OPEN DATA PRINCIPLES AND REQUIREMENTS IN HIGHER EDUCATION

TO HELP INFORM AND SHAPE HESA'S APPROACH TO OPEN DATA PUBLICATION OVER THE NEXT FIVE YEARS

Summary of Consultation Responses
December 2016

HESA

HESA
95 Promenade
Cheltenham
GL50 1HZ
E opendata@hesa.ac.uk
T +44 (0) 1242 255577
W www.hesa.ac.uk

CONTENTS

SUMMARY

In June HESA launched a consultation to help inform and shape its approach and strategy for open data publication over the next five years. 51 responses were received, mainly from higher education providers.

There was overall a high level of support for HESA's open data plans. Some concerns were raised around the open publication of information derived from personal data, or from data that may be considered as commercially sensitive. However, it was recognised that, with its robust approach to data protection, HESA is well placed to lead on this project that will enhance "transparency, efficiency and innovation" across the higher education sector.

BACKGROUND

Open data is "data that anyone can access, use and share". (Open Data Institute)

In practice there are a number of dimensions to the process of releasing open data to optimise access and re-use but the core principle is sound – data that is open promotes transparency for UK citizens, generates economic benefit through commercial re-use, and stimulates innovation in the creation of new tools and services which benefit users and consumers. Open data has been, and continues to be, a key part of UK Government Strategy. The Open Data White Paper – *Unleashing the Potential*[1] (June 2012) in particular set the scene for the emergence of extensive open data planning and actions within government departments covering public data sets.

"Transparency is at the heart of our agenda for government. We believe that opening up will empower citizens, foster innovation and reform public services." Rt. Hon Francis Maude

Subsequently open data aspirations have been expressed and reinforced within a number of government reviews and strategy documents including the *Government Digital Strategy 2013*[2]. More recently, and specifically within the higher education context, Professor Sir Ian Diamond's report on *Efficiency, Effectiveness and Value for Money*[3] highlighted the HE sector's commitment to sharing data, and considered how more value might be derived by opening data held within and outside the sector.

A *Creating Value from Open Data* project to explore the opportunities and challenges of open data in higher education is being taken forward by Universities UK[4] in partnership with the National Union of Students[5], Jisc[6] and the Open Data Institute[7]. HESA has contributed to a number of activities within this project, offering advice and expertise on higher education data.

Although HESA is not a government body and therefore not directly subject to government open data strategy and policy, the data resources we hold describe activities which are supported by significant public funds. When made accessible such data can deliver major benefits to users and

4

interested parties, not least those directly involved in the delivery and consumption of higher education such as students, researchers and staff at higher education provider organisations. HESA has long operated on principles of transparency and accessibility in data handling. We publish widely from the data we collect, we provide extensive on-line access routes to data for higher education providers and not-for-profit bodies and we provide bespoke information to over 1,200 users annually, covering a full range from staff, students, academic researchers, private companies, professional bodies, government bodies, the press and media to international governments and organisations around the world. However, with a few exceptions, up until now most dissemination mechanisms for HESA data have not been technically 'open'; for example, most published material has been subject to copyright, with those wishing to re-use content obliged to seek explicit permission from HESA on an ad hoc case basis. In addition, many types of publication have been made available at a modest charge.

As the next major step in the evolution of HESA as the preeminent official source of higher education data in the UK, we propose an approach aimed at developing dissemination of data which is formally open – meeting the aspirations of the UK government and taking a leadership role on behalf of the UK higher education sector. The over-riding principle is to migrate as much of HESA's non-personal and non-commercially-sensitive data as possible over the next five years to open data publication, reaching a point at which the ongoing business-as-usual assumption for such data is that it will be open by default. The benefits to users of data are clear – improving accessibility and removing barriers to re-use. There are also benefits for HESA in that dissemination of more open data supports our charitable object to advance higher education, ensures that our data attracts maximal usage reinforcing the value of the data we collect, and encourages innovative uses of data which HESA may not have the resources or expertise to undertake itself – resulting in new applications and services meeting the public interest in higher education.

JONATHAN WALLER
Director of Information & Analysis

1 https://www.gov.uk/government/publications/open-data-white-paper-unleashing-the-potential
2 https://www.gov.uk/government/publications/government-digital-strategy/government-digital-strategy
3 http://www.universitiesuk.ac.uk/policy-and-analysis/reports/Pages/efficiency-effectiveness-and-value-for-money.aspx
4 http://www.universitiesuk.ac.uk/ 5 http://www.nus.org.uk/ 6 https://www.jisc.ac.uk/ 7 http://theodi.org/

METHODOLOGY

Nine questions over six weeks.

HESA aimed to consult its data providers, data users and other interested parties on the principles and details of the proposed approach set out in the consultation document and included in each section below. Nine questions were asked, each offering the opportunity for stakeholders to provide quantitative and qualitative feedback to encourage honest and detailed answers to inform outcomes. Respondents were requested to leave questions they did not intend to answer blank, to aid our analysis. The consultation opened for six weeks, from Thursday 16 June 2016 to Thursday 28 July 2016.

RESPONSES

51 responses, mainly from higher education providers.

HESA received 51 complete responses by the consultation deadline. The majority of respondents (two-thirds) categorised themselves as representing higher education providers. Other responses were submitted on behalf of government bodies, HE sector bodies, private individuals, and employer or employer organisations.

Figure 1: categorisation of respondents
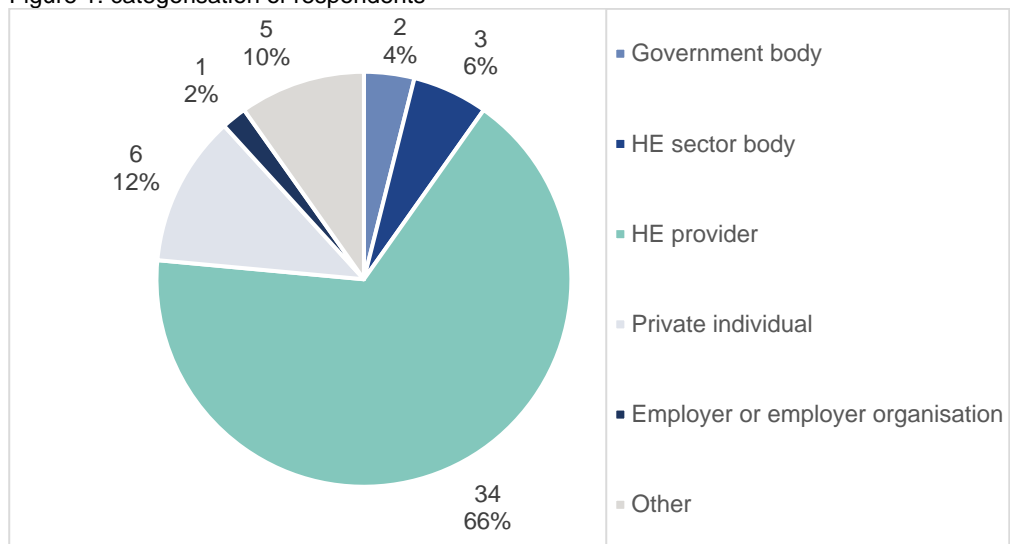
HESA'S APPROACH TO OPEN DATA

Accessibility and the data spectrum

The Open Data Institute publishes *The Data Spectrum* (below) which provides a helpful means of classifying the different types of HESA data resource.

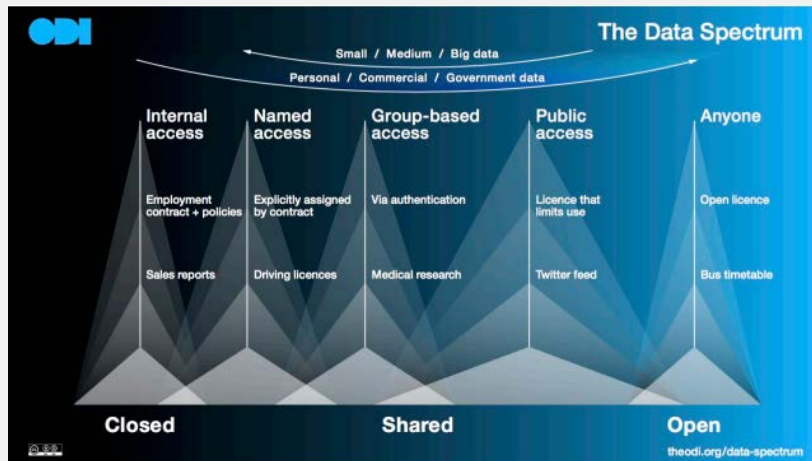Historically, much of HESA's activities in making data accessible have fallen within the 'Shared' zone of this chart, varying between Named access underpinned by legal contracts such as bespoke requests for data, through Group-based access via authentication such as the Heidi online management information service to Public-access as used for a number of HESA publications.

One existing HESA resource – the Unistats data set - is already published as open data. The different points on this spectrum align with concepts of personal data (as defined in the Data Protection Act 1998) or commercially-sensitive data (data which, if improperly used, may have the potential to distort fair competition). It is important to understand that not all data is suitable for publication as open data. In its raw form personal data (including data about individuals which is anonymised but sufficiently detailed as to be potentially disclosive) or commercially-sensitive data published as open data could breach relevant legislation (the Data Protection Act 1998 or the Competition Act 1998 and Enterprise Act 2002).

Open data publication *derived* from such data types is possible, subject to disclosure control processes to preserve confidentiality or to obscure the characteristics of individual organisations in a competitive environment. However, such disclosure control processes can significantly and detrimentally impact on the utility of data. There remains a valid and justifiable need to utilise data sharing processes such as Named access and Group-based access which balance accessibility, utility and legislative compliance in the case of such data types. However, subject to these limitations HESA proposes to make as much of our data as possible available as open data. Annex A in the consultation document (see p25) included a classification of existing HESA data sets according to their suitability for open data publication.

SHOULD HESA MAKE ITS CORE DATA OPEN?

Do you support HESA's aim to make as much of its core data as possible available as open data? (Q1a)

96% of respondents (all except two) said "yes".

Figure 2: answers to Q1a

Pie chart showing 49, 96% (Yes) and 2, 4% (No)
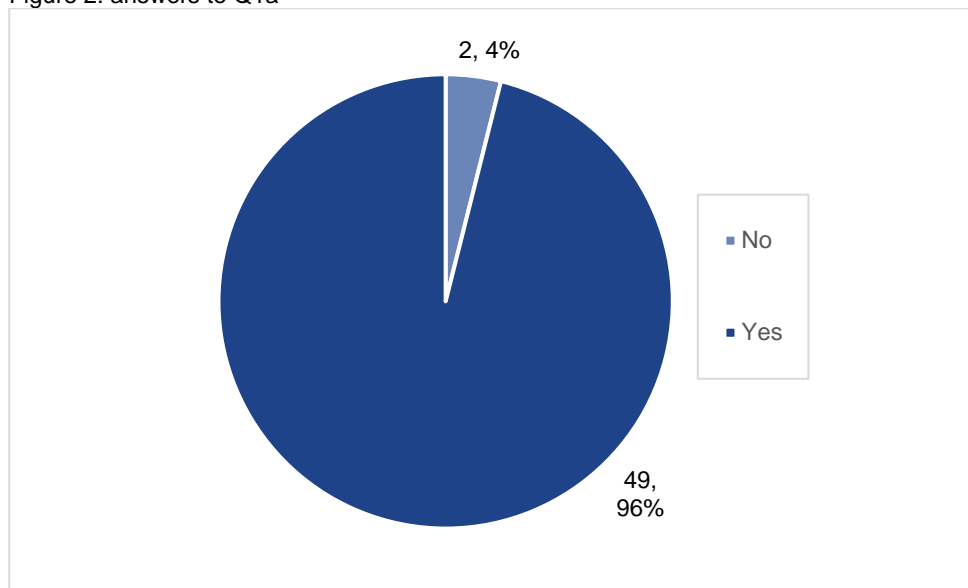
PLEASE EXPLAIN YOUR ANSWER (Q1B)

We received feedback from all 51 respondents in this free-form section. The majority, as expected, reflected the support offered in question 1a for the project to move forward. It was felt that greater transparency was a good thing, supporting evidence-based decision making around policy and funding, and increased accountability around spending.

"making as much data as possible available in a suitable level of detail is a positive move as it will extend the evidence base which can be analysed by both institutions and government."

"A culture of open data will help promote transparency and confidence from the public that funds are being spent wisely."

The freedom to publish data in a re-usable and shareable format was deemed positive. Widening access to HE data will help diversify stakeholder engagement, supporting planning towards advancing the sector. **Publishing comprehensive data definitions and contextual information** was considered essential to reduce risks of misinterpretation. Of specific concern was the scope for misinterpretation of data used in public comparisons of HE providers, such as university league tables.

Notwithstanding the benefits of open data, respondents highlighted data protection as a prominent concern. Alongside ensuring that data about individuals was protected, the disclosure of commercially sensitive data, such as financial information, was highlighted as worrying in an increasingly competitive marketplace. Of the two higher education providers who disagreed with HESA's approach, both outlined their concern that the principles of open data could be contradictory to those of competition law in this context.

Conversely there was recognition that unduly severe disclosure control methods could **compromise the level of detail needed in the data** for it to be useful for research and decision-making. The need to get the balance right is clearly important.

"We have to balance the small risk of working out real un-rounded figures against being able to publish useful statistics."

HESA's plans for moving to in-year data collections was underlined, with concern that in-year statistics, which may present different characteristics to end-of-year statistics, in the public domain could mislead data users. Further consultation in this area was welcomed to clarify outcomes.

Freedom of Information (FoI) Requests were cited as a potential benefit of HESA open data publication, by reducing the number of requests HE providers currently respond to in relation to data that are not in the public domain. However, information on the range and types of FoI requests currently received may need to be considered and fed in to the design of HESA open data resources, to ensure that this benefit is realised effectively.

Other concerns/comments included: Would HESA subscription fees increase to cover income lost from the sale of publications/bespoke data as a result of the move to open data? Will restrictions be provided for companies using the data to generate profit? As alternative providers will benefit from competitor information, should they conform to the same requirements for data submission as other HE providers? Widening access to the data that HESA hold would support the international engagement activities of UK HEIs and other UK education providers.

HESA'S RESPONSE

We received many excellent and thought-provoking responses to this question. There were a number of issues raised:

1. **Risks of misinterpretation of open data**
   This is a concern that applies to most open data published in a variety of subject domains. It is certainly a risk but one that can be mitigated through mechanisms such as publication of thorough definitions and contextual information, and the provision of presentation and support services whereby skilled analysts are available to explain data to users who may be uncertain. In addition to this, HESA will be prepared to comment when we feel data has been

misrepresented – both directly to the source of the material in question and publicly, where appropriate.

2. **Data protection and commercially sensitive data**
   These issues were both addressed in the consultation document, so it was good to see the importance of these reiterated in the responses. HESA aims to operate a gold standard of data protection compliance and we have long-established mechanisms for disclosure control. All open data we publish will be subject to the same level of care on disclosure control as our data publications have been for many years. Commercial sensitivity will also be carefully considered in the case of data that has not previously been published; this may be subject to further consultation as we progress through the period of our open data implementation. This does indeed need to be about balancing the requirements of compliance with the needs of data users for data that is sufficiently detailed for their applications. However, our existing mechanisms for access to detailed data under contract for research will not be replaced by open data – so the most detailed requirements for data derived from personal data collections will continue to be better handled through this process.

3. **In-year (student) data collection (Data Futures)**
   As in-year data collections become available we will need to consider carefully how this new data is reflected within our open data outputs. The point about presenting what may be a partial view based on early data is well-made. As the details of the new data collection processes become clear we shall undertake an assessment of impact on open data publication and publish our proposals. This may involve further consultation.

4. **Freedom of Information (FoI) requests**
   Any reduction in the burden of FoI requests on HE providers will be a welcome benefit of HESA open data publication. We shall, as suggested, explore ways in which intelligence on the nature and type of FoI requests received by the sector might usefully inform our open data plans.

5. **Commercial use of open data and impact on HESA subscription fees**
   Certain types of open data licence provide for the ability to block commercial use of open data resources. However, since one of the aims of open data publication is to stimulate innovation in data use, we feel it would be a shame to block an important source of innovation that could help create useful new products for a range of HE stakeholders, such as prospective students and their advisors, HE providers, policy makers etc. Our preference at this stage is therefore to utilise a variant of open data licence that requires products based on open data to be distributed under the same open data licence, rather than block commercial use entirely. Some respondents correctly noted that HESA generates revenue through the sale of products and services, which serves to ease pressure on the level of subscription fees required of HE providers. This will continue as we implement the open data strategy and beyond. Our current five-year corporate strategy includes both the move to open data and also a range of initiatives aimed at modernising and transforming HESA's other products and services to continue to support our revenue generating activities. The costs of delivering the open data strategy will be met from these income streams and NOT from any increase in mandatory subscription fees.

6. **Remaining issues raised**

   Alternative HE providers: the data reporting requirements for such providers are a matter for the UK HE funding bodies, Department for Education or the devolved administrations to determine; International student recruitment: we agree that provision of open data on international student activity will be helpful in supporting international engagement processes in the sector.

IS HESA PUBLISHING THE RIGHT OPEN DATA?

Do you agree with HESA's assessment of its data sources regarding suitability to publish as open data (see Annex A, p25) (Q2a)

86% of respondents agreed with the data sources HESA plans to publish as open data.

Figure 3: answers to Q2a
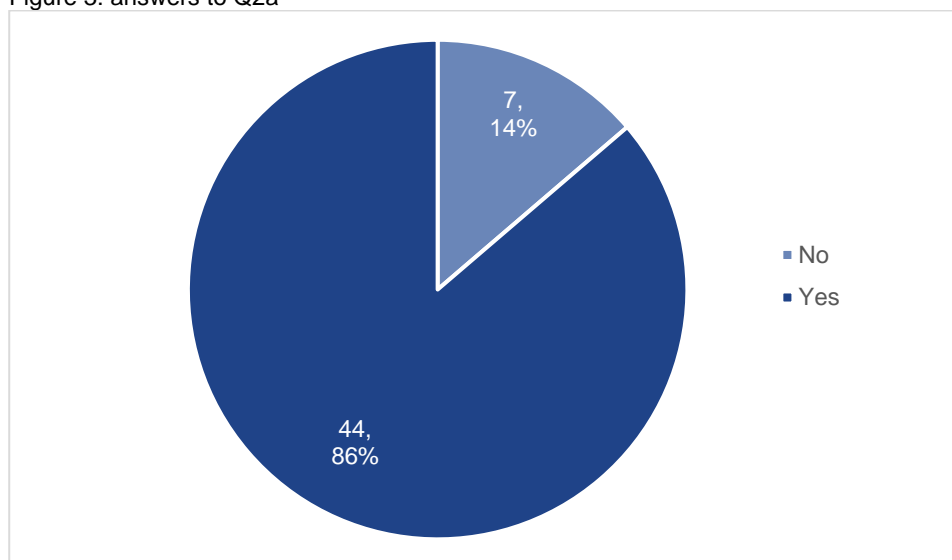
REASONS FOR DISAGREEMENT

If you don't agree with HESA's assessment of its data sources regarding suitability to publish as open data (Annex A, p25), please elaborate. (Q2b)

40% (20) of respondents answered this free-form question; feedback represents those who both disagree and agree in principal with HESA's assessment of its planned open data sources. Positive sentiment about the open data project was evident throughout the majority.

Concerns echo those from question 1b around the need to balance data protection methods with enough detail to make data valuable. Highlighted was the potential for analysis of inter-agency datasets for identification of individual/institution.

Respondents want further and more in-depth consultation to ascertain the content and detail/ level of disaggregation of datasets. It was also noted that as the range and type of HE providers with access to public funding expands, reviews should be made of the appropriate level of publishable information to ensure fairness to the whole sector.

HESA'S RESPONSE

Concerns about protection of individual confidentiality are addressed in the HESA response to Question 1b. The further point made about re-identification of individuals through combination of different data sets is a risk that is already addressed by the existing HESA disclosure control methods that will be used for all relevant open data publication.

With regard to the process of determining the precise content of each open data release, it is envisaged that this process will include further consultation to explore requirements and balance those against any concerns expressed by HE sector stakeholders.

HESA'S PLANNED OPEN DATA RESOURCES

There are two main classes of data and information which HESA proposes to publish as open data:

1. Data drawn from the annual or biennial HESA data collections of HE providers.
2. Meta-data such as coding frames, or information about HE providers.

The publication of HESA data sets as open data will promote their use and reuse, in time leading to new and innovative applications of data which HESA may not have the resources to develop itself. Publication of meta-data as open data allows others to structure their own data using recognised standard coding frames. This supports consistency and interoperability among different data sources, reducing costs and maximizing value from data for end users. Annex B in the consultation document (see p26) included a list of data and meta-data which HESA proposes to publish as open data.

IS THE LIST OF PROPOSED OPEN DATA RESOURCES APPROPRIATE?

Do you feel that the list of open data resources to be published in Annex B is comprehensive, or do you feel there are any other types of open data publication HESA should be planning? (Q3)

49% of respondents agree that the list is comprehensive. 22% feel there are other types of open data publication that HESA should be planning. 4% are unsure and 25% didn't respond, with free-form feedback suggesting they didn't have enough detail to make informed comment.

Figure 4: answers to Q3
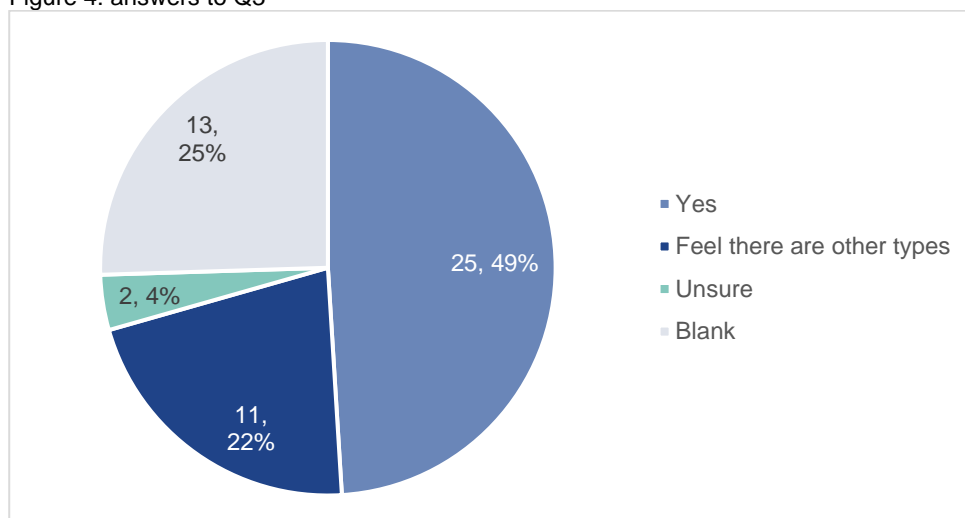
RECOMMENDATIONS FOR OTHER OPEN DATA SOURCES (Q3 CONTINUED)

On viewing Annex B, many respondents (up to 29%, as suggested from figure 4 above) didn't feel they knew enough about other HESA data sources to answer this question. However, almost 50% of respondents were happy with HESA's plans. Of all respondents who answered (over two-thirds), the majority offered further comment, summarised below.

Responses echoed those from questions one and two with reference to the need for appropriate contextualisation of data to avoid misrepresentation. Specific suggestions included: a comprehensive guide to understanding and using the published data; clear and accessible explanations on reasons for withheld data; historical data and documentation of changes made to collections and publications; a list of fields from original datasets that are suppressed/dropped through aggregation.

Specific requests for the clarification of data-table content items included: the exact data fields planned for release; the scope for joining datasets through common fields across HESA and inter-agency data tables; further data tables including data sub-sets on non-UK domiciled students;

There was re-iteration that further consultation is welcome over the detail and timing of data release, as well as the longer-term flexibility of the release schedule to incorporate additional data sets such as TEF.

HESA'S RESPONSE

The suggestions provided for types of supporting information to aid correct interpretation and application of open data are extremely helpful. We will consider these carefully and where possible build these into our publication plans.

Further detail on the specific content of each open data publication cannot be provided at this stage, as it has yet to be determined. Each open data release will be subject to a design process that includes consultation on the specific content; HESA will publish the proposals for content as early as possible. We will ensure that the open data publication schedule retains flexibility to incorporate additional data sets that may become available (and which are suitable for open data publication) over time.

SHOULD HESA PUBLISH META-DATA WITH THE DATASETS?

Do you agree that it is important for HESA to publish meta-data* as open data in addition to the data sets? *Meta-data includes coding frames or information about HE providers ((Q4a)

96% agree that HESA should publish meta-data as open data alongside the data sets.

Figure 5: answers to Q4a
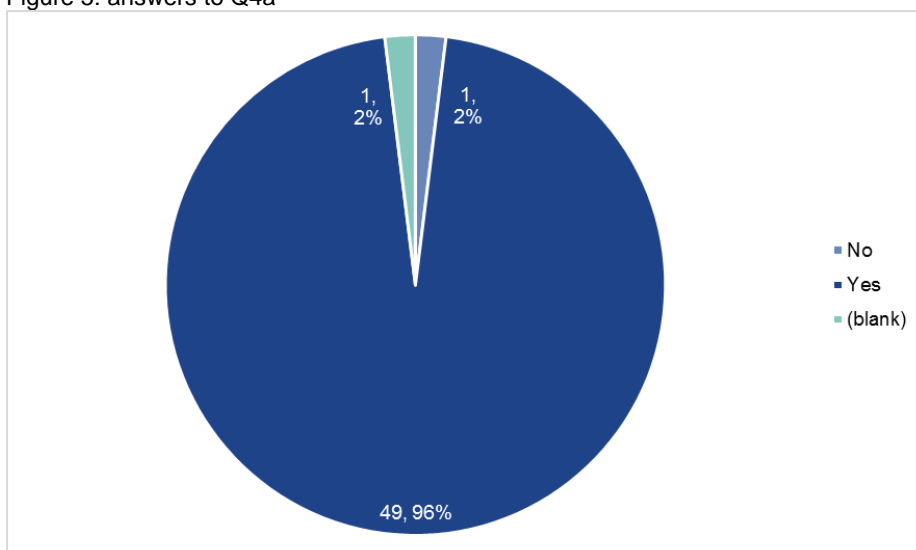
WHAT BENEFITS WILL META-DATA DELIVER FOR USERS? (Q4B)

"Providing meta-data is critical as it provides necessary context for users. Good metadata, alongside clear documentation, can help to reduce the number of queries received and reduce the risk of data being misinterpreted."

It was understood that meta-data will provide a "standardised framework in which to structure analysis" across the sector, increasing ease of analysis, data quality, confidence in the data, and minimising the likelihood of the data being misunderstood and misinterpreted. It will maximise data usage, improve interoperability and cross-mapping. Those not familiar with the HE marketplace would benefit especially and statisticians could link HESA data sets with other relevant data by using unique identifiers. It was stipulated that the meta-data be short, concise and easy to understand.

The respondent who disagreed with it being important for HESA to publish meta-data in this context queried the investment value vs usage for a potentially relatively small group of stakeholders.

HESA'S RESPONSE

It is reassuring to see that the vast majority of respondents supported our assertion that open publication of metadata is a crucial part of our open data strategy and will be important for effective onward use. It will require investment of resources to deliver but we feel strongly that the benefits will more than justify this investment.

EXPERTISE, MATURITY AND CERTIFICATION

Effective open data publication will involve the application of some specialist skills and knowledge in areas such as open data licencing and machine readable data and meta-data formats. HESA will develop the necessary expertise by working with the Open Data Institute (ODI) – utilising their extensive range of training and consultancy support services. We shall also seek ODI Open Data Certification. The certification levels are as follows:

- **Bronze:** data is openly licensed, available with no restrictions, accessible and legally reusable.
- **Silver:** satisfies the Bronze requirements, the data is documented in a machine-readable format, reliable and offers ongoing support from the publisher via a dedicated communication channel.

- **Gold:** satisfies the Silver requirements, is published in an open standard machine readable format, has guaranteed regular updates, offers greater support, documentation, and includes a machine readable rights statement.
- **Platinum:** satisfies the Gold requirements, has machine readable provenance documentation, uses unique identifiers in the data, the publisher has a communications team offering support. This is an exceptional example of an information infrastructure.

HESA will aim to achieve Gold level ODI Open Data certificates in three or more open data resources published within five years.

ARE HESA'S CERTIFICATION AIMS PITCHED CORRECTLY?

Do you feel that HESA's aims on ODI certification are pitched at an appropriate level of ambition? (Q5a)

69% of respondents agree that HESA's aims on ODI certification are pitched at an appropriate level of ambition.

Figure 5: answers to Q5a
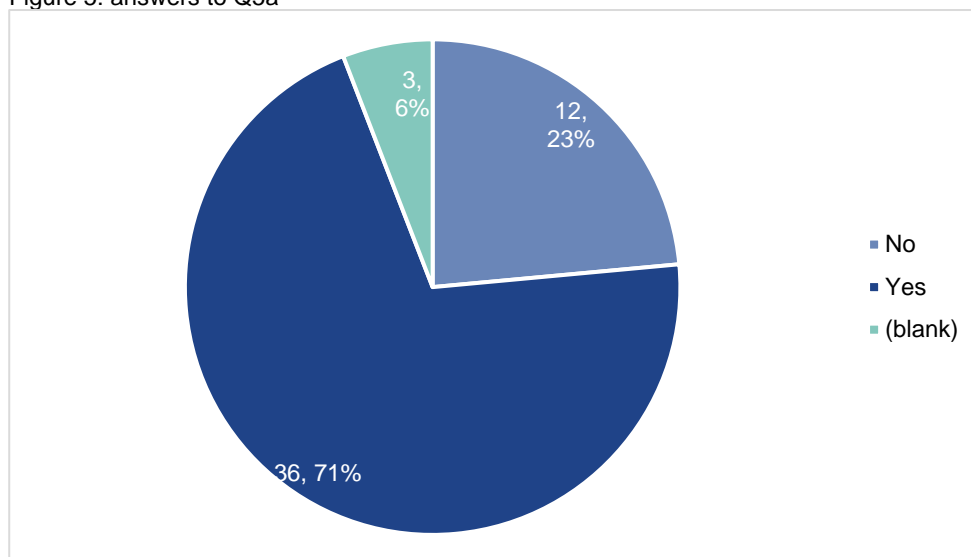
IF NOT, PLEASE ELABORATE ON THE REASONS FOR YOUR ANSWER (Q5B)

This question received responses from 39% (20) of the audience, therefore answers were from a group including those who agreed and disagreed that HESA's certification aims were pitched correctly (Question 5a). Looking at only those who answered "no" to question 5a, many said that that HESA should be aiming for the highest possible standard for all open

data sources, with onus on Gold and Platinum. The features required for the highest level of certification would facilitate better access and linking to other relevant data, as well as ensuring provenance and correct understanding of the data by users.

It was raised that **the data capabilities of users should be taken into account** when selecting accreditation level. To avoid misinterpretation, supporting materials conforming to the certification levels will need to be written for a 'lay' audience. It was also suggested that HE providers should have access to open data publications several months ahead of release to ensure data credibility and resolve issues.

Other queries and challenges included: five years being too long for a review period; why is HESA not proposing Gold certification for all open data releases, what exact data sources the certification will include; the rationale for different certification aims across collections; clarification of the frequency of proposed updates and support offered; development of a contingency plan including extended timelines and lower level certifications.

HESA'S RESPONSE

This question revealed some differences of opinion among respondents, but overall it is reassuring to see a majority in support of the proposed targets.

There were some valid concerns about the extent of the ambitions here. We understand these but would point out that HESA does not have unlimited resources to achieve the open data strategy, so we must balance our level of ambition against the costs of achieving it. Also we have no experience of achieving certification. As we work with the Open Data Institute to implement the strategy and gain certification, the full requirements and scope to reframe our ambitions will become clear. Put simply – if we can achieve more in less time, then we will.

The suggestion to provide HE providers with several months' prior sight of open data releases would not be defensible. Withholding data for such a period after it becomes available could not be justified. HE providers are offered a number of opportunities to review data during the existing data submission processes.

OPEN DATA LICENSING

A clear and recognisable approach to licensing of data using standard open licenses is an integral part of all open data publication. HESA will utilise either the **Creative Commons Attribution 4.0 International license (CC BY 4.0)** or the **Attribution-Share Alike 4.0 International license (CC BY-SA 4.0)**, depending on the data being licensed.

IS CREATIVES COMMONS ATTRIBUTION 4.0 THE MOST APPROPRIATE OPEN DATA LICENCE? (Q6A)

51% of respondents agree that it is the most appropriate open data licence for this project.

Figure 6: answers to Q6a
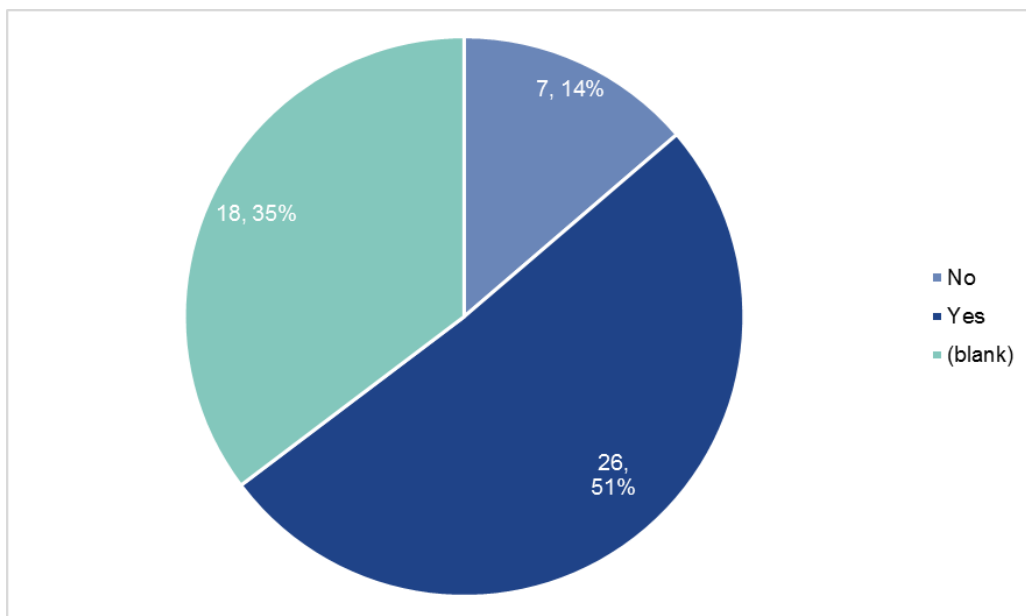
PLEASE EXPLAIN YOUR ANSWER (Q6B)

60% (31) of survey respondents answered this question.

The majority of respondents supported use of the Creative Commons Attribution 4.0 license, feeling that it will ensure users can share and adapt data for both private and commercial purposes, with appropriate attribution. This permits the widest range of re-uses. It will encourage users, both private and public, to make use of the data in a free and open manner. Issues with this included the right for free commercial use of the data; it was felt that commercial interests should pay a fee.

"Producing the data creates or increases costs for HESA and institutions and some recovery of these costs should be sought where others profit from them."

There was some agreement that CC BY-SA 4.0 (the Share-Alike variant) would be more suitable. HESA should seek the Attribution-Share Alike 4.0 International license (CC BY-SA 4.0 Share Alike 4.0) which sets conditions that any derivative work is made available under "the same, similar or a compatible license". The advantage being that it would allow verification of published findings that use HESA data and enable other researchers to build on published results. One comment queried the rationale for choosing this license over others, such as the Open Government Licence.

Other issues included: the implications for the Heidi Plus business intelligence system and current HESA data channels; the use of CC BY-SA 4.0 in terms of defining what a derived work means in this context.

HESA'S RESPONSE

Again, this question received a reassuring set of responses confirming that the proposed Creative Commons attribution licence will be a suitable approach. In terms of why we have selected this licence over others, quite simply, this is the one that has initially been recommended to us by the Open Data Institute. However, we will continue to work with them to explore all related issues including what 'derivative works' means in practice in the context of the CC BY-SA licence. We are open to reviewing the approach to licensing if these further discussions reveal a significant weakness with the Creative Commons variant.

Other HESA data dissemination channels such as Heidi Plus will be unaffected by the open data licensing approach, as these will continue to be delivered under separate licensing that is appropriate to the content and mechanism of dissemination.

CONSULTATION AND USER ENGAGEMENT

It is important to ensure that the needs and expectations of both data users and data suppliers (HE providers) are taken into consideration in the development and implementation of an open data strategy. The consultation document represented the first stage of user engagement. Further engagement will be needed on each data stream in due course to inform the design, content and format of each new open data release.

In most cases in the HE context, the professional communities representing data providers for each data collection are well-defined with clear communication channels. User communities are often less well-defined, although, in most cases, significant key users can be identified and consulted. There are other communication channels available with which to engage with a more diffuse group of users. However, in the case of some HESA data sets the community of users is less well-understood and communications may be more challenging.

In order to maximise positive impact of open data publication, appropriate support mechanisms must be provided for users. We will not require open data users to provide their names and contact details when they use our data releases as this would conflict with the ethos of open data. This suggests the need to establish new communication groups specific to our publication of open data, so that interested parties can optionally choose to register to receive news and updates on our publication progress and provide feedback to us on their user experiences. A number of mechanisms to establish such groups already exist and HESA is exploring other available options.

ADVICE FOR HESA ESTABLISHING OPEN DATA COMMUNICATIONS CHANNELS

Do you have any advice for HESA in establishing communications channels to open data communities and users? (Q7)

70% (36) of respondents offered advice here.

Sharing experiences, ideas and best practice was a key theme, through virtual and face-to-face events, discussions and conferences. The development of an active user community was popular, including news bulletins and updates on the HESA website, as well as regular blogging allowing comments/discussion. It was noted that HESA's website is currently difficult to navigate and understand (technical jargon).

Increasing engagement through real-life examples, demonstrating the value of information gained and analysis techniques. Providing context, guidance and definitions to aid interpretation of data (including limitations) was deemed essential. Clear documentation would make it as easy as possible for stakeholders to engage. Analytics training was also suggested through online tutorials.

Reaching different data-users (markets) was referenced, including senior managers, planners, researchers, operational stakeholders, as well as those working outside HE, including independent researchers and scholars.

In terms of the development of the data sets, use of data champions (organisations) was advised, also valuable for bridging gap between data users and HESA. A data reference group or user tester group was outlined to ensure datasets are fit-for-purpose and meet the needs of key organisations.

Respondents were keen for HESA to maximise the **communications channels and networking opportunities of stakeholders, including:**

- HE agencies - Association of University Directors of Estates (AUDE), BUFDG, Guild HE, HEDIIP, HESPA, Independent Higher Education JISC, Million Plus, NCUB, Nesta, Russell Group, UCAS, University Alliance, UUK;
- Data - Efficiency Exchange, Leeds Institute for Data Analytics, ODI, ONS's Twitter campaigns alongside data releases; Open Knowledge, UK Data Service, Wikidata;
- International agencies - Canada's Open Data Exchange (ODX); European Data Portal, Services for International Education Marketing (British Council);
- Business - Chamber of Commerce, Confederation of British Industry, Institute of Directors, SPARC, Wikimedia;
- Digital – Tableau.

Best practice data user groups were referred to as follows: Community Innovation Survey User Group; HE-BCI (Higher Education-Business and Community Interaction) survey user-group (run by National Centre for Universities and Business); BUFDG; Reference was made to the need for UK-wide provider representation via a user group to ensure that commercially sensitive data is not released.

Other channels mentioned were: Future Learn's training courses, as well as bite-sized get-started tutorials for new users and informal peer-to-peer training (platforms - github, RPubs, or as Jupyter notebooks).

Continued dialogue with the sector was, again, deemed important to clarify project outcomes.

HESA'S RESPONSE

Some excellent suggestions were received for this question. We shall explore all of these carefully and establish as wide a range of communication channels as we can.

ACTIONS AND TIMESCALES

HESA proposed the following actions in support of our open data aspirations:

a. **Build expertise in open data publication, linked data structuring and licensing processes.** We will achieve this primarily by working with the Open Data Institute, utilising their training and consultancy services.

b**. Establish a HESA open data champion.**
We will assign a lead role to a HESA staff member to provide a focus for the development of expertise, to be responsible for the implementation of open data strategy and to act as a champion for open data within the organisation.

c. **Migrate HESA's current range of data publications to open data, free of charge to users within the next five years.**
In so doing we shall consult with users on the content to be migrated, ensuring this meets user needs.

d. **Publish a number of HESA data sets in entirety as open data according to the list and timeline shown in Annexes B and C of the consultation document** (see p26 and p28). We shall consult with users and data providers, taking into consideration their views and preferences.

e. **Publish a range of HE meta-data as open and linked data according to the list and timeline in Annexes B and C of the consultation document.**

f. **Ensure all HESA published open data is available through the two main open data portals relevant to higher education data – data.ac.uk and data.gov.uk.**

g. **Establish a user community for HESA open data.**
This will enable users to feed in their views as an open data strategy is implemented and provides a mechanism through which HESA can provide support and assistance in using the data.

h. **Seek Gold level Open Data Institute certification on at least three open data resources within five years.**

i. **Undertake a review of the impact and benefits of the open data strategy at the end of a five-year period, to inform key stakeholders and the development of strategy for the subsequent period.** This may involve some independent assessment by the Open Data Institute or related organisations.

DO YOU THINK THE LIST OF PROPOSED ACTIONS IS APPROPRIATE AND COMPREHENSIVE? (Q8A)

Almost three-quarters (71%) of respondents agreed that HESA's proposed actions are appropriate and comprehensive. 23% disagreed and 6% left this section blank.

Figure 7: answers to Q8a
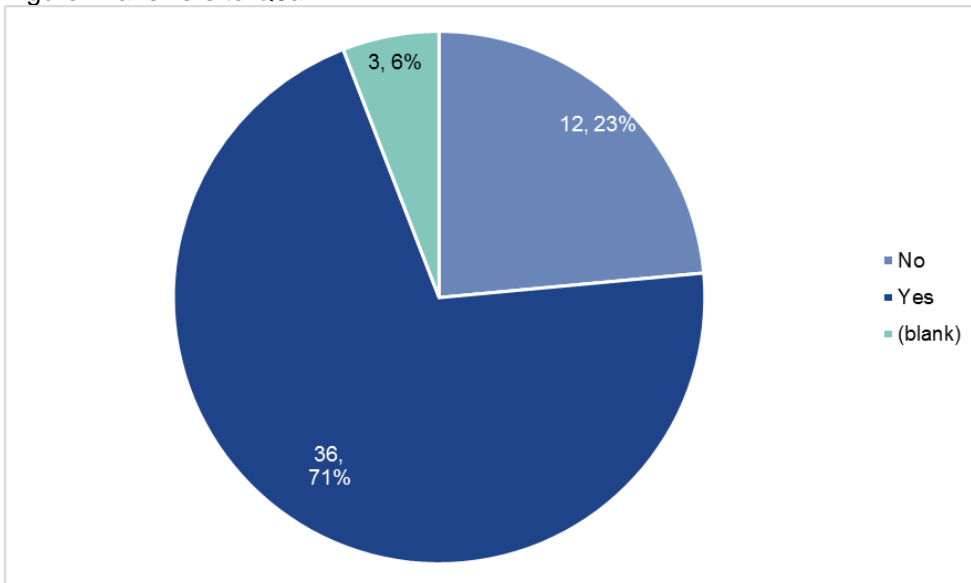
IF NOT, ARE THERE OTHER ELEMENTS THAT SHOULD BE INCLUDED? (Q8B)

15 respondents answered this question; this included three respondents who had answered "yes" in figure 7 as well as those who answered "no".

Concerns raised repeated those from earlier questions. An issue was raised over the rights of businesses to make profit from the data released in an open format.

"Open data should be free at the point of access, but that does not mean it must be free at the point of onward use. Where someone (or some organisation) uses open data to directly make a profit for themselves (e.g. in a service that is charged for at a commercial rate), that profit should be shared with the provider of the open data"

HESA'S RESPONSE

See response to question 1b on commercial use of data. We do not wish to stifle data innovation by the private sector, which could be beneficial for HE stakeholders. Thus, we do not favour a blanket ban on commercial users. However, the form of open data licensing used by HESA will not permit use of HESA open data releases to produce products which will only be made available on a profit-making basis. HESA has alternative licensing models available for this type of onward use, which would fall outside the remit of our open data strategy.

FURTHER COMMENTS

Do you have any other general or specific comments about HESA's proposed approach to open data? (Q9)

65% of respondents provided further comment, summarised as follows:

A significant proportion of responses reinforced the positive sentiment received in question 1 (figure 2). Many reiterated/expanded on concerns, while still crediting the project as being commendable. There was also acknowledgement that in a changing marketplace of significant uncertainty, "great care must be taken to ensure [HESA's open data project does] not conflict with institutional priorities, market forces and financial sustainability."

The future of the project was raised in terms of **"expanding the data provision past the point described in this consultation"** – mainly the assessment of new data sources and the inclusion/exclusion of data fields and sub tables. There was a question posed as to whether providers "should retain a right of veto over their data being included in a particular open data set" and a suggestion that they should have access to the data before it's published in an open format to ensure its validity.

The issue of funding was raised. It's likely the project will result in reduced income through HESA Services. How will this affect **provider subscription costs?** There is potential for devaluing of the subscription to HEIDI if the data is made available through both channels simultaneously.

"Whilst we are supportive of this move towards more open data in the sector we would not support an increase in institutional subscriptions to cover the costs of this work or loss of income."

There was recognition that **datasets need to be comparable across years** to validate the investment in analytics tools.

23

"Changes to the data structure of datasets are inevitable but effort should be made to design dataset structures to extend in future years while still being compatible with tools developed to work with earlier years."

Collaboration with other data services in the sector, such as UUK and UCAS was supported. Aligning identifiers wherever possible would enable the **linking of inter-agency datasets.**

CONCLUSIONS AND NEXT STEPS

The responses to this consultation have provided some excellent comments, suggestions, viewpoints and queries. These are enormously helpful in shaping our open data strategy as we move into the implementation phase. A great deal of detail was provided in response to many questions and we've only been able to summarise much of that in this document. However, we shall review all detail over the coming months as we address the relevant aspects of implementation.

The next stage is the publication of the final strategy document – but final in this context means informed by the feedback from this consultation. Over time the strategy will be reviewed and may change to reflect our evolving understanding of open data requirements and in response to the needs and expectations of our stakeholders.

No significant issues were raised in relation to the planned schedule for open data releases or the relevant topics, so will shall press ahead with the planned releases in Annex B (p26). Many of these will require more detailed consultations to inform their design and content.

A priority will be to establish the effective and relevant communications channels so that all those with an interest in HESA open data are able to keep up-to-date with developments and engage with the process.

ANNEX A

CLASSIFICATION OF HESA DATA SETS

	Data set name	Type	Notes
1	HESA Student Record AP Student Record	Personal data (including sensitive personal data)	Unsuitable for open data publication in raw form: disclosure-controlled data subsets to be published (see Annex B).
2	Aggregate Offshore Record	Non-personal data	Eligible for open data publication
3	HESA Destinations of leavers Early Survey	Personal data (including sensitive personal data)	Unsuitable for open data publication in raw form: disclosure-controlled data subsets to be published (see Annex B).
4	HESA Destinations of leavers Longitudinal Survey	Personal data (including sensitive personal data)	Unsuitable for open data publication in raw form: disclosure-controlled data subsets to be published (see Annex B).
5	HESA Staff Record	Personal data (including sensitive personal data)	Unsuitable for open data publication in raw form: disclosure-controlled data subsets to be published (see Annex B).
6	HESA Finance Record	Non-personal data	Suitable for open data publication
7	Higher Education Business and Community Interaction Survey	Non-personal data	Suitable for open data publication
8	Estates Management Record including Health & Safety information	Non-personal data	May include some elements considered commercially-sensitive. Also includes environmental information which is suitable for open data publication (and already published[11] by HESA).
9	Provider profile record	Non-personal data	Suitable for open data publication
10	Key Information Set (Unistats)	Non-personal data	Disclosure controlled data set already published[12] as open data

11 https://www.hesa.ac.uk/pubs/emr
12 https://www.hesa.ac.uk/unistats-dataset

ANNEX B

OPEN DATA PUBLICATION SCHEDULE

DATA SETS

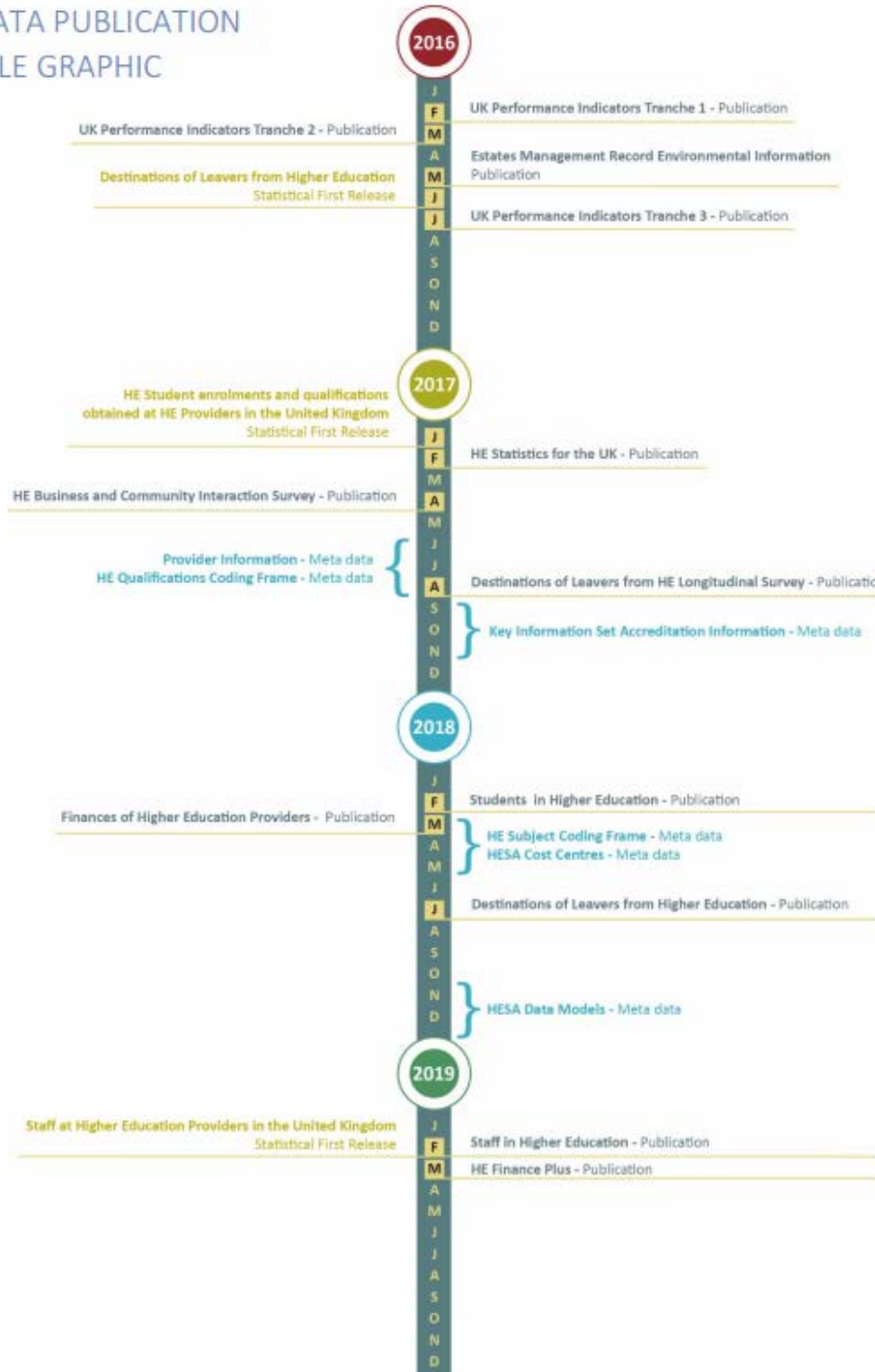|	Data set or publication title	Type	Open data resources to be created	When	Notes
1	Students in Higher Education	Publication	Set of content tables published on HESA website guided by user consultation	February 2018	
2	Statistical First Release - Higher education student enrolments and qualifications obtained at higher education providers in the UK	Publication	Format to be revised followed by publication under open licence	January 2017	National Statistics product
3	Staff in Higher Education	Publication	Set of content tables published on HESA website guided by user consultation	February 2019	
4	Statistical First Release – Staff at HE Providers in the UK	Publication	Format to be revised followed by publication under open licence	February 2019	
5	Finances of Higher Education providers	Publication	Entire HESA Finance Record to be published as open data	March 2018	
6	HE Finance Plus	Publication	Entire HESA Finance Record to be published as open data	March 2019	
7	Destinations of Leavers from Higher Education	Publication	Set of content tables published on HESA website guided by user consultation	July 2018	
8	Destinations of Leavers from Higher Education Longitudinal Survey	Publication	Set of content tables published on HESA website guided by user consultation	August 2017	Dependent on future of Long DLHE survey
9	Statistical First Release – Destinations of Leavers from Higher Education in the UK	Publication	Format to be revised followed by publication under open licence	June 2016	National Statistics product
10	HE Business and Community Interaction Survey	Publication	Entire survey data set to be published as open data	April 2017	
11	Estates Management Record – Environmental Information	Publication	Table of data as per current publication.	May 2016	Completed – latest publication published as open data via website
12	Higher Education Statistics for the UK	Publication	Format to be revised followed by publication under open licence	Commencing February 2017	Split into sections to be released at different points during the year

| 13 | UK Performance Indicators | Publication | Tables of data as per current publication | February, March, July 2016. Agreed that UKPIs for 2016 published under open data licence. |
| 14 | Unistats data set | Data set | Already published as open data | |

META-DATA

	Description	Type	When	Notes
1	Provider information	Meta-data	Summer 2017	Coding frames, geo information, groups
2	HE subject coding frame	Meta-data	Spring 2018	Based on HeCOS with aggregation schemes used for key applications
3	HE qualifications coding frame	Meta-data	Summer 2017	Qualifications on entry, course aims, qualifications obtained.
4	Key Information Set Accreditation information	Meta-data	Autumn 2017	Accreditation body and type of accreditation awarded
5	HESA Cost Centres	Meta-data	Spring 2018	Departmental groupings, potential for mappings to subject coding frames in the future
6	HESA data models	Meta-data	Winter 2018	Providing access to structure and design of HE data models

ANNEX C

OPEN DATA PUBLICATION
SCHEDULE GRAPHIC

2016

UK Performance Indicators Tranche 1 - Publication

UK Performance Indicators Tranche 2 - Publication

Estates Management Record Environmental Information
Publication

Destinations of Leavers from Higher Education
Statistical First Release

UK Performance Indicators Tranche 3 - Publication

2017

HE Student enrolments and qualifications
obtained at HE Providers in the United Kingdom
Statistical First Release

HE Statistics for the UK - Publication

HE Business and Community Interaction Survey - Publication

Provider Information - Meta data
HE Qualifications Coding Frame - Meta data

Destinations of Leavers from HE Longitudinal Survey - Publication

Key Information Set Accreditation Information - Meta data

2018

Students in Higher Education - Publication

Finances of Higher Education Providers - Publication

HE Subject Coding Frame - Meta data
HESA Cost Centres - Meta data

Destinations of Leavers from Higher Education - Publication

HESA Data Models - Meta data

2019

Staff at Higher Education Providers in the United Kingdom
Statistical First Release

Staff in Higher Education - Publication

HE Finance Plus - Publication