Data Management Plan

Supercam'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 Supercam data 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:

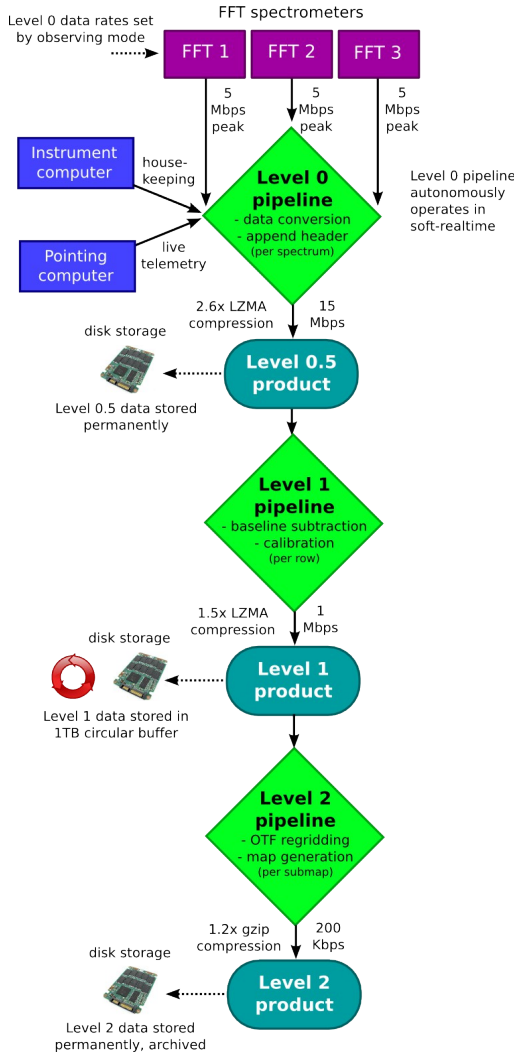

Figure 1: Supercam's demonstrated data pipeline

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 Supercam data computer. A data header is synthesized from streamed data from the HHT tracker and instrument control computers, which together deliver 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 23" resolution with 10" 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 Supercam will deliver. Thus, the highest priority for the observer is to continuously validate the level 2 processing using quicklook versions of the level 0.5 and level 1 data.

This data processing pipeline strategy has demonstrated heritage in the form of the baseline pipeline used with the PI's robotic High Elevation Antarctic Terahertz (HEAT) telescope, and was used with the appropriate changes in the May 2012 Supercam engineering run at the HHT. Relatively few changes are foreseen to the data pipeline; those changes will be managed by PI Kulesa and software lead B. Peters.

Ancillary Products

In addition to the standard level 2 data cubes and calibration data, important ancillary science products include a catalog of spatial and spectral correlated "clumps" in the data, using standard clump-finding routines, and a catalog of all detected outflows, with estimated energy input to the ISM computed. All data analysis tools, stand-alone pipelines, and cloud models will be provided. PI Kulesa is currently performing a systematic calibration of the dataset using high resolution infrared absorption line spectroscopy of CO and H_2 using ARIES at the 6.5m MMT, which will serve to provide precisely measured column densities of the two species. These data will also be provided.

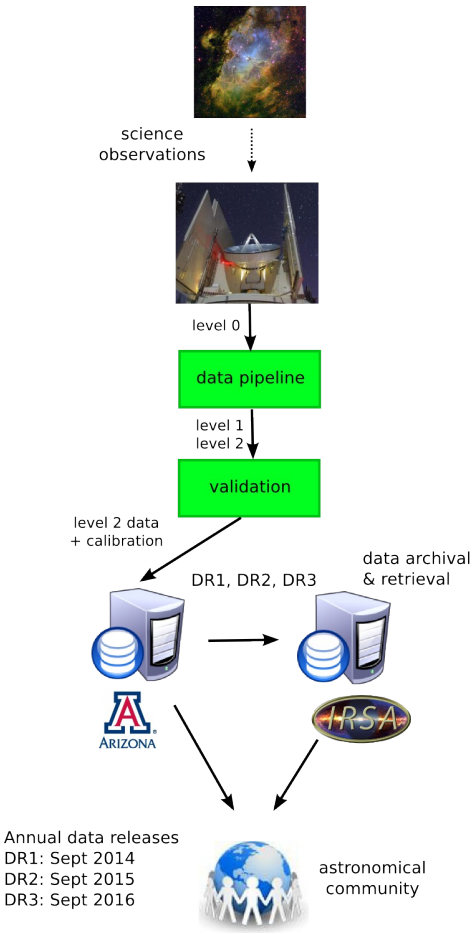

Figure 2: Archival data flow.

Data Archive

The Supercam 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). The data will be released annually in September of 2014, 2015, and 2016 as Data Release 1, 2, and 3 (DR1, DR2, DR3) 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 100 GB in total, including all calibration datasets. The large FITS cubes will be developed within the Supercam 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 Supercam data and extraction of data subsets easy from the astronomer's perspective. PI Kulesa and software lead B. Peters will lead the development of the web interface to the data.