

HET Data Retention Policy – proposed draft 31 Oct 2019

This policy will become effective on January 1, 2020

Until this point the HET has not deleted any images taken from any of the cameras since the new tracker was put in place. Under clear and stable conditions, we regularly collect **over 125 GB per night of science and engineering data per night**, and are looking at >35 TB being stored on site. Given this growth rate of data, it is unsustainable to continue keeping everything indefinitely on site. Toward this end, the HET Board has designated funds to support the use of TACC to provide long-term storage and backup of this large and valuable data set.

The following are the timeframes for keeping the raw data on site in /hetdata/data/:

- VIRUS: keep raw data on site for **400 days**, then decollate*
- LRS2: keep raw data on site for **400 days**, then decollate*
- HPF: keep raw data on site for **400 days** (on /hetdata/data)
 - No data deleted from HPF machine(s)
- Engineering cameras: keep raw data on site **indefinitely** (small footprint)
(ACQ, GC1, GC2, WFS1, WFS2, CWFS, PVW, BIB, HPFACAM)

*After 400 days, VIRUS and LRS2 raw data on site will be decollated (delete the data extension from the FITS files, but keep the headers – reduces footprint by ~100x), which will allow us to investigate possible problems going back >400 days without storing all of the raw spectral data.

All data from /hetdata/data are currently copied to TACC in near-real time via an rsync task. Greg Zeimann is managing the procedure to back up this data to **corral** which keeps them in multiple secure locations and uses the funds provided by the HET Board.