NEWDLHE: THE FUTURE OF GRADUATE OUTCOMES DATA

THE RATIONALE

DAN COOK AND RACHEL HEWITT

MARCH 2017

Dan Cook and Rachel Hewitt
HESA
95 Promenade
Cheltenham
GL50 1HZ
E newdlhe@hesa.ac.uk
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.

RATIONALE

INTRODUCTION AND BACKGROUND

Purpose

In this section, we explain some of the thinking behind our approach to the design, including providing answers to the questions set out in the remit for the review. If you wish to move directly to the proposed model, select here.

About the review

We launched our review in the summer of 2015 by establishing strategic and working groups. They met in the autumn of that year to agree a programme of work. The review had the following high-level aims:

- Future-proofing – a fundamental reconsideration of the kinds of data that will be required for the foreseeable future, taking into account a labour market that is changing at a structural level, and increasing demands for rich information about graduate outcomes.
- Efficiency – taking advantage of new capabilities to link data sources and use modern survey technology to increase value for money and reduce the cost of data acquisition.
- Fit for purpose – ensuring the data collection methodology allows the data to be used in new and emerging contexts, with confidence.
- Supporting legislation – taking into account the legal gateway opened by the Small Business, Enterprise and Employment Act, and supporting developing government purposes for destinations and outcomes data, across the UK.

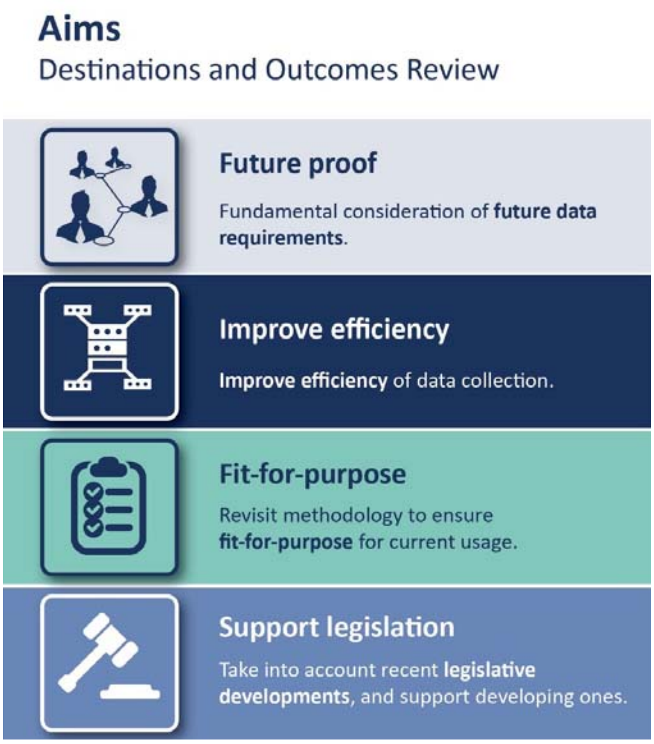

Fig 1: The aims of the NewDLHE review

In the course of our work on the review, we have addressed these core aims. Our proposed model for collecting graduate outcomes data represents a synthesis of the great number of inputs we have received.

QUESTIONS WE SET OUT TO ANSWER

To meet our stated aims, we published a remit,[1] which posed a series of questions that we would seek to answer through the review. We are now in a position to answer all of the questions that we set:

- What data will be made accessible by the Small Business, Enterprise and Employment (SBEE) Act, and in what ways can it be used?
- Given the availability of SBEE Act data, what other requirements do we have for information about what happens to leavers from higher education?
- What student data might be required to contextualise destinations data? For instance, better data on placements or study patterns?
- What are our remaining needs for post-study survey information, if any?
- What methodological improvements can we make to reduce costs, and to improve consistency? Consideration should be given to the benefits of centralisation.
- How complete a picture should the data provide?
- When would data be most timely?
- How should data collection activities relate to HE providers' careers support for graduates?
- What specific measures do we wish to obtain from this data?
- What level of detail is required in order to make the data useful for analysis?

We now explore how we have answered those questions and offer a supporting rationale for the proposal in this latest consultation.

What we learned from consultation, debate and commissioned research

During the review we have engaged in a wide variety of debates and produced a range of documentation that explores the issues surrounding the creation of a new model of collecting graduate outcomes data. We have commissioned original research, held a half-day conference in collaboration with Wonkhe, and undertaken a formal consultation.

We have also been asked to expose a richer set of data on outcomes publicly and raise the profile of data on outcomes and destinations.

We are acutely aware that the publication of Longitudinal Education Outcomes (LEO) data, and the use of DLHE data in the Teaching Excellence Framework (TEF) raises the bar on data collection. We are particularly grateful to have had the ongoing support of key individuals involved in that programme of work at the Department for Education (DfE) in England, as either members of or attendees at, review group sessions. Our proposals create a new model of collecting graduate outcomes data that can remain a trusted source of data for a range of public users, and which

[1] See our document, *Higher Education student outcomes and destinations: A remit for a fundamental review of information needs*, July 2015, https://www.hesa.ac.uk/files/Review_student-outcomes-and-destinations.docx.

offers providers new national measures that may be suitable as a part of institutional TEF submissions.

We learned that we need a universal survey to capture the contextual information about what graduates are doing, and their views. Views are important, as well as more objectively factual information – there was a very high level of support for new measures that capture the graduate's 'voice' and allow their self-assessment to be captured.

We learned that we should make extensive use of linked data – to gather both earnings and further study information.

The survey date should be pushed-out to 12-18 months to align more closely with common expectations regarding transition into a sustained pattern of activity following higher education. We learned that we should have a suite of questions specifically for graduate entrepreneurs, and start to understand the experiences of these individuals better.

We also confirmed that we should collect data on work placements in-year, as part of the HESA Student record (not through a survey).

What data will be made accessible by the Small Business, Enterprise and Employment (SBEE) Act, and in what ways can it be used?

The efforts of the DfE have resulted in publications of LEO data for HE providers in England. The DfE have published their own appraisals of this data.[2]

Individualised salaries can be derived from data held by the HMRC. Utilising an anonymised process, this data can be linked or matched to HESA data with a very high success rate.

Data is available on a tax-year basis. This means that data showing salary is available for the first full tax year following graduation – approximately 21 months after the main cohort of full-time undergraduates finish their courses.

However, the nature of the data also makes it possible to obtain salary data on a longitudinal basis, and also to look at earnings during and prior to study.

The completeness of the data is very good for students working in the UK following graduation, and includes all self-assessment data, even when the assessment is below the threshold for tax. The full range of data made accessible under the Act is still being explored, but we now have confidence that the core requirement of determining earnings from linked data is possible for the majority of students who are working in the UK following graduation.

Responses to our consultation were strongly in favour of utilising this data appropriately within a new data product.

We now anticipate that for graduates who work in the UK, that HMRC-derived data will be used to replace salary in HESA's National Statistics products. Since HESA produces data on a UK-wide

[2] See *Employment and Earnings Outcomes of Higher Education Graduates: Experimental data from the Longitudinal Education Outcomes (LEO) dataset*, 4 August 2016, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/543794/SFR36-2016_main_text_LEO.pdf, pp. 24-32.

basis, this would mean that HESA's statistical releases would be consistent with LEO data published elsewhere where salaries are concerned. HESA are working with DfE to develop a mechanism for secure access to this dataset.

We intend that salary data at course level, and also salary data by protected characteristic at principal subject level, will be made available to individual HE providers, in order to support the intentions of the SBEE Act.

Given the availability of SBEE Act data, what other requirements do we have for information about what happens to leavers from higher education?

We were not able to identify linked data sources for all data of interest. HMRC data on employer and location is not strong, because of the administrative requirements of the data collection process. As a result it is unlikely that we will be able to determine reliable information about employment context (company name, location, industry) for the foreseeable future, and will need to continue to collect this type of data from the graduate, or derive it. There was strong agreement from respondents that we will continue to need a survey of all graduates.

Respondents to our consultation agreed that the broad categories of information currently collected through DLHE remained of interest.

The current focus on employment outcomes, including industry, location, job title and so on, remains of greatest interest, but extra data is needed in a few areas, notably in respect of work-based learning and for graduate entrepreneurs.

We demonstrated that current further study information can be derived from linked data, and there was a strong appetite for this data source to be more fully exploited.

We also saw a strong demand for alternative measures of graduate outcomes, and gathered a great deal of useful feedback on how these should be constructed. Our survey design contains proposals that meet the needs that were expressed to us from a wide range of parties.

What student data might be required to contextualise destinations data? For instance, better data on placements or study patterns?

This question particularly referred to the potential for utilising data from the record of a student's study to add context and value to outcomes and destinations observed later. Not all information is best collected from a survey following graduation, especially if there is an existing high-quality data source that can be utilised.

In this area there was a strong finding in favour of the collection of information about work-based learning and placements. Work-based learning and placements of various types are considered to be an important factor in developing employability. The current HESA Student record collects information on large 'placements', such as during sandwich programmes or study abroad years. However, there is a desire to understand placement activity at a more granular level.

To answer this widely shared preference for data collection, HESA will introduce a mechanism for collecting this information in a future Student data collection, aligning the new requirement with the timetable for Data Futures roll-out in 2019/20. We will develop the specification for the collection inline with existing good practice identified in the sector, by revisiting previously published work by ASET.[3]

Requirements for information on other work-related learning were not conclusive, and we recommend that interested academics and the relevant sector organisations continue to work towards a robust typology.

No other areas were felt to be of sufficient priority to necessitate change.

What are our remaining needs for post-study survey information, if any?

The strongest and most widely held view expressed by a majority of respondents to the consultation was that the current data on salary and SOC code are an inadequate measure of graduate success. While employment is an important outcome following HE, there are other non-economic outcomes and benefits for the individual, the HE provider and wider society. Moreover, these non-economic benefits are seen as increasingly interesting as the structure of the labour market undergoes major change in response to developments in technology, and with the increasingly commonplace view that meaningful or useful activity is not confined to paid work. Nor is there a single or simple pattern of activity relating employment and study. Many students work while learning, and enhanced employability may be a good outcome, even if there is no change in role or salary. Equally, retention, productivity and value are all benefits that employers gain from HE, in addition to a supply of graduate- and postgraduate-level labour. The value created by HE in and beyond the workplace is multi-faceted. There is a requirement for robust alternative measures of graduate outcomes that help us surface this in public information. We have also been asked to raise the profile of destinations and outcomes data.

The overriding requirement is for a highly trusted data source that makes the wide range of outcomes from HE visible and accessible in the same way as occupational classification and salary. We have therefore concentrated on new questions, and on a survey methodology that addresses this requirement.

We also know that there is a need to be able to look at salary, occupational and other outcomes data by city/region, ethnicity and other factors.

What methodological improvements can we make to reduce costs, and to improve consistency? Consideration should be given to the benefits of centralisation.

We undertook an assessment to determine the required level of assurance that users require about the quality of the data (which can be read here), and the measures that need to be taken to meet them. In summary, our assessment was that further work needs to be done to increase the level of assurance for users. We considered the two main methods that could be utilised: either some form of centralisation (i.e. direct oversight or control of parts of the collection process); or a form of audit strong enough to allay concerns. We considered the relative merits of these approaches, and have decided to make a single recommendation for a particular open form of the centralised model. We believe this model meets the practical needs expressed by HE providers and other data users in our first consultation.

[3] See ASET, *Good Practice Guide for Work based and Placement Learning in Higher Education*, September 2013, http://www.asetonline.org/wp-content/uploads/2014/11/ASET-Good-Practice-Guide-2014.pdf, esp. p. 9.

We are grateful to the 114 providers who provided a costing for the running of the DLHE data collection. We published data on this in the summary of consultation responses. Because we received a figure produced on a highly comparable basis from such a large proportion of the sector, we are able to understand the costs of running the DLHE in its current format. Utilising information about the known costs of certain other survey activities, we are also able to assess with some confidence the likely savings that could be gained using a centralised approach. Our proposal takes this evidence into account.

In developing an openly centralised model, we have avoided over-simplification, and ensured HE providers retain a significant stake in the process, while taking advantage of the benefits of technology, economies of scale and combined purchasing power, and removing some of the burdens and limitations of the current methodology. Full details of the open centralisation model are available in the methodology section.

How complete a picture should the data provide?

There was extremely strong support for a universal census survey of all graduates, so as to provide a rich snapshot of information at an important time which can be used for a wide range of purposes.

It is important to note that the concept of a target for ‘completeness’ in future will change. Linked data sources will offer near complete coverage for a defined sub-population (earnings data for all graduates working in the UK; further HE studies for all graduates undertaking further HE study in the UK). The universal census survey data will be one important source among many. However, it is the case that high response rates in excess of 70% will normally be required in order to get data that is usable at course level. Our investigation of the Australian equivalent to the DLHE (which is an online-only survey at 12 months after graduation) leads us to believe that a combined online and telephone survey can continue to deliver the high response rates we collectively require.

Another aspect of completeness is the longitudinal view. While there is strong support overall for a single universal survey, there is significant minority support for other survey activity. In addition, linked data also permits longitudinal views to be built-up, without always necessitating additional surveying. In our design we do not respond to these needs directly, although our design does allow for future adaptation and extensibility, if required. However, at this stage additional surveys are not features of the proposal.

Third party responses still form a part of the methodology, but as now are limited in the scope of their application.

When would data be most timely?

Consultation responses indicate a strong and widely held view that six months is no longer the optimal point at which to survey graduates about their destinations. Various points were made supportive of an extension to 12 or 18 months, but most centred on the changed structure of the graduate jobs market and expectations of longer transitions into a settled pattern of activity. Data from after this transition takes place for the majority of graduates appears to be more useful than faster data on first destinations. There is a widely held view that the utility of the survey will be enhanced substantially by a later date. We have responded to this by proposing a survey at 15 months after graduation, with four survey points through the year.

One of the main risks to a later survey would be a reduction in the response rate. We have investigated a comparator survey – Australia's Graduate Outcomes Survey (GOS). The GOS surveys at 12 months after graduation, and despite this longer time period, manages to obtain response rates of 39% using an online-only method. This is encouraging (it is similar to DLHE, despite the timing difference) and gives confidence that with appropriate attention to maintaining contact details, and a national communications plan, a combined online and telephone survey will continue to achieve response rates in excess of 70% at each provider. There will be a continuing necessity for HE providers to consider what strategies they might use to maintain accurate contact details for this longer period.

To address concerns over losing the current provision of first destinations data, we have incorporated questions that seek extra information about previous work. Although this data will not be available until later than at present, and it has not been designed with absolute continuity in mind, it will offer data that continues to be of interest to many and will extend the longitudinal usefulness of the survey.

How should data collection activities relate to HE providers' careers support for graduates?

We received a wide range of responses on this matter, which reflects the diversity of approaches to the collection of data. The current distributed methodology allows for a range of approaches, with some providers outsourcing the whole process, others working closely with an outsourced collection partner, and some running an in-house call-centre or distributing contact responsibilities to academic staff.

Regardless of the local application of the methodology, HE providers have made it clear how important it is to get contact right. The training and skill of the staff involved matters a great deal. Tenacity and empathy are required in equal measure, and a certain amount of investigative effort is required to locate graduates for whom contact details have become out-dated. While some providers preferred to retain control over the process of data collection, others were keen to extol the benefits of an outsourced approach.

Some factors emerged as highly important to most respondents:

- Data must be perceived as highly trusted and impartial above all else
- Data must be of a high quality and have a high response rate to be useful
- The way data is collected is sensitive to context – this must be respected
- Providers require the latitude to ask their own questions
- Data collection can be integrated with student 'after-care' support
- A high level of training and professionalism is required
- There is intense interest in the survey data, and a near real-time feed is required
- Providers need to have appropriate influence over the collection strategy
- Expertise that providers have developed to maximise responses is important
- The current methodology cannot ensure that strategies for maximisation of responses are equally spread between or within providers.

We took these into account when developing our model, and our proposal for an open centralised methodology addresses these concerns.

What specific measures do we wish to obtain from these data?

Most existing measures remain intact, and we anticipate continuing to produce all of the same measures that are currently produced. By collecting placement/work-based learning data we open-up the possibility of new measures based on these. We have also determined that it will be possible to produce an NS-SEC classification for graduates in work as a by-product of the new approach, if required.

We have concentrated our efforts on developing the new measures of graduate outcomes asked for very strongly by consultation respondents. Each of the outline proposals we put forward had a full-range of responses, but there was no clear pattern of preference to follow.

There was strong support in written comments for a measure that enabled the graduate to self-assess their own outcome – to hear their voice. However, a qualitative measure would be unlikely to offer the kind of response needed for public information purposes – this needs to be a quantitative measure.

An outcome measure based on skills was thought appropriate by many, but the proposed method of determining this was felt to be complex, burdensome and liable to graduates misinterpreting the question. While it was clear that graduates will have a view on how the things they have learned are helping them in their current situation, and that this is interesting and useful data, a blunter tool is required.

Two other areas were recurrent themes in both consultation responses and the two pieces of research we commissioned: one was a sense of meaningfulness or importance of work to the individual; the other was a sense of the fit between current activity and future plans.

We have made proposals for three new questions covering these three dimensions. These take the best of the suggestions offered in consultation, and will be asked of all graduates, regardless of their activity.

Both subjective wellbeing measures and the net promoter score had strong supporters (and detractors). We have proposed these as optional banks of questions, for those that wish to use them. We have also introduced optional question banks using questions from the LDLHE, which ask about the impact of their HE provider on the graduate's future activity, and ask graduates to reflect on the HE choices they made.

We have also proposed new questions for graduate entrepreneurs.

What level of detail is required in order to make the data useful for analysis?

We asked respondents to offer case studies for how they currently use the data, and how they anticipate using it in future. We got a mixture of responses, but when it comes to information about employers and jobs, and about further study and other activities, there is a clear need for individualised data. With this in mind, our proposals ensure survey and linked HESA Student data will be made available at an individualised level.

When considering the availability of salary data, there are limitations that must be borne in mind. Salary information is currently collected by consent, and to move away from this to a position where salary data is available from linked HMRC data, requires careful consideration over appropriateness from a data protection perspective – both from the point of view of the enabling legislation, and from the point of view of the uses that HE providers have for salary data.

While there is always a desire for the most granular salary data possible, the lowest level of granularity required within a provider's own data that we could identify from consultation responses was at the level of an individual course. Other usual views of data were by protected characteristic, by geographical location and by subject.

However, some respondents also indicated a desire to understand salary outcomes in the light of other characteristics held in the HE provider's own system, and a bespoke capability to conduct institutional research of this kind is evidently desirable.

Our proposals therefore offer a mixed approach. First, it is our intention to provide a supply of salary data on an anonymised basis by course, reflecting the main uses that were explained to us as a part of the consultation.

Second, HESA will advocate for, and work with the Department for Education and others to find, a mechanism whereby an HE provider's own data can be linked to HMRC-derived data in a suitably controlled environment, where permitted purposes allow.

OTHER INFLUENCES ON THE DESIGN OF OUR NEW MODEL OF COLLECTING GRADUATE OUTCOMES DATA

Methodological requirements

Following the consultation, we decided to undertake a review of the suitability of the DLHE for retaining National Statistics designation. Methodological considerations had been raised in both the remit for the review and by respondents to the consultation. We knew that the DLHE needed to be perceived as beyond reproach, but we wanted to understand what this would entail in practical terms. The analysis of the suitability of the DLHE for official statistics purposes offered us a framework to think through these issues, and in any case is a recognised good practice to follow. The statistical case can be read here.

The key findings from the statistical case are that a high level of assurance is required, commensurate with retaining National Statistics designation for outputs from the DLHE. We also concluded that a higher level of assurance is required to ensure that the DLHE (and the NewDLHE) retains National Statistics designation. We perceived a risk to retaining this designation if we do not put in place an improvement plan for the quality assurance of DLHE data.

Our key finding was this:

> 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.

We took this into account when deciding how to develop a model. A key decision was how to wrangle with the audit-based and centralisation-based approaches. We offer a rationale for our choice to present a centralised option only, below.

Audit

A rigorous audit system would deliver two notable benefits: it would ensure the necessary statistical robustness, while also allowing HE providers to maintain direct contact with their graduates. To provide confidence in the data, this audit regime would need to go beyond a simple administrative process of record checking. It would have to include enhanced analysis and quality assurance functions at HESA, as well as thorough investigations of local processes and practices. These functions would also need to be complemented by a resurveying of samples of respondents.

We have investigated the implications of developing this kind of system and have determined two significant problems. Firstly, such a system would considerably increase financial and administrative burdens across the HE sector. Graduate outcomes data is typically analysed at course level, which means resurveying would need to be seen to act as a deterrent to malpractice at this level. This would mean resurveying approximately 5%-10% of the total population. Using the costings provided elsewhere in this document, the cost of this resurveying would be in the order of £0.5 million per year. Added to this would be the costs of running the current DLHE (£5.5 million per year) plus the additional costs of auditors.

Secondly, the system would prove tricky to implement and would substantially alter the productive relationship that currently exists between HESA and HE providers. We asked ourselves what happens if an HE provider fails audit? Under emerging conditions this could mean failure to submit data to the TEF, with the consequence of debarring the provider from attaining recognition and the linked right to charge higher fees in England. This is a high-stakes type of audit, and likely open to legal challenge. It is not characteristic of the relationship HESA currently enjoys with the HE sector, nor is it a direction in which HESA would wish to develop.

But audit processes must be seen to be operational and serious to be effective. The converse situation is almost worse: what happens if no HE provider fails audit? There must be public examples or the audit route is not seen as a serious check on behaviour.

These considerations have led us to believe that an audit-approach is expensive and fraught with risks. It is also unclear of the value it brings to the HE sector. A centralised model would deliver the same standard of statistical robustness, while realising substantial efficiency savings – a key aim of the NewDLHE review. It is also the case that HE providers may (and do) retain close contact with alumni in ways that go beyond data collection, and nothing in our proposals limits this. Indeed, we have designed a model of open centralisation which facilitates this interaction, even while delivering efficiencies.

Centralisation

A centralised approach naturally comprises characteristics that promote confidence. There is a clear separation between the data collection activity and the HE provider – it is rapidly apparent that no undue influence can be brought to bear on the data collection process. The HE provider is put in the position of demanding the highest standards of fairness in data collection from the contractor. By placing the data collection in the hands of an impartial body, no public perception of the potential for malpractice can be entertained, and fair dealing is guaranteed. Expectations of fairness and consistency can be built in to the contract, and meeting performance targets can be evidenced, easily. Moreover, the financial risks associated with malpractice are much higher for a single contractor, and appropriate oversight can be built into the contract to reinforce this.

Open centralisation also creates the conditions for reducing costs and leveraging extra value for money. In our financial case, we indicate the substantial savings that can be made by creating economies of scale and pursuing more efficient collection strategies than at present.

Open centralisation can also be agile in responding to changes in the labour market, or in a desire to change a feature of the survey. At present, changes require long notice periods, and updates to student record systems to be written and applied. The considerable effort and expense involved in this is removed, as instead of more than 200 individual implementations of change, only one is required.

Nevertheless, getting centralisation right requires close attention to practicalities, and providers need confidence that they have sufficient control and interaction to ensure their expertise and understanding of their graduates is applied effectively. Our model of open centralisation addresses these concerns, and ensures providers retain a meaningful stake in the approach.

During procurement and implementation the conditions will be set for a fair and consistent application of the methodology. Factors that can affect response rates which HE providers have learned over years (such as the timing of calls) can be built into the contract, and we will be seeking input to ensure this learning is captured.

Providers need to see incoming survey data on a very regular basis, and the proposed model offers this, with near real-time data made available through HESA systems. There will be core questions, plus optional banks that can be asked online only or, for an additional charge, added to the telephone interviews.

HE providers will require a relationship with the survey organisation that enables them to discuss emerging issues with the data, raise concerns and influence the collection strategy to obtain the response levels they need. Open centralisation envisages this relationship as being similar to that currently in place for providers that out-source their DLHE data collection to third party suppliers. There will be a contact at the survey organisation with whom the HE provider can build a relationship.

Providers also need to be able to commission their own questions, and to have the opportunity to define and agree value-adding services with the contractor. This 'bespoke' activity is permitted and enabled by the proposed approach.

The open centralisation model retains the control and oversight that HE providers currently have over the data collection process, while ensuring the methodology is applied rigorously and fairly. It creates efficiencies and agility that can be shared widely, and funded fairly.

Dan Cook and Rachel Hewitt
HESA
95 Promenade
Cheltenham
GL50 1HZ
E newdlhe@hesa.ac.uk
W www.hesa.ac.uk