



QUALITY ASSURANCE SELF-ASSESSMENT OF DLHE AND "NEWDLHE" DATA

DAN COOK AND REBECCA HASLAM
FEBRUARY 2017

HESA



HESA
95 Promenade
Cheltenham
GL50 1HZ
E newdlhe@hesa.ac.uk
T +44 (0) 1242 211 103
W www.hesa.ac.uk

Higher Education Statistics Agency Limited is a company limited by guarantee, registered in England at 95 Promenade, Cheltenham, GL50 1HZ. Registered No. 02766993. Registered Charity No. 1039709. Certified to ISO 27001. The members are Universities UK and GuildHE.

CONTENTS

| | |
|---|----|
| Executive summary | 3 |
| Background | 4 |
| How the DLHE/NewDLHE fits into the UK’s statistical practice and law..... | 4 |
| Our approach to this self-assessment | 5 |
| Determination of the required level of assurance | 6 |
| Quality assurance self-evaluation | 9 |
| Operational context & admin data collection..... | 9 |
| Communication with data supply partners | 12 |
| QA principles, standards and checks by data suppliers..... | 13 |
| Producers’ QA investigations and documentation..... | 13 |
| Summary of quality characteristics | 14 |
| Evidence of embedded quality assurance and review activity | 16 |
| The foreseen impact of NewDLHE on the DLHE SFR and consideration of requirements for National Statistics designation | 17 |
| Principle 2: Impartiality and objectivity | 17 |
| Principle 4: Sound methods and assured quality | 17 |
| Principle 6: Proportionate burden | 18 |
| Protocol 1: User engagement..... | 19 |
| Protocol 3: The use of administrative sources for statistical purposes | 19 |
| Conclusions..... | 20 |
| Recommendations | 20 |

EXECUTIVE SUMMARY

Data about the destinations and outcomes of graduates is used for a number of high-profile purposes. It is therefore vital that there is a high level of confidence in the robustness of the data. In light of HESA's work to develop a new model for the collection of graduate outcomes data, HESA has conducted a self-assessment of the quality of this data using the UK Statistics Agency's toolkit.

HESA seeks to maintain National Statistics designation for its current DLHE Statistical First Release (SFR) and for a future NewDLHE SFR. To maintain this designation, as well as wider public trust in the data, HESA should be aiming for the highest assurance level possible. In terms of the UK Statistics Agency's quality assurance matrix, the aim should be to achieve "A3 – Comprehensive assurance". The current Destinations of Leavers from Higher Education (DLHE) survey does not meet this standard. It is mostly at "A2 – Enhanced assurance" level with some aspects at "A1 – Basic assurance" and some at "A3 – Comprehensive assurance".

This self-assessment makes a number of recommendations to achieve comprehensive assurance. The key recommendation is the need to reconfigure the current DLHE methodology to enhance quality assurance mechanisms. This could either be achieved through a centralised model – which would realise additional efficiencies through economies of scale and greater responsiveness – or through a more substantial audit regime – which would include enhanced analysis and quality assurance functions at HESA and the resurveying of samples of graduates.

BACKGROUND

Higher education providers have historically collected information about the activities of graduates following completion of their studies. In 1994 this activity became centrally co-ordinated by HESA, since which time it has become subject to increasing rigour and quality assurance. During the same period it has also grown in usage and importance to become a part of the UK's critical data infrastructure, and the source of a wide range of public information. HESA periodically reviews each of its data collections, and in summer 2015 announced a wide-ranging review of graduate outcomes and destinations data. For more information about the review see:

<https://www.hesa.ac.uk/innovation/records/reviews/newdlhe>

The HESA review of graduate outcomes and destinations data (referred to by its lively Twitter hashtag “#NewDLHE” for short) envisages the replacement of the Destinations of Leavers from Higher Education (DLHE) and Longitudinal DLHE (LDLHE) surveys with a new data product (referred to here by the informal title “NewDLHE”). As with the current DLHE data, the proposed uses for the NewDLHE data are high profile and require a similarly high-level assurance of the robustness of the data.

As a matter of good practice, and for legal compliance, HESA has undertaken a quality assurance exercise covering the proposed/emerging methodology. This is to guide design decisions for the survey to be put forward to consultation, and to assist respondents in coming to a balanced judgement about the merits of the chosen approach.

HOW THE DLHE/NEWDLHE FITS INTO THE UK'S STATISTICAL PRACTICE AND LAW

The Statistics and Registration Service Act 2007 defined the concept of Official Statistics. This encompassed the previous category of National Statistics which continues to exist as a sub-category of Official Statistics. The Official Statistics Order 2008 first designated the Higher Education Statistics Agency (HESA) as a producer of Official Statistics.

The Code of Practice for Official Statistics (Protocol 2: Release practices) states that statistical reports should be released into the public domain in an orderly manner that promotes public confidence and gives equal access to all, subject to relevant legislation.

HESA produces the following Official Statistics products utilising data on graduate outcomes, drawn principally from the DLHE and Longitudinal DLHE surveys, on a regular (annual or biennial) cycle:

- Destinations of Leavers from Higher Education Longitudinal Survey - Key findings report
- Destinations of Leavers from Higher Education
- Destinations of Leavers from Higher Education in the UK (Statistical First Release)

The Destinations of Leavers from Higher Education in the UK (Statistical First Release) product is additionally designated as National Statistics, signifying that compliance with the Code of Practice for Official Statistics has been formally assessed.

Designation as National Statistics can be broadly interpreted to mean that the statistics:

- Meet identified user needs
- Are well explained and readily accessible

- Are produced according to sound methods
- Are managed impartially and objectively in the public interest.

Once statistics have been designated as National Statistics it is a statutory requirement that the Code of Practice shall continue to be observed.

The Code of Practice for Official Statistics can be found here:

<https://www.statisticsauthority.gov.uk/monitoring-and-assessment/code-of-practice/>

Data collected using the DLHE survey is currently classified as “Administrative Data” for the purpose of evaluation (for more information, see: <https://www.statisticsauthority.gov.uk/monitoring-and-assessment/monitoring/administrative-data-and-official-statistics/quality-assurance-of-administrative-data/>).

The UK Statistics Authority (HESA’s principal regulator) produces a toolkit to assist in its evaluations of administrative data sources. We used the framework offered by this toolkit as the basis for this self-evaluation. We follow the logic offered in the toolkit and encourage this document to be read in conjunction with it. The toolkit is available at:

<https://www.statisticsauthority.gov.uk/monitoring-and-assessment/monitoring/administrative-data-and-official-statistics/>

We also recognise the increasingly high standards being pursued in the Official Statistics sector, and the forthcoming enhancements to the Code of Practice under the Office for Statistics Regulation: <https://statisticsauthority.gov.uk/monitoring-and-assessment/what-we-do/how-we-are-changing/>

OUR APPROACH TO THIS SELF-ASSESSMENT

Following the UK Statistics Authority’s toolkit (see last section) we first determine the appropriate level of quality assurance required for statistics based on the NewDLHE. We provide an explanation of our judgement about this required level of assurance, along with evidence to support the rationale for these decisions.

We also provide evidence of the actions, and the rationale for deciding upon those actions, that will achieve compliance with the chosen level of assurance.

We also provide evidence which demonstrates that HESA has embedded practices for keeping its quality assurance arrangements under review. This document itself comprises part of that practice.

Lastly, in line with the published guidance, we explore the impact of the NewDLHE on the DLHE SFR and consider the impact of changes on National Statistics designation. We also make recommendations for improvement.

DETERMINATION OF THE REQUIRED LEVEL OF ASSURANCE

In this section, we consider the Risk/Profile matrix that is applicable.

In terms of establishing the public interest profile of the data, we reason that this data has the following characteristics:

- 1) Economically important: this data is used in products that 'rank' universities and seek to inform consumer choice.
- 2) Higher education providers invest significant resources in activities designed to improve the outcomes measured by the DLHE.
- 3) There is a complex competitive market for HE within which students make choices, and approaching a half of all young people attend HE. DLHE data is used in many of the state-sector and commercial products that support these decisions, as well as advice and guidance services connected with them.
- 4) There are frequent questions from policymakers regarding graduate destinations in each of the UK's democratic legislative bodies, and HESA (DLHE) data is the most commonly-used source for answering these questions. This data is therefore seen as politically sensitive.
- 5) A hearing on the TEF was conducted by the BIS Select Committee. Metrics to be used in the TEF include measures derived from the DLHE, and HESA data was both mentioned in sessions, and HESA supplied written evidence to the Select Committee.
(<http://www.parliament.uk/business/committees/committees-a-z/commons-select/business-innovation-and-skills/inquiries/parliament-2015/assessing-quality-in-higher-education/>)
- 6) Rankings data (in part based on DLHE) is used by some overseas governments to direct student funding, and these decisions affect the UK market structure and possibly education exports, in what is one of the UK's high-performing export industries.
- 7) The annual release of statistics from the DLHE generates substantial coverage across the popular and specialist press and in television and online media outlets.
- 8) Collection of DLHE data is required by each of the HE Funding Councils to discharge their responsibilities, and the requirement to collect the data is backed by legislation and forms a condition of accountability and assurance for government-backed funding in all parts of the UK.

This analysis is borne out by the following facts:

- 1) Three current publications based on DLHE and LDLHE data are designated as Official Statistics, and one of these (the Statistical First Release) is additionally designated as National Statistics. This level of designation implies a high level of public interest and trust in the data. There are no suitable alternative sources of data on graduate employment to the DLHE, and so this data has a high impact in the public domain.
- 2) DLHE data is utilised in the Unistats information product, backed by governments in each UK administration (see: <http://unistats.direct.gov.uk>).
- 3) The DLHE has been assessed by the ONS as a potential data source for the TEF (see https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/523291/bis-16-269-teaching-excellence-framework-review-of-data-sources-interim-report.pdf).
- 4) DLHE data is used in all major university league tables, published online and in the popular press, including:

- a. The Complete University Guide (<http://www.thecompleteuniversityguide.co.uk/>).
 - b. The Guardian (<https://www.theguardian.com/education/ng-interactive/2016/may/23/university-league-tables-2017>).
 - c. Times Higher Education (<https://www.timeshighereducation.com/student/best-universities/best-universities-uk>).
 - d. Times and Sunday Times (behind paywall).
- 5) News stories relating to “graduate jobs” frequently use and cite DLHE data in the national and international press, for example:
- a. <https://www.timeshighereducation.com/news/graduate-employment-hits-record-high-as-gender-pay-gap-grows>
 - b. <http://www.thehindu.com/features/education/at-the-heart-of-engineering/article9365378.ece>
 - c. <https://www.theguardian.com/business/2015/aug/19/uk-failed-create-enough-high-skilled-jobs-graduates-student-debt-report>
 - d. <http://www.telegraph.co.uk/education/universityeducation/11605297/Top-10-degree-subjects-for-graduate-prospects.html>
 - e. <http://www.universityworldnews.com/article.php?story=20150116090700541>
 - f. <http://www.walesonline.co.uk/news/wales-news/welsh-university-graduates-are-back-4728296>
- 6) Higher education providers in general publish provider-level outputs based on DLHE data, and this is popular to the extent that HESA regularly issues guidance embargoing publication, and delineating carefully the relationship of these publications to the designated outputs produced by HESA. Although not themselves classified as National Statistics or Official Statistics, the existence of HESA’s designated publications based on DLHE appear to have a “halo effect” that extends to publications that cite HESA’s DLHE data as a source. These university-specific press releases are routinely reported in the local press across the country, for example:
- a. <http://www.hulldailymail.co.uk/university-of-hull-in-top-10-for-graduate-employability/story-29496420-detail/story.html>
 - b. <http://www.leicestermercury.co.uk/graduates-leicester-s-universities/story-26877831-detail/story.html>
 - c. <http://thelincolnite.co.uk/2015/09/day-1-in-pictures-students-descend-on-lincoln-for-graduation-ceremonies/>
 - d. <http://www.derbytelegraph.co.uk/a-level-results-15-reasons-to-study-at-the-university-of-derby/story-29629687-detail/story.html>
- 7) In our recent consultation on replacing the DLHE, we gathered 208 responses – a large number for a consultation of this type, drawn from a wide range of interest groups (see https://www.hesa.ac.uk/files/NewDLHE_consultation-synthesis.pdf for details).

We expect that the NewDLHE will have a public interest profile that is as high, if not higher than the DLHE.

In addition, in the recent public consultation, we observed a very high level of agreement that a census survey was needed, and that its profile should be raised further, especially given the development of the Longitudinal Educational Outcomes (LEO) dataset and its application in higher education. In the future, it appears that users perceive this dataset becoming more important.

We conclude that there is a high level of public interest in this data, commensurate with retention of National Statistics designation. Therefore, the highest standards of quality should be assured in its production. In terms of the UK Statistics Authority's quality assurance matrix this means that we should be aiming for the assurance level "A3 – Comprehensive assurance" in order to be fully compliant. This would align with other data with a high level of public interest.

QUALITY ASSURANCE SELF-EVALUATION

Our next step was to consider the level of risk of data quality concerns. Following the Administrative Data Quality Assurance Toolkit, we looked in turn at each of four practice areas associated with data quality:

- Operational context & admin data collection
- Communication with data supply partners
- QA principles, standards and checks by data suppliers, and
- Producers' QA investigations and documentation.

Following this, we summarise our judgement on the current quality characteristics of the DLHE and the required quality characteristics of the NewDLHE.

OPERATIONAL CONTEXT & ADMIN DATA COLLECTION

DLHE data is collected by HE providers in respect of coverage defined by HESA. The full methodology for data collection, enrichment and supply to HESA is available at <https://www.hesa.ac.uk/collection/c15018/> and is available to suppliers and users alike. Within the confines of this extensive methodology and the Code of Practice for HE data collections (<https://www.hesa.ac.uk/about/regulation/provider-info/code-of-practice>), HE providers have latitude to employ practices that work well for them. There is a range of practice, from contracted-out service, to the establishment of temporary call-centres, to the distribution of contact tasks among academic staff.

HESA does not audit the data collection process, but recommends that detailed records are maintained. In England, HEFCE audits the DLHE see <http://www.hefce.ac.uk/funding/da/Data.audit/>). This is a paperwork audit and does not involve the re-contacting of respondents. The results of these audits are not published, and so there is no public log of issues uncovered with the data from this source. In the rest of the UK, the position is less clear, but we are aware that at least one other Funding Council does not audit.

HESA undertakes comprehensive quality checks and the details of all automated checks are made available publicly. Manual checks and responses to queries can result in issues being identified, and when these arise, a note is made within the data supply to indicate any caveats that must be considered when using the data.

HESA does share analysis of the differences in data collection (for instance response rates at different HE providers) when this is requested or relevant, as well as applying criteria for both collection (target response rates) and onward supply of data (aggregation methodologies) that collectively act to ensure the data is used appropriately.

Some data items, such as the Standard Occupational Classification (SOC) code are based on coding by experts at the supplying HE providers. While a great deal of effort goes into producing comparable coding (facilitated both by HESA and the Association of Graduate Careers Advisory Services (AGCAS)) there are concerns around the SOC coding frame itself (not all roles are included), and judgement is required to apply it correctly.

Targets for obtaining valid responses are set with onward uses of the data in mind, and we have no reason to believe that they distort the outcomes, and the high levels achieved contribute to overall confidence in the data.

A helpful analysis has been produced by the English Department for Education's LEO data team, which explores the differences found between the salary data found in DLHE and data in the LEO dataset derived from HMRC tax receipts. Accounting for some of these discrepancies, the team say:

“Human factors may contribute to this discrepancy to an extent; the DLHE is filled in retrospectively, sometimes several months after the census date, so some error in recall can be expected. There may also be a discrepancy in the image that graduates wish to project to their former HEI and the reality of their employment situation. There are also certain types of individuals that are difficult to link to LEO (such as those with earnings below the tax threshold).

Administration error by employers may also create uncertainty around employment start and end dates; for example, it may be that employers are slow to record employees as leaving a job when they are engaged in casual work and would therefore appear as employed on the LEO census date but as unemployed on the DLHE. Similarly, employers who do not provide accurate start date information could be recorded as unemployed on census day despite having secured employment and having declared so on the DLHE. This uncertainty could contribute towards the number of individuals whose LEO records on census date do not corroborate their self-reported employment activity on the DLHE.”

See

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/573831/SFR60_2016_LEO_main_text_v1.1.pdf for full details of discrepancies in observed salaries in the two sources.

HESA has not identified sources of bias or error in the administrative system. However, some respondents to the recent consultation indicated that they held “concerns over the consistency and integrity of the current DLHE” with factors such as the distributed nature of activity and subsequent enrichment, and the potential for high-profile onward uses like league tables and TEF to drive perverse behaviours. Commentators have also indicated concern that in its current form DLHE might be considered overly “reliant on local interpretation. This makes the data inevitably prone to error and potentially liable to gaming” (<http://wonkhe.com/blogs/finding-new-ways-to-measure-graduate-success/>).

Examples of the sorts of alleged and anecdotal issues that may influence user perceptions (including those raised in articles such as this, in the *Times Higher Education*: <https://www.timeshighereducation.com/features/is-employability-data-being-manipulated/2018930.article>) include:

- Focusing on or prioritising calls on courses which generally get positive results, and de-prioritising those that tend to get less positive results
- Focusing calls on students with better academic outcomes, and de-prioritising calls to those who attained worse outcomes
- Not reporting, or mis-reporting, low level salary data or unemployment

- Mis-coding SOC to achieve a classification within Major Groups 1-3, in order to be considered a 'graduate job' for most external uses
- Coding those working away from home as travelling
- Counting someone who has slammed the phone down, after hearing the call is from the university, as an explicit refusal
- Amending responses that are considered to 'not make sense' without re-contacting the respondent.

At present, HESA does not comment routinely on these issues (though we act to safeguard against them). While our own analysis leads us to believe that the effects of any putative 'gaming' are not distorting the figures materially, the existence of this behaviour cannot be ruled-out definitively. There is a clear requirement to increase *perceptions* of confidence in the data either by investigating and publishing detailed analyses, by adopting controls and/or methodological improvements that address or nullify concerns such as these. Despite our technical confidence in the data, perceptions matter, and it is clear that more must be done to reassure data users, and to ensure all parties benefit from high confidence in data standards.

Our self-assessment is that we are currently offering some aspects of each level of assurance: basic, enhanced, and comprehensive. DLHE assurance is closest to the A2 level, but in most respects falls short of the level of comprehensive assurance at level A3.

Broadly, the above concerns are divided into three classes:

1. Making false claims
2. Influencing the survey strategy
3. Problems relating to inconsistent interpretation.

Tackling the issues in points 1) and 3) above requires a substantial reworking of methodology to incorporate either an enhanced audit function or better controls over process, such as through centralisation. When asked if a centralised approach would improve the robustness of data, just over half of those stating an opinion agreed that it would. HE providers were less likely to agree with this proposition, and other respondents very much more likely to agree. Overall, the consultation revealed that there are issues of concern to data users that must be addressed, and since this area of the consultation revealed split opinions, this appraisal document is intended to inform the review's approach to quality assurance.

A more modest approach could be adopted to address 2) above, at least in part: introducing a statistical method for identifying and correcting for non-response bias. This would involve comparing characteristics of the population to the characteristics of respondents. Where, say, a lower proportion of third-class degrees is found in the responses than are present in the survey population, this can be identified, along with any potential distorting effects. In some cases, it may be possible to correct the data for non-response bias, as well as identifying that an effect may be present.

As a minimum and immediate additional level of assurance, we recommend that HESA should follow the advice of ONS's review of DLHE for TEF¹ and investigate and implement an action plan for identifying and controlling for potential non-response bias at provider level for forthcoming DLHE data, prior to NewDLHE. This review notes that "Even with relatively good response rates, significant bias can be present." We therefore consider this an essential step, and if HESA is unable to comply, voluntary de-designation of the DLHE SFR as National Statistics should be considered as an alternative.

To attain the higher, level A3 assurance in this practice area, HESA must consider establishing either:

- a) More extensive investigatory controls, probably through a form of audit that includes re-sampling a percentage of respondents in addition to examining records of collection processes retained by providers/their agents. We would need to analyse and publish guidance on the data as a result of this investigation process, alongside the results of internal quality analysis of the data. We would also need to establish criteria for inclusion/non-inclusion of a provider's data in various outputs.
- b) Alternatively, applying a greater level of process control through increased centralisation of activity would achieve the same ends. The Longitudinal DLHE is currently administered through a centralised methodology.

COMMUNICATION WITH DATA SUPPLY PARTNERS

Communication with supply partners is very good, and HESA has worked hard to establish a collaborative relationship with individual providers. In particular, HESA maintains an effective high-level engagement with AGCAS and the Higher Education Strategic Planners Association (HESPA).

Data is required for statutory purposes, and the processes for collection, supply, transfer, data protection through the fair processing notices, and sign-off arrangements are well-documented and clearly offer an enhanced level of assurance. As part of this there is an ongoing dialogue regarding the IT requirements for the collection system, and advising HE providers of the specification for upgrades with substantial advance notice, through an agreed change management process (and one we have recently enhanced for the DLHE, by increasing our usual notice period by four months). We also specify the content of collection in detail. There is no written agreement with the majority of suppliers, and so these matters rest in part on a collaborative relationship and in part on the mandation of data collection by the sector's regulators and funders, rather than on a contractual duty.

We operate a training programme with basic and advanced chapters, and many suppliers attend this, though it is not compulsory.

We do not engage statistics users in a conference or similar event, though we do meet regularly with some users and suppliers and discuss the data, and have presented at user sessions run by government departments in the past.

¹ See the review of data sources for the TEF at: <https://www.gov.uk/government/publications/teaching-excellence-framework-review-of-data-sources>

We do not offer a formal opportunity for users of our published statistics to feedback to us, but this is a practice we are adopting more routinely, and we will add this to our action plan.

Overall, our assessment is that we are operating an enhanced level of assurance (A2) with some features of the A3 level of comprehensive assurance.

To obtain a firm A3 level of comprehensive assurance, HESA should consider improving the legal/contractual framework for the NewDLHE to ensure the methodological architecture is robust. HESA should also consider running a conference or similar activity to engage users of this high-profile data.

QA PRINCIPLES, STANDARDS AND CHECKS BY DATA SUPPLIERS

HESA provides a full description of its QA framework for the collection of data in a detailed online guide, available at: <https://www.hesa.ac.uk/collection/c15018/>. To the extent that HE providers must use this framework themselves we understand local processes.

HESA publishes guidance on the approach to data collection that should be taken by HE providers. At training seminars we often hear first-hand about the internal quality processes that HE providers undertake.

HESA recommends that HE providers keep extensive records of contact activities and of the rationale for enriching individual records (such as by SOC coding).

HESA publishes a Code of Practice for HE Data Collection, and notes that Funding Councils may audit DLHE collection processes by referring to records. HESA also notes that whistleblowers have prompted investigation by the Funding Council in England. However, HESA does not go further by describing the role of audit (A2) or identifying and documenting the findings of investigations and audits (A3). Neither do any Funding Councils undertake this higher-level activity. HESA neither requires information on, nor keeps a record of, which HE providers subject DLHE to formal internal audit, although we understand anecdotally that this occurs in at least some HE providers.

Most of the activity described above is consistent with a basic (A1) level of quality assurance.

To establish an enhanced or comprehensive level of assurance in this area, HESA would need either to develop an investigatory regime that extensively catalogues supplier-side activity, and consumes suppliers' audit reports, or to establish alternative controls over data collection. The results of this activity should be documented and used to inform users of the data and to improve data collection techniques. At present the level of assurance does not leave HESA in a position to definitively determine that the data is satisfactory for official statistics purposes, in this practice area.

PRODUCERS' QA INVESTIGATIONS AND DOCUMENTATION

HESA has established regular QA checks on the received data. This is a largely automated process. Details of these checks are published in a fully detailed way as quality rules that are

systematically validated. HESA does not publish details of more comprehensive checks such as for sense and consistency, but these do occur regularly and systematically and HESA maintains extensive records of this activity, as well as producing summaries of types of activity that are used to inform quality assurance activity and training activity for suppliers.

HESA undertakes post-collection quality checks and these inform our approach to use of the data. We record any known issues, and if we believe they will have a material effect on an end use we document this and present it on any output that uses the data. We also check all the data (tables and charts) produced specifically for the SFR and publications (and for any bespoke dataset produced from DLHE data).

HESA explains the methods of collection, and the likely degree of risk in using the data is apparent. We publish information about quality indicators, and in some cases suppress data on output if the quality is not high enough (or to control for the possibility of disclosure).

HESA has not undertaken comparisons with other relevant data sources, though DfE in England have done this with respect to salary data (as noted above).

We do not identify or explain the possible distortive effects of performance measures and targets formally, though through this review process we have been made aware of the concerns held by users about this. We have not undertaken a systematic evaluation of this risk, and have not published on it, other than within this document. Please also note the above recommendation that a mechanism for identifying and controlling-for non-response bias at provider level is required.

Overall, DLHE appears to largely pursue an enhanced level of assurance (A2), with some aspects of comprehensive assurance (A3).

In order to operate at A3 level, HESA would need to consider investigating the quality features of the DLHE dataset in substantially more detail than at present, and publishing its findings, including on particular risk factors and their distribution.

SUMMARY OF QUALITY CHARACTERISTICS

Based on the preceding subsections, we summarise our self-assessment in the following table:

| Practice area | Assessment of current level of assurance | Targeted future level of assurance |
|---|--|------------------------------------|
| Operational context & admin data collection | A2 | A3 |
| Communication with data supply partners | A2 | A3 |
| QA principles, standards and checks by data suppliers | A1 | A3 |
| Producers' QA investigations and documentation | A2 | A3 |

Our assessment is that current quality assurance requires an action-plan to:

- Control for non-response bias in the immediate term

- Recommend enhanced quality assurance mechanisms such as audit and resurvey or controls such as centralisation for the NewDLHE
- Provide enhanced mechanisms for educating users and gathering feedback from them about the data.

EVIDENCE OF EMBEDDED QUALITY ASSURANCE AND REVIEW ACTIVITY

HESA undertakes periodic reviews of all its data collections. We review more often where public interest and/or the rate of change in the domain is higher. Details about current and past reviews are made available on our website at: <https://www.hesa.ac.uk/innovation/records>

The #NewDLHE review is the largest review HESA has ever carried out. It has its own substantial web presence at: <https://www.hesa.ac.uk/innovation/records/reviews/newdlhe>

Reviews are conducted openly and all papers and notes from Strategic and Working Groups are made available on the website.

THE FORSEEN IMPACT OF NEWDLHE ON THE DLHE SFR AND CONSIDERATION OF REQUIREMENTS FOR NATIONAL STATISTICS DESIGNATION

The current DLHE SFR is a National Statistic. This section considers whether, with the introduction of NewDLHE we should remove the National Statistic status and publish the SFR as an Official Statistic.

The National Statistics label Official Statistics that are formally assessed by the UK Statistics Authority as being compliant with the Code of Practice for Official Statistics.

Areas for consideration within the Code of Practice for Official Statistics in relation to a future NewDLHE SFR are:

PRINCIPLE 2: IMPARTIALITY AND OBJECTIVITY

Practice 4 – Announce changes to methods or classifications well in advance of the release of the changed statistics.

We are proactively doing this now, with well-publicised user consultations having taken place and detailed information on the destinations and outcomes review available on our website.

PRINCIPLE 4: SOUND METHODS AND ASSURED QUALITY

Practice 3 - Adopt quality assurance procedures, including the consideration of each statistical product against users' requirements, and of their coherence with other statistical products.

A future NewDLHE SFR would not be coherent with previous DLHE SFRs. However, at a time of change in employment patterns, we must identify a new methodology or enhancements to existing methodology that increases the level of assurance offered and secures the widest range of benefits for data users. We are explicitly looking at incorporating data drawn from the HMRC/DWP data made accessible by the Small Business, Enterprise and Employment Act, and will ensure coherence with other public data drawn from this source, including LEO data.

Practice 5 - Seek to achieve continuous improvement in statistical processes by, for example, undertaking regular reviews or releasing statistical work in progress such as experimental statistics.

The preceding sections of this document demonstrate this practice in action, and the recommendations made there should be taken into account.

In addition, survey and data-gathering technology and capabilities have improved, and offer the prospect of making methodological improvements that will improve data quality, while reducing the substantial costs of data collection. These improvements form part of the basis of NewDLHE and provide good reason as to why we need to implement methodological changes.

Practice 6 - Promote comparability within the UK and internationally by, for example, adopting common standards, concepts, sampling frames, questions, definitions, statistical units and

classifications (including common geographic referencing and coding standards). Make the reasons for any deviations from standard models publicly available.

HESA will continue to utilise established ONS coding frames in the form of SIC and SOC.

We have also determined that it may be possible to produce an NS-SEC classification as a ‘by-product’ of other changes.

By linking to HMRC records under NewDLHE we are striving for better comparability in graduate salary information. HMRC are an internationally recognised authority in this field, and using their data would mean we adopt a common standard in relation to salary figures. However, by moving to NewDLHE we are removing comparability with previous DLHE outputs.

One additional issue that has come to light is that the definition of ‘employment’ used in the DLHE SFR and related publications is inconsistent with the International Labour Organisation’s definition of ‘employment’. Given the currency of the ILO definition (including its use in UK government-produced statistics, including LEO) and there being no countervailing argument beyond consistency with previous outputs, which in any case will be affected by the planned changes, HESA should implement a definition consistent with the ILOs for the NewDLHE survey.

Practice 7 – Where time series are revised, or changes are made to methods or coverage, produce consistent historical data where possible.

When NewDLHE is introduced, there will be a break in the time series of DLHE data. It is likely that it would not be possible to produce any consistent historical data in the first year of NewDLHE, due to changing the survey reference date among many other methodological changes. When data has been collected under the NewDLHE approach for two or more years, we will start to be able to build up a time series of data once more.

PRINCIPLE 6: PROPORTIONATE BURDEN

Practice 1: Report annually the estimated costs (for example, on businesses, service providers, or the public) of responding to statistical surveys and strive to develop methods that will reduce the costs to individual organisations or people.

We have not previously costed the production of the DLHE, but took the opportunity of last summer’s consultation to do so. We now have confidence in the figures we published showing overall costs and the range of costs per unit acquisition, thanks to the responses of 114 HE providers who costed their processes for us in some detail.

NewDLHE will be designed offer better value for money and aims to reduce the costs to individual organisations and people. Collecting salary information via HMRC will reduce the burden on respondents/HE providers to provide this information whilst creating no extra work for HMRC who already hold this data.

In terms of providing additional assurance, there appear to be two main options: audit or centralisation.

If the new survey were to be conducted centrally, this would reduce the burden even further on the individual HE providers and benefit from economies of scale and greater responsiveness. An audit function would still be required, but would be limited in scope.

If the new survey were to remain distributed, an enhanced analysis and quality assurance function would need to be established at HESA, alongside an extensive audit regime that would resurvey a sample of graduates as well as conducting audits at HE providers. In this situation we would need to consider what trade-offs would have to be made in terms of meeting user requirements, since a more extensive audit function can be expected to cost more and create additional accountability burden.

Conversely, a centralised approach deals with quality assurance issues with ease, but must be implemented in a way that is workable, cost-effective, and which addresses the legitimate concerns of HE providers.

PROTOCOL 1: USER ENGAGEMENT

Practice 7 - Consult users before making changes that affect statistics (for example, to coverage, definitions, or methods) or publications. Consultations should be:

- *Informed – by relevant central guidance on how consultations should be conducted; and by the views of user groups on the best means of obtaining views;*
- *Efficient – by balancing the importance of the issue and the likely impact of users' views against the time and resources available, so as to obtain good value for money from the consultation process; by liaising and co-ordinating with other producers to avoid duplication of effort and to minimise burdens; and by exploiting different methods of consultation;*
- *Clear – by describing the consultation, and expressing the issues, as simply and concisely as possible; and by publishing the timetable for each consultation; and*
- *Responsive – by publishing the records of decisions and actions following a consultation, together with explanations for them; and by publishing individual responses, unless anonymity is requested.*

A consultation on the principles and future requirements for data about graduates has taken place. This followed the four points above, with one caveat – we did not ask for permission to publish individual responses to our consultation, and have not done so.

PROTOCOL 3: THE USE OF ADMINISTRATIVE SOURCES FOR STATISTICAL PURPOSES

Practice 2 - Only base statistics on administrative data where the definitions and concepts are good approximations to those appropriate for statistical purposes.

Salary information collected for tax purposes by HMRC is a more consistently valid, and hence more accurate, data source than collecting this information via the existing DLHE survey. The definitions and concepts of HMRC data is very appropriate for statistical purposes in this context.

Practice 3 - Maximise opportunities for the use of administrative data, cross-analysis of sources and for the exchange and re-use of data, to avoid duplicating requests for information. Where

possible, use common information technology and information management systems that facilitate the flow of information between producers of statistics.

The Small Business, Enterprise and Employment (SBEE) Act permits the linking of HE student data to national tax and welfare records, for the purpose of educational evaluation. With the arrival of this act, NewDLHE is seeking this as an opportunity to use HMRC administrative data, thereby avoiding duplication of information and reducing respondent burden by no longer asking leavers themselves for this information. We have welcomed the close engagement of colleagues from across the UK's governments in this process.

We also proposed that further study information should be obtained from administrative data rather than self-reported. Further study information collected by HESA will have wider coverage and greater accuracy than self-reported data collected via DLHE. It should be possible to include FE data in time, also. This approach will lead to efficiencies in collection and higher quality data.

We have strong community support for this approach.

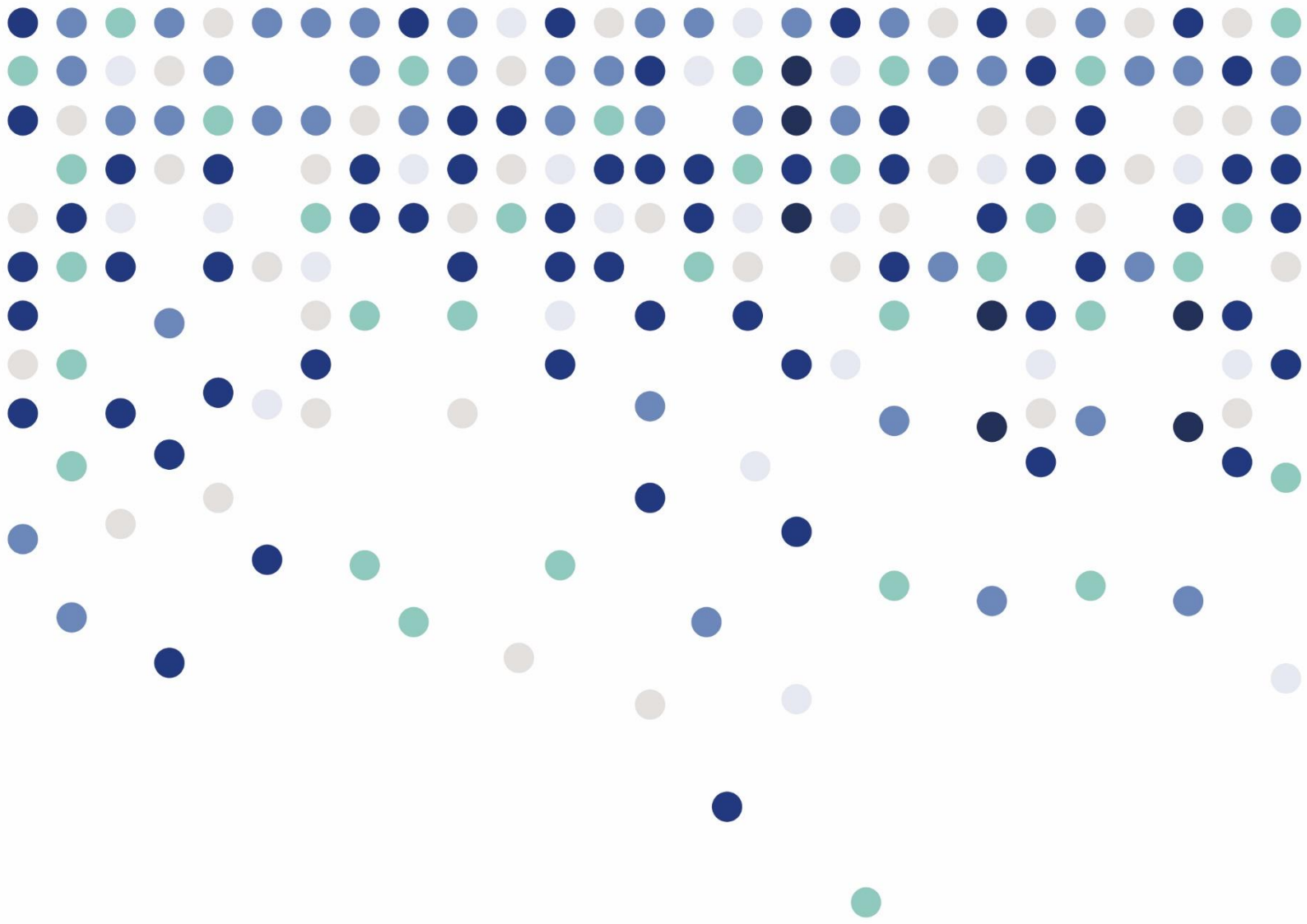
CONCLUSIONS

Our analysis is, subject to satisfactory pursuit of the recommendations for improvement contained in this report, that HESA continues to manage the DLHE SFR in a manner appropriate for designated National Statistics. We can be confident that the NewDLHE can be formulated in a way that retains this designation.

RECOMMENDATIONS

As a result of the above analysis we recommend the following actions:

- 1) HESA should investigate and implement an action plan for identifying and controlling for potential non-response bias at provider level for forthcoming DLHE data, prior to NewDLHE. We consider this an essential step, and if HESA is unable to comply, voluntary de-designation of the DLHE SFR as National Statistics should be considered as an alternative.
- 2) The design of the NewDLHE must deliver a comprehensive level of assurance, and HESA must produce a design that meets this requirement. The necessary methodological improvement could be achieved either through a centralised approach, or through a substantially enhanced audit process which investigates processes and practices, backed up by an enhanced analytical quality function at HESA and the publication of materials generated through these processes. If the collection process is distributed, then this process must include a sample resurvey.
- 3) HESA should adopt a definition of 'employment' compatible with the ILO's, in order to increase the usefulness and comparability of its data.
- 4) HESA should consider running a conference or similar event to regularly engage data users.
- 5) HESA should institute a process for gathering feedback online, from users of our publicly-available DLHE-based statistical products.



HESA
95 Promenade
Cheltenham
GL50 1HZ
E newdlhe@hesa.ac.uk
T +44 (0) 1242 211 103
W www.hesa.ac.uk

