HESA

Heidi Plus
Next Generation Education Analytics

HEIDI STAKEHOLDER GROUP

TUESDAY 28TH FEBRUARY 2017 11:00 – 2:30PM
HESA, 95 PROMENADE, CHELTENHAM

HSG_17_01_05

PRIORITISATION OF NEW DATA FOR HEIDI PLUS

INTRODUCTION

Since Heidi Plus was launched in November 2015, HESA has maintained a record of requests received from users (and other interested parties) for additional data content. This record aims to gauge the growing requirements and aspirations of users.

This paper sets out the requests received, split into three categories;

- Requested additions to existing data sources
- Requests for the inclusion of historical data that is not available within Heidi Plus but is available within Heidi
- Requests for additional new data sources to be included within Heidi Plus

Members of the heidi Stakeholder Group are asked to prioritise the importance of each of the requests, and to make any additional suggestions that are not shown on this list.

Please note that as the suite of information in Heidi Plus grows, so does the level of staff resource required to maintain it. By prioritising this list, this will enable HESA to determine where to focus any resource available to explore new data opportunities/improve existing data sources, but the level of resource available must be balanced against competing requirements.

HESA propose that we will take the recommendations from this group and use these to create a prioritised Heidi Plus data product backlog. When resource becomes available throughout the year, items from the product backlog will be worked on and released.

REQUESTED ADDITIONS TO EXISTING DATA SOURCES

Below is a list of formal requests received from users for improvements to existing data sources, along with supporting notes from HESA, the number of times the request has been formally received. We would ask the group to provide priority for these requests:

No.	Summary of request	HESA notes	Number of times formally requested	Priority (High/ Medium/ Low)
Destination of leavers				
1	Add postcode of employment dimension (only country and region of employment are currently provided)	Due to data protection reasons, this would need to be at County/unitary authority level	3	
2	Add dimensions to reflect domicile postcode district for the student	Due to data protection reasons, this would need to be at County/unitary authority level	1	
3	Add dimension to show actual salary to allow providers to calculate average salary	Actual salary is (de-facto) sensitive data and potentially unique for each individual in a population.	1	
Estates				
4	Some items returned within the Estates Management return are alpha-numeric, and therefore have not been carried over to Heidi Plus (as measures within Heidi Plus must be numeric). We have received multiple enquiries about this data no longer being available (examples are occupancy & frequency rate calculation etc.). We are told that these items are crucial for contextualising Estates Management Data, and users need to be able to access these.	HESA would need to work out how this data could be added to the structure of the current data source without causing confusion for users, however there is a strong case for this.	1	
Finance				
5	HESA have received many requests for improvements to be made to the structure of the HESA Finance data within Heidi Plus. Currently each table of the FSR, for each year, is held in a separate data source. Until recently it was not therefore possible to undertake timeseries analysis on data across years (this has now been addressed by the introduction of data blending in January 2017). However we would like to propose a review of the structure of this data within Heidi Plus to cover the following areas: a) consider whether there is a way of combining data from multiple finance tables into a single datasource b) consider whether there is a way of combining multiple years of data into a single data source, c) consider whether for tables 5a, 5b and 7 Cost centre/Academic departments could be made available as a dimension, d) make it clearer which combination of dimensions and measures need to be used to ensure data displayed is not double counting, e) ensure the data is structured in the best format for onward analysis.		6	

No.	Summary of request	HESA notes	Number of times formally requested	Priority (High/ Medium/ Low)
Performance Indicators				
6	Users have reported that they would like to be able to create Sector or Country averages for indicators within the Performance Indicator tables. This is not currently possible as results are held at provider level, and creating an average of pre-calculated indicators results in an erroneous result.	HESA would like the Stakeholder Groups guidance on this issue. Potentially this type of analysis is of questionable value - users should instead be using the benchmark values to compare each provider against the sector as a whole, however we need to understand in more detail what users are trying to achieve by attempting this type of analysis.	2	
Staff				
7	Add dimension Academic Teaching Qualification	Potentially there are data protection and data quality concerns with this data. Propose a Yes/No/Unknown category.	3	
Student				
8	Add dimension JACS (4 digit) to Student data	Data granularity would be an issue if this was added (rapid collapse to very small numbers likely.	3	
9	Add dimension Unistats subject grouping to Student data	This is possible to include, however the population used within Heidi Plus would differ from that used within Unistats (registering versus teaching institution)	1	
10	We would like the ability to identity distance learners within the student data	Suggest a distance learning/non-distance-learning marker flag for this.	1	
11	Add dimension Student Mobility to Student data	This would be challenging to add to Heidi Plus as it's a one to many relationship, would require a change to the structure of the data.	2	
12	Add dimension COMDATE to Student data	We would not be able to add actual commencement date for data protection reasons. However, it would be possible to provide a grouped version (for example Aug-Dec/Jan-Apr/May-Jul).	1	
13	Add dimension to identity Franchised students	Suggest a franchised/non-franchised marker flag for this.	1	
14	Add dimension LOCSDY to Student data	There are data protection concerns around this, potentially could provide a grouped version of location of study.	1	

No.	Summary of request	HESA notes	Number of times formally requested	Priority (High/ Medium/ Low)
15	Add dimension POLAR to student data		1	
16	Add dimension LPN (Low participation neighbourhood) to student data		1	
17	Add new Tariff band dimension to Student Qualifiers data to match TEF submission		1	

REQUESTS FOR THE INCLUSION OF HISTORICAL DATA THAT IS NOT AVAILABLE WITHIN HEIDI PLUS BUT IS AVAILABLE WITHIN HEIDI

HESA have been focusing over the last year on ensuring that as many data streams as possible are available within Heidi Plus, and on updating existing data streams with annual updates. However, Heidi Plus does not contain the full range of historical data available within heidi, and a small number of data streams have not been migrated across to Heidi Plus. We have received some concerns from users that when heidi is decommissioned in April 2017 they will no longer have access to this data. The most frequently requested historical data is HESA Finance and HE-BCI, with some users expressing a general concern about heidi being decommissioned without either having the opportunity to download historical data and store this locally, or it being available within Heidi Plus.

We would like the views of this group regarding the relative importance of the continued development of historical data, and whether resource should be prioritised to continue this work, and in which areas. With a finite amount of resource for Heidi Plus data development, we would ask the group to provide a steer on the priority of this work compared to the continued development of new data sources, current year updates to existing data sources, and the development of new dashboards. For reference, below is a comparison of data within Heidi Plus as compared to heidi (decommissioned data sources have not been listed);

Data source	Number of years available within heidi	Number of years available within Heidi Plus	Notes	Resource required per historical year added
HESA Student data	2001/02 to 2014/15	2007/08 to 2015/16	Population change between 2006/07 and 2007/08	15 days
HESA Staff data	2001/02 to 2014/15	2004/05 to 2014/15 (2015/16 in progress)	Population change between 2003/04 and 2004/05	15 days
HESA Finance Data	2001/02 to 2014/15	2013/14 to 2014/15 (2012/13 and 2015/16 in progress)		4 days

Data source	Number of years available within heidi	Number of years available within Heidi Plus	Notes	Resource required per historical year added
HESA KFIs	2001/02 to 2014/15	2004/05 to 2014/15 (2015/16 in progress)		1 day
HESA HE-BCI	2008/09 to 2014/15	2014/15	Part A not included in Heidi Plus as Alpha-numeric	4 days
HESA Destination of Leavers data	2002/03 to 2014/15	2011/12 to 2014/15	Population change between 2010/11 and 2011/12	15 days
HESA Estates Management Statistics	2004/05 to 2014/15 (data collected by IPD Global also included for years 2004/05 to 2008/09)	2012/13 to 2014/15	IPD Global collected data until 2008/09, then fundamental collection changes between 2011/12 and 2012/13	3 days
HESA Performance Indictors	2001/02 to 2014/15	2004/05 to 2014/15 (trance 1 updated for 2015/16)	Number of years does vary depending on table	2 days per PI table
HESA Student Staff Ratios	2001/02 to 2014/15	2004/05 to 2014/15		1.5 days
HESA Aggregate Offshore	2001/02 to 2014/15	2007/08 to 2015/16		10 days
HESA Provider Profile data	2012/13 to 2014/15	2012/13 to 2015/16		n/a
National Student Survey data	2004/05 to 2014/15	Coming soon		n/a
SCONUL data	2008/09 – 2013/14	Coming soon		n/a
RAE and REF data	2000/01, 2007/08, 2013/14	None		3 days
UK Patterns data	2006/07 to 2014/15	None		1.5 days

REQUESTS FOR ADDITIONAL NEW DATA SOURCES TO BE INCLUDED WITHIN HEIDI PLUS

HESA have been approached by the TRAC Development group to discuss the inclusion of TRAC data within Heidi Plus.

Following initial discussions, the group have written a proposal which we would like this group to consider (see Appendix 1).

This group is asked to consider this and to advise HESA on whether the proposal is supported, and if so, the relative priority that should be attached when compared with other development items detailed above.

Proposal to add TRAC data to HEIDI+

Summary

The HEIDI+ Stakeholder Group is asked to consider a proposal to include the TRAC sector-level and benchmarking data on HEIDI+ as a silver-level dataset. This would require 3.5-4 days' work from the allocation for developing external datasets.

Background

The Transparent Approach to Costing (TRAC) produces an annual set of sector-level data and benchmarks about costs for teaching, research, and other activities. This process is managed by HEFCE on behalf of the UK Funding Councils.

The TRAC Development Group (TDG) has been engaging with the Higher Education Strategic Planners' Association (HESPA) over the past year to encourage the use of TRAC data by planners. Greater use would increase the informational return on the significant investment the sector makes into compiling TRAC each year.

The data are currently made available to institutions via the HEFCE extranet, via the Finance institutional contact point.

Recommendation

Out of TDG's work with HESPA has emerged a strong recommendation to include the TRAC dataset on the HEIDI+ portal.

The proposal has been made and strongly supported by planners who have attended HESPA-sponsored sessions on "TRAC for Planners", and is being put forward by Olivia Kew-Fickus, the HESPA representative on the TRAC Development Group and Director of Strategic Planning at the University of Birmingham.

A series of conversations between Heather Williams (HEFCE), Jonathan Waller (HESA), and Olivia Kew-Fickus has confirmed that making TRAC data available on HEIDI+ would be in line with HESA's strategy for the development of HEIDI+ and would be acceptable to HEFCE, as the owner of the dataset.

Details

A review of the dataset by HESA has confirmed the technical feasibility. HESA estimates that the initial upload of the data would take 3.5-4 days' effort. This would need to be allocated from the allocation for HEIDI+ development work.

The following parameters have been agreed between HESA, HESPA, and HEFCE:
The dataset should be classified as a Silver Level dataset, in line with other financial datasets.

Heidi Plus is a service provided by HESA.
Higher Education Statistics Agency Limited is a company limited by guarantee, registered in England at 95 Promenade, Cheltenham, GL50 1HZ.
Registered No. 02766993. Registered Charity No. 1039709. Certified to ISO 27001. The members are Universities UK and GuildHE.

Page 6 of 7

The initial upload would include data for 2013/14, 2014/15 and 2015/16 (the latter becoming available in April 2017).

Access would be limited to the Higher Education Providers who make TRAC returns, but not Alternative Providers or other sector bodies.

Process for encouraging use

Through training on "TRAC for Planners", HESPA is developing a core group of planners with an interest in accessing these data. The launch of the dataset would be promoted amongst these individuals, although it is anticipated that the data becoming available should stimulate further interest.

TDG and HESPA will develop guidance for how to use the dataset, as it is not a familiar set to most planners and it differs substantially from HESA-generated datasets in a number of ways (e.g. no institutional data, only benchmarks, available in the Heidi Plus dataset; no cost centre breakdown for non-teaching data).

Timescales

Given the work currently underway, we would like to have the TRAC dataset available alongside the other 15/16 data.

There is a "TRAC for Planners" training session scheduled for 17 May 2017; it would be ideal, although not obligatory, if the dataset could be available for and launched as part of this session, as it will stimulate immediate uptake. However, a later release date could also be accommodated if this were not possible.

Future developments

We expect that as planners use the dataset, they will learn more about the kind of reporting which would be useful. We therefore suggest that in time it may be beneficial to discuss with HEIDI Labs whether further analytical work could add value; however, this should be a second stage rather than part of the initial development.

Olivia Kew-Fickus
Member, TRAC Development Group
Director of Strategic Planning, University of Birmingham
February 2017