VERA C. RUBIN OBSERVATORY

The Rubin Data Products, Abridged (2022)

September 2022

By Melissa Graham
Lead Community Scientist for Rubin Observatory

NSF NOIR Lab AURA SLAC U.S. DEPARTMENT OF ENERGY

Introduction

Rubin Observatory will create the general-use data products and analysis tools to enable the scientific deliverables in the four science pillars: *probing dark energy and dark matter, taking an inventory of the solar system, exploring the transient optical sky, and mapping the Milky Way*.

Producing these deliverables, and pushing into new scientific frontiers, will require the development of specialized algorithms, data products, analysis tools, and cyberinfrastructure that go **beyond** what will be provided by Rubin Observatory and which are best left to the specific expertise of the science community. The independent LSST Science Collaborations are driving this development.

This presentation supplies a brief, informal summary of the planned Rubin Observatory data products and analysis tools, and outlines the boundary between what Rubin will provide and what will be left to the expertise of the science community.

Caveat: The contents of this abridgement are subject to change. The Data Products Definitions Document (DPDD; ls.st/dpdd) and "*LSST: From Science Drivers to Reference Design and Anticipated Data Products*" (Ivezić et al. 2019, ApJ, 873, 111) are the full, original references.

Transients, Variables, and Moving Objects

The data products for transients, variables, and moving objects will be primarily produced by the **Prompt Processing pipelines**, which will perform reduction, calibration, difference image analysis (DIA), source detection and measurement, and alert distribution within 60 seconds of image readout.

Alerts are ascii files containing data for a single detected source in a difference image (DIASource). The alert packets will include catalog data and small cutouts of the difference and template images. Alerts will be distributed to community brokers within 60 seconds of image readout.

Solar System Processing for moving objects will take place during the day (source linking, orbital parameter determination, etc.).

Images and catalogs that result from Prompt Processing will be available after 80 hours, and are fully described in Section 3 of the DPDD. All DIA data products will be re-generated during the annual Data Release Processing (Section 4 of the DPDD). Source detection and measurement on *direct images* (i.e., non-difference images) will *only* be done during the annual Data Release Processing.

Transients, Variables, and Moving Objects

Images	<80 h	yearly
processed visit images (PVIs; “direct images”; calibrated)	x	x
difference images (template-subtracted; calibrated)	x	x
template images (transient-free annual stacks)		x

Catalogs	<24 h	yearly
sources detected with SNR>5 via difference image analysis (DIA), and associated forced photometry: the DIASource, DIAObject, and DIAForcedSource tables	x	x
DIASources linked as moving-objects in the solar system (SS) and their orbital parameters: the SSSource, SSObject, and MPCORB tables	x	x
sources detected with SNR>5 in PVIs, and associated forced photometry: the Source, Object, and ForcedSource tables		x

Transients, Variables, and Moving Objects

Catalog contents will include:

- unique identifiers (IDs)
- measurements (e.g., coords, flux, mag, date/time, shape, size, PSF fit, proper motion, parallax)
- IDs of nearby LSST static-sky catalog objects (i.e., host association)
- orbital parameters derived by the Minor Planet Center (MPC)
- time variability parameters (limited and TBD; see ls.st/dmtn-118)
- pre-discovery ("precovery") PSF photometry in difference-images

Transients, Variables, and Moving Objects

Examples of additional specialized algorithms, data products, and analysis tools that will be left to the expertise of the science community include, but are not limited to:

- photometric and spectroscopic follow-up observations
- object classifications (e.g., light-curve types, astronomical categorization)
- cyberinfrastructure for the large-scale acquisition, processing, and analysis of follow-up
- cross-matching to non-LSST catalogs
- host-galaxy confirmation (e.g., distinguishing faint or blended hosts)
- orbital and/or time-variability parameters beyond what is in the LSST tables
- light-curve parameters (e.g., rise/fall times, peak brightness, asteroid rotation rates)
- shifted-and-stacked images (e.g., to detect faint moving objects)
- multi-night stacks and/or difference-images (e.g., to detect fainter transients)
- physical parameters (e.g., redshift, distance, host extinction, composition, intrinsic magnitude)
- event occurrence rates (e.g., volumetric rates)

Static-Sky Objects (Stars and Galaxies)

The data products for static-sky objects (stars and galaxies) will be primarily produced by the Data Release Processing pipelines, which will reduce, calibrate, and combine (i.e., stack, coadd) all LSST images, and detect, measure, and characterize sources in both direct and “deep coadded” images.

Images and catalogs that result from Data Release Processing will be available annually, and are fully described in Section 4 of the DPDD.

Static-Sky Objects (Stars and Galaxies)

Images	<80 h	yearly
processed visit images (PVIs; calibrated)	x	x
deep CoAdds (stack of all LSST images; one per filter; calibrated)		x

Catalogs	<24 h	yearly
sources detected with SNR>5 in PVIs and CoAdds: the Source and Object tables		x
forced photometry in PVIs at the location of all Objects: the ForcedSource tables		x

Static-Sky Objects (Stars and Galaxies)

Catalog contents will include:

- unique identifiers (IDs)
- measurements (e.g., flux, mag, color, date/time, shape, size, PSF fit, proper motion, parallax)
- centroids and adaptive moments
- Petrosian and Kron fluxes
- deblending parameters (e.g., parent/child associations; priors for crowded fields)
- model fits (e.g., point-source, bulge-disk)
- aperture surface brightness measurements
- photometric redshift (PZ) estimates (general-use algorithm TBD; see ls.st/dmtn-049)
- local shear estimation measures

Static-Sky Objects (Stars and Galaxies)

Examples of additional specialized algorithms, data products, and analysis tools that will be left to the expertise of the science community include, but are not limited to:

- alternative deeply coadded images (e.g., intermediate timescales, multi-band, best-seeing)
- specialized deblending algorithms, or probabilistic catalogs (e.g., for crowded fields)
- stellar types or physical parameters (e.g., metallicity)
- Milky Way component associations (e.g., disk/bulge/halo stars)
- specialized low-surface brightness measurements
- galaxy PZ or physical parameters (e.g., star formation rates) beyond those from the adopted, community-vetted, general-use PZ algorithm
- galaxy shear estimates beyond those provided by the adopted shear algorithm
- other galaxy characterizations (e.g., AGN, group or cluster membership, morphology)
- cyberinfrastructure to support large-scale compute-intensive processing (e.g., wide-area joint pixel analyses with non-LSST data sets)

Compute Resources and Analysis Tools

In order for scientists to access and analyze the LSST data, Rubin Observatory will provide the **Rubin Science Platform (RSP)**. The RSP is a set of integrated web-based applications and services that include tools to query, visualize, subset, and analyze the full LSST data archives in a stable software environment located "next-to-the-data", along with storage space and compute resources.

As defined in the Science Requirements Document (SRD; ls.st/srd), the Rubin Data Management System will provide at least 10% of its total capability for user processing and storage. Scientists will be able to pool their compute resource quotas in order to undertake large processing jobs. If these resources are oversubscribed, a "Resource Allocation Committee" will be established.

Due to the large nature of the LSST data set, it is anticipated that the science community will require significant **external cyberinfrastructure support** in addition to the RSP for, e.g., processing and analyzing follow-up observations for LSST time-domain events; running wide-area joint pixel analyses with non-LSST data sets; building and using frameworks for probabilistic catalogs; iterative development and training for machine learning algorithms; and many other applications.

Rubin Data Rights

The Rubin Data Policy remains the ultimate reference regarding Rubin data rights: ls.st/rdo-013.

All data products generated by Rubin Observatory will have a two-year proprietary period EXCEPT the contents of the alert packets and the catalogs produced by Prompt Processing (real-time image differencing; catalogs of transient, variable, and moving objects), which are public and can be shared with anyone, anywhere.

However, the alerts will only be delivered to seven pre-selected brokers, and the Prompt Processing catalogs will only be available via the Rubin Science Platform (RSP). As RSP access is proprietary to data rights holders, the shareable contents of these data products are not publicly accessible. Most of the brokers currently offer or plan to offer public access to the alert contents.

The LSST Science Pipelines are open source and will be pre-installed for all RSP users.

Ask Questions via the Rubin Community Forum

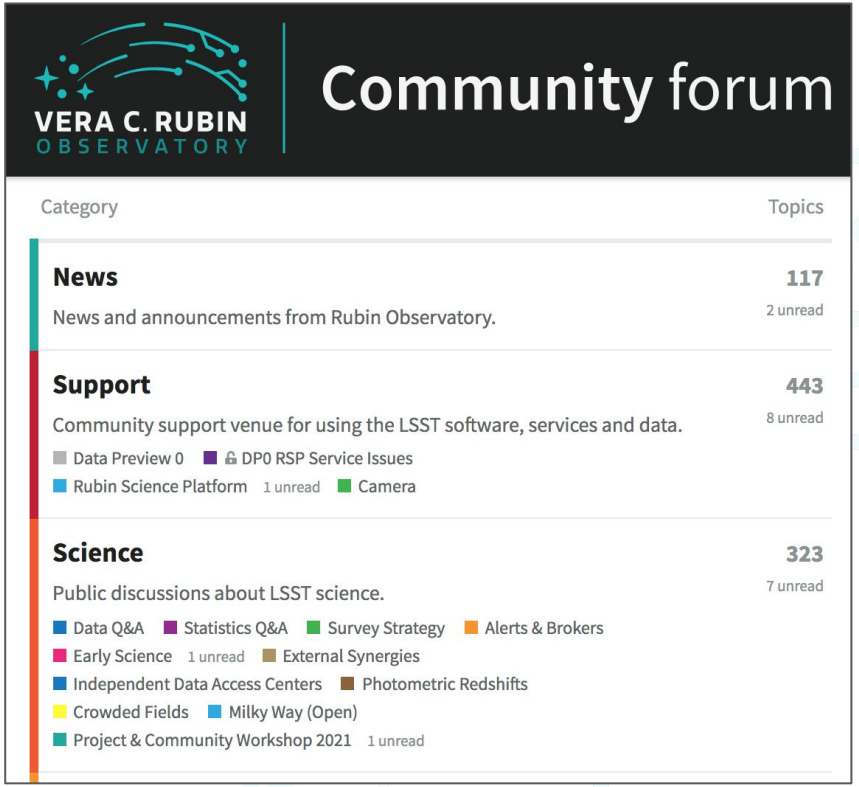
The "Data Q&A" Category of the Rubin Community Forum is dedicated to answering questions about the planned Rubin data products.

Forum membership is open to everyone.

You must sign up for an account in order to post, but most content is viewable without an account.

To ask a question about Rubin data:

- go to community.lsst.org
- under 'Science' click on Data Q&A
- at upper right, click on +New Topic
- compose your question, then +Create Topic
- Rubin staff will answer your question



Acronyms

See also lsst.org/scientists/glossary-acronyms.

DAC	Data Access Center
DIA	Difference Image Analysis
DMTN	Data Management Tech Note
DPDD	Data Products Definitions Document
IDAC	Independent Data Access Center
LSST	Legacy Survey of Space and Time
MPC	Minor Planet Center
MPCORB	a table of orbital parameters for moving objects ingested from the MPC (Section 3, DPDD)
PVI	Processed Visit Image
PZ	Photometric Redshift
RSP	Rubin Science Platform
SNR	Signal-to-noise Ratio
SRD	Science Requirements Document
SS	Solar System
TOM	Target Observation Manager

Glossary of Terms

See also lsst.org/scientists/glossary-acronyms.

Processed Visit Image -- An LSST image from a single visit, including science, mask, and variance arrays, PSF characterization, and metadata. Non-astrophysical background subtracted.

Coadded (Stacked) Image -- A combination of multiple input images, aligned to a common pixel grid and corrected to the same photometric scale and zero-point, with bad pixels and artifacts rejected. Non-astrophysical background subtracted.

Difference Image -- The pixel-by-pixel difference of a visit image and a template image, after matching to the same PSF shape, scaling to the same photometric response, and applying a correction for Differential Chromatic Refraction (DCR).

Template Image -- A co-added, single-band image of the sky that is deep, and created in a manner to remove transient or fast moving objects from the final image.

Glossary of Terms (continued)

Source (DIASource) -- A single detection of an astrophysical object in an image. In LSST catalogs, a signal-to-noise ratio > 5 is necessary for a Source to be considered detected. A "DIASource" is a source detected in a difference image.

Object (DIAObject) -- An association of Sources at the same sky coordinate. A "DIAObject" is an association of DIASources.

SSSource / SSObject -- SSSources are DIASources that are linked as moving objects, and stored as SSobjects, where SS stands for Solar System.

Forced Photometry (ForcedSource) -- A measurement of the photometric properties of a source, or expected source, with one or more parameters held fixed. Most often this means fixing the location of the center of the brightness profile (which may be known or predicted in advance), and measuring other properties such as total brightness, shape, and orientation.

Alert -- A packet of information for each DIASource, containing measurement and characterization parameters based on the past 12 months of LSST observations plus small cutouts of the single-visit, template, and difference images. Alerts are distributed to brokers.

Broker -- Software systems that will ingest, process, and serve astronomical alerts from Rubin Observatory and other surveys to the broader scientific community.