Data maintenance policy   HSG/14/01/01

Background

Since its launch in April 2007, the data content of heidi has grown substantially, at a rate that has outstripped any increases in running costs passed on to subscribers (and therefore any concomitant increase in staffing to maintain the data content). During this time, no data resources have been removed from the system leading to a situation at present where, despite improvements in efficiency of processes, little or no spare capacity exists to explore and develop new data additions to heidi. At the last meeting of the Stakeholder Group it was agreed that a draft policy on data set discontinuation would be presented for approval. This new policy seeks to make best use of the staff resource devoted to maintaining the content of heidi through identification of where their efforts generate benefits for the highest proportion of users. Put simply, staff time spent maintaining data that are seldom used is probably not time well spent. Therefore a structured approach is required in assessing when datasets that are not well-used should be discontinued, so that the heidi Service Team can spend time maintaining and developing those data that are well-used (or could be well-used in the case of entirely new data types). Equally a data maintenance policy must address the circumstances and process by which new data sources are explored and developed for heidi.

Staff resource liberated by data set discontinuation would be diverted into exploring and developing new data sources as listed in Appendix A, in the priority order previously identified by the Stakeholder Group.

This paper contains the proposed policy for consideration and approval, a breakdown of data usage by dataset for the calendar years 2012 and 2013 so that the policy may be applied, and the list of new data sources.

Desired outcome

We invite the Stakeholder group to:

- **Consider the draft policy and suggest any amendments and additions**
- **Approve the data maintenance policy (subject to any agreed amendments and additions)**
- **Apply the policy using the data usage statistics in Appendix B to assess whether any datasets should be discontinued.**

Proposed approach

New data

The overall approach for identifying and developing new data would be to collect suggestions from users throughout the year and to seek prioritisation of these from the heidi Stakeholder Group. This follows previous practice.

Discontinuing data

The suggested approach is based on an annual review of heidi data usage metrics. Definition of a metric or set of metrics which provide an indication of relatively low usage will be agreed by the Stakeholder Group. Such metrics would be designed to 'flag' data sets (or possibly data fields) as candidates for discontinuation. The heidi Stakeholder Group would then be responsible for reviewing the metrics and the flagged data sets and making recommendations to HESA as to which may be discontinued. In this regard the Stakeholder Group may wish to consider wider implications and issues of use beyond the usage metrics in reaching a view. The group must also consider the level of staff resource required to generate data sets and consider to some extent the cost versus benefit.

Precisely what is meant by 'discontinuation' in practice must also be made clear. For any pre-existing dataset the main resource requirements are on the annual updating process. There is also a lesser draw on resources from continued provision of user support. Clearly discontinuation must mean the cessation of any further annual updates to a data set. One might stop there and simply leave 'legacy data' in the system indefinitely but there are other costs and disadvantages associated with this, including the need to maintain user support for the data, the need to carry such data over to any new environment and the performance penalty caused by excessive amounts of data held in the system. Conversely users may have existing reports defined on data scheduled for discontinuation, so immediate removal can be problematic. Therefore in practice, discontinuation must involve a combination of cessation of update followed by later physical removal of legacy data. Of course adequate notice periods must also be provided to users of any major change.

Draft Heidi data maintenance policy

Objectives

This policy sets out to establish a set of principles and practices aimed at ensuring that:
1. the data content of heidi remains appropriate and relevant to the user community,
2. is regularly assessed to ensure continuing relevance
3. the available staff resource for data maintenance is deployed efficiently and provides best value for heidi users.

The regular assessment will be undertaken in consultation with the heidi Stakeholder Group, whose role is to represent the user community and advise HESA on matters of content management.

The policy covers the circumstances under which:

1. New candidate data sets may be identified, assessed and scheduled for addition to heidi and
2. Existing heidi data sets may be assessed, those attracting poor usage identified, and decisions made as to any corrective actions required.

New heidi data sets

1. Candidates for new heidi data sets, or additions to existing data sets may be identified through suggestions from users throughout the year, suggestions from members of the heidi Stakeholder Group or suggestions from HESA staff. All such suggestions will be logged.

2. At each meeting of the heidi Stakeholder Group, the list of suggestions for new data will be reviewed alongside an estimate from HESA of staff time required to implement and the Group will be asked to approve or reject each one. For those suggestions that are approved the Group will then indicate a priority level for each.

3. HESA will then allocate resources as available to further explore and develop each of the candidates for new data in the priority order assigned; referring back to the Stakeholder Group as necessary if any new information on cost or feasibility of implementation comes to light.

Existing heidi data sets

1. At the first annual meeting of the heidi Stakeholder Group an assessment of data set usage levels will be undertaken. This will be informed by a set of standard usage metrics generated by HESA. These metrics will cover each of the previous two full calendar years of usage, and for each calendar year of usage the metrics will cover usage of the most recent two annual updates of each dataset. For example, figures for 2013 would address the usage levels occurring in 2013 of dataset updates for 2013 and 2012 in heidi. The usage metrics will include the following elements:

    a. Total number of reports viewed in period that contain at least one column from each data set
    b. Number of organisations viewing one or more of these reports
    c. Number of users viewing one or more of these reports
    d. An indication of whether the reports in (a) include standard heidi reports (which may influence the likelihood of viewing)
    e. Estimate of person-days of HESA staff time to produce and quality-assure each annual update of data set
    f. Percentage of user organisations that have viewed reports (b)

    g.   Low usage indicator – to flag those data sets where less than 15 users (c) AND fewer than 10% of organisations (b) have viewed reports.

2. The Stakeholder Group will then assess these metrics, in particular those flagged by the low usage indicator (g). The Stakeholder Group will be asked to recommend any data sets for discontinuation, taking into consideration the usage metrics, the staff resource required for annual updates, and any wider issues of perceived strategic or operational importance of data sets, or likely changes in importance through upcoming funding or policy changes.

3. If a data sets is flagged as low usage but is considered strategically or operationally important (or is likely to become so) the Stakeholder Group may wish to propose other corrective actions, such as further promotion or training provided to particular groups of users aimed at increasing usage levels.

4. Consideration for discontinuation may also include the following:
    a. Whether the data set has a niche audience
    b. Whether the maintenance of the data set is funded by a third party and therefore does not impact on the staff resource funded by HESA subscribers

5. Data sets recommended for discontinuation will be subject to the following process:

    a. Wherever practicable, 12 months prior notice of discontinuation will be provided to users
    b. Following the notice period no further annual update of that data set will be undertaken.
    c. Existing data sets in the heidi system will be marked as retired.
    d. In the case of non-HESA data sets, user support for the retired data set will be withdrawn 2 years following marking as retired.
    e. 4 years following marking as retired all editions of the data set will be removed from the heidi system

6. This policy recognises that in some cases the decision to discontinue a non-HESA data set may be made by the data originator and be beyond HESA's control. In such cases the notice period of 12 months may not be practicable.

Appendix A: List of new data for heidi

Following the last Stakeholder Group meeting, a list of suggested new datasets for heidi was circulated and members were asked to prioritise the importance of each for inclusion in heidi.

The results of this exercise are provided below for reference, with the outcome of the prioritisation exercise in the final column.

Category	Notes	Previous priority (H/M/L)	Importance as ranked by Stakeholder Group (H=3, M=2, L=1, higher the number greater the importance)
More data by UCAS subject line		High	11
Domestic & International League Table data		Med/High	10
KIS data	See HSG paper 13/02/05 (please note that this paper was approved by the Stakeholder group on 3 October 2013)	High	10
Fees data (Gross and net Fees)	New data collected by HESA from 2012/13 onwards. Analysis is currently underway to establish robustness of this data for publication.	N/A	10
British Council International data	The BC obtains these data from HESA. There is a difference in population used in the data we supply (theirs includes incoming and visiting exchange students) and there is additional detail in nationality, course aim, age, major source of tuition fees and location of study. Currently checking how much of this is made available from their online tool. Question as to how much of this would be useful to heidi users and whether usage levels would justify the allocation of heidi resource.	High/Medium	9

Research Council data	Initial discussions with the Research Councils have taken place. They are developing a Gateway to Research web portal which will provide access to data on research awards, studentships, fellowships and research outcomes. This is due to be launched at the end of 2013. One of the objectives of this is to have an open data set available for the general public. The preference of the Research Councils is for HESA to obtain this data set when published for inclusion in heidi. The data set will be in CERIF format. The Research Councils do not have the resources to develop an alternative specific set of data for inclusion in heidi in advance of that. Data on research applications is considered to be confidential and could not be included within heidi.	N/A	8
Research Impact data		N/A	8
Research Awards & Research Applications			8
IT data for HE published by UCISA (HEITS)			7
More detailed Aggregate Offshore data	Country in which provision is being delivered appears to be the only element of the Offshore Record that is not currently in heidi.	Medium	6
Home Office (previously UK Borders Agency) International Student data	Initial discussions on this have taken place. The inclusion of this data in heidi is not feasible at this time.	Revisit at a later date	
TRAC data	Significant barriers remain.	Revisit at a later date	
Institution profile information	First collection of data mapping cost centres onto departmental structures for 2011/12. We are planning to develop some proposals in consultation with HEI's.	Revisit at a later date	

Appendix B: heidi data set usage metrics – to follow