Roman Mission Pipeline Processing of Prompt Products with Efficient Data Management and Archive Access

Tyler D. Desjardins, Richard G. Cosentino, Gisella De Rosa, Mark Kyprianou, John Glorioso, Edward Schlafly, Nadezhda Dencheva, Frank Carroll, Brian Charlow, Dario Fadda, Scott Fleming, Jennifer Kotler, William C. Schultz, & The Roman Team

The Nancy Grace Roman Space Telescope (Roman) will be NASA's first petabyte-scale astrophysics mission, producing an unprecedented volume of imaging and spectroscopic data. Processing the data from the 302-megapixel Wide Field Instrument (WFI) will require a robust, automated processing pipeline to ensure the timely release of scientific data products. The Roman Calibration Pipeline (RomanCal) has been developed to calibrate all WFI data for instrumental effects and produce mosaics, catalogs, and other high-level data products for the WFI imaging mode. During operations, rate images from each detector will be made available as prompt products via the Barbara A. Mikulski Archive for Space Telescopes (MAST), shortly after the observations are completed, once the downlinked data are processed by the ground system. Prompt versions of higher-level data products include visit-level mosaics and single-band source catalogs. This presentation provides an update on Roman's prompt data products, their creation, and planned public access to enable astronomical research.

What are "prompt" products?

Roman WFI data products are labeled Level 0 to 5, abbreviated as, e.g., L0. The data levels are defined as:

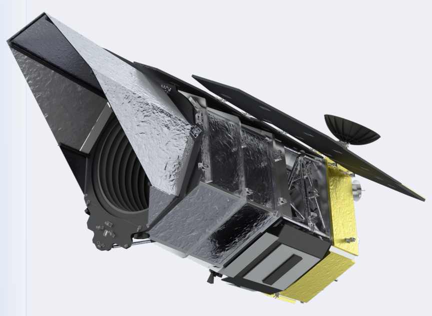
- **L0: raw, packetized data from the telescope.** L0 data are only available in MAST to authorized accounts.
- **L1: uncalibrated WFI exposures.** L1 data contain 3-D cubes that are the non-destructive samples up-the-ramp from the WFI detectors.
- **L2: calibrated rate images.** L2 data have been corrected for detector effects (e.g., dark current, classic non-linearity), and the up-the-ramp cube has been collapsed along the time axis to create a rate image.
- **L3: resampled images.** L3 data have been corrected for geometric distortion by resampling the image onto a rectified grid. In practice, archived L3 products will be co-additions of L2 images.
- **L4: high-level, extracted information.** L4 products contain myriad information types (e.g., source catalogs, light curves, and 1-D spectra) derived from L2 or L3 products.
- **L5: community-contributed products.** L5 products are products created by the science community and stored in MAST.

Prompt products are created as soon as all necessary inputs are available. This is distinct from **data release products**, which are created periodically. Important features of prompt products are:

- The calibration reference files used are the **best available** when the data are processed.
- MAST does not archive versions of prompt products. Reprocessing of a prompt product overwrites the previous version.
- L1 and L2 products are archived as only prompt products.
- L3 and L4 contain a mix of prompt and data release products. Data release versions supersede prompt L3 and L4 products, and prompt L3 and L4 products will be sunsetted from MAST some time (TBD) after a data release.

Prompt data products are available in MAST as quickly as they are processed, with no access restrictions.

How are prompt products made?



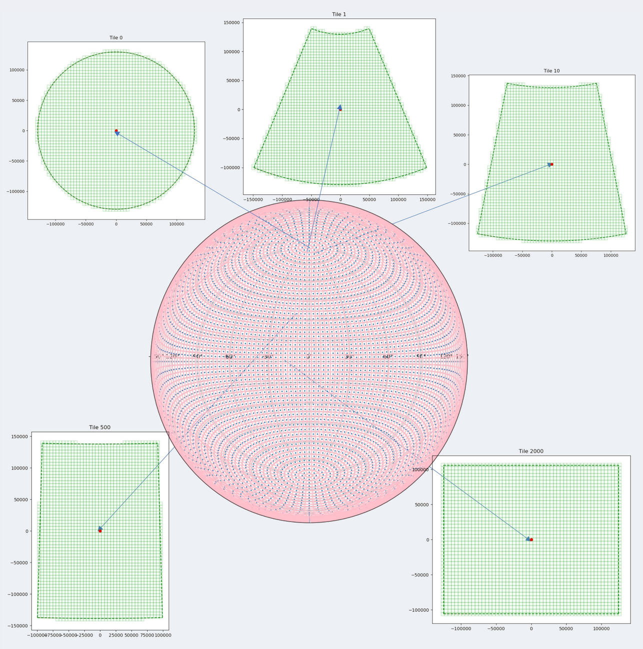
The Roman Science Operations Center (SOC) at STScI generates all L1 and L2 products, and all prompt L3 and L4 products for WFI imaging observations, excluding the Galactic Bulge Time-Domain Survey (GBTDS). L1 products are created and archived within 24 hours of receiving the observations from the telescope.

The L1 data are processed to L2 per-detector through the **Exposure Pipeline**. Processing is paused after creating single-band source catalogs until all 18 WFI detectors have reached this step. Sources in all detectors are compared to the Gaia astrometric catalog to perform a global update to the L2 world coordinate system (WCS). L2 products will be available in MAST within 48 hours of receiving the observations from the telescope. The single-band source catalog and a segmentation map made from L2 inputs are also archived as L4 products. Catalogs are stored as **Apache Parquet** files. L4 products will be available in MAST no later than 7 days after receiving the observations from the telescope, but in practice, may be much sooner.

For spectroscopic and GBTDS observations, processing by the Roman SOC stops before the source catalog is created, and the Roman Science Support Center at IPAC handles subsequent processing.

Non-GBTDS prompt L3 products are created per **visit**. Visits consist of one or more exposures with the same instrument settings. A visit may use a dither pattern (example right) or repeat exposures.

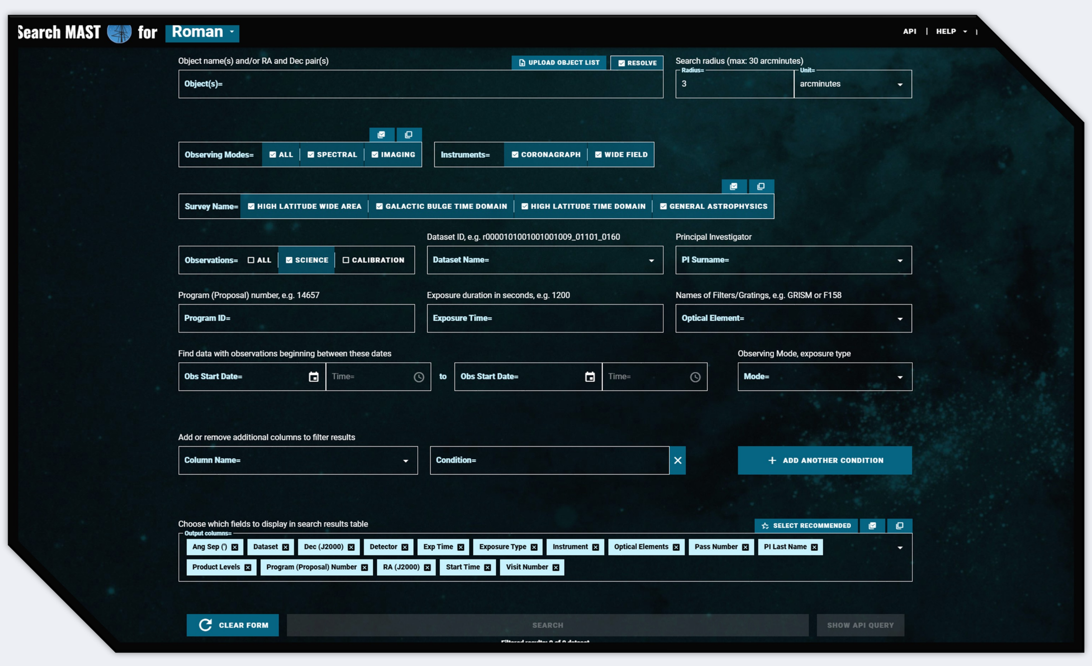
The **Mosaic Pipeline** makes L3 products. It converts L2 images into physical surface brightness units (MJy / sr), rejects outliers, and combines the images using the drizzle algorithm (Fruchter & Hook 2002). L3 products will be available in MAST within 5 days of receiving the observations from the telescope.



Even at the native pixel scale (0.11"/pixel), full-focal-plane L3 products would be prohibitively large at several GB per visit. Therefore, L3 products are stored as **skycell** products using a tessellation of the sky (figure left). Skycells are square and approximately 5,000 pixels on a side. Skycells are components of **projection regions**, which are square near the celestial equator and become less square moving towards the poles. Skycells within a projection region share a common WCS.

L4 single-band source catalogs and segmentation maps are also made from L3 products. They use the same code that makes catalogs and maps from L2 products.

Retrieving prompt products from MAST



Like other missions in MAST, a **user interface** (UI; development version shown left) is available for searching the Roman Archive.

Users will also be able to access WFI prompt data products using:

- The Astropy-affiliated Python ASTROQUERY package
- Curl scripts can be generated from the MAST UI
- The Data Access Application Programming Interface (DAAPI), also known as "Alexandria." The DAAPI is a GraphQL interface to the MAST Archive Catalog.

Archived Roman data will be stored in the Amazon Web Services cloud in the US-East 1 (Northern Virginia) data center. For the best experience, users should access the data in the cloud rather than download the products to a local machine. The **Roman Research Nexus** provides convenient access to compute and storage with proximity to the Roman data in the cloud.

The ASDF metadata will contain engineering telemetry information needed for data processing. Additional engineering telemetry data can be retrieved via another MAST UI or an API.

Prompt data products are archived only as files. Therefore, the MAST catalog service will not contain prompt L4 products. L4 data release products will be made available via the catalog service after the data release is published.

Have Questions?

Visit the Roman Documentation System (RDox) for mission documentation:

roman-docs.stsci.edu

Visit the Roman Help Desk portal to submit questions:

romanhelp.stsci.edu

Take Part in the Mission

Tune In

Sign up for news, listen to webinars, and read documentation.

Interact

Attend conferences, use simulated data and tools, and prepare to submit proposals.

Actively Contribute

Join Wide Field Instrument (WFI) working groups, participate in workshops, and provide feedback to formal committees.

www.stsci.edu/roman/get-involved