*Astronomical Data Analysis Software and Systems X*
*ASP Conference Series, Vol. 238, 2001*
*F. R. Harnden Jr., F. A. Primini, and H. E. Payne, eds.*

# HST Data Flow with On-The-Fly Reprocessing

Daryl A. Swade, Edwin Hopkins, Michael S. Swam

*Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218*

**Abstract.** This paper documents the flow of data through the HST ground system in the OTFR era, both on initial receipt from the spacecraft and from the point of view of archive retrieval. The goal is to provide an understanding of what processing has been performed on the data and to indicate the state of data at various stages of processing.

## 1. Introduction

Traditionally, the Hubble Space Telescope (HST) data flow sequentially processed science observations through an OPUS[1] pipeline (Rose 1998) and archived these data with no further processing in the post-observation ground system. The pre-archive pipeline consisted of receipt of science telemetry, packaging exposures into observations, transforming the images or spectra into a standard format, and calibrating the data. Pre-archive calibration used the reference files available in the pipeline at the time of the observation, and, in almost all cases, this calibration was not optimal. Upon issuing an archive request, HST data users would receive the original processed version of the data, which required them to perform re-calibration on their own with Space Telescope Science Institute (STScI) provided software.

A new paradigm and enhancement to On-The-Fly Calibration (Lubow & Pollizzi 1999, Swam & Swade 1999), On-the-Fly Reprocessing (OTFR) is being implemented at STScI for HST data for most current and future science instruments. With OTFR, data retrieved from the HST archive are completely reprocessed from the raw telemetry (pod files) using the latest data processing and calibration software. Users automatically get the advantage of keyword updates, data processing software upgrades and bug fixes, as well as the latest calibration. Further details about the OTFR system can be found in a companion paper in this volume "Using OPUS to Perform HST On-The-Fly Reprocessing" (Swam, Hopkins, & Swade 2000).

Figure 1 shows the data flow within the HST ground system from the spacecraft and upon retrieval from the archive.

[1] `http://www.stsci.edu/software/OPUS/`

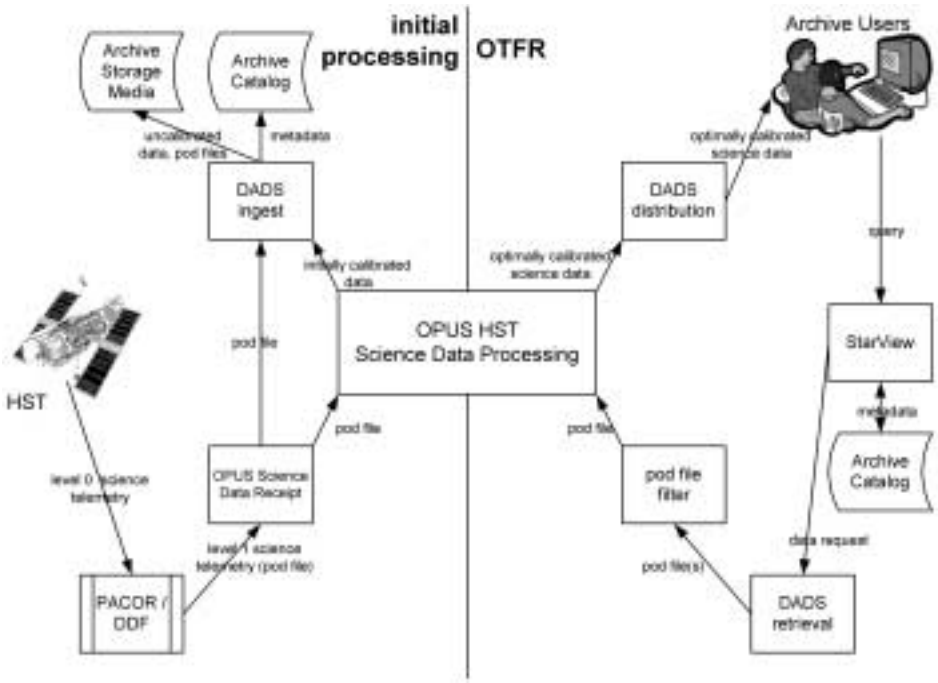


Figure 1. Data flow within the HST ground system from the spacecraft and upon retrieval from the archive.

## 2. Initial Data Processing

HST data is dumped from the on-board science recorder and transmitted as level 0 telemetry through a TDRSS satellite to the White Sands TDRSS downlink facility. From White Sands the telemetry is sent over domestic communication satellite to the PACOR data capture facility at Goddard Space Flight Center in Greenbelt, MD. At PACOR the data are binned into Reed-Solomon corrected packets to generate telemetry pod files (level 1 data). From PACOR the pod files are transmitted over the GSFC network to STScI via ftp.

At STScI, the pod files are received by the OPUS Science Data Receipt pipeline and passed into the OPUS Science Data Processing pipeline (described in section 4 below). OPUS science data processing converts the telemetry into calibrated FITS format data files.

After processing, the data are transferred to the Data Archive and Distribution System (DADS) for archiving. Keyword value information from calibrated headers is incorporated into metadata in the archive catalog. Pod files and uncalibrated FITS format data are written to the storage media.

## 3. Data Processing on Archive Retrieval

An archive researcher uses StarView to interface with the archive catalog. Using StarView, the archive researcher can query the archive catalog on a large number

of parameters to search for specific observations. Once these observations are found, a data request can be submitted to DADS through StarView.

The OTFR system will then retrieve from DADS the relevant pod files for the requested observations. Since a single pod file may contain more than one exposure, or a single exposure may span more than one pod file, a pod file filtering process has been inserted into the data flow to insure that only the necessary science telemetry enters the OTFR pipeline. At this point the pod files that correspond to the data request enter the OTFR pipeline and are processed with the exact same software and input information as the current run-time (pre-archive) data processing pipeline (described in section 4 below).

## 4. Details of the Science Data Processing Pipeline

HST data flows sequentially through a set of processes in the OPUS science data processing pipeline. Here we describe a generic pipeline and the primary operations that occur at each step. For HST, each science instrument has its own specific pipeline that deviates to some degree from the process order described. For example, the data collector process, where the exposures are collected for further processing as an associated unit, can occur at any point in the pipeline.

- Data Partitioning - accepts pod file telemetry as input, sorts data into individual observations, checks for bad data segments, sets data quality flags, and inserts fill into telemetry dropouts when necessary. The output of Data Partitioning is an intermediate data product, the EDT dataset, which is a combination of packetized binary and ASCII files.
- Support Schedule - extracts information from proposal database to populate keywords. This information is saved as ASCII file in the EDT dataset.
- Data Validation - dredges keyword values from telemetry and applies telemetry conversions if necessary, as well as compares actual vs. planned observation parameters.
- World Coordinate System - converts pointing information into standard FITS WCS keywords. The pointing information comes from planning databases and telemetry. This information is also saved as an ASCII file in the EDT dataset.
- Generic Conversion - converts unformatted EDT dataset into FITS data arrays or tables, properly orients the image or spectra, and populates keyword values in FITS headers.
- Data Collector - holds processing for individual exposures that require further processing as a single unit (association).
- Calibration - performs instrument specific calibration.

## 5. System Deployment

Current plans at STScI are to transition STIS and WFPC2 data processing from OTFC to OTFR early in 2001. A NICMOS OTFR system will be deployed soon after, and subsequent upgrades to the NICMOS OTFR pipeline will include temperature dependent darks and automated pedestal correction. OTFR will be developed as a baseline data processing functionality for ACS, COS and WFC3.

Hence, it will be available for these HST science instruments at the time of their launch. There are no plans to develop OTFR systems for the HST legacy science instruments WFPC, FOC, HSP, FOS, and GHRS.

## 6. Implications for Archive Catalog

The archive catalog metadata is populated from the header keywords of the initial data processing within a few hours of when the observation is downlinked from the HST. In the time between the initial processing and a request for archival data, it is likely that various updates will have been made to either the calibration reference files or algorithms associated with the data. Hence, it is possible that fields in the archive catalog will not agree with keyword values in data processed through OTFR. The OTFR keyword values should be considered the current best data. Attempts will be made to keep all common search fields and fields that determine calibration reference files up to date in the archive catalog. The latter is required for the BESTREF utility that is available through StarView[2] and the web[3].

## 7. Future Possibilities

OTFR allows for additional processing steps that cannot be realized in the near real time processing. At the time of archive data retrieval additional information, such as data quality information, engineering telemetry, and more exposures from the same field, will probably be available for the observations. OTFR will be able to provide higher quality and enhanced data products such as combining exposures from different visits or incorporating engineering data parameters in the science data. This also leads to the possibility of OTFR generating new data products that could not be realized in initial, near real-time processing.

## References

Lubow, S., & Pollizzi, J. 1999, in ASP Conf. Ser., Vol. 172, Astronomical Data Analysis Software and Systems VIII, ed. David M. Mehringer, Raymond L. Plante, & Douglas A. Roberts (San Francisco: ASP), 187

Rose, J., 1998, in ASP Conf. Ser., Vol. 145, Astronomical Data Analysis Software and Systems VII, ed. R. Albrecht, R. N. Hook, & H. A. Bushouse (San Francisco: ASP), 344

Swam, M., & Swade, D. A. 1999, in ASP Conf. Ser., Vol. 172, Astronomical Data Analysis Software and Systems VIII, ed. David M. Mehringer, Raymond L. Plante, & Douglas A. Roberts (San Francisco: ASP), 203

Swam, M., Hopkins, E., & Swade, D. A. 2001, this volume, 291

[2] `http://archive.stsci.edu/starview.html`

[3] `http://www.stsci.edu/cgi-bin/cdbs/getref.cgi`