

38. The algorithms used to generate the re-creation are intended to be the same as those to be used for the 'Statistics derived from HESA data for the monitoring and allocation of funding' exercise. We may, however, make changes to improve the algorithms. The CFEE07 re-creation is generated using the methods described in Annex G and the algorithms in Appendix 9.

RAS07 re-creation

39. The web facility generates a RAS07 re-creation from HESA 2007-08 student data. This includes a re-calculation of the 2008-09 research degree programme (RDP) supervision fund. The RAS07 re-creation is generated using the methods detailed in Annex H and the algorithms detailed in Appendix 10.

Derived statistics likely to inform HESES08 audits

40. We will use 2007-08 HESA student data to identify areas of further investigation during HESES08 data audits. The web facility will generate two outputs that will be used to identify areas of potential further investigation relating to completion data. These outputs may be used as part of the HESES08 data audits carried out by the HEFCE Assurance Service. Further details of these tests are given in Annex N, and the associated files are described in Appendix 18.

Derived statistics that may inform funding allocations

Derived statistics that may inform 2009-10 WP allocations

41. 2007-08 HESA student data may be used to inform the following WP allocations for 2009-10:

- widening access for students from disadvantaged backgrounds: full-time and part-time
- improving retention: full-time
- widening access and improving provision for disabled students.

42. The derived statistics used to inform the 2009-10 WP allocations may be generated using the methods detailed in Annex E and the algorithms in Appendix 7.

Derived statistics that may inform the 2009-10 partial completion weighting

43. In 'Review of the teaching funding method: Outcomes of second consultation on changes to the method from 2008-09' (HEFCE 2007/23) we announced that we will use HESA data to count, for funding purposes, the modules completed by students that complete less than their initial study intentions for the year. We will calculate a partial completion weighting for each institution based on their non-completion students who complete at least 20 credits (0.16 FTE). The web facility will include a summary of the students that may contribute towards the 2009-10 partial completion weighting. In addition the standard resource associated with their completed modules will be calculated. The derived statistics used to inform the 2009-10 partial completion weighing may be generated using the methods detailed in Annex F and the algorithms in Appendix 8.

Derived statistics used in HEFCE statistical publications

RDQRs

44. In 2009 we intend to publish RDQRs for full-time students that commenced their research degree programmes in 2001-02, and part-time and mode-switch students that commenced their research degree programmes in 1999-2000.

45. We expect to publish RDQRs for later cohorts in future years; therefore institutions should verify the entire RDQR dataset. This will be the last opportunity for institutions to verify the accuracy of the data underlying their RDQRs for full-time students (starting in 2001-02), and part-time and mode-switch students (starting in 1999-2000) before they are asked to sign off these rates as suitable for publication. This will also be the last opportunity to ensure that 2007-08 HESA data, where these will inform future RDQRs, are accurate.

46. Annex I describes the methods used to produce RDQRs and outlines the steps institutions must take to remedy any issues resulting from underlying errors in the HESA student data. In order for institutions to remedy their RDQRs they must first provide an RDQR action plan. This

action plan must be submitted by **6 October 2008**. Annex L gives details on completing and submitting RDQR action plans, and an example plan is at Appendix 15. We may audit RDQR action plans to ensure that they are supported by the appropriate records and evidence.

47. Errors in HESA data for the years 1999-2000 to 2005-06 will require correction through the submission of an HDE file to enable us to correct the derived fields for the individual students affected. Annex K gives details on producing and submitting HDE files. The deadline for signing-off HDE files is **13 October 2008**.

48. Errors in 2006-07 HESA student data will require institutions to submit a revised return to HESA. Annex J gives details on submitting amendments to 2006-07 HESA data. The deadline for signing-off amendments to 2006-07 HESA data for the purpose of correcting RDQRs is **13 October 2008** This will also be the last opportunity to ensure that 2006-07 HESA data, where these will inform RDQRs, are accurate.

49. Institutions are invited to comment on the sector rates that are provided as part of the RDQR outputs. Institutions should e-mail their comments to hesa_heses_feedback@hefce.ac.uk by **13 October 2008**.

50. Annex I details the full timetable for verifying and publishing RDQRs and Appendices 13 and 14 contain details of the RDQR algorithms.

HEFCE student summaries

HEFCE-fundable student FTEs for TRAC(T)

51. The web facility will generate HEFCE-fundable student FTEs for TRAC(T). Where institutions discover errors with the underlying data used to generate these FTEs, they should remedy these before signing off their data with HESA. The methods used to generate the HEFCE-fundable student FTEs for TRAC(T) are detailed in Annex M, and the algorithms are given in Appendices 16 and 17.

LLN student summaries

52. For institutions participating in an LLN during 2007-08 we intend to identify LLN students using 2007-08 HESA data. These data will be used as part of our evaluation of the LLN initiative. The web facility generates LLN student summaries so that institutions can verify that we identify their LLN students correctly. The methods used to generate the LLN student summaries are detailed in Annex P, and the algorithms are given in Appendix 19.

HESES08 non-completion toolkit

53. The web facility provides institutions with an Excel spreadsheet (toolkit) to assist them in estimating non-completion rates for HESES08. A detailed explanation of how to use the toolkit is given in Annex Q.

HEFCE regional statistics

54. We intend to publish regional statistics derived from 2007-08 HESA student data during 2009. Later this summer we may add another output to the web facility to enable institutions to verify the following regional derived statistics:

- students registered at one institution and taught by another institution
- campuses
- distance learning
- provision by location.

If we add this output to the web facility we will notify HESES contacts via e-mail and provide any necessary documentation to enable verification of the outputs.

Data confidentiality

55. The web facility is a tool for institutions to use. Therefore, HEFCE staff will not access any data submitted to, or derived by, the web facility except where an institution explicitly gives permission. Only those fields listed in the appendices will be retained on our system. We will remove individualised files and any outputs from the 2007-08 web facility when we launch the '2008-09 statistics derived from HESA data' web facility in summer 2009. All other submitted data files will be deleted immediately after the outputs have been generated. Details about our commitments to handling personal data can be found on the HEFCE web-site, www.hefce.ac.uk, under About us/Standards of service/How we do it – our commitments/Release guidelines for personal data.

HEFCE-recognised funding consortia

56. 'HEFCE web facility for 2007-08 statistics derived from ILR data' (HEFCE 2008/25) describes the equivalent web facility for all HEFCE-funded further education colleges and member colleges of HEFCE-recognised funding consortia. The primary purpose of this facility is to help colleges to return accurate 2007-08 Individualised Learner Record (ILR) data to the Learning and Skills Council (LSC). We advise lead institutions of HEFCE-recognised funding consortia to encourage their member colleges to use this facility before submitting their 2007-08 ILR to the LSC.

57. Member colleges' 2007-08 ILR data will form part of the following lead institution's outputs when the 'Statistics derived from HESA data' exercise is launched later in the year:

- HESES07 re-creation
- HESES07 re-creation based on cost centre sector norms
- derived statistics that may be used to inform the 2009-10 WP allocations.

58. When member colleges have uploaded their data to the ILR web facility, these data will be incorporated into the lead institution's web facility outputs (as listed in paragraph 57). We advise leads of HEFCE-recognised consortia to liaise with their member colleges to co-ordinate the submission of these data, to ensure meaningful outputs are produced by the web facility.

Northern Irish higher education institutions

59. For Northern Irish higher education institutions we produce the following outputs:

- HESES07 re-creation
- HESES07 re-creation based on cost centre sector norms
- Derived statistics likely to inform 2009-10 WP allocations
- RAS07 re-creation
- RDQRs
- 2007-08 HEFCE-fundable students FTEs for TRAC(T)
- 2009-10 partial completion weighting
- Derived statistics likely to inform HESES08 audits
- HESES08 non-completion toolkit.

Guidance

Instructions on how to use the web facility

60. Instructions on how to get started with the web facility and how to access the outputs are given in Annex B.

FAQs

61. Frequently asked questions (FAQs) regarding the web facility are on the HEFCE web-site under Learning & teaching/Data collection. We encourage institutions to refer to the FAQs for guidance in the first instance. We will only use our e-mail list of HESES or RAS contacts to notify institutions of significant changes or updates.

HEFCE contacts

62. Support is available for the web facility at HEFCE via telephone and e-mail; see page 3 for contact details. This will be available between 0900 and 1700, Monday to Friday, except during the fortnight before the HESA submission deadline, when we will aim to offer extended support beyond these times.

Annexes and technical appendices

63. The annexes provide guidance on using the web facility and its outputs. The technical appendices describe the algorithms used to generate the derived statistics. The appendices are

available as separate downloadable files on the HEFCE web-site, with this document, under Publications.

Troubleshooting

64. Appendices 2, 5 and 11 will help identify the causes of discrepancies between the 2007-08 HESA student data and the HESES07, HESES07 re-creation based on cost centre sector norms, and RAS07 comparisons respectively.

Generating outputs from individualised files

65. Generally, web facility outputs can be generated from the individualised files provided. Appendices 2, 5, 7, 8, 9, 11, 14, 17, 18, 19 and 20 provide guidance about how to use these individualised files.

Comments

66. All institutions are invited to comment on any of the methods described in this publication. Institutions should e-mail their comments to hesa_heses_feedback@hefce.ac.uk

Annex A

Summary of changes since publication of HEFCE 2007/15

Purpose

1. This annex describes the changes that have been made since the release of last year's version of this report: '2006-07 statistics derived from HESA data: Guide to HEFCE web facility' (HEFCE 2007/15).

New outputs

Derived statistics that may inform the 2009-10 partial completion weighting

2. In HEFCE 2007/23 we announced that we will use 2007-08 HESA data to calculate partial completion weightings for each institution based on the modules completed by students that complete less than their initial study intentions for the year. To contribute to the weighting a student must complete a minimum of 0.16 FTE.

3. To control the extent to which this measure leads to a redistribution of funding between institutions we may cap the extent to which any individual institution can benefit financially from this proposal. Our Board will determine the method for doing this in January 2009 based on 2007-08 HESA data.

4. The web facility will include a summary of the students that may contribute towards the 2009-10 partial completion weighting. In addition the standard resource associated with their completed modules will be calculated. The derived statistics used to inform the 2009-10 partial completion weighting may be generated using the methods detailed in Annex F and the algorithms given in Appendix 8.

CFEE07 re-creation

5. The web facility now generates a CFEE07 re-creation using 2007-08 HESA student data submitted by the institution. An explanation of the re-creation is given in Annex G and the CFEE07 re-creation algorithms are explained in Appendix 9.

6. Any funding adjustments arising from the reconciliation of the CFEE07 re-creation with CFEE07 are likely to affect the funding previously announced for 2007-08, and widening participation (WP) funding and other targeted allocations for 2008-09.

Derived statistics likely to inform HESES08 audits

7. During the HESES08 data audits we will use 2007-08 HESA student data to identify areas that may need further investigation. In particular we will perform two tests on the 2007-08 HESA student data to identify potential non-completions. An explanation of these tests is given in Annex N, and the individualised files are described in Appendix 18.

Changes to outputs

Indicative 2009-10 WP outputs

8. The indicative 2009-10 WP outputs (including an indicative 2009-10 WP allocation calculation) are based on rates and full-time equivalences (FTEs) used for the 2008-09 WP allocation made in July 2008. Where possible, the calculations also incorporate any subsequent changes arising, for example, from amendments to 2008-09 additional student numbers (ASNs) and transfers. During 2008 and 2009 we may update the rates and FTEs used by the web facility for the indicative WP outputs as more current information becomes available.

9. The indicative WP allocations now treat students studying for an equivalent or lower qualification (ELQ) as non-fundable. The provisional assumed FTEs for 2008-09 and the statistics derived from 2007-08 HESA student data which inform the indicative 2009-10 WP allocations have been adjusted to reflect this change.

10. We have made some changes to the measures used to inform the widening access allocations. These changes are described in paragraphs 11 to 14 below; they are a necessary update to the underlying data and do not represent a change in the allocation method.

11. The young participation rates for small areas that underpin the widening access allocation for young full-time students have been updated. Previously we used participation rate quintiles based on the young participation of people who reached 18 between 1997 and 1999. These rates were calculated for the 1991 Census small area statistics wards geography. We have updated the participation quintiles to cover young people who reached 18 between 2000 and 2004, and changed the underlying geography used to calculate the rates to 2001 Census area statistics wards. This classification of areas covers the UK, and the quintiles are formed with reference to the UK young population.

12. We have also made some adjustments to the area groupings based on educational qualification levels used for the widening access allocation for mature full-time and both young and mature part-time students. These groupings now cover 2001 Census area statistics wards throughout the UK, rather than just England and Wales, though the ranges for the quintiles are determined with reference to the English population only.

13. The ‘postcode to 2001 Census area statistics ward look-up’ file that is used in the allocation has been updated to reflect the August 2007 release of the National Statistics Postcode Directory (NSPD), which is maintained by the Office for National Statistics (ONS). This look-up file may be updated as more current information becomes available. Changes to the way we calculate young participation rates have led to a revision of the postcodes that are excluded from the allocation because we cannot obtain robust participation rates. Additionally, for the purposes of the widening access allocations, we have formally excluded postcodes that the NSPD indicates as non-geographic (mostly large user and PO boxes).

14. The assignment of postcodes to the young participation and educational qualification quintile groupings are provided in a file on the HEFCE web-site under Widening participation/POLAR and participation rates/POLAR2. This file includes postcodes which are excluded from the quintile mapping along with the reason for exclusion.

Discontinued outputs

HEFCE regional statistics

15. While we intend to publish regional statistics derived from 2007-08 HESA student data during 2009, we may not be able to include these outputs in this year's web facility. We hope to be able to add these outputs later this summer to enable institutions to verify the following regional derived statistics:

- students registered at one institution and taught by another institution
- campuses
- distance learning
- provision by location.

If we add this output to the web facility we will notify HESES contacts via e-mail and provide any necessary documentation to enable verification of the outputs.

Changes to algorithms

16. Following the redevelopment of the HESA student record for 2007-08 we have made substantial changes to our derived statistics algorithms. The changes are extensive and so we will not detail them here. The algorithms are described in the technical appendices.

HESES07 re-creation

17. Table 7 of HESES07 was introduced to address the possibility that institutions' ability to meet their funding agreement targets for 2007-08 could be affected by a change in the reporting method for non-standard years.

18. The HESES07 re-creation will include a Table 7 re-creation. The return of Table 7 was optional and restricted to institutions that reported non-standard years using the 100:0 method on their HESES06 return. We only re-create Table 7 for institutions that completed the table on HESES07.

19. The derived fields to generate Table 7 are detailed in paragraphs 70-78 of Appendix 1.

Widening participation

20. As a result of the change described in paragraph 12 of this annex, the algorithms for indicating inclusion in the mature full-time widening access population (EDPOPM) and inclusion

in the part-time widening access population (EDPOPPT) have been modified to include students domiciled in Scotland and Northern Ireland (DOMICILE = XH, XG). Details of this change can be found in paragraphs 46 and 47 of Appendix 7.

21. We are now able to map directly from postcode (POSTCODE) to the young higher education participation quintile and adult educational qualification quintile (WAQUIN), so we have removed the derived fields displaying the 1991 census ward (WARD6_C) and 2001 census ward (CASWARD) of the student's home postcode from the individualised file WP07XXXX.ind.

22. As a result of the change described in paragraph 9 of this annex, the derived fields which indicate inclusion in the widening access and full-time improving retention populations (YNGPART, EDPOPM, EDPOPPT and EQPOP) have been modified so that students who are studying for an ELQ are not included. Additionally, for a student in any of these populations where the level of entry qualifications is not known, an appropriate proportion is removed from the field indicating inclusion in the population. Appendix 7 provides further details on the changes to these fields; it also provides information on how we identify which students are studying for an ELQ, and which students have an unknown-level entry qualification (ELQQENT = UNK). It also explains how we have calculated the proportion that should be treated as non-fundable (ELQ_PROP) for students with unknown-level entry qualifications. Note that the field which indicates inclusion in the disability allocation population (DISPOP) has not changed, as home and EC students who are classed as non-fundable (HESTYPE = HOMENF) may be included in this population.

23. Four new values of highest qualification on entry have been introduced to the HESA student record for 2007-08: QUALENT2 = 37 (GNVQ/GSVQ Level 3), QUALENT2 = 38 (NVQ/SVQ Level 3), QUALENT2 = 57 (NVQ/SVQ Level 2) and QUALENT2 = 72 (Diploma in Foundation Studies (Art and Design)) Consequently we have updated our algorithms for the flag indicating whether or not a student has previously studied for their qualification aim, or a higher qualification aim (HIGHQUAL) and the grouping of student's highest qualification on entry (ENTQUAL). Details of this change can be found in paragraphs 43 and 54 respectively of Appendix 7.

Annex B

How to use the HEFCE web facility

Purpose

1. This annex contains detailed instructions on how to use the web facility.

Data preparation

2. The web facility will process the extensible markup language (XML) file structure of the 2007/08 Student Record. Files must therefore be in XML and conform to the relevant XML Schema Definition (XSD) file found on the HESA website (www.hesa.ac.uk) under Student Stream/C07051: Student Collection 2007/08. The web facility can process zip files, provided that the content files conform to the relevant XSD file.

3. While it is currently possible to submit HESA data to the web facility that have not passed all the business rules, it is not possible for the web facility to process XML files that have not passed the schema checks.

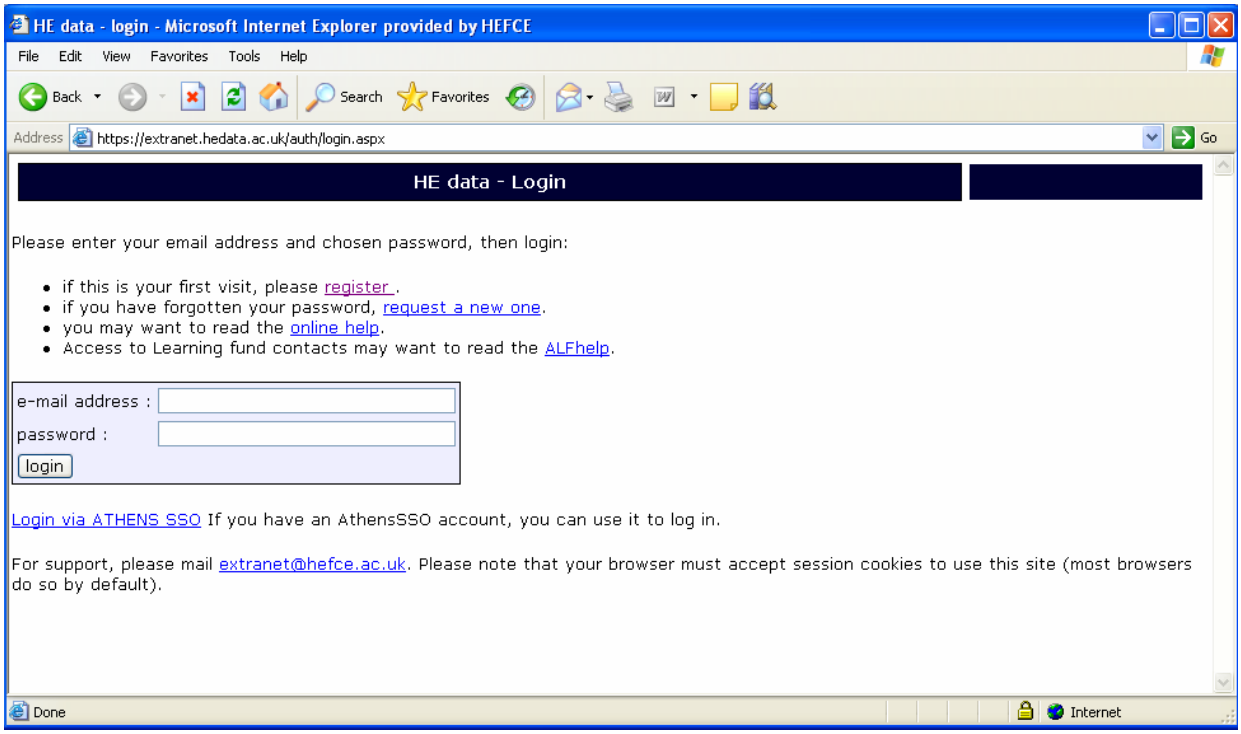
4. It should be noted that although we believe the facility to be robust, it may not function reliably when the input files contain data that have not passed all the business rules. To ensure reliable outputs are generated, institutions should first correct any errors identified by running the HESA validation kit on their data before submission. HESA's validation kit can be found at www.hesa.ac.uk under Data Collection/Submit data to HESA/Validation kits.

5. In addition HESA publishes a number of commit-stage validation checks. If data fail these checks, the web facility may generate unreliable results. In particular, if the combination of HUSID, UKPRN and NUMHUS is not unique for each instance, the re-creations will generate incorrect results.

Access to the facility

6. The web facility can be accessed via the HEFCE extranet at https://extranet.hedata.ac.uk. On viewing the following screen you will be prompted to either log in or register for access to the HEFCE extranet as a new user.

Figure 1 **The login screen**

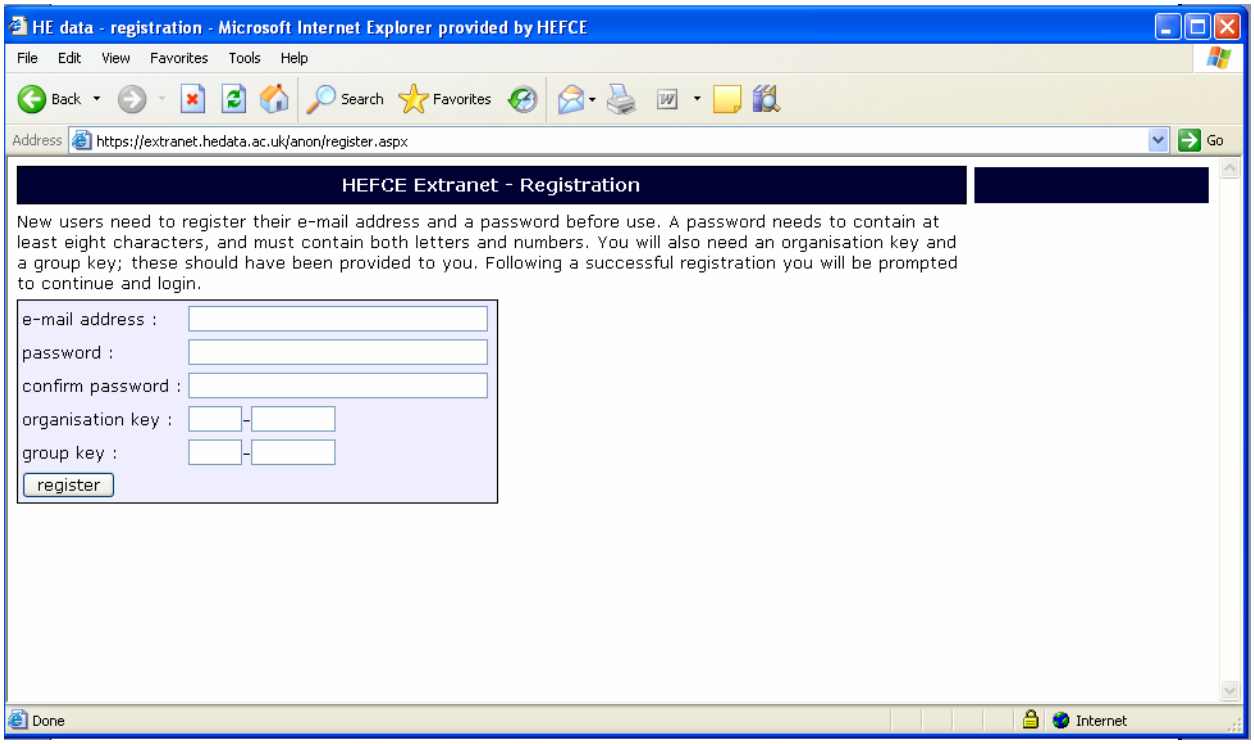


Registering a new account

7. New users of the HEFCE extranet will first need to register an e-mail address and extranet password before use. This can be done by clicking the 'Register' link on the login screen. To register, you will require an 'organisation key' and a 'group key', details of which have been provided in the letter titled 'HEFCE web facility for 2007-08 statistics derived from HESA data', sent to your head of institution and HESES contact by Anthony Ryan in September 2008.

8. Once registered you should be able to log in by entering the e-mail address you used during registration, and the password that you created.

Figure 2 **The registration page**



Existing users of the extranet

9. If you have used the HEFCE extranet for other HEFCE returns (for example HESES07) you will be required to log in and join the HEFCE web facility for 2007-08 statistics derived from HESA data group. Follow the log-in procedure by entering your e-mail address and password. You will be redirected to a page for 'HEFCE extranet – All resources'; under 'Applications' click 'Join a group'. Enter the group key supplied in Annex A of the letter referred to in paragraph 7 and select 'Join group'.

10. If you have registered in the past but your account has expired, you will be required to refresh your account using the organisation key referred to in paragraph 7.

Athens Single Sign On Account

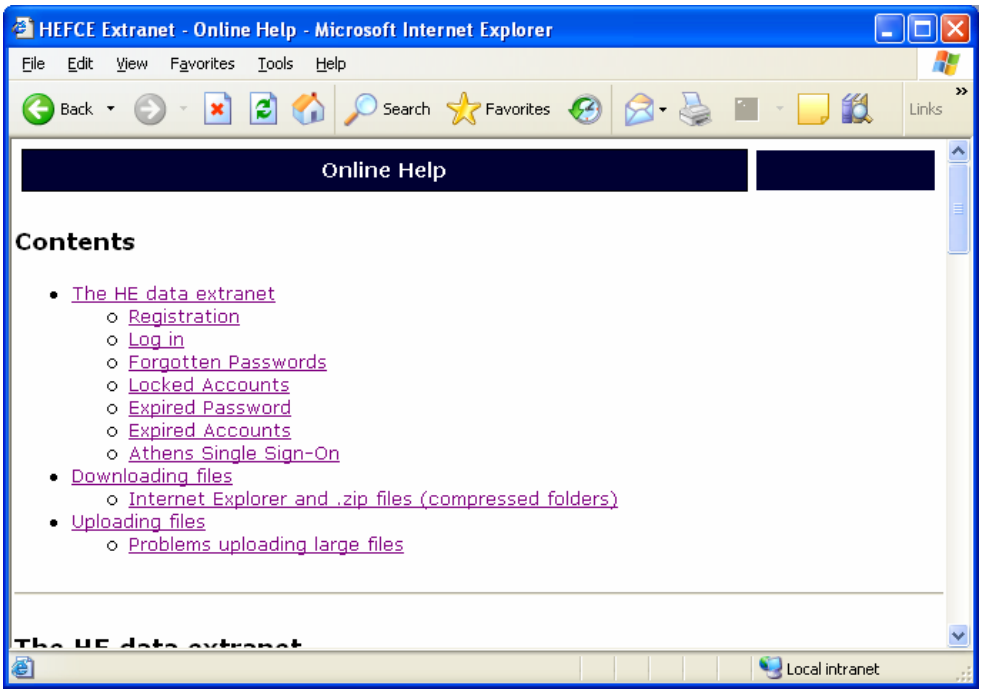
11. You can also log in to the HEFCE extranet using an Athens Single Sign On account (if this is available at your institution).

a. Follow the 'Log in via Athens SSO' link on the HEFCE extranet login page.
b. Log in to Athens as normal (if not already done).
c. When Athens has authenticated you, your browser will be directed to the 'HEFCE extranet - All resources' page.

d. You will be required to join the 'HEFCE web facility for 2007-08 statistics derived from HESA data' group. Under 'Applications' click 'Join a group'. Enter the group key sent to the head of your institution in Annex A of the letter referred to in paragraph 7 and select 'Join group'.

Help

12. An information page can be accessed by clicking the 'Online Help' link, located above the log-in box, or on the right of the 'HEFCE extranet - All resources' page, under 'Navigation'.



Forgotten password

13. If you forget your password, select 'request a new one' on the login page. Enter your e-mail address in the space provided and click 'request password'. A new password will be sent to you. You will be required to change this password before being able to access the HEFCE extranet.

14. Each user has three attempts to log in before the account is locked. If your account becomes locked, or you have other problems logging onto the extranet, contact Lydia Booker (e-mail hesa_heses_stats@hefce.ac.uk) for help.

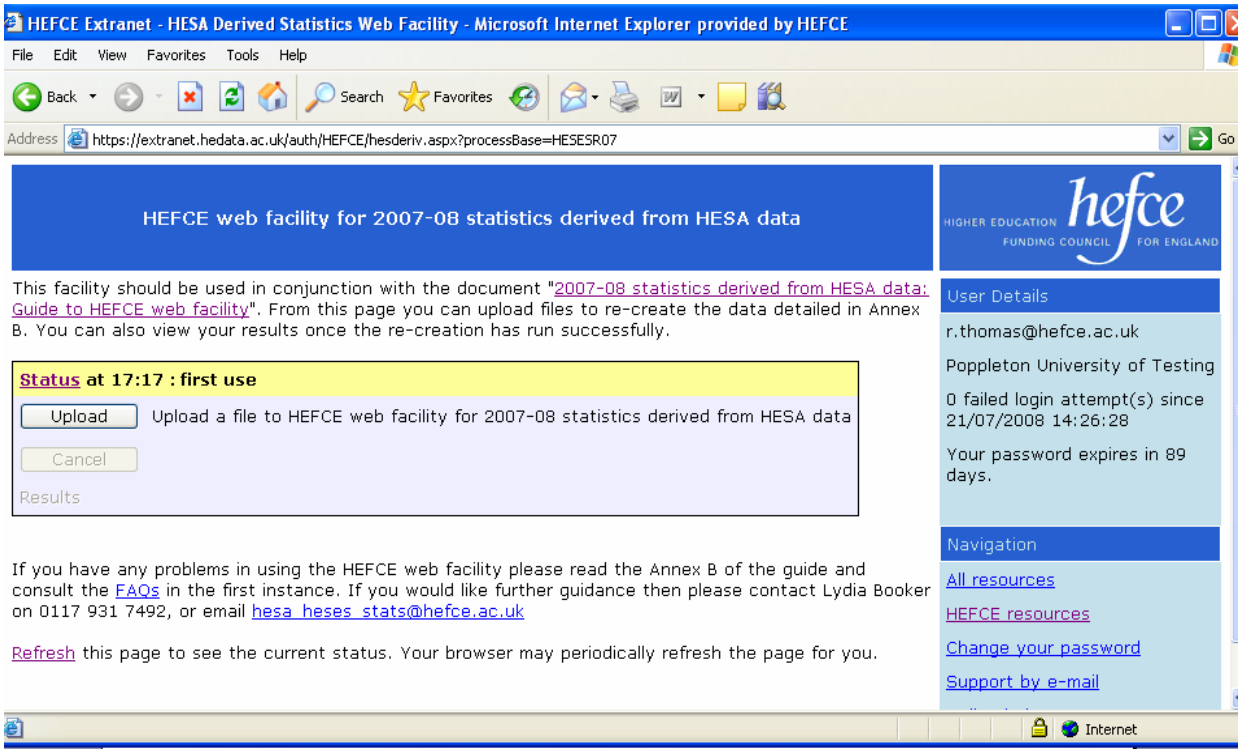
All resources

15. After verifying the username and password, your browser will be directed to the 'HEFCE extranet - All resources' page. Click 'HEFCE resources'.

16. From the ‘Applications’ list, click on the ‘HEFCE web facility for 2007-08 statistics derived from HESA data’ link.

17. This option will allow HESA XML file(s) to be uploaded one at a time. Use the ‘Browse’ button to reach the location of the data file and once the correct file is located, select ‘Upload File’. The browse facility does not allow selection of multiple files.

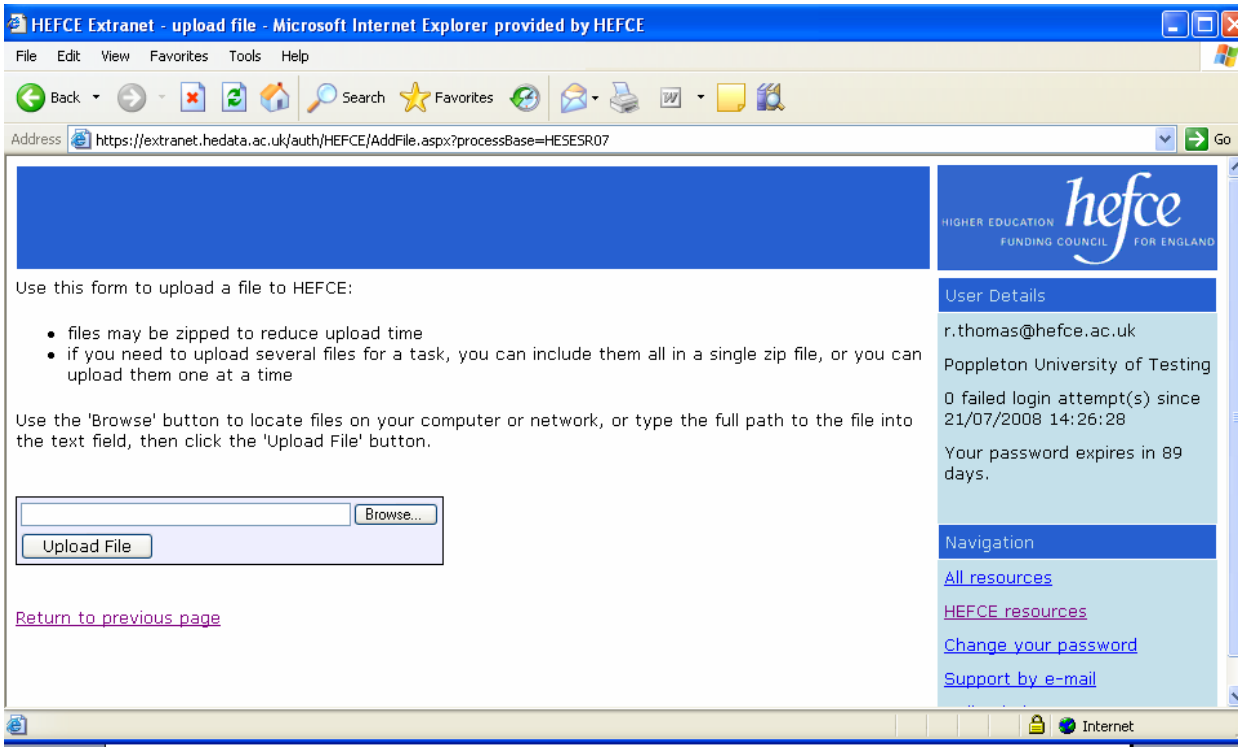
Figure 3 **2007-08 HEFCE web facility main page**



18. Continue to add files until all HESA XML files have been uploaded. All files should be uploaded by repeating the process in paragraph 17.

19. Files which have been uploaded to the web facility are listed under ‘Current files’. An uploaded file may be deleted by selecting the ‘delete’ link next to the appropriate file.

Figure 4 **2007-08 HEFCE web facility upload page**



Process file(s)

20. Once all of the files have been added, users may then select 'Process'. This places the file(s) that were added in the previous step into the processing queue.

Processing

21. You can leave the extranet while the derived statistics are being generated. To do this, exit the web browser in the usual way, or select a new address in the address bar. You may be prompted to log in again when returning to the page, but no information will be lost.

22. Each time a collection of files is submitted for re-creation, you should refresh the page using either the icon in the tool bar or the link on the page. Refreshing the page ensures that the 'History of HEFCE web facility for 2007-08 statistics derived from HESA data submissions' reflects the current stage of processing, as the page may not update automatically.

23. Our systems only allow one submission to be processed at a time; therefore if more than one institution uploads data, the submissions will be placed in a queue. A submission by an average-sized institution typically takes 15 minutes to process (once it reaches the front of the queue). The total time taken to complete the re-creation, including waiting time, will depend on the number of institutions in the queue.

24. If the uploaded files are waiting in the queue it is possible to cancel the submission by selecting ‘Cancel’.

25. If you do not use the extranet for 60 minutes, you will be automatically logged out of the system. To regain entry, select any option and re-enter the e-mail address and extranet password.

Output

26. Once the files have been processed, the outputs can be accessed by selecting ‘Results’.

27. If the outputs were not generated and the reason is known, it will be given under ‘History of HEFCE web facility for 2007-08 statistics derived from HESA data submissions’. If the outputs were not generated because an unspecified error occurred, this will typically be a result of invalid HESA data.

Accessing results

28. The results are made available as a zip archive (HESR07XXXX, where XXXX denotes the HESA institution identifier) containing the following files:

- HESR07XXXX.xls – This is an Excel workbook containing the HESES07 re-creation tables. For leads of HEFCE-recognised funding consortia this includes the combined 2007-08 HESA data uploaded by the lead institution and 2007-08 ILR data uploaded by member colleges. (If member colleges have not uploaded these data they will not be included. We expect institutions within HEFCE-recognised funding consortia to liaise with each other to co-ordinate submission of data to the web facilities.)

- HESR07XXXX.ind – This is the HESES07 re-creation individualised student data file – for leads of HEFCE-recognised funding consortia this file does not include data for the member colleges.

- SNCC07XXXX.xls – This is an Excel workbook containing the HESES07 re-creation based on cost centre sector norms tables. For leads of HEFCE-recognised funding consortia this includes the combined 2007-08 HESA data uploaded by the lead institution and 2007-08 ILR data uploaded by member colleges. (If member colleges have not uploaded these data they will not be included.)

- SNCC07XXXX.ind – This is a comma-separated file containing the derived fields that generate the HESES07 re-creation based on cost centre sector norms.

- WP07XXXX.xls – This is an Excel workbook containing the derived statistics that may be used to inform the 2009-10 WP allocations. For leads of HEFCE-recognised funding consortia this includes the combined 2007-08 HESA data uploaded by the lead and 2007-08

ILR data uploaded by member colleges. (If member colleges have not uploaded these data they will not be included.)

- WP07XXXX.ind – This is the 2009-10 WP individualised student data file – for leads of HEFCE-recognised funding consortia this file does not include data for the member colleges.

- PCMP07XXXX.xls – This is an Excel workbook containing the derived statistics that may be used to inform the 2009-10 partial completion weighting.

- PCMP07XXXX.ind – This is a comma-separated file containing the derived fields that generate the 2009-10 partial completion weighting summary.

- CFEE07XXXX.xls – this is an Excel workbook containing the CFEE07 re-creation.

- CFEE07XXXX.ind – this is a comma-separated file containing the derived fields that generate the CFEE07 re-creation.

- RASR07XXXX.xls – This is an Excel workbook containing the RAS07 re-creation tables.

- RASR07XXXX.ind – This is a comma-separated file containing the derived fields that generate the RAS07 re-creation.

- RDQR07XXXX.xls – This is an Excel workbook containing the RDQRs.

- RDQR07XXXX.ind – This is a comma-separated file containing the derived fields that generate the RDQRs.

- TRAC07XXXX.xls – This is an Excel workbook containing the 2007-08 HEFCE-fundable student FTEs for TRAC(T).

- TRAC07XXXX.ind – This is a comma-separated file containing the derived fields that generate the 2007-08 HEFCE-fundable student FTEs for TRAC(T).

- AUD107XXXX.ind – this is a comma-separated file containing the derived statistics likely to inform HESES08 audits.

- AUD207XXXX.ind– this is a comma-separated file containing the derived statistics likely to inform HESES08 audits.

- LLN07XXXX.xls – This is an Excel workbook containing the LLN student summaries.

- LLN07XXXX.ind – This is a comma-separated file containing the derived fields that generate the LLN student summaries.

- NCTK07XXXX.xls – This is the HESES08 non-completion toolkit pivot table.

29. For leads of HEFCE-recognised funding consortia, in addition to the outputs above the following outputs are also available:

- LEAD07XXXX.xls – This is an Excel workbook containing the HESES07 re-creation tables for the lead institution. These tables are created from the lead institution's 2007-08 HESA data that have been submitted to the web facility.

- HEIFERC07YYYYYY.xls – These are Excel workbooks containing the HEIFES07 re-creation tables for each of the member colleges (YYYYYY denotes the unique provider identification number (UPIN) for the college). These tables are created from the 2007-08 ILR data that the member college(s) has submitted to the ILR web facility. These tables will only become available after the member college has submitted data to the web facility.

- WPLEAD07XXXX.xls – This is an Excel workbook containing the 2009-10 WP allocation tables for the lead institution. These tables are created from the lead institution's 2007-08 HESA data that have been submitted to the web facility.

- IHWPC07YYYYYY.xls – These are Excel workbooks containing the 2009-10 WP allocation tables for each of the member colleges of a HEFCE-recognised funding consortium that have submitted 2007-08 ILR data to the ILR web facility.

Using the individualised files

30. The individualised files show the allocation of students to cells within the equivalent re-creation tables and, where relevant, details of why they were excluded. Generally, the appendices that accompany this document include detailed instructions on how institutions can generate the 2007-08 outputs from the individualised file.

31. For institutions with individualised files that do not contain greater than 65,530 records, the following guidance will assist them in the troubleshooting process. All records with a specific value in a specific field can be found in the following way:

a. Open an individualised file from the list above in Microsoft Excel, and click <File>, <Open>. Users will need to specify 'All Files' in the 'Files of type' box before the individualised file will appear in the file listing. Once the file is selected, the 'Text Import Wizard' will appear. Ensure that 'Delimited' is selected near the top of the window, then click 'Next'. On the next page, uncheck 'Tab' and check 'Comma'. Click 'Finish' to open the file.
b. Select the row containing the field headings.
c. Select <Filter> from the <Data> menu and then <Autofilter>.
d. Click on the arrow in the column containing the data which needs filtering.
e. Either select a specific value or select <Custom> to apply a comparison operator other than equality.
f. To select records using multiple fields, repeat steps d and e for each field.

32. The RDQR action plan template (APRDQR07XXXX.xls) is available from the HEFCE extranet. To access the action plan log in to the extranet and navigate to the 'HEFCE resources' page, as per the instructions in paragraph 15. From here click on '2007-08 statistics derived from HESA data', and then on 'APRDQR07 data collection'. You will need to click the download link to download the template.

33. APRDQR07XXXX.xls is an Excel workbook containing the action plan template for RDQRs. This workbook is available for all institutions.

Using the output

34. The tables are best viewed in Microsoft Excel. When you open the individualised file, some fields (such as HUSID) may look like '9.64E+08'. To obtain the 13-digit code as returned to HESA follow these steps:

a. Highlight the column containing the field you wish to format.
b. Click on <Format>, then <Cells>.
c. In the <Number> tab, select the <Custom> category.
d. In the <Type> text box, enter 13 zeros.
e. Click <OK> to get back to the spreadsheet.

35. The files output by the web facility will be overwritten each time new data are submitted. Note that it is not possible for institutions to see one another's files except for members of HEFCE-recognised funding consortia (for certain outputs only).

36. At this point it may be useful to check whether the tables reflect the number of students expected. If it does not, view the files uploaded and check that they are the files that were intended to be submitted. We are not responsible for the accuracy of the files submitted to the web facility.

Technical issues

37. While there are no limits on the size of files that can be uploaded to the facility, it is common for individual institutions to have limits imposed by their network administrators. Attempts to upload a file beyond the size limit imposed within the institution are unlikely to generate a meaningful error message; it will however cause the browser to 'stall' or 'hang'.

38. If difficulties are encountered uploading files due to file size or any other issue relating to connectivity we will normally be able to advise. We recommend that you involve your network administrator in such discussions.

39. Please note that the internet protocol (IP) address for this year's web facility is 195.194.167.114.

Annex C

Comparison of HESES07 and the HESES07 re-creation

Purpose

1. This annex describes the methods we use to generate a provisional HESES07 re-creation via the web facility, for use by institutions prior to their data submission to HESA. It also describes how we are likely to use the comparison of HESES07 and the HESES07 re-creation in the '2007-08 statistics derived from HESA data' exercise.

Web facility outputs

2. The HESES07 re-creation is generated from HESA 2007-08 student data submitted to the web facility. Annex B describes how to access the provisional HESES07 re-creation Excel workbook (HESR07XXXX.xls – where XXXX denotes the HESA institution identifier). The provisional HESES07 re-creation workbook contains the worksheets listed in Table A below. The outputs are provisional because they are generated before submitting final data to HESA.

Table A **Excel workbook HESR07XXXX**

Page number	Worksheet (see tabs on spreadsheet)	Description
1	Coversheet	Title page
2	Summary	Provisional HESES07 re-creation comparison summary table
3	SummaryPG	Provisional HESES07 re-creation price group comparison table
4	EXCL	Provisional HESES07 re-creation exclusion table
5	FTS	Provisional HESES07 re-creation Table 1a: Full-time counts of years of instance
6	MED	Provisional HESES07 re-creation Table 1b: Medical and dental counts of years of instance
7	SWOUT	Provisional HESES07 re-creation Table 2: Sandwich year-out counts of years of instance
8	PT	Provisional HESES07 re-creation Table 3: Part-time counts of years of instance and FTE
9	FEE	Provisional HESES07 re-creation Table 4: Home and EC fees
10	CONS	Provisional HESES07 re-creation Table 6: HEFCE-recognised funding consortia 2007-08
11	TAB7	Provisional HESES07 re-creation Table 7 (optional): Counts of years of instance and FTE using HESES06 treatment of non-standard years of instance
12	HBK	Provisional HESES07 re-creation grant adjustment report
13	STD	Provisional HESES07 re-creation recalculation of standard resource table

14	F07	Provisional HESES07 re-creation recalculation of assumed fee income table
15	LLN	Provisional HESES07 re-creation grant adjustment report for non-mainstream model 2 LLN ASNs
16	WP	Indicative 2008-09 WP allocation based on provisional HESES07 re-creation assumed FTEs
17	FTSDIFF	Difference between HESES07 Table 1a and HESES07 re-creation Table 1a: Full-time counts of years of instance
18	MEDDIFF	Difference between HESES07 Table 1b and HESES07 re-creation Table 1b: Medical and dental counts of years of instance
19	SWOUTDIFF	Difference between HESES07 Table 2 and HESES07 re-creation Table 2: Sandwich year-out counts of years of instance
20	PTDIFF	Difference between HESES07 Table 3 and HESES07 re-creation Table 3: Part-time counts of years of instance and FTE
21	FEEDIFF	Difference between HESES07 Table 4 and HESES07 re-creation Table 4: Home and EC fees
22	TAB7DIFF	Difference between HESES07 Table 7 and HESES07 re-creation Table 7: Counts of years of instance and FTE using HESES06 treatment of non-standard years of instance
23	hFTS	HESES07 Table 1a: Full-time counts of years of instance
24	hMED	HESES07 Table 1b: Medical and dental counts of years of instance
25	hSWOUT	HESES07 Table 2: Sandwich year-out counts of years of instance
26	hPT	HESES07 Table 3: Part-time counts of years of instance ad FTE
27	hFEE	HESES07 Table 4: Home and EC fees
28	hCONS	HESES07 Table 6: HEFCE-recognised funding consortia 2007-08
29	hTAB7	HESES07 Table 7 (optional): Counts of years of instance and FTE using HESES06 treatment of non-standard years of instance
30	hHBK	HESES07 grant adjustment report
31	hSTD	HESES07 recalculation of standard resource table
32	hF07	HESES07 recalculation of assumed fee income table
33	hLLN	HESES07 recalculation of grant adjustment report for non-mainstream model 2 LLN ASNs
34	hWP	2008-09 WP allocation based on HESES07 assumed FTEs

Note: Only institutions that are the lead of a HEFCE-recognised funding consortium will be able to view data in Table 6 of the HESR07XXXX.xls output, which is a subset of Columns 1 and 2 of Tables 1a, 2 and 3, disaggregated by consortium member and level.

3. The HESR07XXXX.xls workbook also contains a version number that will increment by one following each successful submission of data to the web facility, as well as contact details for any queries relating to data in the workbook.

4. The 'DIFF' sheets (page numbers 17-22 in Table A) will indicate where differences in cell totals between the provisional HESES07 re-creation and HESES07 exceed a given threshold. The size of this threshold can be altered by entering the required value where indicated on the worksheets. These sheets are provided to assist institutions in reconciling differences between HESES07 and the provisional HESES07 re-creation.

5. The information in the HESES07 re-creation tables can be rebuilt from the individualised file, HESR07XXXX.ind, which is an output also generated by the web facility. The file contains details of how each student was classified in the re-creation and hence how the student is represented in the tables or, where relevant, details of why they were excluded. A full description of the data in the individualised file is given in Appendix 1.

Methods

6. The algorithms used to generate the provisional HESES07 re-creation are described in Appendix 1.

7. Appendix 2 contains guidance on troubleshooting the causes of differences between HESES07 and the provisional HESES07 re-creation.

8. In some circumstances we have to make approximations when generating the HESES07 re-creation from HESA data. These 'problems of fit' are described in Appendix 3.

9. In late 2008 we will generate a HESES07 re-creation from 2007-08 data submitted to HESA. We will compare the HESES07 re-creation to HESES07. This comparison takes place after the 2007-08 student data have been finalised with HESA and passed to HEFCE.

10. We will re-calculate grant adjustment reports for the HESES07 re-creation (including the grant adjustment report for non-mainstreamed model 2 LLN ASNs) by applying the same formulae that were used to calculate the grant adjustment reports for HESES07. We will also calculate an indicative 2008-09 WP allocation based on HESES07 re-creation assumed FTEs.

11. We expect to select institutions to explain discrepancies between their HESES07 and HESES07 re-creation on the basis of thresholds. The thresholds are likely to be set in terms of funding differences between the HESES07 and the HESES07 re-creation grant adjustment reports, as well as the 2008-09 WP allocations. The recalculations of 2008-09 WP allocations in the 2007-08 statistics derived from HESA data exercise may differ from those in the web facility

because of subsequent changes to, for example, 2008-09 ASN allocations, transfers and 2006-07 HESA/ILR data. Notwithstanding the thresholds, we may also ask for further information from any institution in respect of this comparison.

12. Each institution required to make a response will be asked to reconcile the two data sources. After reconciling the two data sources and confirming that the HESES07 re-creation reasonably reflects the outturn position for 2007-08, the HESES07 re-creation will supersede HESES07 and any consequent grant adjustments (including 2008-09 WP allocations and other targeted allocations) will be made (subject to the appeals process and the availability of our funds).

Redevelopment of the 2007-08 HESA student record

13. Following the redevelopment of the HESA student record for 2007-08 we have made substantial changes to the HESES07 re-creation algorithms. We may need to further refine these algorithms after the release of the web facility to improve the accuracy of the outputs. However, we expect any changes to be minor.

14. This year we have released this web facility in two phases (see paragraphs 18-19 of the main text)'

15. Rebecca Thomas e-mailed HESES contacts on 21 July 2008 to announce the availability of the HESES07 re-creation outputs via the web facility.

16. We welcome comments on the suitability of the algorithms; these should be sent to Lydia Booker (e-mail hesa_heses_stats@hefce.ac.uk).

Annex D

Comparison of the HESES07 re-creation and the HESES07 re-creation based on cost centre sector norms

Purpose

1. The HESES07 re-creation is generated using cost centre data returned by individual institutions to determine their price group assignments. In addition to this re-creation, the web facility generates a re-creation of HESES07 that uses cost centre sector norms for subjects to determine the price group assignments (rather than the cost centres returned by the institution).

2. This annex describes the methods we use to generate the provisional HESES07 re-creation based on cost centre sector norms for subjects via the web facility. The outputs are provisional because they are generated before institutions have submitted final data to HESA.

3. This annex also describes how we will use the comparison of HESES07, the HESES07 re-creation and the HESES07 re-creation based on cost centre sector norms in the 2007-08 statistics derived from HESA data exercise, which will be released in late 2008.

Background

4. We expect, based on the guidance we provide, that student load has been allocated to cost centres based on the cost centre of the member of staff most directly associated with it. In general, members of staff are assigned to the cost centre of the academic department that they are associated with.

5. Further guidance on assigning academic departments to cost centres is in HEFCE Circular Letter 32/2005, ‘Assignment of departments to academic cost centres: 2005-06’ which was published in December 2005. This is available for download from the HEFCE web-site under Publications/Circular letters.

Web facility outputs

6. For each institution, the HESES07 re-creation based on cost centre sector norms is generated from the 2007-08 HESA student data, together with the sector norm mapping that we have created. Annex B describes how to access the relevant Excel workbook (SNCC07XXXX.xls – where XXXX denotes the HESA institution identifier), which contains the worksheets in Table B below.

Table B **Excel workbook SNCC07XXXX.xls**

Page number	Worksheet (see tabs on spreadsheet)	Description
1	Coversheet	Title page
2	snSummary	Provisional HESES07 re-creation and the provisional HESES07 re-creation based on cost centre sector norms comparison summary table
3	snSummaryPG	Provisional HESES07 re-creation and the provisional HESES07 re-creation based on cost centre sector norms price group comparison table
4	snEXCL	Provisional HESES07 re-creation based on cost centre sector norms exclusion table
5	snDIFF	Differences between the HESES07 re-creation and the HESES07 re-creation based on cost centre sector norms: Price groups by subject area
6	snFTS	Provisional HESES07 re-creation based on cost centre sector norms Table 1a: Full-time counts of years of instance
7	snMED	Provisional HESES07 re-creation based on cost centre sector norms Table 1b: Medical and dental counts of years of instance
8	snSWOUT	Provisional HESES07 re-creation based on cost centre sector norms Table 2: Sandwich year-out counts of years of instance
9	snPT	Provisional HESES07 re-creation based on cost centre sector norms Table 3: Part-time counts of years of instance and FTE
10	snFEE	Provisional HESES07 re-creation based on cost centre sector norms Table 4: Home and EC fees
11	snHBK	Provisional HESES07 re-creation based on cost centre sector norms grant adjustment report
12	snSTD	Provisional HESES07 re-creation based on cost centre sector norms recalculation of standard resource table
13	snF07	Provisional HESES07 re-creation based on cost centre sector norms recalculation of assumed fee income table
14	snLLN	Provisional HESES07 re-creation based on cost centre sector norms grant adjustment report for non-mainstream model 2 LLN ASNs
15	FTSDIFF	Difference between HESES07 re-creation based on cost centre sector norms Table 1a and the HESES07 re-creation Table 1a: Full-time counts of years of instance

16	MEDDIFF	Difference between HESES07 re-creation based on cost centre sector norms Table 1b and the HESES07 re-creation Table 1b: Medical and dental counts of years of instance
17	SWOUTDIFF	Difference between HESES07 re-creation based on cost centre sector norms Table 2 and the HESES07 re-creation Table 2: Sandwich year-out counts of years of instance
18	PTDIFF	Difference between provisional HESES07 re-creation based on cost centre sector norms Table 3 and the provisional HESES07 re-creation Table 3: Part-time counts of years of instance and FTE
19	FEEDIFF	Difference between HESES07 re-creation based on cost centre sector norms Table 4 and the HESES07 re-creation Table 4: Home and EC fees
20	FTS	Provisional HESES07 re-creation Table 1a: Full-time counts of years of instance
21	MED	Provisional HESES07 re-creation Table 1b: Medical and dental counts of years of instance
22	SWOUT	Provisional HESES07 re-creation Table 2: Sandwich year-out counts of years of instance
23	PT	Provisional HESES07 re-creation Table 3: Part-time counts of years of instance and FTE
24	FEE	Provisional HESES07 re-creation Table 4: Home and EC fees
25	HBK	Provisional HESES07 re-creation grant adjustment report
26	STD	Provisional HESES07 re-creation recalculation of standard resource table
27	F07	Provisional HESES07 re-creation recalculation of assumed fee income table
28	LLN	Provisional HESES07 re-creation grant adjustment report for non-mainstream model 2 LLN ASNs

Note: The workbook contains a version number that will increment by one following each successful submission of data to the web facility, as well as contact details for any queries relating to data in the workbook.

7. The information in the HESES07 re-creation based on cost centre sector norms tables can be rebuilt from the individualised file provided. The file (SNCC07XXXX.ind) contains details of how each student was classified in the re-creation, and where relevant, details of why they were excluded. This includes a field that identifies records that are assigned to a different cost centre than the cost centre sector norm for the subject (see paragraph 21 of Appendix 4 for further details). A full description of the data in the individualised file is given in Appendix 4.

Methods

8. The algorithms used to generate the provisional HESES07 re-creation based on cost centre sector norms are described in Appendix 4.

9. The cost centre sector norms represent the most commonly returned cost centre for each subject area on the 2006-07 HESA return. The 2006-07 sector norms will be mapped to version 2 of the Joint Academic Coding System (JACS2) as this will be used for reporting subject information on the 2007-08 HESA record.

10. We identified the cost centre to which most institutions assigned the subject activity. This was calculated as follows:

a. For each institution, the FTE for each subject area was calculated. The first two characters of the JACS2 code were used to assign module activity to subject areas using the HESA subject of module field (MODSBJ).
b. For each institution, if the FTE of a subject within a cost centre area was less than 50 the activity was removed from the analysis.
c. For each institution and each subject area, the cost centre with the largest FTE was assumed to be the institution's cost centre 'preference'.
d. For each subject area, the cost centre with the largest number of 'preferences' was taken to be the cost centre sector norm.

11. Appendix 5 contains troubleshooting guidance to identify the causes of differences between the provisional HESES07 re-creation and the provisional HESES07 re-creation based on cost centre sector norms, especially where they are attributable to errors in HESA data.

12. All known problems of fit with the provisional HESES07 re-creation based on cost centre sector norms algorithms are described in Appendix 6.

2007-08 statistics derived from HESA data exercise

13. This web facility enables an institution to view re-creations based on its HESA student data prior to committing its data to HESA. Once we have received a finalised version of all institutions' data from HESA, in December 2008, we will derive a HESES07 re-creation from HESA 2007-08 student data. We intend to apply the algorithms detailed in Appendix 1 to generate the HESES07 re-creation.

14. At this time we will also derive a HESES07 re-creation based on cost centre sector norms. We intend to use the algorithms detailed in Appendix 4 to generate the HESES07 re-creation based on cost centre sector norms. However, if we identify any errors in our algorithms, we will correct them.

15. We will calculate a grant adjustment report for the HESES07 re-creation and the HESES07 re-creation based on cost centre sector norms by applying the same formulae that were used to calculate the grant adjustment report for HESES07.

16. When the derived statistics exercise is launched in December, it is likely that we will select institutions to explain discrepancies between their cost centre assignments and the sector norm cost centre assignments on the basis of thresholds. The thresholds are likely to be set in terms of funding differences identified from the comparison of the grant adjustment reports for the HESES07 re-creation and the HESES07 re-creation based on cost centre sector norms. Notwithstanding these thresholds, we may also ask for further information from any institution in respect of any part of this comparison. This may ultimately result in adjustments to grant where appropriate.

17. During this comparison, we will incorporate previous approvals that HEFCE has made to specific institutions for the way that they assign subject activity to cost centres (where this differs from the sector norm).

18. Each institution required to respond will be asked to reconcile its HESES07, HESES07 re-creation and the HESES07 re-creation based on cost centre sector norms. After reconciling all three and confirming that the HESES07 re-creation reasonably represents the outturn position for 2007-08, the HESES07 re-creation will supersede HESES07 and any consequent grant adjustments (including 2008-09 WP allocations and other targeted allocations) will be made (subject to the appeals process and the availability of HEFCE funds).

Redevelopment of the 2007-08 HESA student record

19. Following the redevelopment of the HESA student record for 2007-08 we have made substantial changes to the HESES07 re-creation based on cost centre sector norms algorithms. We may need to further refine these algorithms after the release of the web facility to improve the accuracy of the outputs. However, we expect any changes to be minor.

20. This year we have released this web facility in two phases (see paragraphs 18-19 of the main text)'

21. Lydia Booker e-mailed HESES and RAS contacts on 5 September 2008 to announce the availability of the HESES07 re-creation based on cost centre sector norms outputs via the web facility.

22. We welcome comments on the suitability of the algorithms; these should be sent to Lydia Booker (e-mail hesa_heses_stats@hefce.ac.uk).

Annex E

Derived statistics that may inform the 2009-10 widening participation (WP) allocations

Purpose

1. This annex describes how we may use 2007-08 HESA student data to inform the WP allocations for 2009-10. Further details of the algorithms we use on these data are given in Appendix 7.

2. The web facility offers institutions the opportunity to check and, where errors would affect these allocations under current proposals, correct their data prior to submission to HESA.

Web facility outputs and indicative funding calculations

3. The web facility will generate an indicative 2009-10 funding calculation for each of the WP allocations. The calculations use provisional 2008-09 allocation rates (announced in July 2008) applied to provisional 2008-09 assumed FTEs in most cases, and may be updated as more current information becomes available. They may not incorporate all 2009-10 additional student numbers, transfers or mergers. Annex B describes how to access these indicative funding allocations and the derived statistics that may be used to inform the 2009-10 WP allocations in an Excel workbook (WP07XXXX.xls, where XXXX denotes the HESA institution identifier).

4. The indicative allocations for 2009-10 are only provided to highlight potential 2007-08 HESA data errors. They should not be considered to be any kind of funding commitment by us and are without prejudice to what our Board may agree to be the final allocations for any institution. The final allocations for 2009-10 may be higher or lower than the illustrations given in the output from the web facility as a result of changes to data by the institution or to the data provided by any other institution, or to the total sum available for allocation, or as a result of any changes to the funding methods.

5. WP funding allocations are informed by the data provided by institutions. If we find that data errors have resulted in institutions receiving incorrect funding allocations, then we will adjust their funding accordingly. In particular, where reconciliations with 2008-09 HESA data (see Annex C for a description of the equivalent exercise for 2007-08) or HESES08 data audit highlight that the assumed FTEs used to allocate 2009-10 WP funding were incorrect, then we will adjust WP funding accordingly, subject to the availability of HEFCE funds.

6. The WP derived statistics can be rebuilt from the individualised file which we provide. This file (WP07XXXX.ind – see Annex B for how to access this file) contains details of how each student was categorised in the indicative WP outputs and, where relevant, details of why they did not contribute. A full description of the data in the individualised file and instructions on how to rebuild the figures in the WP derived statistics and indicative allocations workbook (WP07XXXX.xls) are given in Appendix 7.

Widening access for students from disadvantaged backgrounds

7. This is a formula-based allocation of funding for teaching to recognise the extra costs associated with recruiting and supporting undergraduate students from disadvantaged backgrounds who are currently under-represented in higher education. The proposed method of allocating funds is as follows.

8. Using postcode information from 2007-08 HESA student data, each student is mapped to a 2001 Census area statistics ward. These wards are themselves assigned to quintiles based on young participation rates (for young full-time students) and quintiles based on the proportion of 16-74 year-olds with an HE qualification (for mature full time, and young and mature part time undergraduates). Each student is weighted according to the relevant quintile assignment of their ward:

Table C **Widening access student weights**

Quintile	Weighting
1 Lowest HE participation (young full-time) or lowest average educational achievement (part-time and mature full-time)	2
2	1
3, 4, 5	0

9. Young students are those aged under 21 on entry to the programme of study; mature students are those aged 21 or over on entry.

10. The young HE participation quintiles are based on an extension to our 2005 report on measuring young participation, 'Young participation in higher education' (HEFCE 2005/03) that will be published in 2008. These updated rates are based on young people who reached 18 between 2000 and 2004 and entered a higher education course in the UK while aged 18 or 19. Young participation rates are calculated for each 2001 Census area statistics ward in the UK and are used to rank the wards into five participation quintiles, each containing 20 per cent of the UK young population.

11. The adult HE qualification quintiles are based on 2001 Census area statistics. We use the national equivalents of the 2001 Census Key Statistics table 13 (KS013, 'Qualifications and students') for 2001 Census Output Areas (subsequently aggregated to 2001 Census area statistics wards). These tables can be obtained from the ONS, the General Register Office for Scotland (GROS) and the Northern Ireland Statistics and Research Agency (NISRA). We calculate the proportion of 16-74 year-olds with an HE qualification for UK 2001 Census small-area statistics wards. These wards are then ranked by this proportion to give the adult HE qualification quintiles, with each quintile covering 20 per cent of the English 16-74 year-old population.

12. We allocate postcodes to 2001 Census area statistics wards using the August 2007 release of the ONS's NSPD. A file containing the allocation of postcode to young participation

and adult HE attainment quintiles is available on the HEFCE web-site under Widening participation/POLAR and participation rates/POLAR2. This file includes postcodes which are excluded from the quintile mapping along with the reason for exclusion (including non-geographic postcodes) and may be updated as more current information becomes available.

13. Part-time and mature students who already hold a higher education qualification at the same level as, or higher than, their current qualification aim, or have unknown entry qualifications[1], are given a weighting of zero, irrespective of their postcode.

14. We calculate a 'widening access average weight' (separately for full-time and part-time students) as follows:

$$\frac{\text{Total weight for all students in the population}}{\text{Total students in the population}}$$

(The population is defined as full-time or part-time (as appropriate) HEFCE-funded UK-domiciled new entrants that generate a Column 4 countable year in the HESES07 re-creation.)

15. Some students are excluded from the population that is defined above:

- those with a postcode that has been identified in our young participation analysis as being associated with an unfeasible number of young entrants in relation to our population estimates. Typically this would be a postcode relating to a boarding school
- those whose postcode is marked as a non-geographic postcode in the NSPD
- those with a postcode that, although valid, is not mapped to the required Census 2001 geography in the NSPD
- those studying for an ELQ.

Additionally, students with entry qualifications of unknown level are partially excluded from the population defined above. The amount we remove for each student is the proportion of their activity which we treat as ELQ. (For further details see the algorithm for ELQ_PROP in paragraph 34 of Appendix 7.)

16. These excluded students (with the exception of those studying for an ELQ and a corresponding proportion relating to students with entry qualifications of unknown level who are assumed to be studying for an ELQ) are counted in the FTEs in the next step (see paragraph 17), and therefore receive an average weight for the purpose of allocating funds.

17. Each average weight derived from paragraph 14 is London weighted (8 per cent for inner London and 5 per cent for outer London) and applied to the appropriate assumed undergraduate

[1] This may differ slightly from students who study for an ELQ, where students with entry qualifications of unknown level are partially removed from the population – see paragraph 15 of this annex.

(including foundation degree) FTEs for 2008-09 (which may not incorporate all 2009-10 additional student numbers, transfers or mergers).

Improving retention

Full-time students

18. As well as allocating funding to widen access, we will allocate funding to improve retention. For full-time undergraduate students, this is likely to be based on their entry qualifications and age, and is likely to be calculated as follows.

19. Using age and entry qualification information from 2007-08 HESA student data, full-time UK-domiciled undergraduate new entrants are assigned to one of six risk categories (see Table E for further information on how students are assigned to risk categories) which are then weighted as shown in Table D below. Students are only part of the population if they generate a HEFCE-fundable Column 4 countable year in the HESES07 re-creation and are not studying for an ELQ. Additionally students with entry qualifications of unknown level are partially excluded from this population. The amount we remove for each student is the proportion of their activity which we treat as ELQ. (For further details see the algorithm for ELQ_PROP in paragraph 34 of Appendix 7.)

Table D **Risk category weightings**

	Young	Mature
Low risk	0	0
Medium risk	1	1.5
High risk	1.5	2.5

20. For this allocation, mature students are those aged 21 or over on entry. The assignment of students to one of the six risk categories based on entry qualifications and age is shown in Table E below.

Table E **Assignment of students to risk categories based on entry qualifications and age**

	Young	Mature
Low risk	• A-levels/Highers/vocational A-levels with more than 260 or unknown* tariff points • Baccalaureate • Degree or higher • Unknown qualifications†	• A-levels/Highers/vocational A-levels with more than 320 or unknown* tariff points • Degree or higher • Unknown qualifications†
Medium risk	• A-levels/Highers/vocational A-levels with between 161 and 260 tariff points • Foundation course • Vocational A-levels only • Other HE qualification (below degree level)	• A-levels/Highers/vocational A-levels with between 1 and 320 tariff points • Other HE qualification (below degree level) • Foundation course • Access course • Vocational A-levels only
High risk	• A-levels/Highers/vocational A-levels with between 1 and 160 tariff points • BTEC • Access course • Other qualifications • No qualifications	• BTEC • Baccalaureate • Other qualifications • No qualifications

* New entrants whose highest qualification on entry is A-levels but who did not enter via UCAS (the universities and colleges admissions body), so do not have tariff points recorded, are allocated to medium risk.

† New entrants with unknown entry qualifications or an unknown UCAS tariff are given a zero weighting, and are identified in a separate category in the individualised file and indicative allocations spreadsheet to aid with data checking. Institutions should ensure that highest qualification on entry is recorded if students are to be weighted appropriately in the allocation method for this stream of funding.

21. We calculate an 'improving retention average weight' as:

$$\frac{\text{Total weight for all students in the population}}{\text{Total students in the population}}$$

22. The average weight derived from paragraph 21 is given a London weighting (8 per cent for inner London, 5 per cent for outer London) and applied to the assumed full-time undergraduate (including foundation degree) FTEs for 2008-09 (which may not incorporate all 2009-10 additional student numbers, transfers or mergers).

Part-time students

23. The part-time allocation is likely to be distributed pro rata to London-weighted (8 per cent for inner London and 5 per cent for outer London) part-time undergraduate (including foundation degree) assumed FTEs for 2008-09 (which may not incorporate all 2009-10 additional student numbers, transfers or mergers).

Widening access and improving provision for disabled students

24. We also allocate funding for widening access and improving provision for disabled students. This allocation is likely to be calculated using HESA 2007-08 student data as follows.

25. Firstly, we calculate for each institution the proportion of eligible home and EC students who were in receipt of the Disabled Student's Allowance (DSA). These proportions are then ranked and split into quartiles. Students are only part of the population if they generate a Column 4 countable year in the HESES07 re-creation.

26. Next, each institution is assigned to a quartile according to the proportion of students in receipt of the DSA as calculated in paragraph 25, although this is smoothed to ensure that no institution falls by more than one quartile since the previous year. Separate weightings are attached to each of the quartiles, as follows:

Table F **Quartile weightings**

Quartile	Weighting
A (lowest proportion)	1
B	2
C	3
D (highest proportion)	4

27. Finally each institution's share of the allocation is pro rata to the assumed FTEs for 2008-09 (which may not incorporate all 2009-10 additional student numbers, transfers or mergers), weighted according to the quartile in which they fall, and given a London weighting (8 per cent for inner London and 5 per cent for outer London). There is a minimum allocation of £10,000 for each institution.

Redevelopment of the 2007-08 HESA student record

28. Following the redevelopment of the HESA student record for 2007-08 we have made substantial changes to the derived statistics algorithms that may inform the 2009-10 WP allocations. We may need to further refine these algorithms after the release of the web facility to improve the accuracy of the outputs. However, we expect any changes to be minor.

29. This year we have released this web facility in two phases (see paragraphs 18-19 of the main text)'

30. Lydia Booker e-mailed HESES and RAS contacts on 18 July 2008 to announce the availability (via the web facility) of the outputs relating to the derived statistics that may inform the 2009-10 WP allocations.

31. We welcome comments on the suitability of the algorithms; these should be sent to Christine Daniel (e-mail hesa_heses_stats@hefce.ac.uk).

Annex F

Derived statistics that may inform the 2009-10 partial completion weighting

Purpose

1. This annex describes how we may use 2007-08 HESA student data to inform the partial completion weighting calculations for 2009-10. Further details of the algorithms we use on this data are given in Appendix 8.

2. The web facility offers institutions the opportunity to check and, where errors might affect the data used to calculate the weighting, correct their data prior to submission to HESA.

Background

3. In HEFCE 2007/23 we announced that we will use 2007-08 HESA data to calculate partial completion weightings for each institution based on the modules completed by students that complete less than their initial study intentions for the year. To contribute to the weighting a student must complete a minimum of 0.16 FTE. We will refer to modules completed by students that complete less than their initial study intentions, but which (when added together) represent at least 0.16 FTE for the year, as partial completions.

4. To control the extent to which this measure leads to a redistribution of funding between institutions we may cap the extent to which any individual institution can benefit financially from this proposal. Our Board will determine the method for doing this in January 2009 based on 2007-08 HESA student data.

Web facility outputs

5. The web facility will generate an output that shows the standard resource associated with the partial completions. Annex B describes how to access these standard resource calculations and the derived statistics that may be used to calculate the partial completion weighting in an Excel workbook (PCMP07XXXX.xls, where XXXX denotes the HESA institution identifier).

6. While the outputs do not attempt to calculate the partial completion weighting, they identify those students and FTE that we may count in calculating the weighting, as well as the associated standard resource (based on 2008-09 prices) of the partial completions.

7. At this stage, it is not possible to calculate (or even provide an indicative value for) the partial completion weighting due to the absence of finalised 2007-08 HESA data or the equivalent historical data which might provide a good approximation.

8. The calculation of standard resource associated with partial completions is only provided to highlight potential 2007-08 HESA data errors. It should not be considered to be any kind of funding commitment by us, and is without prejudice to any funding decisions our Board may make relating to partial completions at any institution.

9. The partial completion derived statistics can be rebuilt from the individualised file which we provide. This file (PCMP07XXXX.ind – please see Annex B for how to access this file) contains details of how each student was assigned to a mode and price group and, where relevant, details of why they did not contribute. A full description of the data in the individualised file is given in Appendix 8.

Redevelopment of the 2007-08 HESA student record

10. Given the extent of the redevelopment of the 2007-08 HESA student record, and as our understanding improves of the module and credit data used to identify partial completions and determine their FTE, we may need to further refine these algorithms after the release of the web facility to improve the accuracy of the outputs.

11. We welcome comments on the suitability of the algorithms; these should be sent to Lydia Booker (e-mail hesa_heses_stats@hefce.ac.uk).

Annex G

2007-08 co-funded employer engagement student numbers (CFEE07) re-creation

Purpose

1. This annex describes the methods we use to generate a provisional CFEE07 re-creation via the web facility. It also describes how we will use the comparison of CFEE07 and the CFEE07 re-creation in the '2007-08 statistics derived from HESA data for the monitoring and allocation of funding' exercise. The outputs are provisional because they are generated before institutions submit final 2007-08 student data to HESA.

Web facility outputs

2. The web facility generates a CFEE07 re-creation. Annex B describes how to access the provisional CFEE07 re-creation Excel workbook (CFEE07XXXX.xls, where XXXX denotes the HESA institution identifier). The provisional CFEE07 re-creation workbook contains the following worksheets.

Excel workbook CFEE07XXXX.xls

Page number	Worksheet*	Title
1	Coversheet	Title page
2	CFEE	CFEE07 re-creation table

* This worksheet reference corresponds to the spreadsheet tabs.

Note: The workbook contains a version number that will increment by one following each successful submission of data to the web facility. There are also contact details for any queries relating to data in the workbook.

3. All the information in the CFEE07 re-creation tables can be rebuilt by categorising and aggregating the data in the individualised file which we provide. The file (CFEE07XXXX.ind) contains details, in the form of HESA and derived fields, of how each student was classified in the re-creation. A full description of the data in the individualised file is given in Appendix 9.

Methods

4. The algorithms used to generate the provisional CFEE07 re-creation are described in Appendix 9.

5. The CFEE07 return will not be collected until late August 2008; however the web facility provides an early opportunity to verify that the 2007-08 HESA student data are correct for these students. By early 2009 we will generate a CFEE07 re-creation from the 2007-08 HESA student data submitted by institutions. We will compare the CFEE07 re-creation to the CFEE07 return. Where appropriate, we will also re-calculate the funding associated with the recruited FTEs as well as any differences from the original co-funded employer engagement funding based on the CFEE07 re-creation by applying the same formulae that were used to calculate the funds for CFEE07. We intend to use the algorithms described in Appendix 9 to generate the CFEE07 re-creation; however, we may change them if we identify any improvements that can be made.

6. The CFEE07 return will include data on the employers involved in the co-funded employer engagement provision, including characteristics of the employer such as sector and size of company. We are not able to re-create these data in the CFEE07 re-creation.

7. We expect to select institutions to explain discrepancies between their CFEE07 and CFEE07 re-creation on the basis of thresholds. These are likely to be set in terms of funding differences arising from CFEE07 and the CFEE07 re-creation. Notwithstanding the thresholds, we may also ask for further information about the comparison from any institution.

8. Each institution required to make a response will be asked to reconcile the two data sources. After reconciling the two data sources and confirming that the CFEE07 re-creation reasonably reflects the outturn position for 2007-08, the CFEE07 re creation will supersede CFEE07 and any consequent funding adjustments (including 2008-09 WP allocations and other targeted allocations) will be made (subject to the availability of our funds).

Redevelopment of the 2007-08 HESA student record

9. Following the redevelopment of the HESA student record for 2007-08 we have made changes to the CFEE07 re-creation algorithms. We may need to further refine these algorithms after the release of the web facility to improve the accuracy of the outputs. However, we expect any changes to be minor.

10. We welcome comments on the suitability of the algorithms; these should be sent to Lydia Booker (e-mail hesa_heses_stats@hefce.ac.uk).

Annex H

Comparison of RAS07 and the RAS07 re-creation

Purpose

1. This annex describes the methods we use to generate a provisional RAS07 re-creation via the web facility. It also describes how we are likely to use the comparison of RAS07 and the RAS07 re-creation in the 2007-08 statistics derived from HESA data exercise. The outputs are provisional because they are generated before institutions have submitted their final data to HESA, from which the re-creation is built.

Web facility outputs

2. The provisional RAS07 re-creation is generated from HESA 2007-08 student data submitted to the web facility. Annex B describes how to access the provisional RAS07 re-creation Excel workbook (RASR07XXXX.xls – where XXXX denotes the HESA institution identifier). The provisional RAS07 re-creation workbook contains the worksheets listed in Table G below.

Table G **Excel workbook RASR07XXXX.xls**

Page number	Worksheet (see tabs on spreadsheet)	Description
1	Coversheet	Title page
2	R1A	Provisional RAS07 re-creation – Form R1a: Fundable home and EC fee-paying full-time research students by year of programme
3	R1B	Provisional RAS07 re-creation – Form R1b: Fundable home and EC fee-paying part-time research students by year of programme
4	FUNDS	Re-created calculation of Research Degree Programme (RDP) supervision funding for 2008-09 using provisional HESA 2007-08 student data and other HEFCE research funding data
5	Summary	Summary comparison of RAS07 and the provisional RAS07 re-creation (fundable home and EC fee-paying students, years 1-3 FT, 1-6 PT)
6	Summary by subject groupings	Summary comparison of RAS07 and the provisional RAS07 re-creation (fundable home and EC fee-paying students, years 1-3 FT, 1-6 PT) by broad subject group
7	rR1A	RAS07 Form R1a: Fundable home and EC fee-paying full-time research students by year of programme
8	rR1B	RAS07 Form R1b: Fundable home and EC fee-paying part-time research students by year of programme
9	rFUNDS	RAS07 calculations of Research Degree Programme (RDP) supervision funding for RAS06 student data and other HEFCE research funding data

10	Broad subject groups	Mapping of units of assessment (UOAs) to broad subject groups used for comparison tables
11	Research cost bands	Assignment of units of assessment to research cost bands

Note: The workbook also contains a version number that will increment by one following each successful submission of data to the extranet, as well as contact details for any queries relating to data in the workbook.

3. The information in the RAS07 re-creation tables can be rebuilt from the individualised file provided. The file (RASR07XXXX.ind) contains details of how each student was classified in the re-creation, and, where relevant, details of why they were excluded. A full description of the data in the individualised file is given in Appendix 10.

Methods

4. The algorithms used to generate the provisional RAS07 re-creation are described in Appendix 10.

5. Appendix 11 contains troubleshooting guidance for the causes of differences between RAS07 and the provisional RAS07 re-creation, especially where they are attributable to errors in HESA data.

6. All known problems of fit with the provisional RAS07 re-creation algorithms are described in Appendix 12.

2007-08 statistics derived from HESA data exercise

7. The '2007-08 statistics derived from HESA data for the monitoring and allocation of funding' exercise will contain a comparison of RAS07 and the RAS07 re-creation. This will be achieved by generating a RAS07 re-creation of forms R1a and R1b, by applying the algorithms detailed in Appendix 10 to HESA 2007-08 student data. This exercise is likely to be launched in late 2008.

8. For the purpose of the web facility, we will use a mapping of UOAs (those UOAs used for the 2001 Research Assessment Exercise (RAE)) to subject groups generated from HESA 2006-07 staff data. If we identify an error in our algorithms, we will correct it.

9. We will assign each UOA to one of eight broad subject groups. The assignments of UOAs to subject groups are given in paragraph 7 of Appendix 10. In the comparison of RAS07 and the provisional RAS07 re-creation by broad subject group we have included only FTE from tables R1a and R1b. In addition we have only included FTE from years 1-3 from the R1a full-time tables, and years 1-6 from the R1b part-time tables.

10. We will compare the RAS07 re-creation to RAS07 after the 2007-08 student data have been finalised with HESA. We are likely to select institutions to make a response to the derived statistics exercise based on absolute differences in postgraduate research student FTEs between RAS07 and the RAS07 re-creation over eight broad subject groups.

Notwithstanding the thresholds, we may also ask for further information from any institution in respect of this comparison.

11. Each institution requested to make a response will be asked to explain the reasons and causes of differences between RAS07 and the RAS07 re-creation. Where we are satisfied with these explanations, and where it is appropriate, we will ask the institution to confirm that the RAS07 re-creation reasonably reflects the outturn position for 2007-08. Upon this confirmation, the RAS07 re-creation will supersede RAS07 and any consequent grant adjustments will be made (subject to the availability of our funds).

Redevelopment of the 2007-08 HESA student record

11. Following the redevelopment of the HESA student record for 2007-08 we have made substantial changes to the RAS07 re-creation algorithms. We may need to further refine these algorithms after the release of the web facility to improve the accuracy of the outputs. However, we expect any changes to be minor.

12. This year we have released this web facility in two phases (see paragraphs 18-19 of the main text)'

13. Lydia Booker e-mailed HESES and RAS contacts on 5 September 2008 to announce the availability of the RAS07 re-creation outputs via the web facility.

14. We welcome comments on the suitability of the algorithms; these should be sent to Lydia Booker (e-mail hesa_heses_stats@hefce.ac.uk).

Annex I

Research Degree Qualification Rates (RDQRs)

Purpose

1. This annex provides details of how we intend to measure the rate at which students undertaking research degree programmes successfully gain their doctorate research degree qualification. We intend to use HESA student data to identify and calculate these rates and the accompanying context statistics.

2. As part of the 2007-08 HESA data collection process, all institutions will be required to confirm that their 2007-08 RDQRs as generated by the HEFCE web facility are suitable for publication.

3. We expect institutions to check that the HESA data used to calculate their RDQRs are suitable for publication. This may involve correction of their 2007-08 student data prior to its submission to HESA, and also correction of any historical HESA data errors that impact the RDQR calculations. Details of the algorithms we have used to produce the RDQRs from HESA student data are in Appendix 13.

4. Institutions are expected to verify all underlying data, even though only certain RDQRs will be subject to publication and use by the Quality Assurance Agency for HE (QAA) in 2009. As the web facility offers the only opportunity to ensure that 2007-08 HESA data (which may impact future published rates) are fit for purpose, institutions are encouraged to check the rates for all years and for all modes of study.

5. The RDQRs that will be published in 2009 are presented in the 'LEVEL1 – ALL' worksheet of the RDQR outputs (see paragraph 30). The 'LEVEL1 – ALL' worksheet contains two sets of tables. The first set presents the data before they have been rounded for publication, to enable institutions to check the underlying data. The second set presents the data adjusted for the rounding strategies that we intend to apply for publication.

Data verification

6. The web facility allows institutions to verify the data underlying their RDQRs. RDQRs are calculated for cohorts of students. The academic year in which a student commences their study towards a postgraduate research degree determines which cohort the student is associated with.

7. We publish data for full-time students once we have seven academic years' HESA data for their cohort. Therefore, the availability of 2007-08 HESA student data will allow RDQRs for the 2001-02 full-time cohort to be published. Use of the web facility will be the final opportunity for institutions to verify RDQRs for full-time students ahead of publication. We provide the outputs described below to enable institutions to identify and remedy any issues with the accuracy of their RDQRs.

8. We publish data for part-time and 'mode-switch' students once we have nine academic years' HESA data for their cohort. Therefore, the availability of 2007-08 HESA student data

will allow RDQRs for the 1999-2000 part-time and mode-switch cohorts to be published. Use of the web facility will be the final opportunity for institutions to verify RDQRs for part-time students ahead of publication. We provide the outputs described below to enable institutions to identify and remedy any issues with the accuracy of their RDQRs. 2009 will be the first year that we will publish RDQRs for part-time and mode-switch students.

9. Where institutions find inaccuracies in their data the following remedial actions should be taken (following approval of the necessary action plan documentation – see paragraphs 11 to 19):

- 2007-08 student data – institutions should correct the error prior to submission to HESA
- 2006-07 student data – institutions should correct the error using the 2006-07 HESA fixed database
- earlier years' HESA student data – institutions should submit HDE files to correct primary derived fields.

10. The timetable for verifying RDQRs is as follows:

Table H **Indicative timetable for RDQR verification process**

Indicative date/deadline	**Process**
6 October 2008	Deadline for submitting an RDQR action plan
13 October 2008	Deadline for signing off amendments to 2006-07 HESA student data or HDE files to correct derived fields
August 2008 – October 2008	Institutions sign off their RDQRs as part of the 2007-08 HESA data collection sign-off process
December 2008	We receive final 2007-08 student data from HESA and produce final RDQRs to be published in 2009
January/February 2009	We pass final RDQRs to QAA to inform its audit activities
Summer/autumn 2009	We intend to publish RDQRs

Action plans

11. Institutions that identify that their RDQRs are not suitable for publication should determine why and submit an action plan that describes the cause and explains how it will be resolved. RDQR action plans will be kept as a permanent record and audit trail for changes that are made to ensure that RDQRs are suitable for publication. We will audit a sample of institutions' action plans to ensure that the changes on the action plans are supplemented by appropriate records and evidence. Guidance for completing and submitting an action plan is in Annex L.

12. We expect any issues with RDQRs to fall into either or both of the following categories:

- errors in HESA 2006-07 student data
- errors in HESA student data for years between 1999-2000 and 2005-06 (historical data errors).

13. If an institution wishes to correct its 2006-07 HESA data for incorporation in the 2007-08 RDQRs for full-time, part-time or mode-switch students to be published during 2009, it should submit an RDQR action plan by **6 October 2008**.

14. Errors in 2006-07 HESA student data should be resolved by submitting a revised 2006-07 HESA student return to HESA. HESA will not accept amendments to 2006-07 HESA student data until it has received confirmation that we are content with the institution's action plan. Amendments should be signed off by **13 October 2008** to ensure incorporation. Annex J describes the process for making amendments to 2006-07 HESA student data.

15. If an institution identifies errors in its HESA student data for years between 1999-2000 and 2005-06 and it wishes to correct the derived fields for the individual students affected, it should submit an RDQR action plan by **6 October 2008**.

16. Errors in HESA student data for years between 1999-2000 and 2005-06 should be resolved by submitting an HDE file to enable us to correct the primary derived fields for the individual students affected.

17. Appendix 13 details all the algorithms and derived fields used in the calculation of the RDQRs. Institutions' HDE files will correct the values for one or more of the following primary derived fields:

- RDQRSTRT
- DOCPOP
- DOCYEAR
- DOCQUAL
- QUALYEAR
- RDQRMODE
- RDQRTYPE
- ACTIVEXX (where XX denotes the academic year in which the student's activity is being assessed)
- RDQRTUIT
- RDQRSBJ.

The procedure for generating and submitting an HDE file is given in Annex K.

18. We link records, across HESA student data returns, using a combination of the fields HUSID, INSTID (UKPRN for 2007-08) and NUMHUS. Where this link has been broken across years through institutional data error, more than one record will be produced within the RDQR derived data for a single student instance. Where this occurs an institution should use HDE files to set one of the instances so that DOCPOP is equal to 0, thus removing it from the doctoral population. The remaining instance should then be corrected using HDE files to set the DOCPOP to 1 to include it within the rates, and then all other primary derived fields should be amended as appropriate to correctly represent the student's experience.

19. HDE files should be signed off by **13 October 2008** to ensure incorporation in the RDQRs to be published in 2009. Annex K describes the process for submitting HDE files.

Overview of the RDQR methodology

RDQR population

20. For each year, starting with 1999-2000, we have identified the population of students that started aiming for a postgraduate research degree (MPhil, PhD and comparable qualifications). Each population is tracked across academic years, up to 2007-08, in order to derive cumulative RDQRs across time. We removed students that leave in the first year of their qualification aim.

RDQR countable

21. Students in the RDQR population become countable towards RDQRs when they start aiming for a doctorate research degree (PhD and comparable qualifications). For example, if a student is aiming for an MPhil they will be in the RDQR population of the year that they started aiming towards the MPhil. However, they will only become countable towards RDQRs if they later change their qualification aim to a doctorate research degree.

Qualifiers

22. Students that become countable towards the RDQR are regarded as qualifiers when they obtain their doctorate research degree.

RDQR calculation

23. RDQRs are calculated by dividing the number of qualifiers by the number of students that become countable towards the RDQR.

Groupings

24. For each year the rates have been grouped into the following categories depending on their mode:

- full-time students
- part-time students
- students that switch mode during the qualification aim.

25. Similarly the students have been grouped as either home and EC, or overseas, depending on their fundability and residential status.

Context statistics

26. Four context statistics are presented alongside the RDQRs: the number of active students, the number on science-based courses, the number receiving funding from the UK Research Councils, and the number that transfer to another institution.

27. Active students are defined as those that are countable towards RDQRs who have not been awarded a doctorate research degree, but who are actively pursuing studies at the point of calculating the RDQRs.

28. For this purpose, science-based courses are defined as medicine, veterinary science, subjects allied to medicine, biological sciences, physical sciences, engineering, mathematics, and agriculture.

29. We include a context statistic on the number of students who transfer to another institution. The methods we will use to identify such students are described in Appendix 13.

RDQR statistics outputs

30. Annex B describes how to access the RDQR derived statistics, in an Excel workbook, that we intend to publish during 2009 (RDQR07XXXX.xls, where XXXX is the HESA identifier for the institution). This workbook contains the following worksheets:

- cover page
- LEVEL1_ALL – RDQRs after seven years for 1999-2000, 2000-01, 2001-02 full-time students, after nine years for 1999-2000 part-time and mode-switch students
- LEVEL1_SECTOR – sector RDQRs after seven years for 1999-2000 and 2000-01 full-time students, six years for 2001-02 full-time students (2007-08 HESA data for the sector is still unavailable) and after eight years for 1999-2000 part-time and mode-switch students (2007-08 HESA data for the sector is still unavailable)
- LEVEL2_ALL – RDQRs for all students in all cohorts
- LEVEL2_HOMEEC – RDQRs for home and EC students in all cohorts
- LEVEL2_ISOV – RDQRs for Island and overseas students in all cohorts
- LEVEL2_SECTOR – sector RDQRs for all students in all cohorts (2007-08 HESA data for the sector is still unavailable).

31. We intend to publish the statistics presented in the worksheet ‘LEVEL1_ALL’. The remaining worksheets allow comparison of RDQRs against the sector (note that 2007-08 HESA data is not available for the sector; the latest sector tables are for 2006-07 data).

32. The information in the RDQR tables can be rebuilt from the individualised file provided by the web facility. The file (RDQR07XXXX.ind) contains details of how each student was classified in the RDQR tables, and where relevant, details of why they were excluded from the population. A full description of the data in the individualised file is given in Appendix 13 and instructions on how to generate the RDQR tables from the individualised file are given in Appendix 14.

33. This is also an early opportunity for institutions to verify data that will be published in the future. Institutions must note that this is likely to be the only remaining opportunity to correct HESA student data for the year 2007-08 for the purposes of RDQRs.

RDQR sign-off

34. All institutions will be required to confirm as part of the 2007-08 HESA student data collection process that their RDQRs, as generated by the HEFCE web facility, are suitable for publication.

Redevelopment of the 2007-08 HESA student record

35. Following the redevelopment of the HESA student record for 2007-08 we have made changes to the RDQR algorithms. We may need to further refine these algorithms after the release of the web facility to improve the accuracy of the outputs. However, we expect any changes to be minor.

36. This year we have released this web facility in two phases (see paragraphs 18-19 of the main text)'

37. Lydia Booker e-mailed HESES and RAS contacts on 5 September 2008 to announce the availability of the RDQR outputs via the web facility.

38. Note that we provided details of the RDQR deadlines (see Table H above) when we announced the release of the RDQR outputs on 5 September 2008.

39. We welcome comments on the suitability of the algorithms; these should be sent to Lydia Booker (e-mail hesa_heses_stats@hefce.ac.uk).

Annex J

Submitting amendments to 2006-07 HESA data for the purpose of correcting RDQRs

Purpose

1. HESA 2006-07 student data were collected from all institutions and committed to HESA's system through the 2006-07 Student Collection (C06011). Where institutions discover through the use of this web facility that post-collection amendments to 2006-07 HESA data are necessary to correct their RDQRs they must follow the procedure laid out in this annex.

2. This is the only opportunity that institutions will be offered to make these exceptional amendments to the fixed database for the purposes of correcting the underlying data for RDQRs. In future years, when we publish RDQRs for future cohorts, we may not provide any further opportunity to amend 2006-07 data (including HDE files).

3. The agreement between us and HESA allows for the costs of processing such exceptional amendments to be recovered from institutions by HESA. It has been agreed that for the student record this charge should be set at 20 per cent of the institution's annual HESA subscription. Amendments submitted via this route will not be used to inform routine publications such as the HESA student volume, the calculation of the performance indicators or the Unistats statistics until the fixed database becomes available in April 2009. HESA has indicated that it will not use these data for ad-hoc analysis before April 2009.

Submitting amendments to HESA data

4. Institutions that wish to make amendments to their 2006-07 HESA student data to correct their RDQRs are required to submit an RDQR action plan to HEFCE before we will consider whether to authorise HESA to receive amendments. Full guidance on submitting and completing an RDQR action plan is given in Annex L.

5. If post-collection amendments need to be made to 2006-07 HESA student data, as a result of our approving an action plan, we will e-mail HESA, authorising them to accept these amendments. We will copy this e-mail to the institution. This authorisation will also summarise the nature of the amendments to be made (for example, the HESA fields that we expect to be amended along with an approximation of the number of records that will be affected) and a date by which we expect the revised data to be submitted. This information will be extracted directly from the approved action plan. HESA will only open its post-collection system when it has received explicit instructions from us to do so, and will only accept those amendments consistent with the summary that we have approved together with any other fields required for validation compliance.

6. HESA's post-collection processing is similar to its standard collection process. HESA will carry out the same data quality assurance processes that apply for the standard collection. A full description of the collection system can be found at

http://submit.hesa.ac.uk/help. Note however that post-collection submissions will only be accepted during a limited period.

7. Shortly after data are ‘committed’ during the post-collection process, we will update our copy of the institution’s 2006-07 HESA student data that are used by the web facility. Once this is completed, we will notify the institution and ask it to resubmit its 2007-08 student data to the facility, to ensure that the amendments have had the expected effect. The RDQRs generated at this point may also include the incorporation of any HDE files that have been submitted for the purposes of correcting RDQRs.

8. Once the institution has confirmed that the amendments have had the expected effect on its RDQRs we will e-mail HESA, copied to the institution, notifying that we are content with the revised data. HESA will continue the exceptional data collection processing (for example, credibility checking) until the process is complete.

9. Throughout this process, if the institution identifies that further 2006-07 HESA data amendments are required to correct its RDQRs, we will ask the institution to submit a revised RDQR action plan, and the procedure described from paragraph 4 onwards will restart. Similarly, if during credibility checking, or at any other point during the processing, HESA identifies that further amendments to 2006-07 HESA data are required, the data must be ‘decommitted’ and any necessary revisions made before the data are once again ‘committed’, at which point the procedure described from paragraph 6 onwards will resume.

10. At the end of the process, HESA will mark the return as ‘credible’ to allow the institution to record a ‘sign-off’ transaction.

11. We will identify and chase institutions that have not submitted valid files by the expected date or where there are other significant delays, for example in achieving a valid ‘commit’ transaction.

Annex K

Submitting historical data error (HDE) files for the purpose of correcting RDQRs

Purpose

1. This annex describes the procedure for creating and submitting historical data error (HDE) files to derived fields for the purpose of correcting RDQRs. An HDE file will only be accepted where errors in HESA student data for years between 1999-2000 and 2005-06 have resulted in incorrect RDQRs. HDE files make changes to the derived fields that generate the RDQRs rather than to the underlying HESA data.

2. This annex details the data structure and format for HDE files. Institutions must only supply HDE files using the file structure and format detailed in this annex. This annex also details which derived fields should be amended to correct a given error in HESA data.

3. Institutions that wish to submit an HDE file are required to submit an RDQR action plan before we will accept the files. Full guidance on submitting and completing an RDQR action plan is given in Annex L.

4. Once we have approved an RDQR action plan we will, where appropriate, e-mail the institution to confirm that we are content for it to submit the HDE file. This confirmation will summarise the nature of the corrections to be made (for example, the RDQR derived fields that we expect to be amended along with an approximation of the number of records that will be affected) and a date by which we expect the HDE file to be submitted. This information will be extracted directly from the approved action plan.

5. HDE files should contain the data structure and format, which are described in paragraphs 10-13. Figure 5 gives an example of an HDE file.

6. These specifications are necessary to ensure we can process changes to derived fields, using HDE files, in a timely and accurate manner. We will require institutions to re-submit HDE files that differ, either in structure or format, from the specifications in this annex.

7. We only allow certain derived fields to be modified by an HDE file. These fields, which we refer to as ‘primary’ derived fields, are those which are directly used to populate the RDQR tables. For example, the value of DOCPOP determines whether a record is included in the RDQR tables. It is therefore a primary derived field. In contrast, the field RDQRPOP (which is used to derive DOCPOP) is not directly used to populate the RDQR tables; we refer to such fields as ‘secondary’ derived fields. A list of primary derived fields is given in Table I.

Table I **List of primary derived fields**

Primary derived field name	Description	Re-creation tables
RDQRSTRT	Academic year of date of commencement for the programme of study	RDQR
DOCPOP	Flag indicating study for a doctoral programme	RDQR
DOCYEAR	Academic year that the student started studying for a doctoral programme	RDQR
DOCQUAL	Flag indicating whether the student was awarded a doctoral qualification	RDQR
QUALYEAR	Academic year that the student was awarded a doctoral qualification	RDQR
RDQRMODE	RDQR: mode of study	RDQR
RDQRTYPE	RDQR: residential status	RDQR
ACTIVEXX (where XX denotes the academic year in which the student's activity is being assessed)	Flag indicating whether the student is actively pursuing studies in the academic year	RDQR
RDQRTUIT	RDQR: source of fees	RDQR
RDQRSBJ	RDQR: subject	RDQR

8. Since each HDE file will only alter primary derived fields, this may give rise to inconsistencies with a record's secondary derived fields. For example, if the value of DOCPOP is altered by an HDE file, its value may be inconsistent with the value of RDQRPOP.

9. To allow institutions to check that an HDE file has had the desired effect, a field called 'HDE', is included in the RDQR individualised file. This takes the value 1 if a historical data error correction has been applied to the record; otherwise its value is 0. This field also allows institutions to determine which records may have inconsistent secondary derived fields, as a result of a historical data error correction.

Format and naming

10. Changes to primary derived fields must be uploaded as a comma-separated file to the 'HESAOVR07 data collection' facility. You can find this facility by logging onto the HEFCE extranet, selecting 'HEFCE Resources' and selecting '2007-08 statistics derived from HESA data'. HDE files must be given a file name in the form rdqXXXXn.amd, where:

- XXXX is the HESA institution identifier for the institution
- n is a sequential number starting at 1.

For example, the first HDE file submitted would be called rdqXXXX1.amd, and the second would be called rdqXXXX2.amd.

Structure

11. Each record must contain complete data for all fields included in the HDE file, even if a particular field remains unchanged in some cases.

12. Only primary derived fields should be included in the change line (line 6) for HDE files. For further information on which primary derived fields should be amended for a given error in HESA data, see paragraphs 18-30.

13. The HDE must contain a header in the following form:

line 1 – HDE file reference in the form rdqXXXXn where XXXX is the HESA institution identifier and n is a sequential number starting at 1; this will be the file name with the '.amd' file extension removed.

Line 2 – creation date of the HDE file in the form ddmmyyyy. For example 01082008 for a file created on 1 August 2008.

Line 3 – a brief description of the change and the reference number of the cohort affected on the action plan. For example 'Corrections to primary derived field DOCPOP for the cohort with start year of 2001-02 (RDQRSTRT = 2001) due to errors in QUALAIM – reference number 3'.

Line 4 – this line must contain the following words separated by a comma: HDE and the start year(s) of the cohort affected. For example, for an error to HESA data for 2002-03 and 2003-04 this line would read 'HDE, 2002, 2003'.

Line 5 – the HESA fields used to identify records on the HDE file, comma-separated. For example, OWNPSD could be used to identify records at the course level or HUSID and NUMHUS could be used to identify records on a student level.

Line 6 – the primary derived fields being changed, comma-separated. For example, 'RDQRSTRT, RDQRMODE'.

Line 7 – number of records in the file, excluding the first 12 lines of header information.

Line 8 – the field used to calculate the file check-sum (see paragraph 14 for an explanation of the file check-sum).

Line 9 – file check-sum value.

Lines 10 and 11 – these lines can contain any notes the institution wishes to include.

Line 12 – fields included in the HDE file. These fields must appear in the same order as each row of the data and must be comma-separated. For example: OWNPSD, RDQRMODE.

Line 13 – correction data must begin on this line.

Check-sum

14. To ensure HDE files have not been corrupted during transit, we use a check-sum. The check-sum is computed by summing the values of the field specified on line 8 over all records in the file. The calculated value should be returned on line 9 of the file. If an HDE file does not contain any numeric fields suitable for calculating a check-sum, an additional field must be included solely for this purpose, for example QUALAIM. Numeric fields that contain values greater than 20,000 (approximately) are unsuitable for calculating the check-sum. If information is not being changed at the student level, then a sequential field called RECNO may be created for the purpose of the check-sum. For example RECNO may contain 1, 2, 3, 4, 5 and so on.

Identifying records

15. To enable us to link HDE files to our derived HESA dataset, we must be able to identify the records on the HESA return where the correction should be applied. The field, or combination of fields, enabling us to achieve this must be listed, comma-separated, on line 5 of the HDE file. We do not prescribe which fields should be used for this purpose, only that they enable us to identify the records on the HESA return.

Saving files

16. Saving HDE files in Microsoft Excel usually results in the loss of leading zeros and the corruption of very large values into exponential form (for example, 9.91E+12). To avoid this, we recommend that HDE files are produced, viewed and saved using a text editor, for example Notepad.

Outcome

17. When we receive an approved HDE file in the structure and format detailed in this annex, we will aim to incorporate the changes into the RDQR calculations within 10 working days. Institutions will be notified by e-mail when they can submit data to the web facility that will generate RDQRs that incorporate the changes.

RDQR derived fields to correct

18. For a given HESA data error the RDQR derived field that should be amended will depend on the type and nature of the error. The guidance that follows is not intended to be exhaustive, but should describe which primary derived fields should be amended for the most common HESA errors that affect RDQRs. If an institution is in any doubt about which RDQR derived field should be amended, it should seek advice from Lydia Booker (e-mail hesa_heses_stats@hefce.ac.uk or tel 0117 931 7338).

19. Only primary derived fields should be corrected using HDE files. For a list of these fields and further details please refer to Table I of Annex K.

Students incorrectly excluded from or included in the RDQRs

20. For students that have been incorrectly excluded from or included in the RDQRs, the historical data error file should amend the derived fields DOCPOP and DOCYEAR. For students incorrectly excluded from the RDQRs, DOCPOP should be returned as 1 and DOCYEAR should reflect the first academic year in which their qualification aim was an RDP. For example, a student that first started aiming for an RDP in 2003-04 should have DOCYEAR = 2003. For students incorrectly included in the RDQRs, DOCPOP should be returned as 0.

21. The algorithms for deriving DOCPOP and DOCYEAR are given in paragraphs 22 and 23 respectively of Appendix 13. Most commonly, historical errors in the HESA field QUALAIM will lead to the derived fields DOCPOP and DOCYEAR requiring correction. However, under certain circumstances, errors in any of the following HESA fields can cause students to be incorrectly excluded from or included in the RDQRs: COMDATE, DATELEFT, MODE (where a student is incorrectly recorded as dormant, writing-up or on sabbatical) and RSNLEAVE.

Students in an incorrect research degree cohort start year

22. For students that appear in an incorrect research degree cohort start year, the historical data error file should amend the derived field RDQRSTRT. RDQRSTRT should reflect the first academic year in which the qualification aim for the student was a research doctorate, research masters or research bachelorship. For example, a student that first started aiming for a research masters in 2004-05 should have RDQRSTRT=2004.

23. The algorithm for deriving RDQRSTRT is given in paragraph 21 of Appendix 13. Only historical errors in the HESA field COMDATE will lead to the derived field RDQRSTRT requiring correction.

Students with an incorrect research doctorate start year

24. For students with an incorrect research doctorate start year, the historical data error file should amend the derived field DOCYEAR. DOCYEAR should reflect the first academic year in which the qualification aim for the student was a research doctorate. For example, a student that first started aiming for a research doctorate in 2002-03 should have DOCYEAR = 2002.

25. The algorithm for deriving DOCYEAR is given in paragraph 23 of Appendix 13. Only historical errors in the HESA field QUALAIM will lead to the derived field DOCYEAR requiring correction.

Students incorrectly treated as either qualified or not qualified

26. For students incorrectly treated as either qualified or not qualified, the historical data error file should amend the derived fields DOCQUAL and QUALYEAR. For students incorrectly treated as not qualified, DOCQUAL should be returned as 1 and QUALYEAR should reflect the academic year in which their research doctorate was awarded (for example, a student awarded their research doctorate in 2002-03 should have QUALYEAR = 2002). For students incorrectly treated as qualified, DOCQUAL should be returned as 0.

27. The algorithms for deriving DOCQUAL and QUALYEAR are given in paragraphs 26 and 27 respectively of Appendix 13. Most commonly, historical errors in the HESA fields QUAL1 or QUAL2 will lead to the derived fields DOCQUAL and QUALYEAR requiring correction. However, under certain circumstances, errors in the HESA field MODE (where a student is incorrectly recorded as dormant, writing-up or on sabbatical) can cause students to be incorrectly treated as not qualified.

Students with an incorrect mode

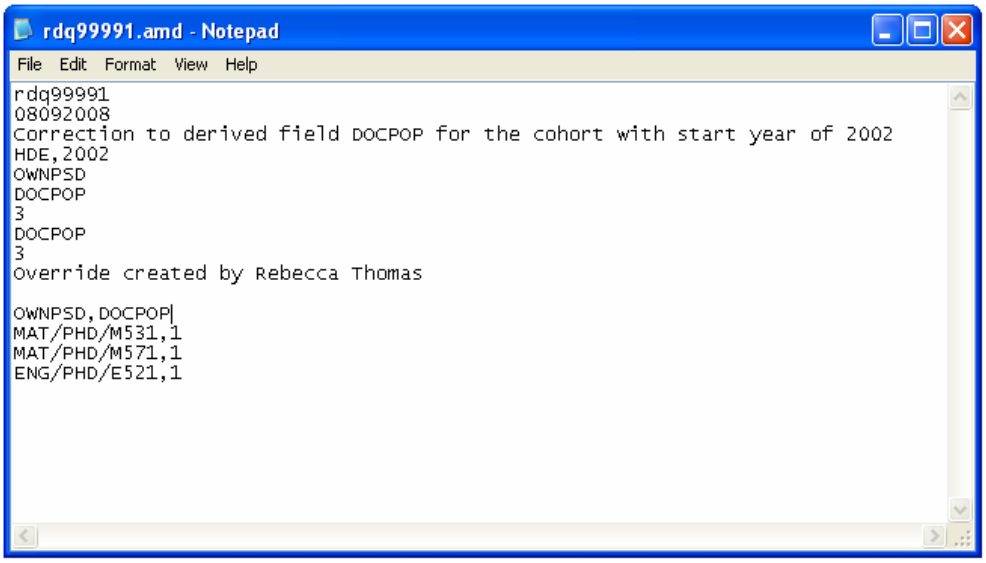
28. For students with an incorrect mode, the historical data error file should amend the derived field RDQRMODE. RDQRMODE should reflect the student's mode either in 2006-07, or in the last academic year in which they were active (FT = full-time, PT = part-time and SWTCH = switched from part-time to full-time or vice versa during the programme of study). For example, a student that started as full-time but switched to part-time study during the programme should have RDQRMODE = SWTCH.

29. The algorithm for deriving RDQRMODE is given in paragraph 34 of Appendix 13. Most commonly, historical errors in the HESA field MODE will lead to the derived field RDQRMODE requiring correction. However, under certain circumstances, errors in the HESA fields LOCSDY or FEEBAND can cause students to be assigned an incorrect mode.

Students incorrectly returned with more than one HIN (HUSID X INSTID or UKPRN X NUMHUS) across years

30. Please see paragraph 18 of Annex I for guidance on correcting these issues.

Figure 5 **An example RDQR historical data error file**



Annex L

Guidance for completing and submitting RDQR action plans

Purpose

1. This annex provides guidelines for completing RDQR action plans. It also describes how to submit plans to the HEFCE extranet. Paragraph 32 of Annex B describes how to access the RDQR action plan templates.

Approval

2. Institutions that identify that their RDQRs are not suitable for publication should determine the cause of the issue and submit an action plan. Action plans must be submitted before either HESA will accept amendments to 2006-07 data, or we will accept HDE files to correct the derived fields for the individual students affected for the years between 1999-2000 and 2005-06.

3. Where an institution wishes to amend its 2006-07 HESA student data, we will use the action plan to gain an understanding of the reasons why amendments are being made, and to which HESA fields, so that the impact of incorporation can be checked. In addition the action plan provides us with dates when we can expect the institution to submit the data to HESA. Similarly, where an institution wishes to correct the derived fields for individual students affected by errors in data for years between 1999-2000 and 2005-06, we will use the action plan to gain an understanding of the reasons why changes to the derived fields are being submitted, and to which derived fields, so that the impact of incorporation can be checked.

4. RDQR action plans will be kept as a permanent record and audit trail for changes that are made to ensure that RDQRs are suitable for publication. We may audit RDQR action plans to ensure that they are supported by the appropriate records and evidence.

Detailed requirements for RDQR action plans

5. Below are detailed instructions about the information that we require in each column of the action plan. The RDQR action plan template is available for download from the HEFCE extranet. An example RDQR action plan is given in Appendix 15 (an Excel file).

6. We require the following information in the RDQR action plan template, APRDQR07XXXX.xls (where XXXX denotes the HESA institution identifier).

Column 1

7. This column should contain a sequential reference number starting at 1 which is used to reference each area of difference identified on the action plan. This is provided automatically in the action plan template.

Column 2

8. This column should describe the cohort affected. We would typically expect the cohort to be identified by the cohort start year (RDQRSTRT – see Appendix 13 for a detailed description of RDQRSTRT); however it may be possible to be more specific if the issue relates to a particular group within the cohort, for example '2002-03 cohort start year (RDQRSTRT = 2002) full-time Research-Council-funded'.

Column 3

9. This column should contain the error type for the issue. The error type can be attributed to either:

- errors in HESA 2006-07 student data
- errors in HESA student data for years between 1999-2000 and 2005-06.

10. Paragraphs 18-30 of Annex K give a further description of these broad types of error and the action that should be taken in each case.

11. The only information that should be entered into Column 3 are the words '2006-07 HESA' (for errors in 2006-07 HESA student data) and 'HISTORIC' (for errors in HESA student data for years between 1999-2000 and 2005-06).

Column 4

12. This column should contain a brief description of the cause of the error. The description should be sufficient to allow us to understand why this discrepancy occurred. For example: 'The script in our student record system for generating MODE on the HESA return incorrectly assigned all students that became dormant during 2002-03 on the 2002-03 HESA return as code 64 "Dormant – previously part-time".'

Column 5

13. If the error type identified in Column 3 is '2006-07 HESA', either 'CHANGE', 'ADD' or 'DELETE' should be selected to highlight the type of amendments that are being made to the 2006-07 HESA record.

Column 6

14. Where Column 5 is returned as 'CHANGE' and Column 3 is '2006-07 HESA', a full list of the fields that the institution expects to correct must be identified and included. If Column 5 is 'ADD' or 'DELETE' then the words 'All fields' should be entered.

Column 7

15. If the error type identified in Column 3 is '2006-07 HESA', the number of records that are being amended should be included.

Column 8

16. If the error type identified in Column 3 is '2006-07 HESA', the date by when amended data will be submitted to HESA must be returned in Column 8. Guidance on how to submit amendments to 2006-07 HESA data is given in Annex J. Amended data must be signed off with HESA no later than **13 October 2008**.

Column 9

17. If the error type identified in Column 3 is 'HISTORIC' then a full list of the derived fields that require change, using historical data error files, must be identified by the institution and listed, for example 'RDQRMODE'. Annex K provides further guidance on how to identify the derived fields that require correction for certain errors in HESA student data for years between 1999-2000 and 2005-06.

Column 10

18. If the error type identified in Column 3 is 'HISTORIC' then the number of records to be overwritten should be included.

Column 11

19. If the error type identified in Column 3 is 'HISTORIC', the date by which historical data error files will be submitted to HEFCE must be returned here. Guidance on how to submit historical data error files is given in Annex K. Historical data error files should be signed off no later than **13 October 2008**.

Submitting RDQR action plans

20. Log onto the HEFCE extranet by following the guidance at Annex B. Then click 'HEFCE Resources' under the 'Folders' heading to be directed to the 'HEFCE extranet – HEFCE resources' page. Next click '2007-08 Statistics derived from HESA data' to be directed to the 'HEFCE extranet – 2007-08 Statistics derived from HESA data' page. If this link is not visible, it is possible that you do not have the appropriate access. To obtain this, you will need the appropriate group key, details of which are supplied in Annex A of the letter titled 'HEFCE web facility for 2007-08 statistics derived from HESA data' sent to your head of institution by Anthony Ryan in September 2008 (see paragraph 7 of Annex B for further details). Click the 'APRDQR07 data collection' link. Next click on 'Upload', browse to the location where the action plan is saved and click 'Upload'.

Annex M

2007-08 HEFCE-fundable student FTEs for TRAC(T)

1. This annex provides details of how we will use 2007-08 HESA student data to generate HEFCE-fundable student FTEs for TRAC(T). These FTEs are presented by HESA academic cost centres for each institution.

2. The web facility produces these derived statistics so that institutions can verify that the underlying data is correct prior to submitting the final data to HESA. Where institutions discover errors with the underlying data used to populate these tables, they should correct it before signing off their data with HESA.

3. These data should be used by all institutions in completing their annual TRAC(T) return. Details of the algorithms we have used to produce these statistics are provided in Appendix 16, while additional guidance that highlights the links between the HESA student data and the HEFCE-fundable student FTEs for each cost centre is given in Appendix 17.

Background

4. Since January 2006 we have been working on developing a national framework for costing teaching in English HEIs, based on Transparent Approach to Costing (TRAC) principles. The framework is known as TRAC(T). In 2006 we invited institutions to submit data on the full costs of teaching.

5. While participation was voluntary in the first two years, there was a good response rate from institutions. Data has now been submitted for academic years 2005-06 and 2006-07. In 2007 and 2008 we also held benchmarking workshops on TRAC(T) data. More information on TRAC-T is at www.hefce.ac.uk/trac-t.

6. From 2009 it will be mandatory for institutions to submit TRAC(T) data. Institutions will need to submit data on subject-related costs of HEFCE-fundable provision against 2007-08 HEFCE-fundable student FTEs for each cost centre. This is referred to as Subject-FACTS (Subject-related Full Average Costs of Teaching a Student). As with the previous two years we will pre-fill Section B of the TRAC(T) benchmarking return with HEFCE-fundable student FTEs for each cost centre from 2007-08 HESA student data

7. This annex and Appendices 16 and 17 describe the methods we have used to generate the 2007-08 HEFCE-fundable student FTEs for TRAC(T) by HESA academic cost centre. Guidance for completing the TRAC(T) benchmarking return is available at www.jcpsg.ac.uk/guidance/draft.

2007-08 HEFCE-fundable student FTEs for TRAC(T) outputs

8. Annex B describes how to access the Excel workbook (TRAC07XXXX.xls, where XXXX is the HESA identifier for the institution) that contains the 2007-08 HEFCE-fundable student FTEs for TRAC(T) tables. This workbook contains the following worksheets:

- cover page
- LEVEL1 – TRAC07 (this worksheet contains the 2007-08 HEFCE-fundable student FTEs for TRAC(T))
- LEVEL2 – FTS (more detailed summary of the full-time student data used to generate the 2007-08 HEFCE-fundable student FTEs for TRAC(T))
- LEVEL2 – SWOUT (more detailed summary of sandwich year-out student data. These data are not used to generate 2007-08 HEFCE-fundable student FTEs for TRAC(T), but may be used to help remove the costs of students on sandwich years-out from the departments that support them)
- LEVEL2 – PT (more detailed summary of the part-time student data used to generate the 2007-08 HEFCE-fundable student FTEs for TRAC(T)).

9. We will pre-fill the section B of the TRAC(T) benchmarking return with the statistics in worksheet 'LEVEL1 – TRAC07'. The remaining worksheets may assist institutions in rebuilding the figures in 'LEVEL1 – TRAC07' and highlight the close links to the HESES07 re-creation (see Annex C for further details).

10. The information in the 2007-08 HEFCE-fundable student FTEs for TRAC(T) tables can be rebuilt from an individualised file. The file (TRAC07XXXX.ind) contains details of how each student was classified in the 2007-08 HEFCE-fundable student FTEs for TRAC(T), and, where relevant, details of why they were excluded from the tables. A full description of the data in the individualised file is given in Appendix 16, while additional guidance on generating the 2007-08 HEFCE-fundable student FTEs for TRAC(T) output from the TRAC07 individualised file is given in Appendix 17.

Overview of the 2007-08 HEFCE-fundable student FTEs for TRAC(T) methodology

11. The level 2 tables (LEVEL2 – FTS, LEVEL2 – SWOUT, LEVEL2 – PT) are generated by the same algorithms used for the HESES07 re-creation, with modifications to allow for the fact that the student FTEs are summarised by cost centre rather than price group. For leads of HEFCE-recognised funding consortia the data for member colleges are not included.

12. Some of the cost centres have been split further by price group so that Subject-FACTS can reflect the different levels of cost attached to the different types of provision returned under them. The cost centres affected are: clinical medicine (cost centre 01); clinical dentistry (02); veterinary science (03); media studies (30); education (34); and sports science and leisure studies (38).

13. The level 1 table (LEVEL1 – TRAC07) is generated from the full-time and part-time level 2 tables restricted to HEFCE-funded and independently funded student FTEs, as shown in column 4 (assumed countable years) for full-time and column 4a (assumed countable years) for part-time. Sandwich year-out FTEs are not included in the LEVEL1 table. They are provided in the LEVEL2 – SWOUT table to assist institutions in removing the costs of this provision from each academic cost centre. Further guidance is in Chapter F.2 of part VIII of the TRAC for teaching guidance (see HEFCE Circular Letter 22/2006).

Redevelopment of the 2007-08 HESA student record

14. Following the redevelopment of the HESA student record for 2007-08 we have made substantial changes to the 2007-08 HEFCE-fundable student FTEs for TRAC(T) algorithms. We may need to further refine these algorithms after the release of the web facility to improve the accuracy of the outputs. However, we expect any changes to be minor.

15. This year we have released this web facility in two phases (see paragraphs 18-19 of the main text)'

16. Lydia Booker e-mailed HESES, RAS and TRAC contacts on 5 September 2008 to announce the availability of the 2007-08 HEFCE-fundable student FTEs for TRAC(T) outputs via the web facility.

17. We welcome comments on the suitability of the algorithms; these should be sent to Lydia Booker (e-mail hesa_heses_stats@hefce.ac.uk).

Annex N

Derived statistics likely to inform HESES08 audits

Purpose

1. We will use 2007-08 HESA student data to identify areas of further investigation during the audits of HESES08. As part of the HESES08 data audits carried out by the HEFCE assurance service the outcomes of two tests on 2007-08 HESA data will be used to identify areas of potential further investigation during the audit.

2. The two tests are as follows:

- Test 1 – potential non-completions: we will produce a list of Column 4 HEFCE-funded students returned as completions in the 2007-08 HESA return that, according to HESA data, left during 2007-08 without obtaining a qualification.

- Test 2 – potential non-completions: we will produce a list of Column 4 HEFCE-funded students returned as completions in the 2006-07 HESA return that either do not appear on the 2007-08 HESA student record or appear in the 2007-08 HESA return but have an FTE of zero.

3. During 2009 we may also develop tests to identify students that have potentially been returned with inflated FTE on the 2007-08 HESA return. As we develop these tests we will share them with the HEFCE assurance service for piloting during the HESES08 data audits with an aim to embed the use of such tests as a standard part of the data audit process in future years.

4. The lists generated from the two tests can be found in two individualised files: AUD107XXXX.ind and AUD207XXXX.ind respectively. The files contain details, in the form of HESA and derived fields, of how each student was classified for the purpose of the tests, and therefore why they are included in the list. Descriptions of the data in the individualised files are given in Appendix 18.

Redevelopment of the 2007-08 HESA student record

5. Following the redevelopment of the HESA student record for 2007-08 we have made substantial changes to the non-completion test algorithms. We may need to further refine these algorithms after the release of the web facility to improve the accuracy of the outputs. However, we expect any changes to be minor.

6. We welcome any comments on the suitability of the algorithms; these should be sent to Lydia Booker (e-mail hesa_heses_stats@hefce.ac.uk).

Annex O

Information for lead institutions of HEFCE-recognised funding consortia

Purpose

1. This annex provides leads of HEFCE-recognised funding consortia with information about the outputs available to them and member colleges via the web facilities for HESA and ILR data respectively. It also contains more detailed information about the worksheets within the re-creation outputs that are available for member colleges.

Data collection arrangements for HEFCE-recognised funding consortia

2. Under HEFCE-recognised funding consortia arrangements, each student is recognised as a student of the appropriate consortium member, rather than of the lead institution. Therefore lead institutions of HEFCE-recognised funding consortia must submit data to HESA only for students that are registered at their institution. Likewise, students that are registered at a member further education college will be returned on that college's 2007-08 ILR.

3. All member colleges' data for students funded under the consortium arrangement should have been returned by the lead institution on its HESES07 return. Therefore, for the lead institution of a HEFCE-recognised funding consortium, the HESES07 re-creation will incorporate the 2007-08 HESA data that it submits to the web facility together with the ILR data that each member college has submitted to our equivalent web facility for further education colleges: 'HEFCE web facility for 2007-08 statistics derived from ILR data'. A similar arrangement is in place for the indicative 2009-10 WP allocations outputs.

4. Annex C provides further details about the HESES07 re-creation. Annex E provides further details about the indicative 2009-10 WP allocations outputs.

5. Further information about the HEFCE web facility for further education colleges is in '2007-08 statistics derived from ILR data: Guide to HEFCE web facility' (HEFCE 2008/25). We encourage leads of HEFCE-recognised funding consortia to liaise with their member colleges and co-ordinate the way their members are returning data to the web facility in order to maximise the usefulness of the combined re-creations.

Web facility outputs

HESES07 and the HESES07 re-creation

6. The web facility generates a provisional HESES07 re-creation. This is described in Annex C. Annex B contains a description of how to access the HESES07 re-creation workbook from the HEFCE extranet. The information in the HESES07 re-creation will vary depending on whether the institution is the lead of a HEFCE-recognised funding consortium.

7. Lead institutions of HEFCE-recognised funding consortia will be able to access the following files:

- LEAD07XXXX.xls – This is an Excel workbook containing the HESES07 re-creation tables for the lead institution, which is constructed from the 2007-08 HESA data that the lead institution has uploaded to the web facility.

- HESR07XXXX.xls – This is an Excel workbook containing the HESES07 re-creation tables that are constructed from the 2007-08 HESA data that is uploaded by the lead institution as well as ILR data that is uploaded by member colleges to the HEFCE web facility for 2007-08 statistics derived from ILR data.

- HEIFERC07YYYYYY.xls – These are Excel workbooks containing the HEIFES07 re-creation tables for each of the member colleges (where YYYYYY denotes the UPIN for the college). These workbooks contain the 2007-08 ILR data that each member college has submitted to the ILR web facility. Note that these tables will only become available after member colleges have submitted data to that web facility.

- HESR07XXXX.ind – This is the individualised file that supplements the provisional HESES07 re-creation tables for the lead institution (LEAD07XXXX.xls). All of the information in the HESES07 re-creation tables for the lead institution can be rebuilt by categorising and aggregating the data in this file. It contains details, in the form of 2007-08 HESA fields and derived fields, of how each student was classified in the re-creation. A description of the data in the individualised file is given in Appendix 1.

8. Member colleges of HEFCE-recognised funding consortia will be able to access the following files from the HEFCE web facility for 2007-08 statistics derived from ILR data:

- HEIFERC07YYYYYY.xls – These are Excel workbooks containing the HEIFES07 re-creation tables for the member college.

- HEIFER07YYYYYY.ind – This is the individualised file to supplement the provisional HEIFES07 re-creation tables for the member college (HEIFERC07YYYYYY.xls). All of the information in the HEIFES07 re-creation tables for the member college can be rebuilt by categorising and aggregating the data in this file. It contains details, in the form of 2007-08 ILR and derived fields, of how each student was classified in the re-creation.

<u>Derived statistics likely to inform the 2009-10 WP allocations</u>

9. The web facility generates indicative 2009-10 WP allocations. These are described in Annex E. Annex B provides instructions on how to access the WP07 re-creation Excel workbook from the HEFCE extranet. The information in this re-creation will vary depending on whether the institution is the lead of a HEFCE-recognised funding consortium.

10. Lead institutions of HEFCE-recognised funding consortia will be able to access the following files:

- WPLEAD07XXXX.xls – This is an Excel workbook containing the 2009-10 WP allocation tables for the lead institution, derived from the 2007-08 HESA data that the institution has uploaded.

- WP07XXXX.xls – This is an Excel workbook containing the indicative 2009-10 WP allocation tables built from the HESA data uploaded by the lead institution. It also contains the 2007-08 ILR data that has been uploaded by member colleges to the HEFCE web facility for 2007-08 statistics derived from ILR data.

- IHWPC07YYYYYY.xls – These are Excel workbooks containing the indicative 2009-10 WP tables for each of the member colleges built from the 2007-08 ILR data they have uploaded to the HEFCE web facility for derived statistics derived from ILR data. Note that these tables will only become available after member colleges have submitted data to the web facility.

- WP07XXXX.ind – This is the individualised file that supplements the indicative 2009-10 WP allocations (WPLEAD07YYYY.xls). All of the information in the WP07 re-creation tables for the lead college can be rebuilt by categorising and aggregating the data in this file. It contains details, in the form of 2007-08 HESA fields and derived fields, of how each student was classified in the re-creation. A full description of the data in the individualised file is given in Appendix 7.

11. Member colleges of a HEFCE-recognised funding consortium will be able to access the following files from the HEFCE web facility for 2007-08 statistics derived from ILR data:

- IHWPC07YYYYYY.xls – This is an Excel workbook containing the indicative 2009-10 WP allocation tables for the member college.

- IHWP07YYYYYY.ind – This is the individualised file that is intended to supplement the indicative 2009-10 WP allocation table (IHWPC07YYYYYY.xls). All of the information in this allocation table for the member college can be rebuilt by categorising and aggregating the data in this file. It contains details, in the form of 2007-08 ILR and derived fields, of how each student was classified in the re-creation.

Access to outputs

12. Lead institutions will be given automatic access to the HEIFERC07YYYYYY.xls and IHWPC07YYYYYY.xls outputs for each of the member colleges. Each member college will receive access to its own HEIFES07 re-creation and WP07 output, separate HEIFES07 re-creation and WP07 individualised files containing only the 2007-08 ILR data that the college has uploaded.

13. Where a member college gives us permission to grant the lead institution access to its HEIFES07 re-creation or WP07 re-creation individualised files, we will write to the lead institution to describe how it can access the individualised files via the HEFCE extranet.

Other outputs available to the lead institution

14. In addition to the HESES07 re-creation and member college HEIFES07 re-creation and WP07 outputs listed above, each lead institution will have access to the other outputs relating to its own data listed in paragraph 28 of Annex B.

Other outputs available to the member college

15. In addition to the HEIFES07 re-creation and WP07 outputs listed above, each member college will have access to other outputs relating to its own data listed in the publication 'HEFCE web facility for 2007-08 statistics derived from ILR data' (HEFCE 2008/25).

Worksheets

16. The provisional HESES07 re-creation workbook for lead institutions contains the following worksheets:

Table J **Provisional LEAD07XXXX.xls description**

Page number	Worksheet (see tabs on spreadsheet)	Title
1	Coversheet	Title page
2	FTS	Provisional HESES07 re-creation Table 1a: Full-time counts of years of instance
3	MED	Provisional HESES07 re-creation Table 1b: Medical and dental counts of years of instance
4	SWOUT	Provisional HESES07 re-creation Table 2: Sandwich year-out counts of years of instance
5	PT	Provisional HESES07 re-creation Table 3: Part-time counts of years of instance and FTE
6	FEE	Provisional HESES07 re-creation Table 4: Home and EC fees
7	CONS	Provisional HESES07 re-creation Table 6: HEFCE-recognised funding consortia 2007-08
8	TAB7	Provisional HESES07 re-creation Table 7 (optional): Counts of years of instance and FTE using HESES06 treatment of non-standard years of instance
9	Excl	Students excluded from the provisional HESES07 re-creation

17. The provisional HEIFES07 re-creation workbook for member colleges, HEIFERC07YYYYYY.xls, contains the following worksheets:

Table K **Provisional HEIFERC07YYYYYY.xls description**

Page number	**Worksheet (see tabs on spreadsheet)**	**Description**
1	Coversheet	Title page
2	FTS	Provisional HEIFES07 re-creation Table 1: Full-time counts of years of programme of study for the member college
3	SWOUT	Provisional HEIFES07 re-creation Table 2: Sandwich year-out counts of years of programme of study for the member college
4	PT	Provisional HEIFES07 re-creation Table 3: Part-time counts of years of programme of study and FTE for the member college
5	FEE	Provisional HEIFES07 re-creation Table 4: Home and EC fees for the member college
6	Excl	Provisional HEIFES07 re-creation exclusion table for the member college
7	Credibility	Identifies areas for recognised HE qualification aims on 2007-08 ILR where data is potentially inaccurate

Annex P

Lifelong Learning Network (LLN) student summaries

Purpose

1. This annex describes how we intend to identify students that are part of an LLN using their 2007-08 HESA data. Institutions participating in an LLN in 2007-08 should use this opportunity to check that we are correctly identifying their LLN students from their HESA data. This annex only applies to institutions that have registered students during 2007-08 at their institution as part of an LLN initiative.

2. We produce an indicative 'calculation of notional grant associated with LLN students' report based on the number of LLN students that we identify from the HESA data. As part of the teaching funding methodology we monitor LLN ASNs following model 1 as part of an institution's total ASN targets. These are shown in the HBK worksheet of HESR07XXXX.xls, as described in Annex B. Similarly we monitor LLN ASNs following model 2, using the LLN worksheet of HESR07XXXX.xls. Therefore the indicative report is provided solely to highlight potential 2007-08 HESA data errors and should not be considered to be any kind of funding commitment by HEFCE.

Background

3. In John Selby's letter of 2 May 2006, we asked institutions to provide details of how we can identify LLN students on their HESA returns. This letter also explained that, wherever possible, we will use HESA data to evaluate the operation of the LLN initiative. If 'LLN students' are identified on the HESA record we can follow them through their HE study and measure their achievement within HE, their evaluation of their HE experience (through the National Student Survey) and, where appropriate, their employment outcomes.

Web facility outputs

4. Annex B describes how to access the derived statistics, in an Excel workbook, that are used to generate the LLN student summaries (LLN07XXXX.xls).

5. To investigate the underlying data, institutions should use the individualised file (LLN07XXXX.ind) that contains details of how each student was classified and, where relevant, details of why they were excluded from the LLN summaries. A full description of the data in the individualised file is given in Appendix 19.

Redevelopment of the 2007-08 HESA student record

6. Following the redevelopment of the HESA student record for 2007-08 we have made substantial changes to the LLN student summary algorithms. We may need to further refine these algorithms after the release of the web facility to improve the accuracy of the outputs. However, we expect any changes to be minor.

7. We welcome comments on the suitability of the algorithms; these should be sent to Lydia Booker (e-mail hesa_heses_stats@hefce.ac.uk).

Annex Q

Using the HESES08 non-completion toolkit

Purpose

1. This annex contains details on how to use the HESES08 non-completion toolkit as part of an institution's estimation of non-completion rates on HESES08. This annex should be read in conjunction with and used in accordance with the forthcoming HESES08 publication in which full guidance relating to the return of non-completion forecasts will be given.

Description of the non-completion toolkit

2. The non-completion toolkit is a spreadsheet that allows institutions to generate estimated non-completion rates for all students by course and year for 2007-08, from submitted HESA 2007-08 student data, using fields that they consider influential at the forecast level. The estimated values generated may be used by institutions to inform the forecasts of non-completions returned in HESES08.

Reliability of estimated non-completion rates

3. Institutions should note that the toolkit is provided purely as an aid in generating non-completion rate estimates in HESES08. We expect institutions to use their own judgement in assessing whether the rates generated are applicable.

4. Estimates may be unreliable if the numbers of students involved are small, or if the underlying HESA data contain errors or are not completed in line with the HEFCE guidance on non-completion status. Annex E of 'HESES07' (HEFCE 2007/26) gives definitive guidance on non-completions.

5. The toolkit will not treat non-exempt ELQ students that are returned as HEFCE-fundable on the 2007-08 HESA return as non-fundable. Therefore colleges should make an adjustment for such students if applying the results of the toolkit to their HESES08 return.

Using the non-completion toolkit

6. Users should follow the instructions in Annex B to access the file NCTK07XXXX.xls. The file should be opened in Excel 2000 or a later version. (Due to the technology used the toolkit will not function correctly in earlier versions of Excel.) The creation of tables uses macros, so these should be enabled to use the facility. The worksheet called 'Ncdata' contains the raw data used to generate the tables, and the 'Pivot Table' worksheet displays the results. When you open the file you will be prompted to select the fields you want to use to generate the non-completion rates – we advise summarising by course in the first instance. To create the table click 'OK', and the screen will close, enabling you to view the table. Click on 'Create a new table' to repeat the process – this will delete the previous table from the worksheet.

7. Non-completion rates based on small numbers of students may be unreliable, so any rate based on a population of less than 100 is coloured red and noted as unreliable. In these cases we advise using less detailed data in order to obtain a more reliable estimate.

8. In order to save a table you have created, open the <File> menu and select <Save As>.

Redevelopment of the 2007-08 HESA student record

9. Following the redevelopment of the HESA student record for 2007-08 we have made changes to the algorithms that populate the non-completion toolkit. We may need to further refine these algorithms after the release of the web facility to improve the accuracy of the outputs. However, we expect any changes to be minor.

10. We welcome comments on the suitability of the algorithms; these should be sent to Lydia Booker (e-mail hesa_heses_stats@hefce.ac.uk).

Annex R

Submitting overrides to primary derived fields

Background

1. We will only apply an override where the data submitted on the HESA return are correct but there is a problem of fit with the HEFCE algorithms. In these instances it is the derived fields that generate the re-creations which require amendment rather than the underlying HESA data.

2. Problems of fit occur where the derived field that we generate is based on an assumption which may not necessarily fit with the institution's actual position. All known problems of fit with the HESES07 re-creation algorithms are described in Appendix 3. All known problems of fit with the HESES07 re-creation based on cost centre sector norms are described in Appendix 6. And all known problems of fit with the RAS07 re-creation algorithms are described in Appendix 12.

3. We will only apply overrides where we agree that they are appropriate, and (in the case of an override to the sector norm cost centre mapping) where we have made a decision based on evidence provided. Therefore we may seek further information where necessary. For example, in the case of the sector norm cost centre mapping we may seek further module information. It may not always be possible to determine whether an override is appropriate until we have examined the students affected. Therefore we may not accept an override once submitted, or we may seek further clarification.

Purpose

4. This annex details the data structure and format for overrides to derived fields. Institutions must only supply override files using the file structure and format detailed in this annex.

5. Override files should contain the data structure and format described in paragraphs 10-20. These specifications are necessary to ensure we can process overrides to derived fields in a timely and accurate manner. Any override files submitted by institutions must not differ, either in structure or format, from the specifications detailed in this annex..

6. This year we will only accept overrides to primary derived fields. Primary derived fields are those which are directly used to populate the re-creation tables. For example, the value of HESEXCL determines whether a record is included in the HESES re-creation. It is therefore a primary derived field. In contrast, the fields EXCL1-EXCL64 are not directly used to populate the re-creation tables; we refer to such fields as 'secondary' derived fields. For example, HESEXCL is determined by the values in the secondary derived fields EXCL1-EXCL64. A list of primary derived fields is given in Table L.

7. Since each override will only alter primary derived fields, this may give rise to inconsistencies with secondary derived fields. For example, if the value of HESEXCL is altered by an override file, its value will be inconsistent with the values of EXCL1-EXCL64.

8. To allow institutions to check that an override file has had the desired effect, a new field, 'OVERRIDE', is included in the individualised file. This takes the value 1 if an override has been applied to the record; otherwise its value is 0. This field also allows institutions to determine which records may have inconsistent secondary derived fields as a result of an override.

9. We also require that certain primary derived fields are returned as a group, in order to maintain data integrity. If any field in the group requires an override, all members of that group should be included in the override file, even if the values of some fields in the group remain unchanged. Details of the primary derived fields which must be returned as a group in override files are given in Table L. For example, if an override to amend price group information is being submitted, we require that the file contains FTEB-D, FTEMEDIA, FTEITT and FTEINSET, even if the value in some fields remain unchanged.

Table L **Primary derived fields**

Primary derived field name	**Description**	**Re-creation tables**
HESCOMP	HESES completion of year of instance indicator	HESES re-creation
HESEXCL	Reason for exclusion from the HESES population	HESES re-creation
HESFEELV	Fee level	HESES re-creation
HESLEVEL	Level of study	HESES re-creation
HESMED	Table 1b inclusion flag	HESES re-creation
HESMODE	Mode of study	HESES re-creation
HESREG	Column 1or 2 indicator	HESES re-creation
HESTYPE	Fundability status	HESES re-creation
LENGTH	Flag indicating long or standard length year of instance	HESES re-creation
FTEA-D, FTEITT, FTEINSET, FTEMEDIA, FTEPSYCH	FTE in each price group (these fields must be returned as a group)	HESES re-creation
SNPRGA-D, SNINSET, SNITT	Proportion of countable year in each sector norm price group (these fields must be returned as a group)	HESES re-creation based on cost centre sector norms
MSUB	Submission identifier for UOAs where multiple submissions were made to the 2001 RAE	RAS re-creation
RASFTE	FTE consistent with RAS definitions	RAS re-creation
RASMODE	Mode of study for research degree	RAS re-creation
RASTYPE	Fundability status	RAS re-creation

RASUOA1-3	Units of assessment (these fields must be returned as a group)	RAS re-creation
RASYEAR	Year of programme of study as returned to RAS06	RAS re-creation
RSTUEXCL	Reason for exclusion from RAS student population	RAS re-creation
UOAP1-3	Proportion of time spent in each subject area (these fields must be returned as a group)	RAS re-creation

Format and naming

10. Overrides to primary derived fields must be uploaded as a comma-separated file to the 'HESAOVR07 data collection' facility. You can find this facility by logging onto the HEFCE extranet, selecting 'HEFCE Resources' and selecting '2007-08 statistics derived from HESA data'. Override files must be given a file name in the form ovrXXXXn.amd, where:

- XXXX is the HESA institution identifier for the institution
- n is a sequential number starting at 1.

For example, the first override file submitted would be called ovrXXXX1.amd, and the second would be called ovrXXXX2.amd.

Structure

11. Each record must contain complete data for all fields included in the override file, even if a particular primary derived field remains unchanged in some cases.

12. Only primary derived fields listed in Table L should be included in the change line (see line 6 below) for override files.

13. Override files must contain a header in the following form:

line 1 – override reference in the form ovrXXXXn where XXXX is the HESA institution identifier and n is a sequential number starting at 1; this will be the file name with the '.amd' file extension removed

line 2 – creation date of the override file in the form ddmmyyyy. For example 01082008 for a file created on 1 August 2008

line 3 – a brief description of the override. For example: 'Overrides to primary derived field HESFEELV'

line 4 – this line must contain the words OVERRIDE and either the word TEMPORARY or PERMANENT. If the override is temporary then the last academic year that it applies to should be entered. For example: 'OVERRIDE, TEMPORARY, 2007' indicates that the override will be applied in academic year 2007-08 but not in 2008-09 onwards

line 5 – the fields used to identify records on the override file, comma-separated. For example OWNPSD could be used to identify records on a course level; HUSID, NUMHUS, and RECID can be used to identify records on a student level

line 6 – the primary derived fields being changed, comma-separated. For example: HESFEELV, HESEXCL

line 7 – number of records in the file, excluding the first 12 lines of header information

line 8 – the field used to calculate the file check-sum (see paragraph 20 for an explanation of the file check-sum)

line 9 – file check-sum

lines 10 and 11 – any notes the institution wishes to include

line 12 – fields included in the override file. These fields must appear in the same order as each row of the data and must be comma-separated. For example:
OWNPSD, HESFEELV on one line

line 13 – override data must begin on this line

end of file – there must be a single blank line following the final record in the override file.

Special cases

FTE in each price group

14. If overrides are being applied to the FTE in a particular price group then we require information to be provided about all of the price group fields FTEA, FTEB, FTEC, FTED, FTEMEDIA, FTEITT, FTEINSET (even if a particular price group FTE is not being changed).

Proportion of FTE in each sector norm price group

15. If permission has been granted by HEFCE, and overrides are being applied to the sector norm price groups, then the override file should contain information about all of the sector norm price group fields SNPRGA, SNPRGB, SNPRGC, SNPRGD, SNMEDIA, SNITT, SNINSET (even if a particular sector norm price group is not being changed). We recommend that a field called SBJ that contains the JACS2 code of the subject of the area of study is used as the linking field. If required, we will construct the sector norm price group override file on behalf of the institution. This will be implemented when the institution has checked and confirmed that the override file that we construct is correct.

UOAs

16. If permission has been granted by HEFCE, and overrides are being applied to RASUOA1-3, then we would recommend that the subject of qualification aim fields SBJQA1-3 are used as the linking fields. If required, we will construct the UOA override file on behalf of

the institution. This will be implemented when the institution has checked and confirmed that the override file we construct is correct.

Proportion of time spent in each subject area, used to scale FTE

17. If overrides are being applied to UOAP1, UOAP2, UOAP3 then we require information about all of these fields. We recommend that the subject of qualification aim fields SBJQA1-3 are used as the linking fields.

Identifying records

18. To enable us to link override files to our derived HESA dataset, we must be able to identify the records on the HESA return where the override should be applied. The field, or combination of fields, enabling us to achieve this must be listed, comma-separated, on line 5 of the override file.

Saving files

19. Saving override files in Microsoft Excel usually results in the loss of leading zeros and the corruption of very large values into exponential form (for example, 9.91E+12). We recommend that override files are viewed and saved using a text editor, for example Notepad.

Check-sum

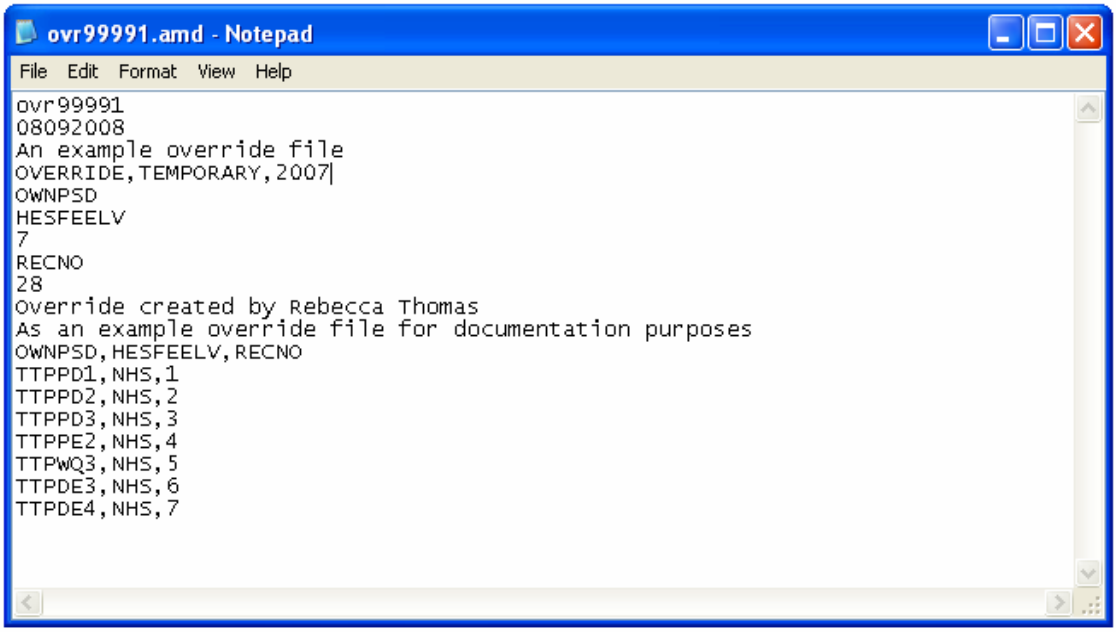
20. To ensure override files have not been corrupted during transit, we will check that the sum of values in the field specified on line 8 of the override file matches the value returned on line 9. If an override file does not contain any numeric fields suitable for calculating a check-sum, an additional field must be included solely for this purpose, for example QUALAIM. Numeric fields that contain values greater than 20,000 (approximately) are unsuitable for calculating the check-sum. If information is not being changed at the student level, then a sequential field called RECNO may be created for the purpose of the check-sum. For example RECNO may contain 1, 2, 3, 4, 5 and so on.

Outcome

21. When we receive a valid override file in the structure and format detailed in this annex, we aim to provide feedback within five working days. Institutions will be notified by e-mail when the revised re-creation tables and individualised file will be available for generation via the web facility.

Example of an override file

Figure 6 **A typical override file**



List of abbreviations

ASNs	Additional student numbers
CFTE	Contract full-time equivalent
DSA	Disabled Student's Allowance
EC	European Community
ELQ	Equivalent or lower qualification
FAQ	Frequently asked questions
FE	Further education
FEC	Further education college
FES	(Scottish) Further Education Statistics survey
FTE	Full-time equivalent
FTS	Full-time
HDE	Historical data error
HE	Higher education
HEFCE	Higher Education Funding Council for England
HESA	Higher Education Statistics Agency
HESES	Higher Education Students Early Statistics
HIN	HUSID X INSTID or UKPRN X NUMHUS
INSET	In-service education for teachers
ILR	Individualised Learner Record
ITT	Initial teacher training
JACS2	Joint Academic Coding System (version 2)
LLN	Lifelong Learning Network
LSC	Learning and Skills Council
PGCE	Postgraduate Certificate of Education
POLAR	Participation of local areas
QAA	Quality Assurance Agency for higher education
QR	Quality-regulated
QTS	Qualified teacher status
RAS	Research Activity Survey
RAE	Research Assessment Exercise
RDP	Research degree programme

RDQR	Research degree qualification rate
SIVS	Strategically important and vulnerable subjects
UOA	Unit of Assessment
UPIN	Unique provider identification number
WP	Widening participation