

HEAT data retention schedule

This document sets out the long-term retention and use of data collected on the HEAT database. The data entered into the HEAT database remains the property of individual institutions, but this model will ensure the largest cohort possible for the duration of the longitudinal research. HEAT will work to facilitate this plan on behalf of members, but it will remain the responsibility of individual member institutions to implement it. Should a member choose to use their own retention policy, they will also be responsible for implementing it.

This document aims to clarify for members:

- The law with regards to data retention
- The retention period needed to fulfil HEAT's long-term research aims
- Any financial economies of data management that can be made
- How we are going to do it
- What the next steps are

Legality

The UK is now subject to the General Data Protection Regulation (GDPR), which has superseded the Data Protection Act of 1998. Under these new regulations, data which permits identification of data subjects (outreach participants in this case) can be kept for **no longer than is necessary** and only for the purposes for which the data was collected and processed in the first place.

HEAT recognises that member universities may be bound by the interpretation that their individual institution places on the GDPR. We also know that members need to feel reassured that HEAT data retention is necessary and that research purposes will not change. Therefore, this document sets out a retention schedule informed by HEAT's long-term research aims.

Research aims

HEAT's overarching research programme aim (or GDPR 'purpose for retaining data') is to investigate the effect of outreach on individuals in terms of their social mobility, attainment, progression and graduate outcomes, including employment.

The table below illustrates the longitudinal nature of this endeavour using an example of a Year 9 'Student A'. If this student has **permission to track** and takes a straightforward and orthodox path through pre- and post-16 education into HE and on into employment, it would take 17 years to track them into their 10th year of employment.

EXAMPLE

School/Education Year	Cumulative years on database	Tracking eligibility
Student A is first recorded on the database in school Year 9	Academic year recorded on database	
Student A reaches Year 11	=Academic year recorded on database +2 years	KS4 data
Student A reaches Year 13	=Academic year recorded on database +4 years	UCAS data
Year 1 HE (Entry)	=Academic year recorded on database +5 years	HESA Track
Year 3 HE (Graduation)	= Academic year recorded on database +7 years	HESA Track
Year 1 of employment or further study	= Academic year recorded on database +8 years	HESA Track or Longitudinal Educational Outcomes (LEO)
Year 3 of employment or further study	= Academic year recorded on database +10 years	LEO
Year 5 of employment or further study	= Academic year recorded on database +12 years	LEO
Year 10 of employment or further study	= Academic year recorded on database +17 years	LEO

Proposed retention schedule

The HEAT retention schedule will allow us to decide annually which students should be:

1. Permanently anonymised
2. Deleted
3. Or remain on the database

Deletion

Students who have not given their permission to track will remain on the database for 10 years after the academic year their record was created. This 10-year period allows members to use the student's data for monitoring (where permission exists) or gives members time to gain permission to track from these students if they work with them as part of a sustained programme of activities. At the end of this 10-year period, they will be deleted (since we are unable to track them and we would be storing their details for longer than is necessary).

Students who have given **permission to track** but don't have a complete record (valid **DOB**, valid **Postcode**, **First Name**, **Last Name**) will also be deleted 10 years after their record was created.

Anonymisation

Where a student is less than 18 when they are added to the database, they will be tracked through HESA when they reach 18 and we will continue to track them for 10 years. If they have not been found in the HESA data within this 10-year period they will be anonymised. If they are found to have enrolled in HE, they will be given seven academic years to graduate. If HESA data shows that they have not graduated in this seven-year period, their record will be anonymised. Those that have graduated will be retained for 15 further years in line with LEO data (which currently plots 15 years after graduation). At the end of 15 years, the data will be anonymised.

Where a student was 18 or over when they were added to the database, they will be tracked through HESA for 10 years from their **HE Ready Year** or the date their record was created, whichever is later. If they are found to have enrolled in HE, they will be given seven academic years to graduate. If HESA data shows that they have not graduated in this seven-year period, their record will be anonymised. Those that have graduated will be retained for 15 further years in line with LEO data.

Under our current research programme, we propose that retaining personal data longer than 15 years from graduation – or longer than 10 years where individuals have not been found in HESA data – cannot not be justified. Although the new GDPR allows for retention of personal data for ‘longer periods’ for the purposes of research, it also states that if research purposes can be fulfilled by a data set which has been anonymised, then it should be anonymised.

The data would need to be anonymised to the extent that it would be deemed we had taken all necessary measures to prevent any reasonable likelihood of an individual being identified, after which the data is no longer covered by data protection law. We will investigate secure storage and access to this extract, and we will contact the UK Data Service as the potential archivist. The HEAT team would work with members to decide the specification for this secondary research extract.

Incomplete records or individuals with ‘no permission to track’

The administrative datasets to which HEAT matches individuals require a complete record to facilitate matching. Going forward, any individuals who do not have a complete record on the database will not be tracked by HEAT, even if they have given their **permission to track**. They will remain on the database for 10 years from the academic year in which they were entered onto the database. After this 10-year point, their record will be deleted.

Those individuals who are entered onto the database who **have not** given their **permission to track** – but **have** given consent to retention of their personal data on the HEAT database for **permission to monitor** – will also remain on the database for 10 years from the academic year of entry onto the database. After 10 years, their record will be deleted.

Financial economies

We must bear in mind that the GDPR requires us to assess the proportionality of how long it is appropriate to retain personal data. We must, therefore, be realistic about the chances of ‘finding’ individuals in the HESA Track after a significant passage of time, and of drawing conclusions about the effect of outreach on progression when a long period of time has elapsed between the two.

There is also a financial incentive in choosing to stop ‘looking’ in the HESA data in line with an agreed schedule. HESA Track costs grow exponentially as more individuals are added to

the database and become eligible for tracking. Removing individuals from the Track in line with the proposals above would help us to manage these costs.

Managing data subjects' expectations

A new requirement of the GDPR is that privacy notices include data retention periods. We remind members of the need to review and update privacy notices to ensure their compliance with this, whether they decide to use the HEAT retention periods or different periods specified by their data protection officers.

Review

This is the second version of this document, in line with our commitment to reviewing this policy annually.

This schedule does not fundamentally change the original purpose of collecting and processing the data or the agreed HEAT retention periods.

