Data Management Plan

HEAT's extensive 3D FITS spectral line data cubes of the Galactic Plane, and targeted deep surveys will be acquired, reduced, analyzed, and distributed to the broader astronomical community via publications and permanent data archives.

Data Pipeline

The rate at which raw (Level 0) data is collected from the spectrometer is substantial in OTF mode (see Project Description, Section 2.3 for a description of the mapping strategy) and not in the form desired for scientific distribution; therefore data processing is performed on the HEAT instrument control computer. The data flow is depicted in Figure 1 and can be operated autonomously after basic verification. The steps undertaken in each data processing level are described as follows:

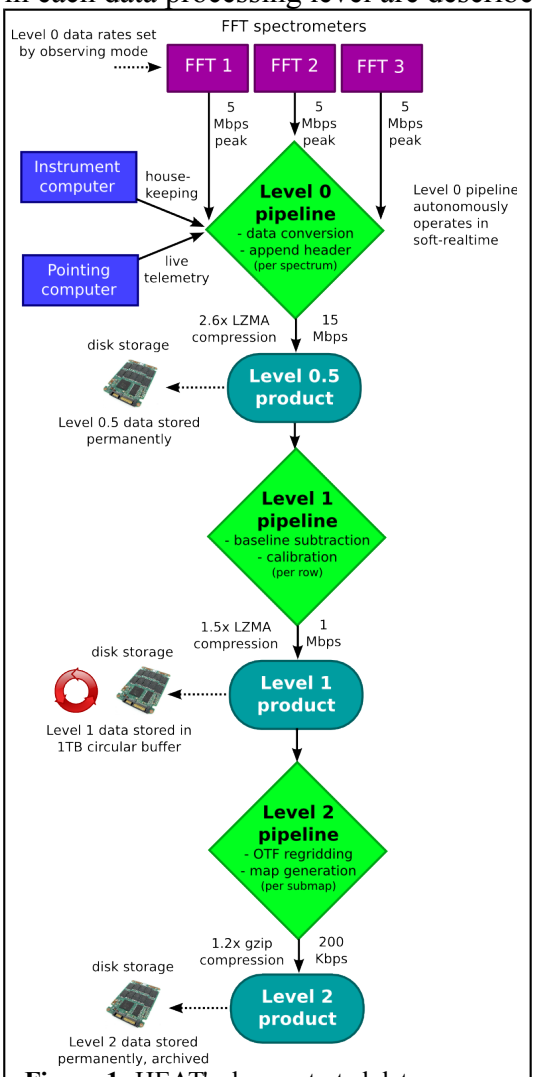

Figure 1: HEAT's demonstrated data pipeline

Level 0.5 (data conversion, header tagging) Each of the spectrometer data files is time-tagged upon being written to a RAM disk on the HEAT control computer. A data header is synthesized from streamed data from the HEAT tracker, which delivers telescope telemetry, and instrument housekeeping data. The data payload is rescaled from 64-bit words to 32-bit integers and written as a single-dish FITS file. The archival disk storage holds the LZMA-compressed (.xz) file, while a 'scratch' disk maintains the uncompressed file for follow up processing for a limited time. After validation, the 'raw' level 0 files are removed from memory.

Level 1 (baseline subtraction and calibration): After the conclusion of a single OTF scan, the map data can be preliminarily processed. The reference scan is subtracted from the source scans acquired during drift mode. If poor results are obtained, the best adjacent reference scan is used instead. Residual artifacts are masked from the resulting spectrum, and the data are flux calibrated using the ambient temperature chopper wheel method. Based on the antenna pointing and the time, the spectra are frequency calibrated onto a V_{LSR} velocity scale.

Level 2 (OTF regridding & map production): Once a submap has been repeated a sufficient number of times that the desired sensitivity has been achieved, the highly oversampled data are regridded and convolved to 90" resolution with 45" pixels. Optionally, spectral smoothing and additional spatial smoothing can be applied at this

stage. The numerical methods used during regridding follow that used by the 32-beam Sequoia array used at FCRAO to deliver outstanding high fidelity maps of CO emission in the first Galactic quadrant. These level 2 FITS cubes represent the baseline science products that HEAT uploads over Iridium modem. Thus, the highest priority for the observer is to continuously validate the level 2 processing using sparse quicklook versions of the level 0.5 and level 1 data that are streamed using available bandwidth.

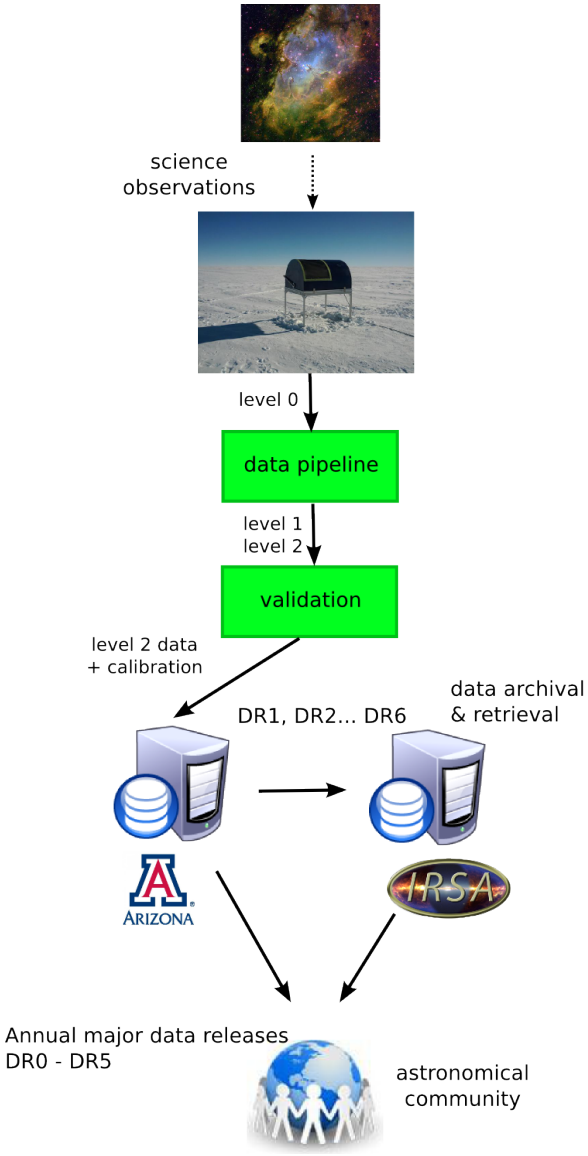

Figure 2: Archival data flow.

Data Archive

The HEAT data products will be in the form of FITS data cubes provided to the community from the University of Arizona and registered to the National Virtual Observatory (NVO). Major data releases will be released annually in April, with preliminary "early" releases of very recent data each August, from 4/2013 to 4/2018 as Data Releases 0, 1, … 6. (DR0, DR1,... DR6) as soon as calibration and formatting is complete, with **no proprietary period**. The archival data flow is diagramed in Figure 2. The maximum data volume is expected to be 5 GB in total, including all calibration datasets. The large FITS cubes will be developed within the HEAT team and hosted both at the University of Arizona and at the Infrared Science Archives (IRSA) at the Infrared Processing and Analysis Center (IPAC), as was done with the BGPS survey. The FITS headers will be stored in a SQL database to make a web-based relational queries of HEAT data and extraction of data subsets easy from the astronomer's perspective. PI Kulesa will lead the development of the web interface to the data.