



CONSULTATION ON OPEN DATA PRINCIPLES AND REQUIREMENTS IN HIGHER EDUCATION

To help inform and shape the Higher Education Statistics Agency's approach and strategy for open data publication over the next five years

June 2016

#OpenDataHESA

Higher Education Statistics Agency Ltd (HESA) is a company limited by guarantee, registered in England at 95 Promenade, Cheltenham, GL50 1HZ. Registered No. 2766993. Registered Charity No. 1039709. Certified to ISO 9001 and ISO 27001. The members are Universities UK and GuildHE.

CONTENTS

Introduction.....	3
How to respond	5
Summary of consultation questions.....	6
HESA approach to open data.....	7
Accessibility and the data spectrum.....	7
Expertise, maturity and certification.....	9
Open data licensing	10
Consultation and user engagement	10
Actions and timescales	11
Annex A Classification of HESA data sets	13
Annex B Open data publication schedule.....	14
Annex C Open data publication schedule graphic.....	16

INTRODUCTION

OPEN DATA IS DEFINED SIMPLY BY THE OPEN DATA INSTITUTE AS “DATA THAT ANYONE CAN ACCESS, USE AND SHARE”.

In practice there are a number of dimensions to the process of releasing open data to optimize 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 2013². 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 UK⁴ in partnership with the National Union of Students⁵, Jisc⁶ and the Open Data Institute⁷. 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 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.

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/>

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, this consultation proposes 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

HOW TO RESPOND

This consultation document was released publicly on 16 June 2016. The consultation period runs until 23:59 on Thursday 28 July 2016.

Please respond to the formal consultation by using the online facility at <https://hesa.onlinesurveys.ac.uk/opendata>

We are pleased to engage with other communications on an informal and formative basis, but *only those responses submitted using the online facility will be recognised as formal submissions to the consultation* for the purposes of detailed analysis.

Respondents are requested to leave questions they do not intend to answer blank, to aid our analysis.

The online survey software will allow you to securely save your partially-completed response, and to download a copy of your submitted responses for future reference if you wish.

If you have questions about the process, please contact us using: opendata@hesa.ac.uk

SUMMARY OF CONSULTATION QUESTIONS

HESA wishes to consult our data providers, data users and other interested parties on the principles and details of the proposed approach. We seek responses to the following questions, which are set in context within the body of this document:

Question 1

- a. Do you support HESA's aim to make as much of our core data as possible available as open data?
- b. Please explain your answer

Question 2

- a. Do you agree with HESA's assessment of its data sources regarding suitability to publish as open data (Annex A)?
- b. If not please elaborate on any areas in which you disagree

Question 3

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?

Question 4

- a. Do you agree that it is important for HESA to publish meta-data as open data in addition to the data sets?
- b. What benefits will this deliver for users?

Question 5

- a. Do you feel that HESA's aims on ODI certification are pitched at an appropriate level of ambition?
- b. If not please elaborate on the reasons for your answer

Question 6

- a. Do you agree that Creative Commons Attribution 4.0 is the most appropriate open data licence for HESA to use?
- b. Please explain your answer

Question 7

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

Question 8

- a. Do you think the list of proposed actions is appropriate and comprehensive?
- b. If not, are there other elements which should be considered?

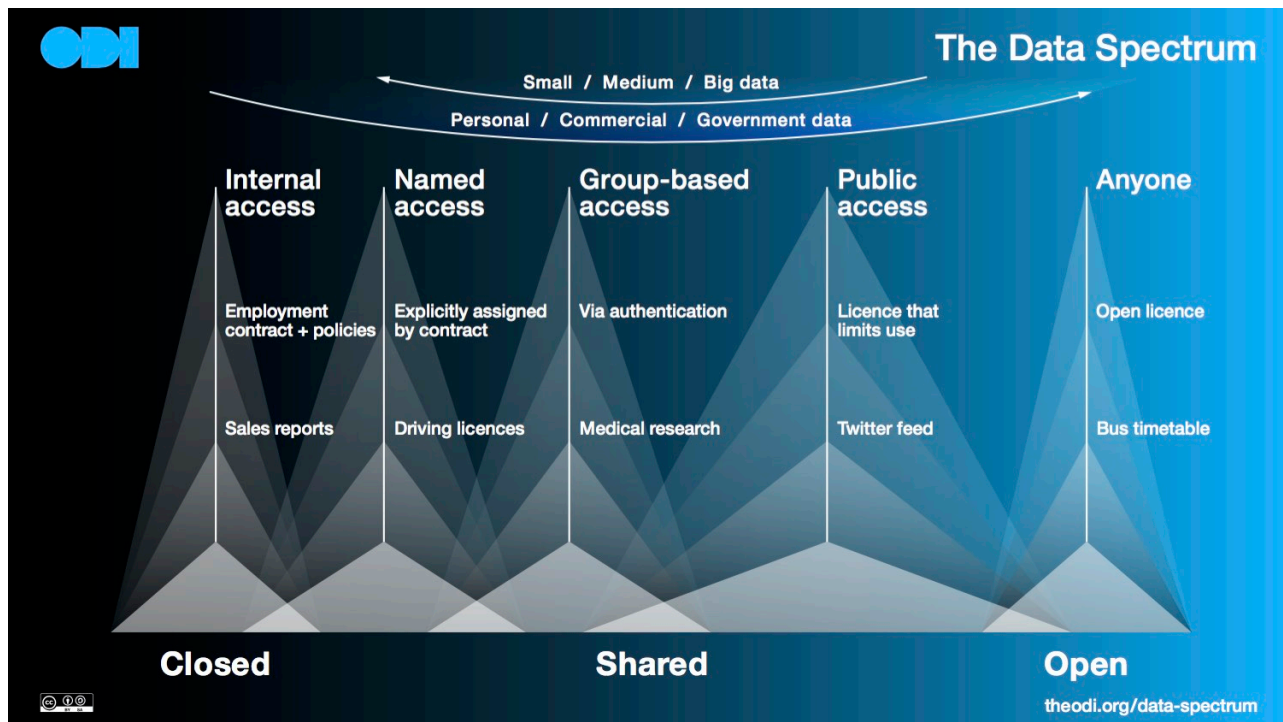
Question 9

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

HESA APPROACH TO OPEN DATA

ACCESSIBILITY AND THE DATA SPECTRUM

The Open Data Institute publishes 'The Data Spectrum' as shown 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 includes a classification of existing HESA data sets according to their suitability for open data publication.

Question 1

- a. Do you support HESA's aim to make as much of our core data as possible available as open data?
- b. Please explain your answer

Question 2

- a. Do you agree with HESA's assessment of its data sources regarding suitability to publish as open data (Annex A)?
- b. If not please elaborate on any areas in which you disagree

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 includes a list of data and meta-data which HESA proposes to publish as open data.

Question 3

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?

Question 4

- a. Do you agree that it is important for HESA to publish meta-data as open data in addition to the data sets?
- b. What benefits will this deliver for users?

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.

Question 5

- a. Do you feel that HESA's aims on ODI certification are pitched at an appropriate level of ambition?
- b. If not please elaborate on the reasons for your answer

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 Commons⁸ 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.

Question 6

- a. **Do you agree that Creative Commons Attribution 4.0 is the most appropriate open data licence for HESA to use?**
- b. **Please explain your answer**

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. This consultation document represents 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 maximize positive impact of open data publication appropriate support mechanisms must be provided for users. Since any form of compulsory registration in order to access data conflicts with the ethos of open data, HESA will not automatically possess user contact details in order to provide proactive support. 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.

Question 7

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

⁸ <https://creativecommons.org/licenses/>

ACTIONS AND TIMESCALES

ACTIONS

HESA proposes 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 Annex B.**
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 Annex B.**
- f. **Ensure all HESA published open data is available through the two main open data portals relevant to higher education data – data.ac.uk⁹ and data.gov.uk¹⁰.**
- 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.

⁹ <http://www.data.ac.uk/>

¹⁰ <https://data.gov.uk/>

Question 8

- a. Do you think the list of proposed actions is appropriate and comprehensive?
- b. If not, are there other elements which should be considered?

Question 9

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

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

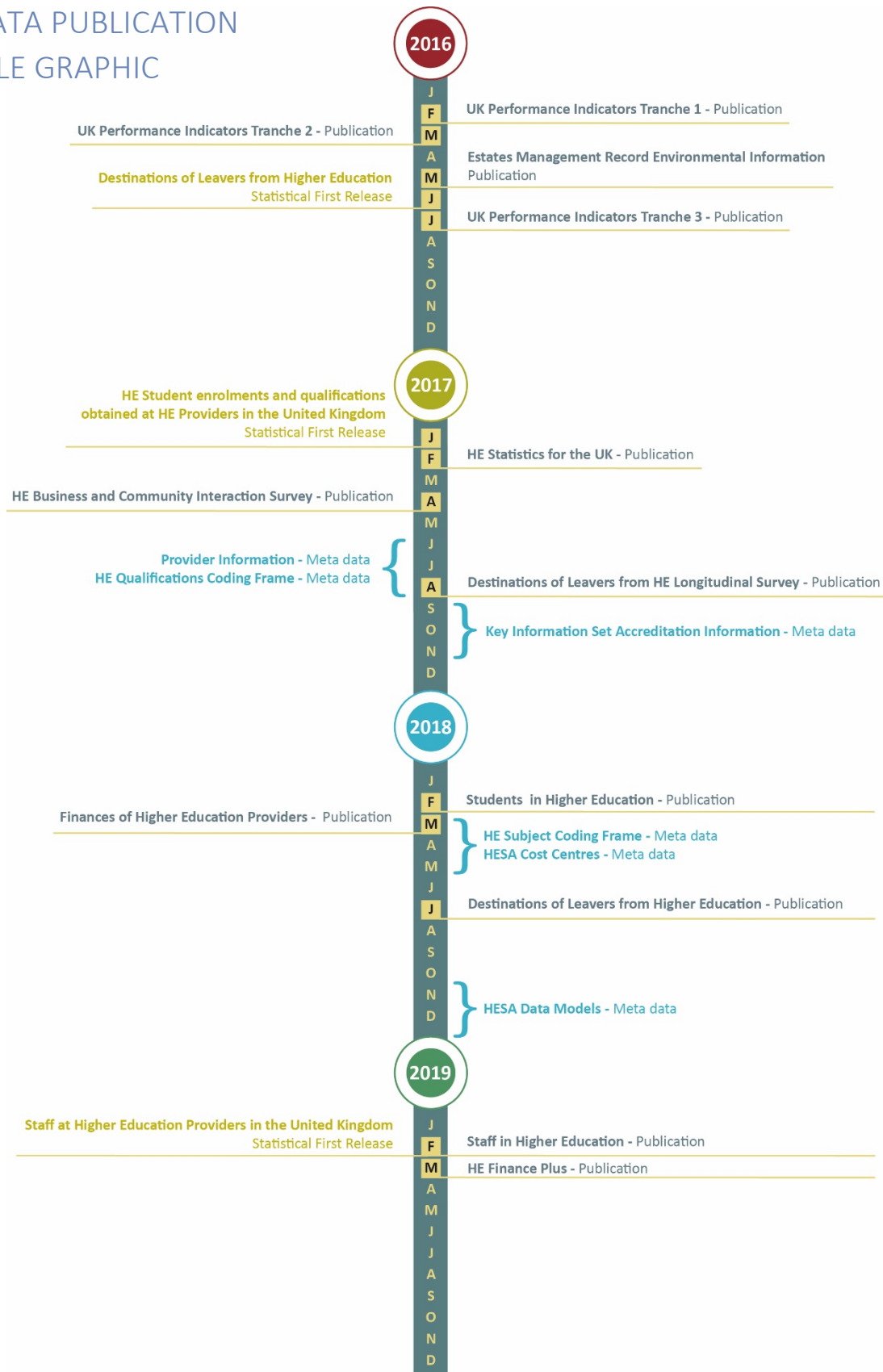
13	UK Performance Indicators	Publication	Tables of data as per current publication	February, March, July 2016. Agreed that UKPIs for 2016 published under open data licence.
14	Unistats data set	Data set	Already published as open data	

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





HESA
95 Promenade
Cheltenham
GL50 1HZ
www.hesa.ac.uk
+44 (0)1242 255577