Consultation

Office for Students

Consultation on Data Futures and data collection

This consultation runs from 16 December 2021 to 3 February 2022.

The Office for Students is the independent regulator for higher education in England. We aim to ensure that every student, whatever their background, has a fulfilling experience of higher education that enriches their lives and careers.

Our four regulatory objectives

All students, from all backgrounds, and with the ability and desire to undertake higher education:

- are supported to access, succeed in, and progress from, higher education

- receive a high quality academic experience, and their interests are protected while they study or in the event of provider, campus or course closure

- are able to progress into employment or further study, and their qualifications hold their value over time

- receive value for money.

Contents

About this consultation

The Office for Students is consulting on how we can ensure that we have the necessary data to perform our functions while ensuring that our data requirements from registered higher education providers are proportionate. In particular we are seeking views on the approach to the collection of in-year individualised student data as part of the Data Futures programme. We would like to hear your views on the proposals in this consultation.

Timing	Start: **16 December 2021** End: **3 February 2022**
Who should respond?	Anyone with an interest in the data collected about higher education providers and their students. We are particularly interested in views from individuals responsible for managing higher education data and providing it to the designated data body and the Office for Students.[1] We are also interested in the views of individuals who make use of the data provided by the designated data body under section 65 of the Higher Education and Research Act 2017.
How to respond	Please respond by **Thursday 3 February 2022**. Use the online response form available at https://survey.officeforstudents.org.uk/s/datafuturesdatacollection/
How we will treat your response	We will summarise or publish the responses to this consultation on the Office for Students website (and in alternative formats on request). This may include a list of the providers and organisations that respond, but not personal data such as individuals' names, addresses or other contact details. If you want the information that you provide to be treated as confidential, please tell us, but be aware that we cannot guarantee confidentiality in all circumstances. We will not regard an automatic confidentiality disclaimer generated by your IT system as a confidentiality request.

---

[1] The designated data body performs the duties set out in sections 64 and 65 of the Higher Education and Research Act 2017, including data collection, data processing, data storage, data publication and provision. This body is designated by the Secretary of State following consultation and a recommendation from the OfS. The Higher Education Statistics Agency was selected as the designated data body in 2018.

The Office for Students will process any personal data received in accordance with all applicable data protection laws. We may also share responses with the Higher Education Statistics Agency as the designated data body, given the impact of these proposals on its work. A privacy notice for this consultation is available to view on our website.[2]

We may need to disclose or publish information that you provide in performing our functions, or to disclose it to other organisations for the purposes of their functions. Information (including personal data) may also need to be disclosed in accordance with UK legislation (such as the Freedom of Information Act 2000, Data Protection Act 2018, and Environmental Information Regulations 2004).

Next steps

We will publish or summarise the responses to this consultation in spring 2022. We will explain how and why we have arrived at our decisions, and how we have addressed any concerns raised by respondents. We will then set out next steps in the policy and implementation process.

Enquiries

Email dfa@officeforstudents.org.uk

Alternatively, call our public enquiry line on 0117 931 7317.

For more information about our approach to data, please visit the OfS website.[3]

---

[2] Available at www.officeforstudents.org.uk/ofs-privacy/.

[3] See www.officeforstudents.org.uk/publications/office-for-students-data-strategy-2018-to-2021/.

Introduction

1. This consultation seeks views on how the Office for Students (OfS) can ensure it has the data required to perform its functions and how that data can be collected in a way that is proportionate. By 'proportionate', we mean that the requirements that we place on higher education providers are necessary to achieve our regulatory aims.

2. The OfS's regulatory framework sets out our broad aim to use reliable data to underpin our regulatory functions.[4] We consider the regular collection and analysis of standardised data a proportionate way to support the delivery of many of our regulatory activities. We have set out, in Annex B, the activities that currently rely on the use of regular student data and how we consider that they could be improved in their effectiveness or efficiency by collecting individualised student data within the academic year. The proposals in this consultation are thus a proportionate response to the improved uses of data identified in Annex B, which will enable us to better achieve our regulatory aims.

3. The focus of this consultation is how the Higher Education Statistics Agency (HESA), as the designated data body in England, takes forward the Data Futures programme of work designed to provide in-year student data.[5] In particular it is on the nature, frequency and timing of data collections by HESA to enable us to better achieve our regulatory aims. We are also consulting on a number of other areas we have identified where the amount of data collected by HESA or the OfS could be reduced with limited impact on the performance of the OfS's functions.

4. We are not consulting on whether HESA should, as now, collect individualised student data at least once a year. We are also not consulting on the approach to the collection of data from further education colleges and sixth form colleges, as this data is already collected in-year by the Education and Skills Funding Agency.

5. In September 2020, we decided to review the approach to Data Futures that HESA was proposing to take, as part of a wider review of regulatory burden.[6] This reflects the OfS's commitment to ensuring that the requirements we place on providers are proportionate. It also acknowledges the concerns of some providers that the increased frequency of data collection proposed in the Data Futures programme could result in an unjustified increase in burden.

6. We have delayed undertaking this work, as our approach to the frequency of data collection is to some extent dependent on the nature of the data we require for several of our functions, including:

    *   our approach to the regulation of quality and standards

---

[4] See paragraphs 28 to 32 of our regulatory framework at www.officeforstudents.org.uk/publications/securing-student-success-regulatory-framework-for-higher-education-in-england/.

[5] See https://www.hesa.ac.uk/innovation/data-futures.

[6] See 'Strategic guidance to the Office for Students: Additional teaching grant and funding/reducing the bureaucratic burden on providers', available at www.officeforstudents.org.uk/advice-and-guidance/regulation/guidance-from-government/.

- the Teaching Excellence and Student Outcomes Framework (TEF)

- the National Student Survey (NSS).

7. We expect to consult on aspects of these issues while this consultation is open.

8. We are consulting now as we need to decide on an approach for the 2023-24 academic year to allow providers and HESA time to prepare. We are also aware that Scotland and Wales plan to move to the collection of in-year data from 2023-24. Therefore, aligning changes is likely to reduce costs for HESA and the firms that provide software to providers, which is likely to reduce the costs to providers in turn.

9. The proposals in this consultation primarily relate to the information duties under the Higher Education and Research Act 2017 (HERA) (section 64), and parts 2 and 3 of this consultation represent a consultation under section 64(6). In addition, the OfS collects data itself, and several proposals in this consultation will affect the data the OfS collects directly.

10. The consultation questions are listed in full in Annex E.

11. We are running this consultation in accordance with the government's consultation principles.[7] Annex F sets out the matters to which we have had regard in developing these proposals. These include our general duties under HERA, the Regulators' Code, the Code of Practice for Statistics, and the Public Sector Equality Duty.

---

[7] See https://www.gov.uk/government/publications/consultation-principles-guidance.

Background

12. The OfS is an evidence-led regulator. This means we use data and information to inform our effective, efficient, and intelligent regulation in the student interest. Data plays a critical part in our regulation and in the delivery of what students, the public and the government expect of the OfS. Good use of data allows us to anticipate where there is a risk that our regulatory objectives may not be delivered, and helps us mitigate those risks.[8]

13. To be an effective regulator, we require up-to-date data so we can anticipate challenges to the higher education sector or to individual providers and take timely regulatory action. Much of the available data about higher education providers and their activities is at least a year out of date.

14. This consultation is not seeking views on whether HESA should collect in-year student data. We consider this to be a given: in-year data equips us with timely information for effective regulation in the student interest. Rather, our proposals explore ways this data might be collected, and the advantages and disadvantages of each approach.

15. We are committed to meeting our need for in-year student data in a way that is proportionate and that delivers the widest benefits to the OfS and providers. We recognise that both the frequency and timing of data collection affect the burden on providers, and we are mindful that any benefits to collecting in-year data should be proportionate.

16. This consultation sets out how we propose to have regard to the Minister of State for Universities' statutory guidance in reviewing the proposed termly collection of data.[9] Through our proposals we aim to ensure that the burden of in-year student data collection is proportionate to its use in regulation. We have proposed some changes to the approach to collecting data and the removal of a small number of data items from the student record. We also aim to explore new ways to make further use of linked and third-party data.

17. In responding to this consultation, we are particularly interested in any unintended consequences of the proposed approaches and their effect on:

- particular types of courses and on course design and delivery

- particular types of students, including those who share protected characteristics.

Scope and structure of consultation

18. This consultation has three parts.

    Part one: Approach to in-year student data collection. In this section we set out the background to the Data Futures programme. We describe the current higher education data landscape and methods of student data collection. We propose three alternative approaches to

---

[8] The OfS's regulatory objectives are at www.officeforstudents.org.uk/advice-and-guidance/regulation/the-regulatory-framework-for-higher-education-in-england/.

[9] See 'Strategic guidance to the Office for Students: Additional teaching grant and

funding/reducing the bureaucratic burden on providers', available at www.officeforstudents.org.uk/advice-and-guidance/regulation/guidance-from-government/.

collecting individualised student data, all of which involve returning some individualised student data during the academic year to reduce the delay between the majority of students starting their courses and having data on them. We also include the current Data Futures approach as a fourth, discounted proposal.

19. **Part two: Other changes to data collection.** In this section we set out a number of proposals where we believe improvements to efficiency can be made, including the removal of a small number of student and staff data items that we no longer have a compelling reason to collect, possible clarification of data quality expectations, and changes to provider profile data.

20. **Part three: Use of linked and third-party data.** In this section we set out our proposals to continue to make use of data linking between individualised student data and data from third parties, and to continue to explore the potential of applying data science techniques to data collected through social media and other online sources.

21. There are a number of annexes:

- Annex A: Methods of student data collection

- Annex B: Value and potential uses of individualised in-year data

- Annex C: Roundtable discussions

- Annex D: Summary of possible approaches to collecting individualised student data

- Annex E: List of consultation questions

- Annex F: Matters to which we have had regard in developing our proposals.

- Annex G: Abbreviations

Other data reviews

22. On 30 November 2021 we published the outcomes of the KPMG review of the Transparent Approach to Costing on behalf of the UK Higher Education Regulators and Funders (Financial Sustainability) Group.[10] We are currently considering the recommendations of that work. We are also working with UK Research and Innovation, HESA and bodies in the devolved administrations to review the Higher Education Business and Community Interaction (HE-BCI) survey to ensure that it meets our needs and is proportionate. We are therefore not explicitly considering these data returns in this consultation.

23. In developing this consultation, we have considered alternative options for securing our objectives. We have identified one such option and set out the reasons why we do not propose to take it forward in paragraphs 81 to 89.

Consultation proposals and questions

---

[10] See www.officeforstudents.org.uk/publications/review-of-trac-2021/.

Part one: Approach to in-year student data collection

Background and current higher education data landscape

24. Currently the higher education data landscape in England for most providers, other than further education colleges, consists of an in-year aggregate data return (the Higher Education Students Early Statistics (HESES) survey), and for some the medical and dental student survey, as well as a comprehensive year-end collection of individualised student data by HESA.

25. Aggregate data returns collect the total numbers of students in defined groups without linking them to details about each individual student. Individualised data returns, as collected in the HESA student record, link the data to details about each individual student, their course, and their outcomes. Annex A provides further information about the advantages and disadvantages of aggregate and individualised data collections.

26. These data collections are supplemented by administrative data from bodies such as the Student Loans Company (SLC) and UCAS. In addition to these data collections, the OfS runs a process to add and remove students from the NSS target list. This process is necessary as the initial target list is generated using data from the previous academic year, which will not correctly capture students whose circumstances change, for instance because they repeat a year or choose not to take a planned sandwich year.

27. As set out in the current OfS data strategy, we aim to use individualised data because it is inherently flexible and allows for much richer analysis, for example enabling understanding of the intersectional nature of disadvantage.[11] It is also easier to quality assure than aggregate student data. This is because individual data items clearly relate back to individual students, can be checked for internal consistency, and are typically subject to less manipulation, which can lead to error. Our experience of reconciling HESES to HESA data is the most significant issues are more often found in HESES.

28. HESA currently collects individualised student data retrospectively at the end of an academic year, with data made available to the OfS and other statutory customers in December. This means that data is initially available 14 months after most students start their academic year.[12]

29. The Data Futures programme, led by HESA, emerged in 2015 from the Higher Education Data and Information Improvement Programme, and predates the OfS. When the blueprint for the new data landscape was drawn up in 2015 there were 97 different organisations collecting over 520 separate data returns, the vast majority of these relating to students.

30. The most cited reason why organisations collected their own data, rather than using the HESA individualised student data, was timeliness. Higher education is the only educational stage in England for which learner level data is collected only once per year. Schools' data is collected on a termly basis and further education data is collected from colleges up to 14 times per year.

---

[11] For our strategy see www.officeforstudents.org.uk/publications/office-for-students-data-strategy-2018-to-2021/, and for our work on associations between characteristics www.officeforstudents.org.uk/data-and-analysis/associations-between-characteristics-of-students/.

[12] Over 70 per cent of the student intake occurs in late September and October.

The Data Futures programme proposed replacing many of the aggregate data returns with more frequent collection, assurance, and dissemination of individualised student data.

31. Data Futures was originally envisaged by the higher education sector as a burden reduction exercise in terms of the volume of data collected and the rationalisation of data collections. The proposition was that providers would not need to maintain data relationships with multiple organisations, nor would they need to submit different sets of data, with different definitions, at different times. Organisations, including professional, statutory, and regulatory bodies (PSRBs), requiring regular higher education data collections to fulfil their business requirements, would benefit from the collection of in-year data which could replace their own collections, removing burden on them and providers. In addition to meeting existing data needs, PSRBs would be able to use the wider HESA data collection to benchmark accredited provision in their professional area or have access, subject to privacy considerations, to additional course or demographic details where these were already collected.

32. HESA's experience in designing the Data Futures model, and feedback from providers, is that adapting the Data Futures programme to meet the needs of a wide range of stakeholders is likely to add considerably to the burden of data collection. We therefore conclude that in-year collection of individualised student data through Data Futures needs to be justified predominantly on the basis of the needs of the current users of HESA data, rather than any potential future uses by PSRBs.

33. The need for more timely data collection, which prompted the initial investment in the Data Futures programme, is as relevant now as it was in 2015. However, the context has now changed. HERA, and the OfS's regulatory framework, have put in place a new regulatory approach. The sector has expanded, with a broader range of registered providers now offering higher education in England. The economic environment has also changed, particularly as a consequence of the pandemic, with some providers facing considerable financial and other pressures. In addition, the OfS is committed to ensuring that the regulatory burden that it places on providers is proportionate and risk-based, and is committed to reducing regulatory burden where appropriate.

34. Annex B sets out ways in which in-year individualised student data could be used to improve the effectiveness of our regulatory activities.

35. The current Data Futures programme sets out an approach by which HESA would make three discrete collections of in-year individualised student data from all providers via a new data collection platform. This has led to concerns from some providers that an increase in the frequency of data collections from the existing annual HESA student record could result in a substantial increase in administrative burden.

36. Given these concerns the OfS decided in September 2020 to review the Data Futures proposals, to ensure that they are still fit for purpose and are not disproportionate.

37. Reporting student data earlier, as proposed in the current Data Futures programme, does not change the need to collect that data and assure its quality, both of which should be one-off tasks. It does, however, constrain when those activities can take place, and this may increase burden because the work cannot be scheduled efficiently. We appreciate that returning data earlier may require significant systems and process improvements at providers. However,

some providers have told us that they expect these improvements may lead to longer-term resource savings, as cumbersome and manual processes are replaced with greater automation.

Data collected on the Individualised Learner Record

38. Our current data strategy sets out our intention to use the Individualised Learner Record (ILR) already collected by the Education and Skills Funding Agency to collect data on students studying at further education colleges and sixth form colleges. This approach reduces the burden on these providers.

39. However, this approach is not without challenges, as it requires additional work for us and HESA in combining data from both sources. This approach also requires a compromise as it is necessary to work within the ILR data model, primarily designed to support the funding and regulation of further education courses which are typically shorter in cycle and more structured. While seeking in general to work within the confines of the ILR, we need to include within it a small number of fields to collect data that is not otherwise collected, for example data on student full-time equivalent (FTE). We seek to keep such additional data items to a minimum and to use native ILR data wherever possible.

40. We recognise that for some further education colleges with large numbers of higher education students, there may be a benefit in making HESA student returns so that their data is fully comparable. However, we do not consider that the advantages of having data collected against the HESA specification outweigh the additional costs to colleges. This is especially the case as we seek to ensure that the HESA and ILR data are compatible and that data from the ILR is reported alongside that from HESA.

41. We are aware that providers that offer higher apprenticeships currently need to return these students to both the HESA and ILR records, reflecting the joint regulatory and funding responsibility for these students. We have considered whether it would be possible for us to use the data collected on the ILR for these students in the same way that we propose to do for further education colleges. Our current view is that regulating a single provider based on data in two student records is likely to increase burden for providers as they will need to understand how we are using both sets of data. We therefore intend to retain apprenticeships in the HESA record and to use HESA data in preference to ILR where both are available.

Roundtable discussions

42. In July 2021 we held a series of roundtable discussions with stakeholders to help us understand more about the relative burden of different types of higher education student data collections and different activities required to complete a student data return. Information about these discussions is provided in Annex C.

Our proposals

43. We are proposing three alternative approaches that we could take to collecting individualised student data. We think there are considerable regulatory benefits to significantly reducing the delay between students starting their courses and having data on those students. We have therefore focused on options that meet this aim, which inherently means returning some data during the academic year.

44. As all proposals include an element of in-year individualised student data that will deliver the benefits set out in in Annex B, we have only drawn out particular benefits or disadvantages where there are significant differences between the approaches.

45. Approach 1 is our preferred approach, for the reasons set out in our summary of its benefits. For completeness, we have also included the current Data Futures model as a fourth, discounted, approach, as we originally considered it as an alternative proposal.

46. Annex D provides a summary of the four different approaches.

Approach 1: Two individualised student data collections a year with reduced data requirements in the first data return

Summary of approach

47. This approach would reduce the frequency of collecting individualised data from the current Data Futures approach of three times a year to twice a year.

48. The first in-year data return would have a reduced number of data items, with a second, more comprehensive data collection at the end of the academic year.

49. The timing of the first data return would need to be carefully considered, but our expectation is that it would capture activity between 1 August and mid-November, although the precise dates would need careful consideration and may change. We have assumed that the timing of the in-year return would be the same for each provider, as varying the timing would undermine any uses that required comparability between providers. We would expect the second return to cover activity between mid-November and 31 July, with data submitted over the summer and early autumn.

50. Having differential requirements for in-year and end-of-year returns could create additional complexity. This would also create some redundancy, as all data would need to be returned at the end of the year, even for students who had completed their activity by the mid-November cut-off. To avoid this issue, we are proposing that we would use the student course session concept within the Data Futures model. This would require data at the end of the student's year rather than the end of the reporting year, thereby ensuring that students only need to be returned in a period where they are active. This would avoid the need for different specifications for the two returns while in practice reducing the data requirements in the first return for the majority of students.

51. Part two sets out our proposals for changes to the content of the individualised data. Subject to decisions about these, we would expect the in-year return to include the following subset of data from the year end return:

- student course data

- student names and characteristics

- study start and end dates and reasons for leaving

- qualifications awarded

- fundability data

- mode of study and student FTE

- details of subcontractual arrangements.

52. Approach 1 is our preferred approach, as it will allow us to deliver most of the benefits of in-year data identified in Annex B, while allowing providers more time to assure the quality of data that is not required at the start of the year.

Advantages

53. This approach would reduce the number of returns for each provider from the currently planned three returns while retaining most of the immediate benefits of in-year data.

54. The universal collection of data in this way will also tend to increase the value of the data collected as it will allow us, the government, and other stakeholders to have a more timely understanding of the regulated sector. We also expect comprehensive data to deliver increased value to providers.

55. Most providers would see a decrease in other data collection activity with this approach. Making one in-year data return late in the calendar year would mean that we could remove the need for the NSS additions and removals process, while improving its targeting and reporting. This approach may also allow the removal of the HESES return, although this would require wider changes to our funding approach. Focusing the early return on a smaller number of data items that are needed for key purposes, such as funding, would help to further minimise the burden.

56. This approach would mean we were able to understand recruitment within three months, and full-year continuation within 15 months, of a student's autumn start date. For recruitment at other times of the year the time lags would be larger but would not exceed nine and 21 months respectively.

57. This approach allows providers to time their data quality work for fields not required until the end of the student course session to suit them, thus reducing the administrative burden of assuring the quality of certain data items early in the academic year.

Potential disadvantages

58. This approach requires a compromise over the timing of the first return. If the first return was before the end of the calendar year, it would miss the second biggest intake of the year in January and leave a gap of around nine months between updates. However, moving the cut-off date until after the January intake would mean data was not likely to be available until late March or early April, which could hinder some onward uses of the data, for example in funding and to generate a definitive NSS target list. We therefore propose that, if this approach is taken forward, the reference period for the in-year collection would end on the 15 November to align with the current specification for reference period 1, with data signed off by mid-January.

59. There is a small number of providers in the Approved category with no undergraduate students where the increased burden of making two returns is not offset by a reduction in other data collection. However, we consider that the collection of a dataset covering all providers is a

proportionate response as it delivers considerable value, in particular for data to inform student choice and to meet the HERA section 65 requirements. The reduction in the number of returns from the previously planned three returns will benefit these providers.

Approach 2: Cumulative in-year individualised student data collection with potential for differential reporting

Summary of approach

60. This approach cumulatively builds up collection of student data over a year. Figure 1 shows the reference periods covered by each of the collection points under this approach. Every student active in the period would have all their activity for the period returned.

Figure 1: Summary of Approach 2

Collection point	Reference period covered by collection point
1	August to November
2	August to March
3	August to July

61. The cumulative model means that students who complete or leave part way through the academic year would continue to be returned until the end of the academic year (31 July). For example, a student completing a course in October would be returned in all three reference periods under the cumulative model. The final reference period in the cumulative approach would be a complete data return for the entire academic year (1 August to 31 July). The cumulative model should not be confused with the original 2015 Data Futures approach, where returns were built up from a series of changes and additions.

Advantages

62. Whatever pattern of recruitment a provider has, we will be able to understand recruitment within six months and full-year continuation within 18 months, compared with 14 and 26 months respectively with annual end of year collections.

63. This approach allows some flexibility over which of the two in-year returns a provider makes. It would be possible for the OfS not to require one or more of the in-year returns as a provider would return all the data at the end of the year. This flexibility means that it may be possible to have different in-year returns for different providers to match the main dates in their academic calendar. For example, a provider that does not have a January intake may not need to make a return for the second reference period. It would also be possible, once the new in-year collection approach was established, to vary collection frequency, but not timing, to reflect the OfS's view of regulatory risk for a provider.

64. The flexibility of this approach is also likely to make the transition to collecting in-year data easier, as there is less pressure to get data right for the first reference period of the first year. Indeed, it may be possible to use only the second and third reference periods in the first year.

65. This approach may create opportunities to simplify the current data amendments process. In some cases, we may be able to use the data from a later reference period and so not require a correction.

Potential disadvantages

66. This approach would be a more fundamental change to the current proposed Data Futures model and is similar to the approach HESA used for in-year data prior to 2002-03. This change may therefore have a more significant effect on the timetable for implementation and may increase the complexity, and therefore the cost to HESA, of managing differential reporting.

67. This approach would introduce an element of redundancy into the data collection, with some activity returned more than once. For example, unless there was an error that needed correcting, a student who left in September would have the same data returned up to three times. HESA and providers would need to check that there were no unexpected changes between data returns.

68. It would reinforce the primacy of the current academic year structure of 1 August to 31 July. While it would still be possible to form a view across other time periods, this is likely to require a level of assumption or the return of additional data. In practice we expect that for many statistics an annualised cycle of reporting will still be wanted. While there may, in time, be a desire to move away from the current reporting year, given its convenience and long history we consider that the current 1 August to 31 July academic year will retain its importance for many years to come.

69. If all providers are no longer required to make all returns, the uses to which the OfS can put the resulting partial data will be more limited. The wider benefits of the data to providers and others are also likely to be reduced, as it is unlikely that we would publish potentially misleading partial data in-year.

Approach 3: Changing the timing of collection

Summary of approach

70. Given the uses of individualised data identified in Annex B, one way to meet many of the needs for in-year data while retaining a single collection of individualised student data would be to change the timing of the reporting periods, so that instead of covering an academic year (1 August to 31 July) they cover a different period, for example 1 November to 31 October. Under this approach, the data from this single return would be available for use early in the next calendar year.

Advantages

71. This approach would capture the most recent autumn student intake, allowing timely reporting and year-on-year continuation for that group. It might also allow timely reporting of qualifications for many masters' students, where these awards are made before November.

72. This approach means that there would be only one collection of student data per year, which, if the complexity could be managed, might be more proportionate for providers than returning data more frequently.

73. Changing the timing of the reporting periods for single data collections would also resolve one of the tensions of the current data collection timing, which relates to the target population for

Graduate Outcomes cohort D.[13] Cohort D covers students who complete between 1 May and 31 July. Under the proposed timing for Data Futures the final reporting period for each year would see data collected in September. This would potentially miss results for some students who had undertaken resits in September, or had their awards considered by exam boards in the autumn which would need to be added later. This is not an issue if there are multiple collections in the year, but is an issue if there is a single collection in September or early October. This would likely affect the response rate and could result in some students being added to a cohort based on the date of their resit.

Potential disadvantages

74. While this approach is likely to meet the current uses of data, it could not be extended if a need for more in-year data was identified at a later point, as it is fundamentally designed for a single collection.

75. The transition to this approach would be extremely complex and would most likely require the first return of data to cover a 15-month period as we move to the new reporting period.

76. While we think that a single return of data covering a November to October year could meet many of our requirements, we are concerned that there may be hidden complexity for providers. This is because the data return would, for many students, cover two academic or course years. This complexity of the course year and reporting year already exists for students with a year of study that spans 31 July, but for most providers this represents a minority of their students. The increased complexity of the collection to allow reporting against the academic year and reporting years may increase burden more than making multiple returns.

77. This approach would delay reporting of outcomes for most first degree students by about three months.

78. As this approach would mean a fundamental move away from aligning data collections with the traditional concept of the academic year, it might have unintended consequences for some providers. Changing the reporting period is likely to affect comparability with staff and HE-BCI data, which would still be collected in-line with the current academic year. This might also affect comparability with providers' financial returns, which for most providers are currently aligned with the 1 August to 31 July academic year. We note, however, that not all providers adopt this financial year, so this issue already exists for a number of providers. Where a requirement to report data against the current reporting periods remains, this is likely to require collection of additional data items, for example student FTE split according to academic year.

79. At the roundtable discussions we heard that a move to a single November to October collection may require restructuring of data collection systems for some providers; for other providers this may not be an issue. We welcome views on this approach and would be particularly interested in whether there are any unintended consequences of which we should be aware.

80. Although this approach may resolve an issue with the target population for Graduate Outcomes cohort D, it may create a similar issue for cohort A, those students leaving between 1 August and 31 October. Depending on when results are confirmed for these students, and the precise cut-off date for the collection, the results might be missed from the collection immediately after

---

[13] See https://www.hesa.ac.uk/collection/c19071/coverage.

these students complete. Under this approach, the Graduate Outcomes survey window for these students starts before the next collection of student data, meaning any such students would need to be added part way through the survey fieldwork.

Approach 4: Discrete in-year individualised student data collection three times a year

Summary of approach

81. Approach 4 is the current Data Futures approach, which we originally considered as an alternative proposal but have now discounted.

82. The original plan for Data Futures was for HESA to make three discrete collections of in-year individualised student data. Under this model, providers would make three returns in each academic year covering the reference periods August to November, December to March, and April to July. Each return would cover all the students who are active during the period covered by the return, and each would operate independently, with no overlap in the reporting of activity between the returns.

83. In practice, discrete collection means that a student completing or leaving a course in October would be returned in the first reference period but would not be returned in the second or third returns, whereas a student starting in October and completing in June would be returned in all three returns even if the information was unchanged.

Advantages

84. Whatever patterns of recruitment a provider has, we would be able to understand recruitment within six months and full-year continuation within 18 months, compared with 14 and 26 months respectively with annual end of year collections. One of the considerations in developing Data Futures has been to accurately reflect more varied academic cycles within providers.

85. It would be easy to construct three different year-long views starting at the beginning of a reporting period each year. This may help in providing more regularly updated counts of students.

Potential disadvantages

86. The primary reason why we are now discounting this approach is that uses for the data from the second reference period would be more limited than for the first and third reference periods, as for most providers there would be limited change between November and March. We therefore consider that in the short to medium term the burden of making the second return each year will significantly outweigh the benefits gained from it for the regulation of most providers.

87. This approach would require each provider to make a return for each reference period, even when there was little new activity such as new starters or qualifiers to report. There is no opportunity under this approach to vary the collections required of providers.

88. This approach requires consistency in reporting structures and data in each reporting period as it assumes that reporting periods are interchangeable.

89. Taking this approach would usually require amendments to a reporting period where errors were found, as data is not automatically superseded.

Consultation questions: Part one

Question 1

Do you have any comments about our preferred Approach 1: Two individualised student data collections a year with reduced data requirements in the first data return?

Question 2

Do you have any comments about Approach 2: Cumulative in-year individualised student data collection with differential reporting by provider risk?

Question 3

Do you have any comments about Approach 3: Changing the timing of collection?

Question 4

Do you have any comments about our discounted Approach 4: Discrete in-year individualised student data collection three times a year?

Question 5

Please rank the approaches in order of preference. What are the reasons for your preference, particularly the approach you have ranked in first place?

Part two: Other changes to data collection

Changes to student data content

90. At our roundtable discussions providers identified several data items that they considered created high burden.[14] These tended to be data that they did not need for their own operations or that changed throughout the year, meaning that they needed to update it and potentially explain why items had changed.

91. The HESA student record data specification includes a directory of all data items collected on the student record.[15] We propose that we keep the current Data Futures record structure and most of these data items as set out in the current data specification, but are seeking views on the possible removal of a small number of data items for England as set out in paragraphs 92 to 104.

Qualifications on entry

92. Currently, detailed information on qualifications on entry is collected for UK-domiciled students entering with Level 3 qualifications. This data is important as prior educational attainment has been one of the strongest indicators of students' likely success. Analysis that attempts to understand the relative performance of different groups will usually need to take qualifications on entry into account. We currently use this data in TEF and our consideration of grade inflation.

93. Prior entry qualifications are most relevant for young students. For English domiciled students, this data is already available from the National Pupil Database and can be linked for statistical and regulatory purposes. It would therefore be possible to remove this requirement for English-domiciled students and, if access to the other UK nations' data could be secured, for all UK-domiciled students. However, this may not allow individual providers to access the data and may limit the uses to which HESA could put the data, as access to the National Pupil Database data is, rightly, strictly controlled.

Module data

94. In specifying the Data Futures model, we and HESA had assumed that, as with the current model, providers would hold data on the individual modules studied by students. We have, however, removed the requirement to attach subject data to modules, recognising that this was not something providers needed for their own purposes and was therefore burdensome. Even though most providers maintain records of the modules studied by students we understand that the fluid nature of student module choice makes returning modules especially burdensome, particularly in-year when module choices can change.

95. We have reviewed the data items collected on modules and while we can identify some previous uses for the data, the only current use of this data in England is the collection of cost centres. We recognise that it is not essential to have data on modules to collect cost centres;

---

[14] See Annex C for further information.

[15] See the Data Futures specification at https://codingmanual.hesa.ac.uk/21056/dataDictionary/.

they could be collected as a simple pairing of cost centre and the proportion of activity in that cost centre for each student.

96. We are, however, aware of government proposals in respect of the Lifelong Loan Entitlement, which may require the collection of detailed module data for students studying via this route. However, any additional data requirements are not yet known. We are expecting individual students' loan entitlements to be determined, as now, by the SLC. We do not wish to prejudge the final nature of the scheme or the reporting requirements while it is still in development, so are not consulting on detailed changes currently but will review this once more detail is known on its implementation.

Other data fields

97. It is desirable to ensure that all data that is collected has a clear purpose that is obvious to those providing the data. We intend to work with HESA to ensure that the reasons why each data item is required are clear and remove any items that we no longer have a compelling reason to continue to collect. We have identified a small number of fields where the current use cases are less clear and where it may therefore be desirable to cease collection. In considering the case for removing items, we have given weight to the fact that it is often change that generates burden, so where a data item already exists, we have tended to require a lower degree of justification than we would for adding a new item.

98. Some data items within the Data Futures model relate specifically to our current approach to funding, most notably the notion of funding completion captured in the FUNDCOMP field. We are committed to reviewing our teaching funding method and will seek to remove burdensome concepts wherever possible.

99. Providers have indicated that certain types of off-venue activity, for example placements, are difficult to capture and return accurately, particularly where they are not a formal part of a student's course. We use off-venue activity to identify funding populations.[16] We do not require details of off-venue activity unless this is a formal part of the course, and do not consider that the Data Futures model goes beyond the requirements of the current HESA student record.

100. Providers at our roundtables raised concerns about the collection of term-time accommodation and term-time postcode. Both data items are useful for us, the Department for Education and the Office for National Statistics in understanding student mobility. However, this data is not currently critical to our regulation of providers so could potentially be removed.

101. Currently the HESA record collects data on the gross and net fee levels for students not in receipt of student support. This data allows us to monitor compliance with statutory fee limits and to understand how fees vary by course type at the sector level. Understanding this can, together with cost information, help us understand the extent to which fees cover costs. However, the use of this data is infrequent and has previously been collected through ad hoc surveys.

102. The current data model collects details of the financial support offered to students. While financial support remains an important feature of many providers' access and participation

---

[16] https://www.officeforstudents.org.uk/advice-and-guidance/funding-for-providers/annual-funding/technical-guidance-and-funding-data/

plans, a focus on expenditure was largely a feature of the previous regulatory system. Therefore, it may be possible for us to regulate access and participation matters, and for providers to undertake their own evaluation work, without collecting all of this data separately or in a way that is linked to individual students. However, we may still require some data to understand the level and effectiveness of bursaries.

103. Currently the HESA record collects data about whether a student with a disability is in receipt of Disabled Students' Allowance or not, captured in the DISALL field. It may be possible for the OfS to get this data from the SLC directly rather than asking providers to submit this information in the HESA record.

What are we proposing?

104. Removing the requirement to collect data items that we no longer need or can source from elsewhere will help to ensure that the requirements that we are placing on providers are proportionate. We propose that we remove the requirement to collect data on:

   a. Qualifications on entry for English-domiciled students, as this data is already available from the National Pupil Database and can be linked for statistical purposes.

   b. Term-time accommodation and term-time postcode, as this data is not currently critical to our regulation of providers.

   c. Gross and net fee levels for students not in receipt of student support, as this data is infrequently used and could be collected through ad hoc surveys as in the past.

   d. Most data on financial support offered to individual students, as it may be possible to regulate access and participation matters, and for providers to undertake their own evaluation work, without collecting this data.

   e. Students in receipt of Disabled Students' Allowances.

Data quality

105. The OfS and HESA have previously required that data be fit for purpose. While this remains our overall approach to data quality and regulation, we recognise that this principles-based approach can be hard for providers to interpret in practice. Therefore, to support providers in judging whether their data is likely to be fit for purpose, we intend to clearly articulate quality expectations for their individualised student record. This might be set as an accuracy expectation for each field, or group of fields. We would expect to work with HESA to define a quality expectation which would be set out in HESA's data collection guidance. In general, we expect these quality expectations to be high, often requiring accuracy of over 99 per cent, but they will vary according to the likely impact of data errors.

106. We accept that providers will typically take time to clean data and assure its quality. Therefore, we will expect data quality to improve over time: we would not expect the data quality for a student who started shortly before the end of a reference period to be the same as that for one who had been registered for many months. While the quality expectations we will set are designed to aid providers in ensuring their data is fit for purpose, they will not determine our judgements, as we would always consider the materiality of errors in data. These may be

greater in some circumstances, for example where they might systematically cause us to allocate additional funding.

107.   We will use data verification and data assurance exercises to check that providers are meeting the required quality threshold for their data returns. To gain assurance over the data we use, we will, on occasion, audit providers. These audits are usually risk-based and targeted at those providers where we have reason to believe that data quality may be an issue. When we audit a provider, we seek to understand all the errors in its data so that we can form a comprehensive view of data quality. However, this does not reflect the standard we expect in providers' returns where we expect the data to be fit for purpose. Providers that submit data within our quality expectations will still be able to correct any material inaccuracies of which they become aware.

What are we proposing?

108.   We propose that we set an accuracy expectation for each field in the individualised student record, which we will clearly articulate in any data collection guidance. Given the importance of consistency, we expect to work with HESA and colleagues in the devolved administrations in setting these expectations and will seek views from providers as part of this process.

109.   We propose to continue the current practice of checking that individual providers are meeting the required quality threshold for their data returns using data verification and data assurance exercises.

110.   We propose to continue to gain assurance over the data we use through targeted data audits of individual providers.

Changes to staff data content

111.   We currently require providers to return information about members of their governing bodies to the HESA staff record. While we consider the diversity and makeup of governing bodies to be important, it is not something that we normally regulate and is therefore not necessary to meet our regulatory objectives. We are also aware that some providers find the inclusion of governors on the staff return burdensome. In light of our objective to be proportionate in our use of data, we do not consider it necessary to continue to collect this data. We are therefore proposing that we cease collection of governor data on the staff return from the 2021-22 academic year.

112.   The primary statutory users of the staff return in England are the Department for Education and UK Research and Innovation, with the coverage of the return currently restricted to providers in the Approved (fee cap) category. We recognise that there has not been a substantial review of the staff record in many years. We have therefore asked HESA, in conjunction with the devolved administrations, to initiate a fundamental review of the content and coverage, including provider coverage, of the staff record, with a view to ensuring that the data collected is proportionate, with a clear rationale for why data is required. We would expect the outcome of that review to be a consultation led by HESA, in relation to which the OfS would consider the responses and make any final decisions over the coverage and content of the return for English providers.

What are we proposing?

113.   We propose to cease collection of governor data on the staff return from the 2021-22 academic year.

Changes to provider profile data

114.   The provider profile is a separate data return designed to collect metadata to support the other HESA collections. In particular, it collects data on providers' campuses and the mapping of internal academic departments to cost centres. Given the use of campus data in both student and staff collections, we do not consider it appropriate to subsume collection of the provider profile into other data returns. This is because there may be cost centres or campuses with staff but no students. However, we have made limited use of data on the mapping of departments to cost centres. While theoretically this data is useful to understand the student, staff, and finance data it is not critical to our day to day uses of the data. It could therefore be collected as part of an infrequent bespoke collection if necessary, although we consider that continuing to collect the list of cost centres a provider intends to use can add value by ensuring there are links between student, staff and finance data.

What are we proposing?

115.   We propose that data on mapping departments to cost centres is removed from the provider profile.

Consultation questions: Part two

Question 6

Do you have any comments about our proposals on student data content?

Question 7

Do you have any comments about our proposals on data quality?

Question 8

Do you have any comments about our proposals on changes to staff data content?

Question 9

Do you have any comments about our proposals on changes to provider profile data?

Question 10

Are there any other data items where the collection does not appear to be justified?

Question 11

Do you have any other comments on our proposals to make changes to data collection?

Part three: Use of linked and third-party data

Our proposals

116.   To be an effective regulator, the OfS needs access to student data that allows us to perform our functions. It is essential that we can access data that is fit for purpose in terms of its coverage, definition, and timeliness, and that we can change these should our requirements change. Although other sources of student data exist, for example data from the SLC, UCAS, and the Home Office, these all have limitations. Data from different sources may differ in their:

- coverage of students and activity, as well as the content of these data fields

- level of detail

- alignment of dates.

117.   It would be ineffectual for a regulator to rely solely on data from third parties, although this can play an important role as part of a wider data ecosystem. Although we might be able to influence some of the data collected, the requirements of the third party would always take precedence. We would not be able to demand access to the data from third parties and would always need to have access to alternative data collection sources to ensure we could continue to regulate effectively.

118.   Individualised student data can be linked to other data sources, for example the National Pupil Database data on free school meals, UCAS data and SLC data, and can be enhanced without imposing an additional burden on providers. We will continue to make use of data linking between individualised student data and data from third parties, and explore how this can be further enhanced and used where appropriate to reduce the amount of data that providers are required to return to us or HESA.

119.   We will continue to explore the potential of building analytical tools and applying data science techniques to data collected via social media and other online sources to make better use of near-real-time data and information in our regulation. For example, we could systematically review providers' social media accounts to monitor the timely reporting of reportable events.

What are we proposing?

120.   We propose that we will continue to use data linking between individualised student data and data from third parties, and explore how this can be further enhanced and used where appropriate.

121.   We propose that we will continue to explore the potential of applying data science techniques to data collected via social media and other online sources to make better use of near-real-time data and information in our regulation.

Consultation questions: Part three

Question 12

Do you have any comments about our proposals to make use of linked and third-party data?

Consultation questions: General

Question 13

Are there aspects of the proposals you found unclear? If so, please specify which, and tell us why.

Question 14

In your view, are there ways in which the objectives of this consultation (as set out in paragraph 2) could be delivered more efficiently or effectively than proposed here?

Annex A: Methods of student data collection

1. There are two broad mechanisms that can be used to collect student data: individualised or aggregate data returns. Both can be used alongside analysis of administrative data such as that held by UCAS or the Student Loans Company, although it is only the collection of individualised data that allows for these data sources to be effectively linked.

2. Collecting individualised student data at least once per year is essential to understand the nature of higher education. We would be unable to generate many of the key statistics that we and the government rely on without individualised data that allows us to follow students through and beyond higher education.

Individualised student data collection

Advantages

- Is inherently flexible, as once collected the data can be repurposed and reanalysed including in ways which were not originally envisaged.

- Makes it easy to assure quality, as the data relates to individual students so can be readily checked back to source.

- Can be linked with other data to answer new questions and give new insights, for example into the barriers to student success and the journeys students take through education and into employment.

- Provides a robust evidence base for policy development and research by the Office for Students, the government, the Higher Education Statistics Agency's statutory customers (including in the devolved administrations), and the research community.

Disadvantages

- Relies on higher education providers having fully fledged student record systems to streamline returns. Such student record systems may go beyond some providers' needs. This is particularly likely to be relevant for small providers, where it may be more appropriate to manage student data in spreadsheets or other desktop tools.

- Leads to providers focusing on small issues, such as the quality of one data item, rather than the overall quality and completeness of their data.

- Requires a complete record for every student at one time. While it is possible to establish collections where only some data is returned in each collection, these add complexity to the systems at both data providers and receivers.

Aggregate student data collection

3. Aggregate data collections involve collecting summary data at a level that will typically group many students in each data cell. For example, in the Higher Education Students Early Statistics (HESES) survey we collect data on the total number of students on a sandwich year, rather

than details of each student. Each additional dimension or variable tends to add complexity to the return, and this complexity can become unmanageable for those making aggregate data returns. The nature of aggregate returns means that they can only be used for a limited number of purposes and usually no ancillary benefit can be gained from them. Aggregate data returns do not generally require providers to hold any less data, as both individualised and aggregate returns will need to draw from the same underlying student data.

Advantages

- Can reduce burden, as it collects specific data items for given purposes.

- Makes it easier to see the big picture.

- Is likely to need less complex systems development where a return is simple, as data can be extracted in a way that is easy for the provider. As the amount of data is typically small it may be possible, in some cases, to simply input the numbers into a template or web form.

- Can also include estimates and other adjustments to figures in a way that individualised data returns cannot. While this can reduce burden in some cases, it may instead increase it, as there is a need to validate and justify estimates that are included. Our experience is that estimates are often materially wrong.

- May impose a lower burden on providers where they use data and systems that already exist, at least in the short term.

Disadvantages

- Needs to be tailored to specific purposes to manage its complexity. This means that multiple collections may be necessary if there are a range of uses for in-year data.

- Normally only enables data to be used for the planned purpose.

- Presents challenges in quality assuring and ensuring the accuracy of the data. Historically the OfS, and previously the Higher Education Funding Council for England, have had to invest significant resource in quality assuring the HESES return, both at the point of collection and when reconciling to individual student data. Our experience of reconciling HESES to HESA data is that, even disregarding forecasts, it is more often HESES where the most significant issues are found. Therefore, where aggregate returns are used to monitor compliance with regulatory expectations, we would expect to need to expend considerable effort (and therefore generate significant burden for providers) in assuring ourselves that the data is reliable. This quality assurance activity also needs to be repeated for each return.

- Tends to mask complexity and the cost of change. While in theory it is simple to change the layout or content of aggregate returns to address new or emerging issues, every change of this type will require systems change at each provider making the return. For example, adding a price group to HESES could be viewed as simply adding a few rows to the relevant table, but in practice this will mean every provider changing the code which

extracts student data for HESES to reflect the new structure. The same is true of adding new data dimensions to aggregate returns. In order to complete an aggregate return, providers will normally need to already hold, and assure the quality of, that data on their student record system for all students covered by the return. Therefore, the data collection and quality assurance burden on providers is broadly equivalent to that of individualised student data collections.

- Normally needs to be designed with a specific use in mind. Therefore, we would need to anticipate regulatory issues or policy requirements in advance to collect relevant specific targeted data.

- Requires redesign if uses change.

Annex B: Value and potential uses of individualised in-year data

1. The main ways in which the Office for Students (OfS) currently uses individualised student data are:

   - general monitoring of registered higher education providers including degree classifications

   - regulating student outcomes, including continuation, completion and progression (as part of decisions about compliance with conditions of registration, and as part of the Teaching Excellence and Student Outcomes Framework (TEF))

   - approving and monitoring access and participation plans

   - providing information to students about providers and courses

   - allocating funding to providers

   - research and analysis

   - policy development.

2. We currently use aggregate student data, such as the Higher Education Students Early Statistics (HESES) survey and the medical and dental student survey, to support our funding allocations, as this data is returned in-year.

3. The OfS is not the only body to use Higher Education Statistics Agency (HESA) data. It is used by a wide range of other bodies and regulators, as well as informing policy development and research on school and higher education performance by the government, academics, and policy specialists. Under section 64(8) of the Higher Education and Research Act 2017 (HERA), HESA has a duty to cooperate with other organisations that collect information from registered higher education providers. This includes professional, statutory, and regulatory bodies (PSRBs), which may set the standards for, and regulate the standards of entry into, particular professions.

4. HESA data, and the Data Futures programme, also have critical importance in Scotland, Wales and Northern Ireland. The review of Data Futures therefore needs to consider the impact of any changes across a wide range of uses. We recognise the benefits of UK-wide data, for example in benchmarking, to colleagues and providers in the devolved administrations.

5. The individualised student data collected by HESA can be used to answer a wide range of regulatory and policy questions. It can also be linked to other data sources, for example the National Pupil Database data on free school meals, UCAS and Student Loans Company (SLC) data. Therefore, this data collection can be enhanced without imposing an additional burden on providers in a way that aggregate returns such as HESES cannot.

6. We are aware that providers also gain benefit from the data they provide to HESA. Currently most providers subscribe to the HEIDI Plus service, which allows them to extract data for all UK

providers that return data to HESA for benchmarking, course development or other purposes.[17] Where universal data is collected, we would expect this to continue, with data being released shortly after each collection.

7. In paragraphs 9 to 35 of this annex, we explore some ways in which in-year individualised data could be used to improve the effectiveness of our regulatory activities. We have started from the current data landscape, which in some cases relies on in-year aggregate returns. We have also illustrated how the activity might work without aggregate in-year data.

8. Figure B1 shows a summary of when we would expect to be able to use in-year data in steady state under the four approaches, and under the current HESA collection timings. These timings should be treated as illustrative, as they assume that it is possible to progress all uses immediately after data signoff, and that data collection windows are in line with the current Data Futures approach and shorter than the current HESA collections.

---

[17] See https://www.jisc.ac.uk/heidi-plus.

Figure B1: Indicative timelines of uses of steady state data under approaches 1 to 4 and the current position

| Use of data | Year one | | | | | | | | | | | | Year two | | |
---	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar
Continuation		①② ③④											C ②		
National Student Survey target list	①② ③④														
Funding		①② ④											C		③
Access and participation dashboards (assuming update timed to minimise lag for autumn starts)				①② ③④											C ②
Unistats		①② ③④											C ②		
Grade inflation													C ① ② ④		③
Full-time equivalent for fees, degree awarding powers etc		①② ④											C ②		③
HERA section 65 publication (first)		①② ③④											C ②		

① Approach 1   ③ Approach 3
② Approach 2 (earliest)   ④ Approach 4
② Approach 2 (latest)   C Current

Note
'Approach 2 (earliest)' represents when the activity could occur if every provider is required to make every return. 'Approach 2 (latest)' represents when the activity could occur if we decided that not every provider is required to make every return.

Approval of, and monitoring of progress on, access and participation plans

9. Currently the OfS can undertake early monitoring of some, but not all, changes in the demographics of students recruited to providers by analysing UCAS data. However, it is not possible to conduct a comprehensive assessment because of gaps in this data.

10. The lack of universal in-year data also means that the access and participation dashboards cannot be updated until March each year. The dashboards, therefore, reflect the offers that were typically made two years previously.

11. While providers will have internal data that allows them to understand changes in their own recruitment, they will not have the access to demographic characteristics that the OfS dashboards make available to them, since the OfS uses linked data to examine other underrepresented groups. Nor will they be able to place their data in the context of other providers, and additional effort will be needed to calculate comparisons with population statistics.

12. In-year individualised data would allow us to update the dashboards with each submission of student data. This would mean that we, and providers, would be better able to see whether providers were making the progress to which they have committed. This would allow for more timely monitoring of progress, public communication, and information for providers.

Finalisation of NSS target list

13. The National Student Survey (NSS) aims to target all students in their final year of undergraduate study. The current process uses year-end student data from the previous academic year to predict which students will be in their final year. This approach means that students whose plans change because of breaks in study or repeating years may be included before their final year, or in some cases excluded. To reduce the impact of this we operate an additions and removals process that includes checks to ensure that changes made are legitimate. In 2021, around 27,000 changes were made to the NSS target list via this process. Three-quarters of the 251 HESA providers with students on the NSS target list had changes approved through the additions and removals process, representing approximately 5.5 per cent of the 463,000 students on the final NSS target list.

14. While most providers engage with the process, as they have large numbers of students who change status, some providers that have small numbers do not, which could bias the survey results.

15. The NSS is published by linking the results back to the course, study and demographic details from the previous year, meaning that where students change course near the end of their study they are often reported against different courses. This also leads to missing data for some students who return from a break in study to take their final year.

16. With in-year data, the target list would be drawn directly from the first return of the year, completely removing the need for an additions and removals process and ensuring that data published reflected the most recent course, study and demographic characteristics. There would be no additional effort over and above ensuring data quality. The use of a consistent process should also improve consistency between providers, as we will not be reliant on their

identifying students that they need to add or remove. This would lead to more accurate data for students and the removal of a time-consuming process for providers and the OfS.

Completion rates and between-year continuation rates

17. We use completion rates and between-year continuation rates for general monitoring purposes, to regulate student outcomes, to operate the TEF and to populate the Discover Uni website.

18. We plan to consult on the indicators for these purposes in early 2022. Our current indicators look at continuation one year and 14 days after a full-time student starts their course (and two years and 14 days after a part-time student starts). Most students start their courses in September and October, but data is not received by us until December in the following year. This means that, for students who start in late September or early October, the continuation rate cannot be calculated until 14 months after the anniversary of them starting.

19. Completion rates calculated for shorter courses, or those running on non-standard academic years, cannot be calculated until some time after the students have completed their course. For students completing a taught masters' course in October, for example, the completion rate cannot be calculated until 14 months after they finish their studies.

20. With in-year data it would be possible to calculate the between year continuation rates for students no more than five months after the anniversary of their starting. This would be shorter for the main September and October starters based on the timing of the first return in November. Completion rates could also be calculated within five months of courses ending. This would enable more responsive regulation allowing earlier intervention where necessary.

21. Depending on the timing of the TEF submissions, it is likely that TEF could use data that is one year newer, which would mitigate one of the significant timing issues relating to continuation data. Students preparing UCAS applications in the autumn would have data available on Discover Uni that is one year more current.

In-year continuation rates for use for general monitoring purposes and as contextual information

22. We are considering including in-year continuation rates in our general monitoring approach as a more timely indicator of quality and as contextual information which could be used to inform our regulation more broadly. This would be in addition to partial indicators we are developing based on SLC data.

Monitoring changes in recruitment at providers in the Approved registration category

23. Currently the only in-year recruitment data available for providers in the Approved registration category is SLC and UCAS data. While this data gives a good indication of recruitment for the core student populations that UCAS and the SLC cover, the coverage is limited, so it is not possible to gain a holistic view of the recruitment at each provider. Data on providers in the Approved (fee cap) category is collected via the HESES survey in December.

24. With in-year data it would be possible to monitor the recruitment of all providers on the same basis, giving up-to-date data on both the number of new students and the total number of students. This would enable more timely monitoring and intervention, when necessary.

Data for use in strategic priority funding

25. Currently data used to allocate strategic priority funding is collected in-year through the HESES survey, with further detail on student characteristics provided by the previous years' end of year individualised student return. The HESES survey is complex, comprising six separate tables and over a thousand data items for each provider, with significantly more for those offering pre-registration health courses. The HESES return undergoes a significant verification exercise each year and is reconciled to year-end individualised data returns. Despite the extensive verification work, we routinely find material errors in HESES returns in both the forecast and actual elements of the return. Providers with medical or dental schools are also required to make an annual medical and dental return to allow us to monitor intakes in these subjects.

26. Without HESES or in-year individualised data we would need to base funding entirely on the student numbers and characteristics from the previous year. This would mean that funding would be based on the students from two years prior to the year in which it is paid.

27. With in-year individualised data it would be possible to use the first in-year student data to calculate funding that would reflect changes in the main intake for virtually all providers in the Approved (fee cap) category. It would also facilitate the removal of the HESES and medical and dental survey and associated verification and reconciliation exercises.

Full time equivalents used to calculate student numbers for OfS registration fees and other regulatory purposes

28. Currently our student volume measures use individualised year-end data. With in-year data the process would be similar, but we could instead use more timely data from the most recent 12-month period, meaning that OfS registration fees would be more responsive to change – although we would not anticipate changing fees once they had been set. Further, assessment of applications for degree awarding powers and university title could use data from the current year, allowing providers to demonstrate that they meet the FTE requirements without any need for further data collection.

HERA section 65 publications

29. Currently the HERA section 65 publications produced by HESA as official statistics use the end of year data. The value of these publications for a wide range of users is likely to be significantly enhanced by some content being available earlier. For example, publications that draw on this data to support student choice would be able to draw on more recent data, which should lead to better decisions. These publications also facilitate inclusion of data in tools like HEIDI Plus and research and evaluation activities, in a way that is consistent with the Code of Practice for Statistics.

Research and evaluation activities and policy development

30. Currently most research and evaluation activities either need to wait until the individualised year end data is available or need to conduct ad hoc data collections. This process means that

research and evaluation activities that are inherently lagged suffer further lags, and necessary changes and enhancements to programmes can be delayed. These delays in research and evaluation activities are felt most acutely in widening access, where it is often not possible to conduct impact work on schemes like Uni Connect until another cohort of students has already started. Similar issues are found with initiatives such as artificial intelligence conversion courses. In-year data would enable research and evaluation activities to take place earlier, enabling changes to programmes' effectiveness sooner.

31. The lack of timely student recruitment and retention data has been acute during the coronavirus pandemic. Similar issues occurred during the 2008 financial crisis and following the change to fees in 2012. Each time one of these significant events occurs it reduces the ability and flexibility to analyse and understand the impact on the higher education sector. In-year data would allow key events to be analysed more swiftly and responsively.

Curriculum and course planning, course benchmarking, and marketing

32. UCAS provides UK-wide data on the recruitment to most full-time undergraduate programmes, which allows providers to assess the likely demand for future courses. For courses not covered by UCAS, there is no data available, which means that any course and curriculum planning must use HESA data that reflects recruitment at least one year previously.

33. If in-year data were available, we would expect it to be published in the same way as the end of year data is now. This would mean that providers would have more timely data to plan course offerings leading to students having a better choice of courses.

Third parties including professional, statutory, and regulatory bodies

34. Currently most PSRBs that need data from providers make specific requests for it. The most often cited reason why HESA data cannot be used is that it is too lagged for their purposes.

35. The collection of universal in-year data should make it possible to meet the needs of many PSRBs. In some cases, this may require the addition of a small number of additional data fields, for those students and the relevant courses to be explicitly identified. In many cases this could be done through the 'Curriculum/Accreditation' entity in the Data Futures data model. This might remove the burden of numerous collections for individual PSRBs, and allow richer information for those bodies than it might be reasonable to collect.

Annex C: Roundtable discussions

36. In July 2021 we held a series of roundtable discussions with stakeholders to help us understand more about the relative burden of different types of higher education student data collections and different activities required to complete a student data return.

37. Senior managers and data practitioners in universities and colleges, together with student data software providers, were invited to sign up for one of the roundtable discussions. In total we heard from approximately 60 different stakeholders.

38. The discussions focused on exploring:

   • the nature and causes of burden in student data collections

   • options for managing burden and ensuring proportionality

   • elements of the current Data Futures data model that may be particularly burdensome, whether collected annually or more frequently.

39. The discussions considered the key drivers of burden for activities associated with different types of student data returns, including:

   • method of student data collection

   • one-off activity compared with repeated activities

   • frequency of new data requests or changes to the content and frequency of collection

   • timing

   • data content

   • quality expectations.

40. Other issues the roundtable discussions considered included:

   • the use of linked and other data sources

   • duplication of data requests from different bodies

   • the proportion of data activities that would be undertaken by providers themselves, regardless of the Office for Students' (OfS's) requirements

   • alternative models of collecting in-year individualised data.

What we learned from the roundtable discussions

Timing and quality expectations of data items

41. We heard from higher education providers that, while data is often collected as students register, significant data cleansing often occurs throughout the year. This means that data returned early in the year may well be incomplete or incorrect. While providers recognised the

value of quality assurance at source, the reality for most was that some retrospective cleansing was required. Currently the fact that data is returned at the end of the year means that cleansing can be fitted around the work of the data teams.

42. There is a perception that the OfS expects all data items to always be correct and that there are no materiality thresholds applied. This perception is amplified through the quality assurance processes that currently apply to the Higher Education Statistics Agency (HESA) student record, and can often lead to significant effort being expended to correct data for a single student. Providers felt that this focus on particular items for individual students often distracted from the bigger picture.

Frequency of data requests

43. The current Data Futures plan proposes three discrete collections of individualised student data, each covering a period of approximately four months. We heard from some providers, particularly those with a small student data team or no dedicated team, that an increase in the frequency of data collections from the existing annual HESA student record would result in a substantial increase in administrative burden.

44. Student data collection systems within providers are complex. Our understanding is that any changes to the frequency of data collection would likely result in these systems needing to be adjusted, resulting in a significant time lag in the data and significant burden each time the frequency of collection changed. Providers also observed that a planned schedule of returns was easier to resource than responding to ad hoc or changing requests for data.

45. For providers in the Approved (fee cap) category, the removal of the Higher Education Students Early Statistics survey and the National Student Survey (NSS) additions and removals process were generally seen as a proportionate benefit of returning individualised data, provided the data required in-year (or shortly after students' registration) was minimised. The direct benefits to providers in the Approved category were less clear, especially for those that made very few changes to their NSS target list or did not have any students covered by the NSS.

Burden of additional data requirements

46. Providers were conscious that many of the data items required by professional, statutory or regulatory bodies or other organisations were not held on the student record systems that are used for the HESA returns. Including this data on the HESA record would either need a significant expansion of central student records or create complex systems integration issues. While some providers saw some benefit in this for internal management purposes, most took the view that the costs would outweigh any benefits, as the data is primarily of interest to individual academic departments.

47. We received similar feedback regarding the detail on placements included in Data Futures, as the management of placements is often devolved to departments. This highlights a challenge where information is known about students by the provider, but is not necessarily consolidated into the student record system.

Annex D: Summary of possible approaches to collecting individualised student data

1. Table D1 provides an overview of the four approaches to the collection of in-year individualised student data alongside the advantages and disadvantages of each approach. To aid users in understanding how the approaches differ we have also included Figure D1, showing which periods returns would cover under each proposal.

Table D1: Summary of approaches to the collection of in-year individualised student data

Approach	Number of data collections	Timeliness of data	Advantages	Limitations
Approach 1 (preferred): Two individualised student data collections a year with reduced data requirements in the first data return	2	For students who are recruited in the autumn we would be able to understand recruitment within three months and full-year continuation within 15 months.	Reduces the number of returns for each provider from the current proposal of three per year. More timely understanding of the regulated sector and increased value to providers. Most providers would see a decrease in other data collection activity. Reduces administrative burden of assuring the quality of certain data items early in the academic year.	Compromise over the timing of the first return. Direct benefit of in-year collection of data is lower for a small number of providers, particularly those in the Approved category.

Approach 2: Cumulative in-year individualised student data collection with potential for differential reporting	up to 3	For all students we would be able to understand recruitment within six months and full-year continuation within 18 months.	Allows some flexibility over which of the in-year returns a provider is required to make. It may be possible to have different in-year returns for different providers to match the main dates in their academic calendar. Transition to collecting in-year data is easier given less pressure to get data right for the first reference period of the first year. May create opportunities to simplify the current data amendments process.	Fundamental change to the current proposed Data Futures model. May have more significant effect on the timetable for implementation and may increase the complexity, and cost to the Higher Education Statistics Agency. Introduces an element of redundancy into the data collection, with some activity returned more than once. Reinforces the primacy of the current academic year structure. The uses to which the OfS can put the data will be more limited, and the wider benefits of the data to providers and others are also likely to be reduced.
Approach 3: Changing the timing of collection	1	Data from this single return would be available for use early in the next calendar year.	Timely reporting of intakes and year-on-year continuation Timely reporting of qualifications for many masters' students where these awards are made before November.	Could not be extended if a need for more in-year data was identified at a later point. Transition would be extremely complex and would most likely require the first return of data to cover a 15-month period May involve hidden complexity for providers because the data return would, for many students, cover two academic or course years. Changing the reporting period is likely to affect comparability with staff and Higher Education Business and Community Interaction

			Only one collection of student data per year. May resolve one of the tensions of the current data collection timing which relates to the population for Graduate Outcomes cohort D.	survey data, which would still be collected in-line with the current academic year. Delay in reporting outcomes for most first degree students of about three months. Potentially misses results for some students in Graduate Outcomes cohort A. May require restructuring of providers' data collection systems.
Approach 4 (discounted): Discrete in-year individualised student data collection three times a year	3	Whatever patterns of recruitment a provider has, we would be able to understand recruitment within six months and full-year continuation within 18 months.	Accurately reflects more varied academic cycles within providers. Easy to construct three different year-long views starting at the beginning of a reporting period each year.	Each provider required to make a return for each reference period even where there is little new activity to report. Uses for the data from the second reference period would be limited, as for most providers there would be limited change between November and March. Requires consistency in reporting structures and data in each reporting period. Usually requires amendments to a reporting period where errors are found, as data is not automatically superseded.

Figure D1: Summary of reporting periods

		2023					2024										
	Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec

Current: 2023-24 Reporting year | 2024-25 Reporting year

Approach 1: 2023-24 Reference Period 1 | 2024-25 Reference Period 1 | 2023-24 Reference Period 2 | 2024-25

Approach 2: 2023-24 Reference Period 1 | 2024-25 Reference Period 1 | 2023-24 Reference Period 2 | 2024-25 Reference Period 2 | 2023-24 Reference Period 3 | 2024-25 Reference Period 3

Approach 3: 2023-24 Reporting year | 2024-25

Approach 4: 2023-24 Reference Period 1 | 2024-25 Reference Period 1 | 2023-24 Reference Period 2 | 2024-25 | 2023-24 Reference Period 3

Annex E: List of consultation questions

Question 1

Do you have any comments about our preferred Approach 1: Two individualised student data collections a year with reduced data requirements in the first data return?

Question 2

Do you have any comments about Approach 2: Cumulative in-year individualised student data collection with differential reporting by provider risk?

Question 3

Do you have any comments about Approach 3: Changing the timing of collection?

Question 4

Do you have any comments about our discounted Approach 4: Discrete in-year individualised student data collection three times a year?

Question 5

Please rank the approaches in order of preference. What are the reasons for your preference, particularly the approach you have ranked in first place?

Question 6

Do you have any comments about our proposals on student data content?

Question 7

Do you have any comments about our proposals on data quality?

Question 8

Do you have any comments about our proposals on changes to staff data content?

Question 9

Do you have any comments about our proposals on changes to provider profile data?

Question 10

Are there any other data items where the collection does not appear to be justified?

Question 11

Do you have any other comments on our proposals to make changes to data collection?

Question 12

Do you have any comments about our proposals to make use of linked and third-party data?

Question 13

Are there aspects of the proposals you found unclear? If so, please specify which, and tell us why.

Question 14

In your view, are there ways in which the objectives of this consultation (as set out in paragraph 2) could be delivered more efficiently or effectively than proposed here?

Annex F: Matters to which we have had regard in developing our proposals

The general duties of the Office for Students

1. In formulating these proposals, we have had regard to our general duties as set out in section 2 of the Higher Education and Research Act 2017. We consider that the proposals in this consultation are particularly relevant to general duty (g), which relates to best regulatory practice, including the principles that regulatory activities should be transparent, accountable, proportionate, and consistent. This has been particularly relevant in developing proposals for the number of data collections each year and the amount of data collected in each collection. We have considered how the method of collection should be proportionate, so that it does not place unnecessary burdens on providers and allows the Higher Education Statistics Agency and the Office for Students (OfS) to access and use data effectively. Furthermore, our proposals to set clear quality expectations are designed to ensure that our expectations are proportionate, clear and that we apply consistent quality expectations to all providers. Throughout we have considered whether data requirements are necessary and whether data can be obtained in other ways, for example by linking to other data sources although we recognise that this may reduce transparency.

The Regulators' Code

2. The Regulators' Code requires us, in 1.1 and 1.2, to consider the burdens that our activities place on regulated entities. As outlined in paragraph 1 this has been central to our considerations throughout the formulation of these proposals.

The Code of Practice for Statistics

3. As an official statistics producer we have had regard to the Code of Practice for Statistics in formulating these proposals.[18] The Code of Practice for Statistics aims to ensure that the statistics produced by the government and public sector bodies are trustworthy (impartial and free from political influence), high quality and of public value and that effective governance structures are in place to protect transparency and accountability. In forming our proposals, we have considered the benefits to students, potential students and other in data on regulated providers and their activities being available earlier. The proposals surrounding clear quality standards will aid us and the Higher Education Statistics Agency in demonstrating compliance with the quality pilar of the code.

The Public Sector Equality Duty

4. Under section 149 of the Equality Act 2010, the OfS must have due regard the public sector equality duty. This requires the OfS to have due regard to eliminating unlawful discrimination, foster good relations between different groups and take steps to advance equality of opportunity. In general, we expect that having detailed data available earlier will allow us, and others, to understand the impact of policies on different students earlier and thus mitigate any

---

[18] See https://www.statisticsauthority.gov.uk/code-of-practice/.

unintended consequences of those policies earlier, thereby minimising any overall negative impacts.

5. Through this consultation we are seeking views on any unintended consequences of our proposals, for example on particular types of providers or students. We are also seeking views about the potential impact of our proposals on individuals on the basis of their protected characteristics. Responses to this consultation will assist us in having regard to those matters under the Public Sector Equality Duty.

Strategic guidance from the Minister of State for Universities

6. In strategic guidance to the OfS dated 14 September 2020 the Minister of State for Universities welcomed our decision 'to review the proposed termly collection, with a view to making data collection more proportionate and […] look at how data can be collected more quickly as the systems are improved'. This consultation is the culmination of the review.

7. In our review we have considered the benefits that each of the three returns would have delivered for the regulation of English higher education providers, and have considered how we could make the burden of each return more proportionate to the uses to which it will be put. We considered whether varying the number of returns required from providers based on risk or other factors would be a more proportionate way to meet our needs. We concluded that we would need in-year data from most providers and that there was significant public benefit to comprehensive data. Furthermore, we learnt through our round tables that the burden on providers of responding to varying collection frequency would likely be significant.

Annex G: Abbreviations

FTE   Full-time equivalent

HE-BCI   Higher Education Business and Community Interaction survey

HERA   Higher Education and Research Act 2017

HESA   Higher Education Statistics Agency

HESES   Higher Education Students Early Statistics

ILR   Individualised Learner Record

NSS   National Student Survey

PSRB   Professional, statutory, and regulatory body

SLC   Student Loans Company

TEF   Teaching Excellence and Student Outcomes Framework