The Roman Coronagraph Community Participation Program: Data Reduction and Simulations

Maxwell Andrew Millar-Blanchaer[a], Jason Wang[b], Ellis Bogat[c], Jürgen Schreiber[d], Marie Ygouf[e], Kevin J. Ludwick[f], Alexandra Z. Greenbaum[g], Vanessa Bailey[e], Sergi Hildebrandt Rafels[e], Dmitry Savransky[h], Matthias Samland[i], Lisa Altinier[j], Ramya Anche[k], Beth Biller[l], Amanda Chavez[m], Elodie Choquet[j], Julien H. Girard[n], Justin Hom[k], James G. Ingalls[g], N. Jeremy Kasdin , Oliver Krause[i], John Livingston[p,q], Johan Mazoyer[r], Laurent Pueyo[n], Taichi Uyama[s], Robert T. Zellem[t], and Neil T. Zimmerman[u]

[a]University of California, Santa Barbara, CA 93106, USA
[b]Center for Interdisciplinary Exploration and Research in Astrophysics (CIERA) and Department of Physics and Astronomy, Northwestern University, Evanston, IL 60208, USA
[c]University of Maryland College Park, Center for Research and Exploration in Space Science and Technology II (CRESST II), NASA GSFC, 8800 Greenbelt Rd, Greenbelt, MD, USA
[d]Max Planck Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany
[e]Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA
[f]University of Alabama in Huntsville, 301 Sparkman Dr, Huntsville, AL 35899, USA
[g]IPAC, Caltech, 1200 E. California Blvd., Pasadena, CA 91125
[h]Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY, 14853, USA
[i]Max Planck Institute for Astronomy, Koenigstuhl 17, 69117 Heidelberg, Germany
[j]Aix Marseille Univ, CNRS, CNES, LAM, Marseille, France
[k]Steward Observatory and the Department of Astronomy, The University of Arizona, 933 N Cherry Ave, Tucson, AZ, 85721, USA
[l]Institute for Astronomy, University of Edinburgh, EH9 3HJ, Edinburgh UK
[m]Northwestern University, 633 Clark Street, Evanston, IL 60208
[n]Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
[p]Astrobiology Center, NINS, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan
[q]National Astronomical Observatory of Japan, NINS, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan
[r]LESIA, Observatoire de Paris, Université PSL, CNRS, Sorbonne Université, Université de Paris, Meudon, France
[s]Department of Physics and Astronomy, California State University Northridge, 18111 Nordhoff Street, Northridge, CA 91330, USA
[t]NASA Goddard Space Flight Center, Exoplanets and Stellar Astrophysics Laboratory (Code 667), Greenbelt, MD 20771, USA
[u]NASA's Goddard Space Flight Center, 8800 Greenbelt Rd., Greenbelt, MD 20771

ABSTRACT

The Nancy Grace Roman Space Telescope's Coronagraph Instrument will for the first time demonstrate active wavefront sensing and control for a space-based coronagraph, and may image the first planet in reflected light.

Further author information: (Send correspondence to M.A.M.-B.: maxmb@ucsb.edu)

Space Telescopes and Instrumentation 2024: Optical, Infrared, and Millimeter Wave, edited by Laura E. Coyle, Shuji Matsuura, Marshall D. Perrin, Proc. of SPIE Vol. 13092, 1309256 · © 2024 SPIE · 0277-786X · doi: 10.1117/12.3020478

The Community Participation Program has been initiated to engage members of the broader scientific community in the preparation for its planned launch in late 2026/early 2027. Here we will present the on-going work of the Data Reduction and Simulations working group, one of the four working groups within the Community Participation Program. The working group is charged with the development of the data reduction and post-processing pipeline for the on-sky data and the development of a simulation suite to aid in the preparation and planning of Roman Coronagraph observations.

Keywords: Roman Space Telescope, Data Analysis, Exoplanet Direct Imaging, Coronagraphy

1. INTRODUCTION

The Nancy Grace Roman Space Telescope, set to launch in late 2026 or early 2027, is being equipped with a coronagraphic instrument, as a technology demonstration of critical technologies for the Habitable Worlds Observatory.[1] A primary goal of the technology demonstration is to be able to measure the brightness of a flux ratio $\geq 10^{-7}$ astrophysical point source with SNR$\geq$5 located $6-9$ λ/D from an adjacent star with $V_{AB} \leq 5$, where the bandpass shall have a central wavelength $\leq 600nm$ and a bandwidth $\geq 10\%$ (a.k.a. the TTR5 requirement). In 2023, an international Community Participation Program (CPP) was initiated to engage the broader high-contrast imaging community in order to maximize the return of the mission. The CPP has created four working group (see paper 13092-54 by Savransky et al., this conference): the observation planning (see paper 13092-127 by Wolff et al., this conference), data reduction and simulations, hardware and polarimetry working groups. Here, we focus on the work being carried out by the data reduction and simulations (DR&S) working group. The work we present here builds off of extensive previous and on-going work by Roman's Project Science team, Coronagraph Technology Center, Science Support Center (SSC), Science Investigation teams, and many more.

The DR&S working group is currently focused on two main initiatives: the development of a Python-based open-source data reduction pipeline and the expansion of existing simulation tools. The data reduction pipeline will be used to produce official archival data products from on-sky data, but is being developed to also be used by the broader exoplanet direct imaging community at-large. It is currently under active development and is openly available via a GitHub repository *. The expanded simulator tools will be designed to simulate on-sky data and will be used for data reduction pipeline functional testing, algorithm development and observation planning.

In this manuscript we describe the overall architecture of the Roman Coronagraph Instrument Data Reduction Pipeline, "CorGI-DRP," in its current form. We emphasize that the pipeline is still under development. However, although we expect changes in some of the details, we don't anticipate any major changes to the overall architecture moving forward. We first describe the definitions of the various data "levels" in § 2, which we follow with a description of the DRP's overall architecture in § 3. In § 4 we describe the calibration framework, before discussing future work and our conclusions in § 5.

2. DATA FORMAT AND LEVELS

Data from Roman CGI will go through several defined levels, with each level having a set of pre-defined processing and analysis steps. Analog data has higher flux and is taken at low-to-moderate EM gain and, other than the extra step of dividing by EM gain, is processed much like any classical CCD dataset. Photon-counting data has average flux $<\sim$0.1 photon/px/frame and is collected at very high EM gain to raise the signal above the detector noise.[2] Photon-counting data has a rather different processing method, where it's processed into a binary True/False detection per pixel per frame.

Different types of data will have slightly different definitions of each level (e.g. photon-counting EMCCD data vs. analog mode EMCCD data), and some data (e.g. calibration data) is not defined at all levels. Figure 1 displays the data levels as defined for a typical analog exoplanet imaging dataset (i.e. a coronagraphic imaging sequence where the star is not observed in photon-counting mode). The DRP will be responsible for converting L1 data products into Technology Demonstration Analysis (TDA) data products. Calibration data will be processed to different levels as needed for the different specific calibrations.

*https://github.com/roman-corgi/corgidrp

Example of Data Levels for an analog Dataset

L1 -> L2a	L2a -> L2b	L2b -> L3	L3 -> L4	L4 -> TDA
• Separate image from prescan/overscan • Subtract bias • Mask cosmic rays • Correct non-linearity	• Convert DN to electrons • Divide EM gain • Apply dark & flat field • Apply master bad pixel (BP) map • Desmear and CTI Correction	• Normalize by exposure time • Construct World Coordinate System transformation & distortion correction	• Correct distortion/BP • Measure target star position • PSF subtraction • Combine images in observation sequence	• Calculate target star apparent mag. • Flux ratio noise vs. separation • Companion apparent mag, flux ratio, SNR

Figure 1. An overview of the various data levels for an exoplanet imaging dataset, including the processing steps required for each level.

Output L2a-L4 data and calibration data will be saved as FITS files unless otherwise specified. The 0th extension shall be reserved for just a header populated by the telescope, and the 1st extension will include imaging data and a header populated by instrument and DRP parameters. All extensions require the EXTNAME header keywords with a descriptive name. All auxiliary data shall be added to the header or as additional named extensions, as appropriate.

Each L2a-L4 fits file will have an associated error and data quality extension. The error extension will contain a 3D datacube of the form [n_error_terms+1, size_Y, size_X], where the first image in the stack will hold the error terms added in quadrature. The subsequent image slices optionally hold image arrays of the individual error terms, but this option is turned off by default in order to conserve space. The units of the error extension will be adjusted as the data is adjusted (e.g. dividing by the detector gain in e-/ADU), and the units will be contained in the image extension header. Errors will be tracked for any operation on the data that includes noisy terms, on a best-effort basis. For example, bias subtraction and flat-fielding can introduce errors and should introduce additional error terms, but dividing by the exposure time does not. The values in the data quality extension will record different causes for bad pixels and other pixel-level effects that will be tracked. The data quality of each pixel will be tracked using individual bits, such that each bit corresponds to a specific issue.

3. DRP ARCHITECTURE

For each distinct use case of the DRP (e.g. coronagraphic image processing, calibration creation, polarimetric imaging, spectroscopy, etc.), the DRP will have a recipe template that outlines the data reduction steps and associated parameters for each step. Each step will be a Python function, called a "step function", with the first input of the function being the input Dataset object and the output being a new Dataset object (independent in memory) containing the processed data. Each step will perform a single operation (or a small number of closely related operations) on a single Dataset. Datasets will be processed by defining a specific recipe, based off a recipe template for a given type of data, with details filled in for the specific data at hand.

A recipe file will be a list of step functions to run, and associated arguments/keyword arguments for each of the steps. The recipes will save the output data after predefined steps (corresponding to data levels L2a, L2b, L3, L4 and TDA). Optionally, data can be saved after any individual pipeline step, although these products will not be formally delivered. When running the DRP for independent analysis, users will be able to define new recipes that add additional optional steps or remove steps as needed.

A "production" version of the DRP will be run on the Data Analysis Environment server at the Roman Science Support Center and will be responsible for generating archival data products that will be immediately available to the public with no proprietary period. The DRP shall have a "Walker" program (the CorGI walker), which, when activated, will ingest a list of files and associated observation scenario information, identify the correct recipe template and populate an instance of it with appropriate argument values, saving the dataset-specific recipe file. The walker will automatically identify appropriate calibration files and include them in the recipe file. It will then execute the steps as defined in the recipe. The recipe will be written to the header of each output data product. Figure 2 provides a graphic overview of the primary DRP components.

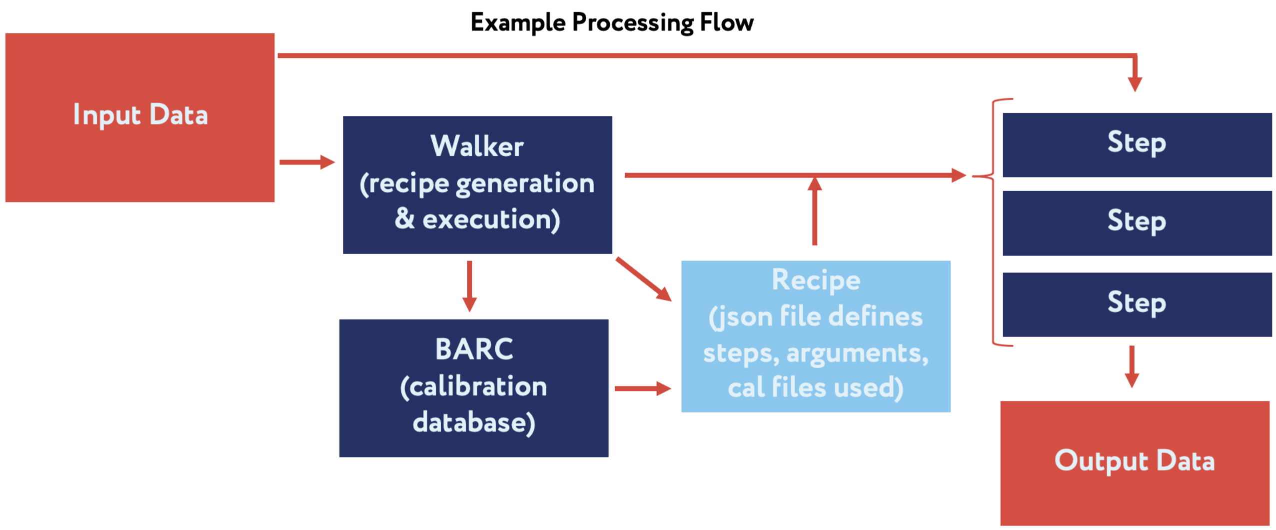

Figure 2. A diagram demonstrating the interactions between the various components of the DRP. Input data is analyzed by the Walker, which then generates a recipe from a recipe template after identifying appropriate calibrations from the BARC. The Walker then processes the input data in a series of modularized step functions saving as output data when appropriate.

4. CALIBRATION FRAMEWORK

The DRP shall maintain an internal calibration database to track all possible calibrations that it could use (the Bank of Archival Roman Calibrations - CorGI, BARC). In the production version of the pipeline this shall include both the official directory of calibrations maintained and vetted by the SSC and should also have the option to include calibration produced by the DRP that have yet to be vetted. By default the production DRP shall only use vetted calibrations, but for development purposes an option should remain to include un-vetted calibrations. Each instance of the BARC will be local and users can easily build their own local database based on the files on their computer, either with calibrations files they have generated themselves or those obtained from the public archive.

The BARC should include calibration filenames, file types, and other relevant information, such as exposure times and dates. The full extent of information stored in the BARC is to-be determined. The BARC will be saved in an easily accessible format (CSV) that can be examined in isolation. Upon completion of a calibration recipe, the walker shall add newly created calibrations to the BARC (noting the status as unvetted, if necessary).

For a given dataset, the CorGI walker will automatically select appropriate calibration files from the BARC and include them in the recipe file. The selection rules shall depend on the step (e.g. for darks, exposure times will be matched, but not for flats) and will likely evolve as the instrument is better understood. If a user should need to manually specify a calibration file, they may modify a recipe before it is executed.

Calibration data, to the extent possible, shall be stored as FITS files with headers detailing the input files and the DRP settings used to generate them. Where possible, error and DQ extensions will be used to track calibration errors and calibration data quality.

5. CONCLUSIONS AND FUTURE WORK

The DRP is under active development by a team of $\sim 8-10$ scientists. A total of 10 step functions have been written to-date, with over 245 commits to the main branch since Jan. 2024. The first formal delivery by the CPP of the DRP L1→L2b analog mode data product generation software is schedule for Sept. 23, 2024, with new deliveries every ∼6 months (Table 1). In the near future the DR&S working group will begin working closely with the newly formed Polarimetry working group to define the DRP processing steps for polarimetric data, as well as the team at Goddard Spaceflight Center to define the spectroscopic data processing.

In the summer of 2024 the working group will begin the development of a new simulation tool. The tool will be developed to support the on-going DRP development and testing, and will also be used for observation planning. The new simulator will be able to generate L1 data products for a wide variety of scientific scenes, as well as different calibration types. These new capabilities will leverage the existing simulation packages `CGISIM` and `emccd_detect`.

Table 1. DRP delivery schedule for the CPP

Release	Date	Release Content
R3.0.1	Sept. 23, 2024	L1→L2b analog mode processing with associated calibration generation, "Walker" infrastructure with recipe generation, Boresight Calibration
R3.0.2	Apr. 1, 2025	L1→L2b photon-counting mode processing with associated calibration generation, Flux calibration, ND Filter calibration, Distortion Map generation, core throughput calibration, plate scale calibration, L1→TDA Processing
R3.1	Sept. 25 2025	Polarization Data Processing and Calibration
R3.2	Mar. 3 2026	Launch Version

ACKNOWLEDGMENTS

This material is based upon work supported by NASA under awards 80NSSC24K0097, 80NSSC24K0216 and 80NSSC24K0087.

REFERENCES

[1] Bailey, V. P., Bendek, E., Monacelli, B., Baker, C., Bedrosian, G., Cady, E., Douglas, E. S., Groff, T., Hildebrandt, S. R., Kasdin, N. J., Krist, J., Macintosh, B., Mennesson, B., Morrissey, P., Poberezhskiy, I., Subedi, H. B., Rhodes, J., Roberge, A., Ygouf, M., Zellem, R. T., Zhao, F., and Zimmerman, N. T., "Nancy Grace Roman Space Telescope coronagraph instrument overview and status," in [*Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*], *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* **12680**, 126800T (Oct. 2023).

[2] Nemati, B., "Photon counting and precision photometry for the Roman Space Telescope Coronagraph," in [*Space Telescopes and Instrumentation 2020: Optical, Infrared, and Millimeter Wave*], Lystrup, M. and Perrin, M. D., eds., *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* **11443**, 114435F (Dec. 2020).