



GRADUATE OUTCOMES DISSEMINATION POLICY

JONATHAN WALLER, DIRECTOR OF DATA AND INNOVATION
REBECCA MANTLE, OFFICIAL STATISTICS MANAGER
MAY 2020



Official Statistics

HESA

95 Promenade

Cheltenham

GL50 1HZ

E official.statistics@hesa.ac.uk

T +44 (0) 1242 388 513

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

| | |
|---|----|
| Introduction | 3 |
| Key users and uses | 3 |
| Legal and Compliance considerations..... | 4 |
| Data Protection Act 2018/GDPR | 4 |
| Official Statistics | 5 |
| Freedom of Information (FOI)..... | 6 |
| Competition Law..... | 6 |
| Ethical considerations | 6 |
| Misrepresentations of Graduate Outcomes data | 7 |
| Products and services..... | 8 |
| Supporting information for users of the data..... | 10 |
| Data definitions..... | 10 |
| Quality and data intelligence..... | 10 |
| Methodology statement | 10 |
| Content and presentation of data in HESA data releases..... | 11 |
| Operational policy on Tailored Datasets | 11 |
| Provider Preview service..... | 12 |
| Comparisons with DLHE | 12 |
| Key data concepts and standards | 13 |
| Introduction | 13 |
| Classes of usage | 13 |
| Response rates and incomplete responses | 14 |
| Weighting of survey responses..... | 15 |
| Graduate activity | 15 |
| Employment and work | 16 |
| Activity between qualifying and survey census week | 17 |
| Salary data | 18 |
| Subjective wellbeing | 19 |
| Opt-in question banks..... | 19 |
| Further data standards and derivations | 19 |

INTRODUCTION

This document sets out HESA's policy, approaches and standards for dissemination of data from the Graduate Outcomes survey. The goal of the policy is to support users and stakeholders by explaining the products and services through which this data will be disseminated, the processes and approaches used by HESA, and some of the key data standards we will use (i.e. how we use the data to express key concepts in Graduate Outcomes). This information is important because it allows stakeholders, such as HE providers, to understand how data will be used in HESA statistical products and disseminated to third party-users. It also assists data users in making informed choices about how they utilise the published data and statistics.

This document will be updated and extended over time, as further decisions are made about data standards and to provide further detail on any areas which are subject to significant numbers of queries from stakeholders.

Please refer to our [glossary page](#) for definitions of terms used in this document.

KEY USERS AND USES

The primary users of data on the outcomes or destinations of graduates from higher education can be classified at summary level as being from the following interest groups:

- 1) Press/journalists (national, local, and specialist publications/websites)
- 2) Governments (UK and overseas)
- 3) HE providers and sector bodies
- 4) Academic researchers
- 5) Graduate employers

The general public, prospective and current HE students and their parents, guardians, and advisers, are also users. However, a relatively small proportion of these will seek information from HESA directly. The majority of users in this category will use HESA data in a format that is mediated by one of the interest groups above.

The main observed uses of previous data on destinations of graduates (and therefore likely future uses of data from the Graduate Outcomes survey) can be split into two broad categories:

- 1) Market information. In the form of news stories, league tables and rankings; information, advice, and guidance services; and regulatory/performance information for public consumption. The main intended consumers are the public, particularly prospective students, their parents, guardians, and advisers.
- 2) Business uses. Expert-led data analysis to support research, advocacy, benchmarking, planning, and operations within HE providers, and to inform national-level research and analysis to support service development and delivery, funding and procurement, governance and regulation, graduate recruitment, and other system-level activities.

These uses call for a certain degree of expertise and understanding on the part of users. The extent to which such expertise is pre-existing among data users will vary, although a majority are likely to have at least some experience of using data and a reasonable grasp of the subject domain described by the Graduate Outcomes survey. This document explains the choices HESA has

made in creating and disseminating its survey outputs, and describes the advice that HESA will be providing to users, where appropriate, on interpretation of these outputs.

Our main goal is to support interpretation of the data and encourage appropriate and responsible usage by a diverse range of communities.

This document is not intended to directly form part of the suite of information intended for end users of the data – such information will be delivered as part of the data releases. However, for those users who are interested it provides some information in advance of the data releases on data standards that will be used.

To supplement this, and other, written information, HESA has also organised a series of webinars to coincide with the launch of the Graduate Outcomes data. These are intended to offer similar information in a different format, and also offer a degree of interaction unavailable in a written document. These webinars will be recorded and made available via the HESA website.

This policy and the webinars are offered in addition to the standard support arrangements for all HESA data, including a substantial suite of definitional materials and descriptions of the methodology, along with advice by telephone and email from our analysts and statisticians.

In the lead up to the statistical releases, HESA has also been sharing information about Graduate Outcomes in order to build understanding and awareness of the distinctive characteristics of the survey and the methodology that has been employed. Content includes blogs and news items from key HESA experts and guidance on the releases themselves. All of this content is available on the dedicated [graduates Open Data](#) page.

We welcome feedback from users on our materials and welcome all suggestions for improvements and refinements.

LEGAL AND COMPLIANCE CONSIDERATIONS

The HESA dissemination policy and practices for Graduate Outcomes data operate within the following compliance environment. We explain these compliance obligations because they set the terms within which our policies on publishing and otherwise disseminating data must operate. They therefore have a profound influence on the approaches we take.

DATA PROTECTION ACT 2018/GDPR

HESA collects identifiable personal data about students and graduates in and from UK HE. The majority of the data and information HESA expects to disseminate around Graduate Outcomes is drawn from both of these personal data collections.

HESA takes its responsibilities under data protection extremely seriously and recognises that, beyond basic compliance with the law, HESA data collection operations depend upon maintaining the trust and confidence of data subjects. Within our data dissemination activities, we focus a great deal of effort on reducing any risks to the confidentiality of individuals to the absolute minimum. We deploy tried and trusted approaches in achieving this, such as data minimisation, pseudonymisation, anonymisation (deploying statistical disclosure control) and contractual controls. [Find out more about approaches used by HESA.](#)

Where more detailed data, such as microdata (i.e. data at a very detailed level that could pose risks to confidentiality of individuals), is required to support particular uses (such as academic and other research), data extracts will only be supplied under a strict data sharing agreement which specifies how the data may be used, including the precise purpose, the permitted user(s), protections for data subjects, and the timeframe over which it is allowed to be retained. Data supplied under such terms must be destroyed once the permitted use has been completed. The agreement under which such data is supplied includes stringent requirements on data recipients to maintain information security and protect confidentiality of individuals. All uses must be compliant with the terms of the [Data Collection Notices](#) issued to data subjects.

OFFICIAL STATISTICS

HESA is a designated producer of Official Statistics under the [Statistics and Registration Service Act 2007](#) and associated [Official Statistics Orders](#).

HESA publishes a wide range of Official Statistics, such as [Statistical Bulletins](#). In respect of our National Statistics outputs specifically (and by implication our wider Official Statistics outputs), we are under statutory obligation to comply with the [Code of Practice for Statistics](#), which sets out a range of expectations on how producers should plan, design, construct and disseminate statistics based around core themes of trustworthiness, quality and value.



One of the key requirements of the Code is that organisations **should commit to releasing their statistics in an open and transparent manner that promotes public confidence**. Access to statistics prior to general publication is strictly controlled under the terms of secondary legislation – the Pre-release Access to Official Statistics Orders¹ established by each home nation of the UK.

Under the Code, together with these Orders, producers of Official and National Statistics must restrict access to unpublished statistics to those directly involved in the production and quality assurance process. The Orders permit Official Statistics to be shared with a minimal number of 'eligible persons' (primarily Government Ministers and Senior Official together with their advisors) no more than 24 hours prior to general publication. HESA will normally accept requests to provide pre-release access to eligible persons where such requests are fully compliant with the Pre-release

¹ <http://www.legislation.gov.uk/ukxi/2008/2998/contents/made>,
<http://www.legislation.gov.uk/ssi/2008/399/contents/made>,
<http://www.legislation.gov.uk/nisr/2009/71/contents/made>,
<http://www.legislation.gov.uk/wsi/2009/2818/contents/made>

Access Orders, subject to the requirement to keep the number of such persons to an absolute minimum. Access prior to publication is not made available to any other parties under any circumstances, and this includes members of the UK press and media.

Information on HESA's compliance with the Official Statistics framework is available via our [website](#).

FREEDOM OF INFORMATION (FOI)

HESA is not covered by this legislation as we are not a Public Authority under the Freedom of Information Act 2000. However, HESA aims to support openness and transparency in its operations and we publish extensive information via our website as well as responding promptly to all reasonable requests for information.

COMPETITION LAW

HESA evaluates all dissemination practices to ensure compliance with UK Competition Law. As part of HESA's compliance, where information is approved for sharing, we seek to make it equally accessible to all potential competitors.

ETHICAL CONSIDERATIONS

HESA takes its compliance responsibilities under the Code of Practice for Statistics seriously and aims to act in ways that promote trust in its work. This means that we have skilled staff operating appropriate methodologies that result in high-quality statistics that offer value to users and data suppliers.

HESA also takes its responsibilities under the [Code of Practice for HE data](#) seriously and we aim to act with honesty, impartiality, and rigour in all that we do. We believe this helps secure better outcomes for HE with data, and it supports our [charitable mission](#).

HESA has considered a range of specific ethical issues in the development of the Graduate Outcomes survey, including an evaluation of graduate perceptions of the survey undertaken as part of cognitive testing, and provision of support for respondents who are affected by the survey (including training for interviewers to recognise such issues in respondents). Our [Graduate Outcomes Steering Group](#) offer us a range of valuable perspectives that help us to determine an appropriate course of action in the management of the survey.

HESA also has access, on request, to the [National Statistician's Data Ethics Advisory Panel](#), a new body set up to improve ethical practice in statistics. While we have not yet had opportunity to ask for scrutiny by the Panel, HESA is currently evaluating the materials produced by that group. This is with a view to developing a specific Data Ethics policy or framework. This work has not yet progressed far enough to issue a definitive statement about how ethical considerations will be evaluated as a part of the Graduate Outcomes dissemination policy. However, all decisions about data are taken in accordance with HESA's prevailing protocols on data governance, and existing policies.

MISREPRESENTATIONS OF GRADUATE OUTCOMES DATA

Misrepresentation is the action of giving a false or misleading account of the nature of something. Misrepresentation may occur unintentionally and can arise through misunderstanding. Intentionality does not materially alter the need to intervene in response to misrepresentation but should be taken into account in determining the form of any intervention. HESA maintains an extensive range of user support to promote appropriate and responsible use of its data, but misrepresentations do sometimes occur.

HESA cannot take responsibility for the decisions of external parties over their use of our statistics. Neither do we seek to be an arbiter of debates over policy and practice in the domains where we publish, nor do we possess the resources to proactively police and assure all external uses of our data. HESA does, however, recognise the potential risks to (and serious impacts on) the perceived trustworthiness, quality, and value of its statistics, in cases where they are cited as supporting invalid conclusions or interpretations.

With regard to HESA's statistics, misrepresentation is deemed to have occurred if any of the following statements is true:

- a. The advice of statistics professionals has been ignored.
- b. HESA statistics that are used in a document or statement by an external party, are presented in such a way that (in the Agency's opinion) they are liable to mislead the public or undermine the integrity of official statistics.
- c. Information has been circulated which is falsely claimed to derive from HESA, or might reasonably be misinterpreted as HESA data by users.

When we become aware of such cases of misrepresentation, we will evaluate the risks, and may choose to intervene. We may choose to do this in a variety of ways, depending on the circumstances. The first step will normally be to discuss the issue with the party believed responsible for the misrepresentation. Where the organisation recognises that a mistake has been made, undertakes to clarify an aspect of the statistics promptly, and does so in line with our advice, then we can decide that one (or more) of the following courses of action applies:

- 1) No further action at all is required.
- 2) No public letter need be written, but a private letter will be written.
- 3) A simple public message welcoming the clarification is published by HESA.

Where initial engagements do not achieve the desired outcomes, the following further options are open to HESA:

- 4) Publicising the fact that a suspected misrepresentation has come to our attention, and that we will be looking into the matter.
- 5) Publishing a clarification of our own, referencing the misrepresentation.
- 6) Writing a public (published) letter to the person or organisation responsible for the misrepresentation, informing them of the error, and explaining both why this is a misrepresentation, and the seriousness of the matter.
- 7) Writing to the regulator (or other authority) most relevant to the professional domain of the organisation held responsible for the misrepresentation, or of the forum in which the misrepresentation occurred.

- 8) In the most serious cases, where the actions above have been exhausted and are still deemed insufficient, HESA's Board can decide to take further actions, for example:
- Seeking advice from the National Statistician, or their Data Ethics Committee.
 - Referring the matter for investigation by the UK Statistics Authority's Office for Statistics Regulation.
 - Where permissible and feasible, imposing sanctions on future access to data.
 - Any other lawful and reasonable measures.

PRODUCTS AND SERVICES

Our current suite of outputs and dissemination routes derived from analysis of the Graduate Outcomes survey data are as follows:

| | |
|---|--|
| <p>Graduate Outcomes: Statistical Bulletin</p> | <p>This represents the first annual release of Graduate Outcomes statistics. The publication date for the annual Graduate Outcomes Statistical Bulletin is announced via the Upcoming data releases page on the HESA website. The exact publication date may vary from year to year and will be agreed in advance with the UK's education departments, devolved administrations and the Office for Students. Release dates are listed on the webpage advised above on the HESA website and on the UK Statistics publication hub at least four weeks prior to release. Release happens at 09:30am on the chosen day.</p> <p>This release includes highlights and summary analysis drawn from the data, with associated commentary to give context to the data presented. It contains interactive tables and charts with accompanying data downloads. It provides high level findings and characteristics of graduate outcomes at a national level. It provides a useful overview of the survey results alongside a summary assessment of quality.</p> <p>The first Statistical Bulletin to be published in 2020 will be published as Experimental Statistics under the terms of the Code of Practice for Statistics. The reasons for and implications of this are explained in a HESA website blog.</p> |
| <p>Graduate Outcomes: open data</p> | <p>This release provides a more extensive suite of open data tables drawn from the Graduate Outcomes survey. This will build on the Statistical Bulletin providing a much wider range of more in-depth interactive tables and charts with accompanying data downloads. A key difference in content compared to the Statistical Bulletin is that the Open Data will include data disaggregated by individual HE Providers. The schedule for release is announced in a similar way to the Statistical Bulletin and via the same channels. Release is also at 09:30 on the chosen day. The open data publication normally follows publication of the statistical release by a few days – however, this is under review and in future these two products may be merged to form a single new combined product.</p> <p>As with the Statistical Bulletin, the 2020 edition of this product will be published as Experimental Statistics.</p> |

| | |
|---|--|
| <p>Ad-hoc blogs and research papers</p> | <p>HESA undertakes specific research and analysis activities from time to time to inform and support internal operational improvement and to deliver insights for our users. We anticipate that research outputs relating to Graduate Outcomes data will be published on an ad-hoc basis, serving purposes such as:</p> <ul style="list-style-type: none"> • Quality assessment of survey data (e.g. understanding quality in the context of survey methods or comparisons with other related data sources such as LEO, to support survey improvement) • Addressing novel insights into areas of policy or topical interest. <p>Any such materials will be available on our Open Data – Graduates area of our website.</p> |
| <p>A new UK Performance Indicator (UKPI) covering Graduate Outcomes</p> | <p>We will develop a new UKPI for Graduate Outcomes to replace the employment and study indicator previously published using data from the Destinations of Leavers from Higher Education (DLHE) survey (UKPI Table E1). A new external UKPIs advisory group will be tasked with advising on the development of the new indicator.</p> |
| <p>Microdata for Statutory purposes undertaken by HESA’s statutory customers</p> | <p>HESA makes full curated files of quality-assured microdata, including derived fields, available to its statutory customers according to an agreed schedule, and subject to the processing notice under which the data was collected. The transfer of this data is subject to the legal agreements between HESA and its statutory customers, which specify <i>inter alia</i> the lawful grounds for processing the data under the GDPR/Data Protection Act 2018.</p> <p>Each statutory body receives microdata covering constituencies of HE providers and a range of data fields that aligns with their statutory powers and public functions.</p> |
| <p>Microdata for HE providers on whose behalf data is collected</p> | <p>HE providers that have subscribed to HESA and that have graduates eligible for the survey receive a file comprising the microdata for their own survey respondents on an individual level basis (except for subjective wellbeing data which is delivered in aggregated and anonymised form).</p> <p>All data is provided through the ‘Graduate Outcomes provider portal’ – an online portal collecting graduate contact details and delivering access to Graduate Outcomes survey data.</p> |
| <p>ONS Secure Research Service</p> | <p>HESA’s intention is to provide access to microdata for academic researchers through the Office for National Statistics (ONS) Secure Research Service in due course. Discussions between HESA and ONS to enable this are underway. This is not expected to be in place with regard to Graduate Outcomes data until the latter part of 2020.</p> |

| | |
|---|---|
| Heidi Plus | This online analytics service is delivered for the primary benefit of HE providers subscribed to HESA, and other not-for-profit organisations operating in the UK HE sector. HESA has appointed Jisc to be its data processor in operating Heidi Plus. The service will include new datasets comprising a range of data drawn from the Graduate Outcomes survey. The intention is for these datasets to be available to users as soon after publication of the open data release as possible. |
| Tailored Datasets | The Tailored Datasets service provides data and analysis to customers' exact requirements. Data drawn from the Graduate Outcomes survey will be made available via this service, as soon after publication of the open data release as possible. As with Heidi Plus, HESA has appointed Jisc as a data processor in delivering the service. |
| Jisc 'interactive insights' products | Jisc is developing new interactive data dashboards that will aim to deliver insights for staff in HE providers responsible for careers guidance and strategic planning. More information on these will be available by contacting Jisc at data.analytics@jisc.ac.uk . |

SUPPORTING INFORMATION FOR USERS OF THE DATA

Graduate Outcomes products and services will be accompanied by a range of resources designed to guide users with varying levels of skill and experience with data.

DATA DEFINITIONS

All products and services which incorporate Graduate Outcomes data will be accompanied by comprehensive definitional documentation. This will explain in detail the data standards which are utilised and how these have been derived.

QUALITY AND DATA INTELLIGENCE

In addition to full definitional information, all Graduate Outcomes data dissemination will be accompanied by a published quality report. This report will include the most comprehensive current assessment of the strengths and weaknesses of the data as well as providing information on any known specific quality issues. This report will be a 'live' document updated to include any new quality issues as they are discovered through dissemination and utilisation of data by users.

METHODOLOGY STATEMENT

HESA has published a two part [Graduate Outcomes methodology statement](#). This covers the history and formulation of Graduate Outcomes, together with a detailed account of the implementation and operation of the survey. This will be a 'live' record of the methodology, at least in the first few years as we make improvements to our data collection, processing, analysis, and dissemination policies. We aim to revise it as changes to existing processes are made and new initiatives are introduced.

CONTENT AND PRESENTATION OF DATA IN HESA DATA RELEASES

As part of HESA's [Open Data Strategy](#) we are working to publish as much of our data as possible as open data, using recognised open data licencing approaches to encourage the widest range of re-use. However, the desire to publish as much as possible must be balanced against the risks to confidentiality of individuals for many of HESA's data collections which are drawn from individualised personal data, including the Graduate Outcomes survey. We deploy tried and tested approaches to preserving confidentiality of individuals in our published data releases through use of [statistical disclosure control](#) and appropriate design of data tabulations and visualisations.

Content and presentation of data within HESA published releases is designed to deliver a wide range of data and information to meet the most commonly expressed requirements of users. User requirements are captured through a number of different mechanisms including engagement with user associations, user surveys, user feedback channels via the HESA website and through identification of areas of policy and topical interest.

HESA undertakes regular monitoring of the general and specialist press as well as studying emerging HE policy, regulatory and funding matters in order to inform this work. Our aim is to enable content and presentation of HESA releases to continuously evolve and improve, as well as being sufficiently well-informed and flexible to address topical issues as they arise.

OPERATIONAL POLICY ON TAILORED DATASETS

The [Tailored Datasets service](#) delivers access to specific extracts of data or additional analysis of data that are 'tailored' to the needs of each user. This service acts alongside HESA's range of published materials to enable flexible third-party access to as wide a range of data as possible which can be configured for particular applications.

The Tailored Datasets service is operated by Jisc analysts on behalf of HESA, according to a clear set of procedures and protocols which are defined and monitored by HESA. A trained analyst discusses the data and analytical requirements with each third-party requesting data through the service, and determines the most appropriate data or analysis specification which meets the requirements, whilst also complying with any necessary constraints such as data protection. Data and/or analyses are typically provided under an appropriate data sharing agreement which defines the purposes for which data may be used over a pre-agreed period of time, after which data must be destroyed.

Data delivered through this service will typically include requests by producers of university league tables and guides for prospective students. It is HESA's policy to engage with such users and to enable them to access appropriate data to deliver information resources for prospective students. Use of data in league tables represents a contentious area and HESA's approach is to restrict our advice to objective matters. We will therefore advise on issues such as data quality and coverage and may also comment on data fitness for purpose for a given application. However, HESA does not comment on subjective matters such as selection of measures for use in league tables, nor the ways in which measures are combined to create university rankings. It is HESA policy to utilise our 'Provider Preview' service for HE Providers in respect of all data requests for use in league tables and university guides.

Provider Preview service

This service informs HE providers about third-party requests received by HESA for Tailored Datasets to be used for high-profile and significant purposes. It is used primarily for league tables, press enquiries and the calculation of provider subscriptions, along with a few other examples of high-profile uses. The service provides advance notice of each such request along with a copy of the data requested for each HE provider. This is normally made available two weeks prior to the data being made available to the third-party requester.

It represents an opportunity to understand the range of uses and for the HE provider to supply any supporting or contextual information that is pertinent to the intended onward uses. HESA will forward any such contextual information alongside the data to promote appropriate and well-informed usage.

COMPARISONS WITH DLHE

The former Destinations of Leavers (DLHE) survey was the predecessor to the Graduate Outcomes survey. It operated for many years and, subject to some changes in content and coding frames, provided a good time-series for data users over the period of operation. The Graduate Outcomes survey replaces it and the review process that led to the formation of Graduate Outcomes together with the key differences in the two surveys are explained in detail in part one of the [Graduate Outcomes methodology statement](#).

It is very important that users of data from the Graduate Outcomes survey understand that it is a fundamentally different survey from DLHE, since this has significant implications for those who wish to compare data from the surveys and create time-series analysis.

HESA has taken the decision that we will not undertake, publish or otherwise disseminate any comparisons of data between the Graduate Outcomes survey and the DLHE survey. This is due to the fact that the data from the two surveys are not directly comparable.

HESA advises all data users against attempting to directly compare data between Graduate Outcomes and DLHE. Any such comparisons are likely to generate highly questionable results that are open to misinterpretation.

The two surveys represent a discontinuity in time-series which, although inconvenient for data users, is necessary for the reasons explained in the methodology statement.

The key survey differences which explain the lack of comparability are:

1. Timeframe change - the census point in Graduate Outcomes is at approximately 15 months after gaining qualifications whereas for DLHE the census points were at 6 months and 3.5 years (for the follow-up sample). Graduates surveyed are therefore at fundamentally different stages in their progression after gaining their qualification(s) between the surveys.
2. Centralisation - the contact centre, fieldwork and coding of Standard Industrial Classification (SIC) and Standard Occupational Classification (SOC) are undertaken centrally, no longer by the provider, allowing for greater consistency.
3. The survey instruments (questionnaires) used for DLHE and Graduate Outcomes are **different by design**. The set of questions established for Graduate Outcomes were designed to address similar themes to DLHE but in many cases are worded slightly differently or cover more detail. An example is that there is much more detail on graduates

who are self-employed or running their own business in Graduate Outcomes than was the case in DLHE, with greater insight into the inter-relationships between self-employment and other activities. Graduate Outcomes questions are designed to provide information about graduates' experiences that is relevant in the current graduate labour market, as compared with DLHE questions which were designed to address the labour market and associated data needs many years ago. Lack of direct comparability between the surveys was accepted as necessary in order to achieve this and information requirements for Graduate Outcomes were defined independently of DLHE. Comparability of graduates' perceptions and responses to questions that are worded differently between DLHE and Graduate Outcomes were not cognitively tested because this was not a design criterion for Graduate Outcomes. Data users therefore cannot assume that graduates' responses to such questions are comparable between the surveys.

4. Graduate Outcomes includes a range of new questions such as 'Graduate Voice' questions which provide a novel understanding of success in outcomes as judged by graduates themselves. There are new questions to understand the work and study pathways graduates may have taken between gaining their original qualifications and the survey reference week. Banks of opt-in questions give providers the opportunity to ask graduates additional questions.

KEY DATA CONCEPTS AND STANDARDS

INTRODUCTION

HESA has long been a strong advocate for data and metadata standards. These help us to improve consistency across disparate processes or sources, to reduce the costs of integration between sources, and to manage data quality. They help our users through stable and predictable approaches that are applied across different years and by supporting harmonised definitions between different datasets.

There is a longstanding practice at HESA to utilise existing national and international data standards where possible and appropriate in all of our data collection and dissemination activities. Graduate Outcomes is no exception and we aim to apply externally defined standards where we can. However, HESA data tends to cover specific areas where existing data standards may not exist or where existing standards may be unsuitable for a range of reasons (e.g. sometimes different data standards may exist for different industrial sectors but have poor applicability to the HE sector).

In such cases, HESA may establish its own approach by creating a data definition or derivation. Of course, that doesn't automatically result in a new data standard, since data standards only result from widespread adoption of a definition – but often we have seen HESA definitions being adopted for a range of purposes.

Classes of usage

In the published [methodology statement part one](#), we explain the wide range of expected uses of Graduate Outcomes data. These will necessitate different sub-sets of survey data, 'cut' and presented in different ways. There are various ways to categorise the range of uses but one categorisation in particular influences the derivation of data definitions or standards – that is between uses for broadly descriptive purposes and uses for regulatory or performance assessment applications:

- By 'descriptive purposes', we mean uses where the primary aim is to understand and generate insight into the experiences of graduates after they gain qualifications.
- By 'regulatory and performance uses' we refer to applications of data to assess the characteristics and performance of HE providers in delivering on national policy imperatives or providing high quality education for their students.

It is reasonable to expect that in some cases, choices of data definitions for descriptive statistical uses may differ from those used for regulatory and performance assessment applications – a definition that may be helpful in explaining the experiences of graduates may not provide a fair and reasonable basis for assessing HE provider performance.

An example of this may be in the use of survey questions about 'main activity' versus 'all activities' when assessing percentages of graduates entering 'highly-skilled' jobs. From a descriptive statistics perspective it is interesting for users to be able to understand the job types of graduates in all circumstances but as a measure of HE provider performance in preparing graduates with the skills to undertake specific professions, it may (arguably) be fairer to only consider the jobs that graduates report as their 'main activity'.

Recognising the differences between these main categories of use, HESA has choices to make about how we approach data definitions and presentation of our statistics. One choice could be to narrowly prescribe definitions driven primarily by expected uses in regulatory and performance measures – accepting that the external organisations that are largely responsible for such measures may choose not to adopt HESA definitions in any case.

The alternative approach, and our preference, is to structure definitions and data within our main statistical products in flexible ways that are capable of being used for a wide range of purposes.

We believe that users of our data products should be empowered to make informed choices about how they utilise the data we publish. We give them *choices* through use of interactive and dynamic data publication (e.g. use of data filters on tabulations and charts) and through services supported by highly trained analysts. We ensure they are *informed* to make the right choices through the delivery of comprehensive explanation of data and definitions, along with information to help them understand the options they have in utilising the data.

The following sections explain some of the key data definitions and approaches that HESA plans to take in use of Graduate Outcomes data.

RESPONSE RATES AND INCOMPLETE RESPONSES

It is important to understand that no single graduate would be expected to answer all available survey questions given that a routing structure directs respondents to particular sets of questions that are most relevant to their circumstances (a survey routing diagram is provided on the [Survey Results coding manual](#) for the appropriate collection year). In addition, some questions are optional to answer. The survey has been designed with a target minimum number of core questions for different types of graduate respondents that result in designation as a 'survey completed' response. A proportion of graduate respondents have not answered sufficient core questions to meet that standard and have instead chosen to answer varying subsets of those questions. These are designated with a status of 'partially completed'.

For more details on this, please refer to the [methodology statement](#) or learn more about how we encourage graduates to fully complete the survey on a blog on the HESA website: '[From partial to complete](#)'.

HESA statistical releases and other forms of dissemination will use data from responses marked as 'survey completed' and 'partially completed'.

In the latter case, responses will be included in dissemination if they satisfy the criteria to enable derived activity to be calculated. i.e. graduates who have completed the first two questions of the survey (all activities in the census week or 'ALLACT' and most important activity in the census week or 'MIMPACT'). Graduates' responses which do not address these two core questions will be considered non-responses.

HESA believes that response rates provide important contextual information for users of statistics derived from the Graduate Outcomes survey. Although they do not provide a reliable measure of data quality alone, they are useful alongside other quality measures such as sample sizes and representativeness when compared with population characteristics. As such, HESA will publish response rates within the statistical releases, calculated for a range of demographic variables and by HE provider. These response rates will be published separately as two different rates to cover 'survey completed' responses only and the combination of 'survey completed' and partially completed' responses.

It is important to note that, since the Graduate Outcomes survey is collected centrally from graduates and not via HE providers, response rates by provider do not represent any kind of performance measure for those HE providers. Such response rates, together with other descriptive statistics about the survey, merely enable users to understand quality and coverage of Graduate Outcomes data and thereby make informed choices about how to utilise it.

WEIGHTING OF SURVEY RESPONSES

Our findings on weighting to date have led us to conclude that there is no compelling case for the application of weighting to the first year of Graduate Outcomes data at this time. Information regarding the weighting methodologies that we have explored and the findings from the work that led us to this conclusion are described in summary in the 'data analysis' section of the [methodology statement](#) and described in full within a [technical paper](#).

This does not preclude the introduction of weighting in future and it is possible that a more nuanced approach could be developed that would have some value in improving the quality of estimates derived from the Graduate Outcomes data. We are therefore planning a further programme of analysis to be undertaken over the coming months up to the point at which year two Graduate Outcomes data becomes available. More details of this are also available from the methodology statement and the technical paper.

GRADUATE ACTIVITY

The Graduate Outcomes survey instrument (questionnaire) invites respondents to select their 'activities' within the census week in two main ways which recognise the changing nature of work in modern societies. These support the recording of multiple simultaneous activities, e.g. combinations of different types of work alongside study. The first activity question (ALLACT) asks respondents to indicate the full range of relevant activities they were involved in. The second

question (MIMPACT) asks respondents to indicate which one of these activities they consider to be most important. 'Most important' in this context is intentionally not strictly defined but guidance is provided to suggest that the most important activity might be the one which is most related to future plans, the one which pays the most money or the one that the respondent spends the most time doing. It is for each graduate to determine what 'most important' means to them.

For convenience within HESA data releases, it is helpful for general purposes to define a summary list of activities that condenses the results of the two questions into a single set of categories. This is expressed as a 'derived field' (entitled XACTIVITY) which generates the summary categories using an algorithm which is determined by HESA. The derivation of this has required careful consideration and consultation with the Graduate Outcomes Steering Group and other HE sector bodies.

The principles that underpin this derivation are as follows.

1. The concept of an overall summary activity categorisation is one that is very commonly used in data dissemination and the approach is similar to one used for DLHE data, with which experienced users of graduate statistics will be familiar. It is intended to provide a general purpose and convenient categorisation but is not intended to meet the requirements of every application of data. For many applications it is expected that users will wish to utilise the source fields of most important activity and all activity, and indeed HESA will be doing that for some data tabulations and visualisations in our statistical releases.
2. The derivation of the summary activity is not designed to convey or imply any value-judgements about what may subjectively be 'better' or 'poorer' outcomes. A good example of this is in paid versus voluntary or unpaid work. There are many cases where voluntary or unpaid jobs can be helpful stepping stones into successful (paid) careers or where voluntary jobs are considered by graduates to represent excellent outcomes.
3. The summary activity is not intended to form any standard definition of activities to be used for all other uses, such as those related to regulation or HE provider performance assessment. Different treatments of activities may be more appropriate for such uses.
4. The selection and definition of categories within the summary activity is designed to provide good flexibility for data users to select single categories or to combine groups of categories. However, as stated above there will remain applications that require use of the source activity fields.
5. In most cases the summary activity field respects the choices made by graduates in determining what their 'most important' activity is. However, in a small number of cases there are compelling grounds to consider the choice of most important activity to be anomalous in the context of the wider range of activities reported. An example of this is when a graduate reports unemployment as their most important activity when they simultaneously report paid employment one of their other activities. In such cases the summary activity field 'overrides' the most important activity in favour of the relevant other activity (in this example paid employment). This approach is intended to align with national and international standards on definition of work and employment described below.

EMPLOYMENT AND WORK

In accordance with HESA's stated desire to align with national and international data standards where appropriate, we have referred to existing standard definitions of work and employment. Within the UK the Office for National Statistics use standard definitions for surveys such as the

Labour Force Survey and publish a [Guide to Labour Market Statistics](#). These standards largely align with international standards set by the [International Labour Organisation \(ILO\)](#).

The ILO define a category of 'work' as "any activity performed by persons of any sex and age to produce goods or to provide services for use by others or for own use"² and a sub-category of 'employment' defined as "work performed for others in exchange for pay or profit"³, which clearly excludes voluntary or unpaid work. HESA will apply these definitions to statistics published and disseminated from the Graduate Outcomes survey and they have informed the derivation of the summary activity field described above.

Many of the published and disseminated tabulations and visualisations of Graduate Outcomes survey data will focus on those graduates in work or employment – for example, when reporting on the Standard Industrial Classifications or Standard Occupational Classifications of jobs held or salary levels. Consequently, decisions are required as to the use of work versus employment in such cases i.e. whether voluntary and unpaid jobs are included. Additionally, the manner in which work or employment has been recorded within the most important activity or as part of the wider range of all activities needs consideration. It is likely that different applications will require different definitional approaches. For example, data on mean salaries would clearly need to be restricted to those in paid employment.

Whether or not the work in question was considered by the graduate to be their most important activity may also have an impact on how resulting data are used e.g. if a graduate is undertaking a temporary job simply to augment income whilst actively looking for a job that is more appropriate to his or her qualifications, it may or may not be appropriate for data users to consider the characteristics of that temporary job.

HESA believes that, given the range of approaches which may be appropriate to different applications, we should publish statistics in a manner which provides flexibility for users, and users should be able to make informed choices as to the approaches they wish to utilise. In practice this means that in most cases, tabulations and visualisations published by HESA concerning work or employment will be based on the widest population of graduates undertaking any type of work, within any one of their activities. These tabulations and visualisations will provide users with interactive filter options that will allow them to further restrict their analyses to those for whom work was their most important activity and, independently, to choose whether to include or exclude those graduates in voluntary or unpaid work.

ACTIVITY BETWEEN QUALIFYING AND SURVEY CENSUS WEEK

The structure of the survey allows us to report on activity undertaken during the 15-month period between a graduate qualifying from HE and the survey census week. Those graduates who, for example, have undertaken further study at postgraduate level (such as a one year master's or PGCE course) between graduating and the census week may be at very different points in their career and had different experiences compared to graduates who have been undertaking different activities such as employment during the 15 months. Similarly, graduates who have been engaged

² http://www.ilo.ch/wcmsp5/groups/public/---dgreports/---stat/documents/normativeinstrument/wcms_230304.pdf, paragraph 6.

³ http://www.ilo.ch/wcmsp5/groups/public/---dgreports/---stat/documents/normativeinstrument/wcms_230304.pdf, paragraph 7.

in further study at postgraduate level and who are currently in employment may have different salary expectations than their peers who went into employment directly after graduating.

As with other categorisations of data such as work and employment, decisions as to whether it is appropriate for graduates who have been employed for the 15-month period (or undertaking other non-study activities) to be compared with graduates who have spent most of the 15 months in full-time study are dependent on the purpose of the statistics. Once again, HESA aims to provide users with the flexibility to make informed choices about what is appropriate in their context. We will therefore introduce optional interactive filters to relevant published tabulations and visualisations so that users can choose to include or exclude graduates who are likely to have spent most of the 15 months undertaking study.

Given the data that the survey provides on this, HESA has taken the decision to interpret this characteristic as those graduates who have undertaken study during the 15 months on at least one full-time course at one of the following levels of study:

- Higher degree, mainly by research (e.g. PhD, DPhil, MPhil, MRes).
- Higher degree, mainly by taught course (e.g. MA, MSc, MBA).
- Postgraduate diploma or certificate (including PGCE / PGDE).
- Professional qualification.
- Undergraduate degree (including integrated master's degrees) (e.g. BA, BSc, MBChB, MEng).
- Other undergraduate diploma or certificate not specified above.

SALARY DATA

As with previous presentations of graduate salary data, HESA expects to show data only for graduates reporting themselves as in full-time paid UK employment where the currency paid is British pounds.

Preliminary analysis of the data on salaries submitted by respondents in the first year of the Graduate Outcomes survey reveals a small number of salary outliers which are suggestive of data quality issues, such as misinterpretation of the salary question. Whilst HESA will wish to address this in future years of the survey through mechanisms such as enhanced guidance and quality rules, decisions must be made on the treatment of salary outliers for dissemination of the first year of survey data. In assessing what may be reasonable, HESA has taken the decision to exclude outlier salaries that fall below the UK national minimum wage equivalent.

At the opposite end of the salary range we see a small proportion of very high salaries reported which seem implausible. Statistical analysis of the data suggests that if the top 1.5% of reported salaries are excluded the remaining data would more closely fit a 'normal' statistical distribution (which would be the usual expectation for data such as this drawn from a very large sample). HESA has therefore concluded that it would be appropriate to exclude the top 1.5% of salaries as outliers. All published and disseminated salary data will be based on reported salaries excluding these outliers.

HESA statistical releases will show numbers of graduates by salary bands which start at the national minimum wage equivalent and are divided into £3,000 divisions within the most common range of graduate salaries. Published salary data will also show medians and upper and lower quartile values.

SUBJECTIVE WELLBEING

We have explored the survey data on subjective wellbeing from the first year of Graduate Outcomes and have researched how other organisations use and report on this data. The results of this research have aided us in determining our choice of breakdowns and aggregations within outputs, which align with the approaches used by the Office for National Statistics for similar survey data. Whilst we expect to publish summary information on subjective wellbeing at national level, we are not expecting to publish or release this information by individual HE providers during our first year of dissemination. A later assessment of data quality will help us to decide whether provider level statistics could appropriately be disseminated in future years of the survey.

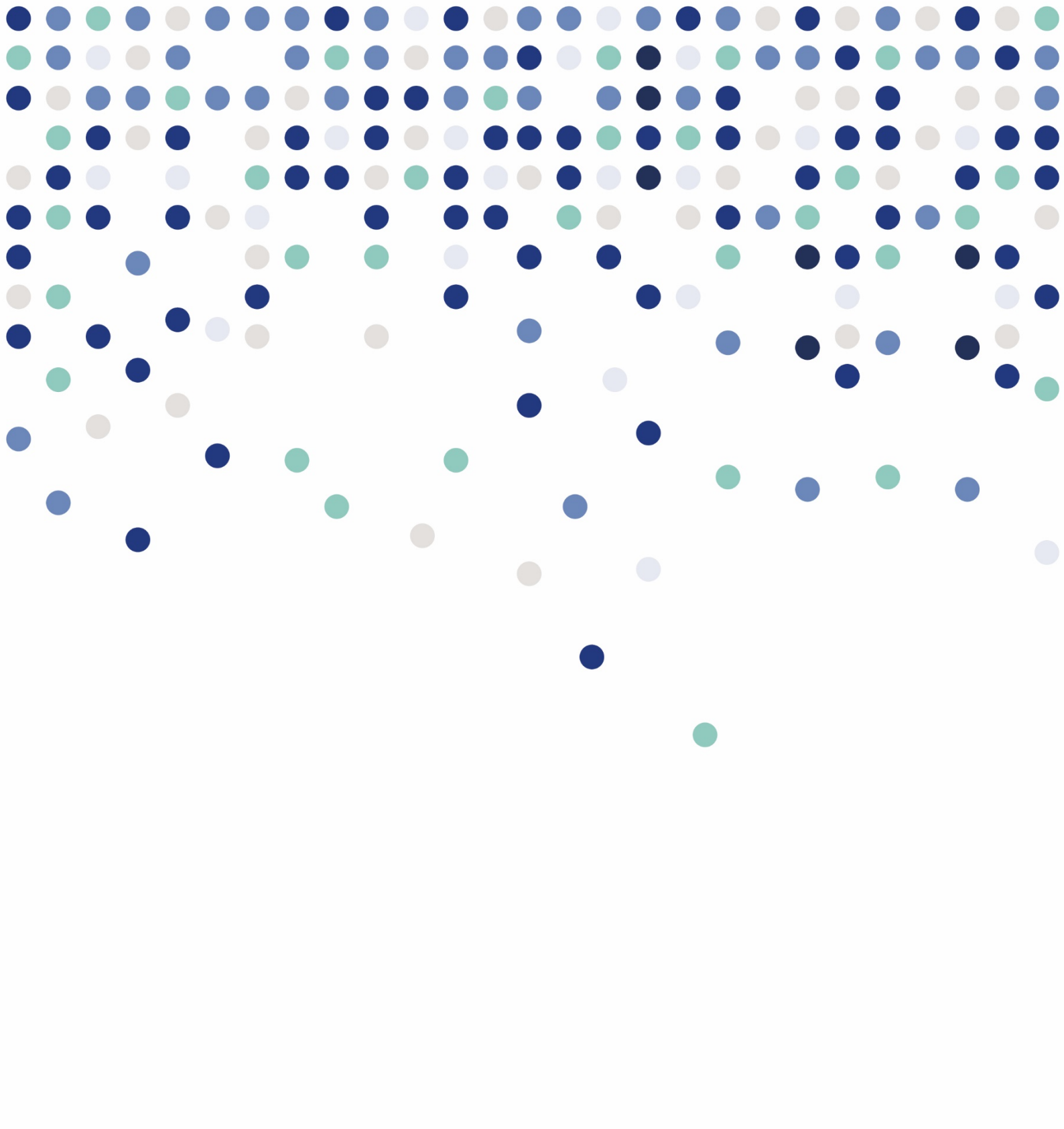
OPT-IN QUESTION BANKS

Opt-in question banks are questions that may be asked in addition to the core question bank contained within the Graduate Outcomes survey. They are selected from an approved set of questions (approved by the Graduate Outcomes Steering Group) by the HE provider or HE regulatory/funding bodies and are asked of the entire relevant population.

For the first year of outputs, we do not intend to present any data on opt-in question banks as take-up of these questions by HE providers was low. We will review this position for future years as we are able to determine the level of take-up and proportion of responses received.

FURTHER DATA STANDARDS AND DERIVATIONS

The above list of data standards and derivations address the areas on which HESA receives frequent queries from HE sector stakeholders. Full explanations of a range of additional derivations and data standards will be published as supporting information alongside the statistical releases.



Official Statistics

HESA

95 Promenade

Cheltenham

GL50 1HZ

E official.statistics@hesa.ac.uk

T +44 (0) 1242 388 513

W www.hesa.ac.uk

