

IAA-PDC-23-04-04  
Automated Data Processing and Image Quality Analysis Pipelines for the DART  
DRACO Instrument

C. Dany Waller<sup>(1)</sup>, Raymond C. Espiritu<sup>(1)</sup>, Hari Nair<sup>(1)</sup>, Carolyn M. Ernst<sup>(1)</sup>, R.  
Terik Daly<sup>(1)</sup>, Olivier S. Barnouin<sup>(1)</sup>, and the APL DART Team<sup>(1)</sup>

<sup>(1)</sup>Johns Hopkins University Applied Physics Laboratory, 11100 Johns Hopkins Rd., Laurel, MD 20723,  
USA, Corresponding author: Dany.Waller@jhuapl.edu

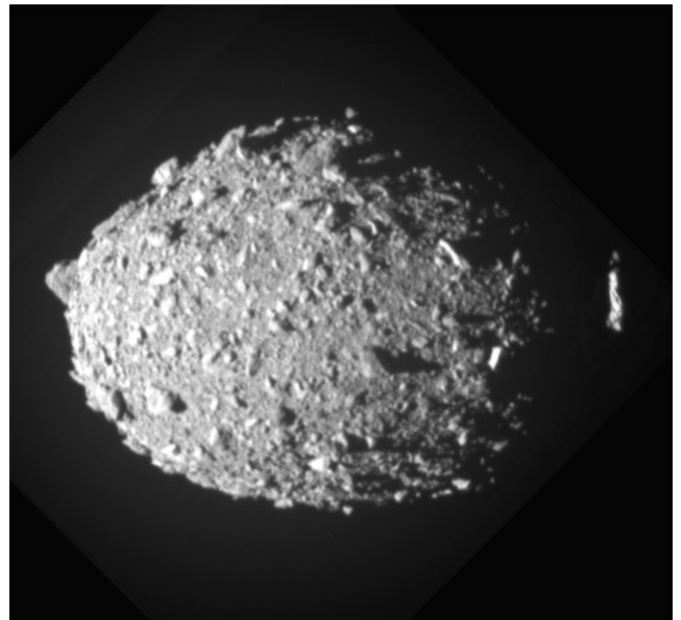
**Keywords:** DART, data processing, data visualization, image analysis

**Extended Abstract** — The Double Asteroid Redirection Test (DART) was the first successful demonstration of a kinetic impactor for planetary defense. The lone instrument onboard DART was the Didymos Reconnaissance and Asteroid Camera for Opnav (DRACO) [1], which obtained 250,000+ images during its charge. The DART mission utilizes a novel automated data processing pipeline to quickly produce raw and calibrated images for analysis and visualization, and a semi-automated pipeline to track image quality metrics and collate statistics from all images taken by DRACO over the course of the mission.

The processing pipeline was fed images of interest identified by the Science Operations Center during the impact event to prioritize calibration of the most visually and scientifically rich images (e.g., Fig. 1). Optimization and parallelization of the DRACO data processing pipeline allowed the final 150 images prior to impact to be calibrated and delivered to the Investigation Team and the press within 15 minutes after the impact. The DART data processing software architecture and automation implementation are outlined here as a framework for future missions, with an emphasis on planetary defense applications. The DRACO instrument produced a large volume of data at an extremely high rate with variable imaging modes and targets, which required rapid processing and analysis and “scientist in the loop” decision making.

**Data Processing Pipeline:** Downlinked images and housekeeping telemetry were unpacked by the Mission Operations Center (MOC) and passed to the Science Operations Center (SOC). The autonomous DRACO data processing pipeline registered incoming MOC products to a database with a table that tracked the associated raw product, raw product creation time, calibrated product, and calibrated product creation time. The processing pipeline queried the database on an adjustable time delta to find MOC products that did not have associated raw products and produced SOC-formatted files. These records were fetched in batches where the number of files ingested could be modified in

a configuration file used by the pipeline to set processing parameters that also included allocation of computing resources for parallel processing.



**Figure 1:** A view of the asteroid Dimorphos, oriented with its north pole toward the top of the image. DART’s onboard DRACO imager captured this image from a distance of 68 kilometers. Dimorphos is roughly 160 meters in length. This figure was modified from an image identified in real time as the last frame to contain all of the lit body of Dimorphos in the field of view [2]. The body is oriented with the north pole toward the top of the image.

Software tools were developed to efficiently format MOC products into raw SOC data products, which are Flexible Image Transport (FITS) format files with metadata information stored in the header. The metadata stored in each FITS header includes calibration information such as detector temperature, spacecraft geometry calculated from SPICE, and information from DART’s autonomous navigation software, SMARTNav [3]. Once the new SOC raw products were created and registered to the DRACO

database with their creation time, the processing pipeline queried the table again to find any raw product records that did not have associated calibrated products.

Calibration tools were developed to create either radiance or I/F calibrated files from the input raw product based on the image target and mission phase keywords. The SPICE-derived geometry values in the raw FITS headers were used for I/F conversion when applicable. The processing pipeline rejected files where the observation type was identified as a dark, bias, or test pattern image based on raw FITS header keywords, as well as images affected by detector reconfiguration and partial header information. A second table in the DRACO database was used to correlate calibration information based on timestamp, detector temperature, exposure time, and other key metrics that the pipeline fetched from the raw FITS header metadata. Upon processing pipeline completion, calibrated images were registered to the DRACO database with their creation time, and the pipeline would idle until the next time delta passed.

In addition to searching for unprocessed MOC files, the processing pipeline also queried the database for records where the raw product creation time was newer than the calibrated product creation time. A newer time indicated that a newer version of a raw image existed, and the calibrated image should be reprocessed. The DRACO data processing pipeline was extremely efficient and virtually hands-free from end-to-end, although the pipeline could still be manually run using command line scripts for troubleshooting and calibration verification.

The automated data processing pipeline was implemented in Nextflow [4, 5], which allowed for parallelization of Python and Java tools developed by the SOC. Calibration algorithms were vectorized where possible during the commissioning, cruise, and approach phases of the mission. Optimization of the Python and Java tools made the Nextflow pipeline capable of producing approximately 25 raw images per second and 15 calibrated images per second. The raw and calibrated data products were then passed to a semi-autonomous analysis pipeline for rapid identification of detector misconfigurations, missing data, and other adverse events.

**Image Quality Analysis Pipeline:** The vast number of images taken by DRACO made it impractical to rely on humans to identify image quality issues. For example, some observation sequences acquired more than 16,000 images in the span of several hours. The DRACO image quality analysis pipeline was a collection of Python scripts that semi-autonomously collected statistics on image quality metrics requested by the instrument scientists. Header metadata from raw and calibrated FITS files were parsed to determine metrics such as missing keywords, timing inconsistencies based on SPICE-derived values, partial images, and images split between two or more files. Files that were identified as “split images”, where one or more files contained subsets of data for a whole image, were then reprocessed by the MOC to stitch together a single full

image. For raw products, the image data were analyzed for maximum and minimum DN levels above or below user-specified thresholds to identify detector reconfigurations and bad images.

Image quality statistics were stored in CSV format alongside housekeeping data and used to fill out an HTML template that created an internally accessible webpage on the SOC (Fig. 2). A Python script was used to compare the most recent raw and calibrated product creation dates to the image quality stats file creation date and determine if the CSV file needed to be updated. This pipeline is referred to as “semi-autonomous” as the image analysis Python scripts were run in job manager bash scripts using the cron command-line utility, however, individual Python scripts were frequently run in the command line to implement new analyses and refresh the HTML webpage.

year	day	mission phase	total images	# partial images	# bad images	# of files not 01.fits	# of Radiance files	# of I/F files
2022	269	terminal	13399	0	0	0	0	13399
2022	269	approach	2789	3	0	0	2789	0
2022	269	final	236	1	0	0	0	236
2022	267	approach	2508	1	0	0	2508	0
2022	266	approach	1237	0	0	0	1237	0
2022	265	approach	1237	0	0	0	1237	0
2022	264	approach	1237	0	0	0	1237	0
2022	263	approach	989	0	0	0	989	0
2022	262	approach	248	0	0	0	248	0
2022	260	approach	1237	0	0	0	1237	0
2022	259	approach	1196	1	0	0	1196	0

**Figure 2:** An excerpt of the image quality analysis webpage for calibrated data products used by the SOC to identify partial images, bad images, and potential split images. These statistics were collected from common keywords in DRACO science product FITS headers and organized by Day of Year (DOY) and mission phase.

Engineering housekeeping data were also collected by the SOC to track daily, weekly, and monthly trends in both the Field Programmable Gate Array (FPGA) and detector temperature, voltage, frame counts, and error counts. These data were collated using a Python script to identify new SOC raw products in a database table, then housekeeping packets and raw FITS file headers were parsed to fetch relevant information. Trend plots were created and updated daily for the Investigation Team to track instrument health.

**Future Work:** The DART Investigation Team continues to analyze and reprocess images post-impact to provide the best possible set of raw and calibrated products to the community. The initial DRACO archive delivery has been made available via NASA’s Planetary Data System (PDS), including documentation of calibration performed by the data processing pipeline. The final archive delivery will be submitted to the PDS by mid-2023, including all data from DART’s companion CubeSat, LICIACube, and ground observatories [6].

The processing and analysis software architecture and automation framework developed during this mission will be generifed and adopted for future

missions, to enable rapid product generation and “scientist in the loop” decision making for semi-automated analysis tasks. A subset of these tools will be publicly released as part of the Small Body Mapping Tool [7]. The ability to analyze products in near real time while continuously downlinking and processing data is particularly relevant for planetary defense, where both efficiency and accuracy are critical for mission success.

**References:**

- [1] Fletcher Z.J. et al. (2022) *Proc. SPIE*, 12180.
- [2] Daly R.T. et al. (2023) *Nature*. doi: 10.1038/s41586-023-05810-5
- [3] Chen M.H. et al. (2018) *Advances in the Astronautical Sciences*, 164, 347–359.
- [4] Di Tommaso P. et al. (2017) *Nature Biotechnology*, 35, 316–319.
- [5] Heyd R.S. et al. (2019) *4<sup>th</sup> Planetary Data Workshop*, Abstract #2151.
- [6] Espiritu R. et al. (2023), *8<sup>th</sup> IAA Planetary Defense Conference*, this session.
- [7] Steele R.J. et al. (2023), *54<sup>th</sup> Lunar and Planetary Science Conference*, Abstract #2728.