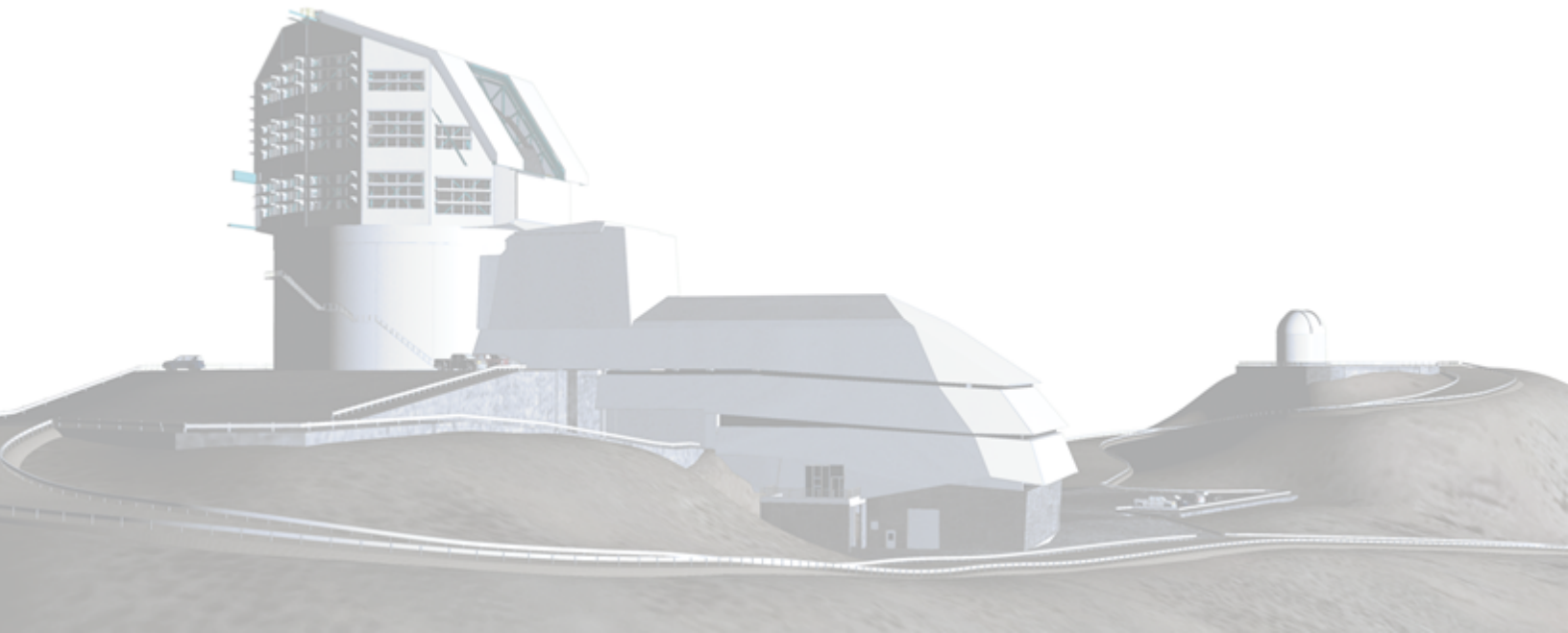
Vera C. Rubin Observatory
Data Management

Plans for Early Science

Leanne P. Guy, Eric Bellm, Bob Blum, Željko Ivezić, author list incomplete ... add your name as you contribute.

RTN-011

Latest Revision: 2021-11-09



Abstract

Science during the first year of Rubin Observatory / LSST operations is a high priority. Prior to Data Release 1 (DR1), scientific investigations may be undertaken with data processed in Alert Production and Solar System Processing as well as Data Previews of the commissioning data. Alerts of transient, variable, and/or moving objects are the only data product that will be immediately available (within 60 seconds of image readout) and publicly shareable, i.e not subject to a proprietary period. Alert Production requires high-quality template images of the sky, and thus full volume and full fidelity alerts will not be available until after Data Release 1 (DR1). However, Rubin Observatory is planning a process of incremental template generation to maximize the time-domain and solar system science achievable in the first year of operations. The worldwide community is actively preparing to process the LSST alert stream and use it to generate groundbreaking scientific results. This note describes the Rubin Observatory plan for ensuring Early Science. It is a living document that will evolve over the course of the remainder of the construction project and up until Data Release 1 (DR1).

Change Record

Version	Date	Description	Owner name
1	2020-10-30	First draft	Leanne Guy
2	2020-12-16	Draft 1.1	Bob Blum
3	2021-10-08	Rework structure	Leanne Guy
4	2021-10-21	Add timeline	Leanne Guy
5	2021-11-05	Edits throughout	Eric Bellm

Document source location: `https://github.com/rubin-observatory/rtn-011`

Contents

Plans for Early Science

1 Introduction

This note describes the plan for ensuring the Rubin community will have sufficient data access, data products, and data analysis tools to produce early science when the full Legacy Survey of Space and Time (LSST) begins. The start date for full survey operations is planned for April 1, 2024 with considerable uncertainty. As of November 2021, the construction project schedule shows completetion between December 2023 and July 2024 (i.e. planned finish and finish with contingency). It is unlikely the project will finish in December 2023 and highly likely it will finish by July 2024.

1.1 Definition of Early Science

Given that a first release of catalog data in operations requires significant progress in year 1, incuding generation of templates for alert production, Early Science (ES) is defined as any science enabled by Rubin for its community that happens before Data Release 1 (DR1) at the end of the first year of full survey operations. Community expectations for early science are high due to the transformative nature of the Rubin data. Additionally, for many science goals, time-sensitive follow-up observations are needed to take full advantage of the Rubin data.

1.2 Motivations for Early Science

The motivation for an Early Science programme is that in the current *baseline* there will be no *baseline* or regular science data products before Data Release 1 (DR1). This is due to the fact that *baselined* templates needed for Difference Imaging are produced in Data Release Production (DRP) and Data Release 1 (DR1) is not until one year after the start of operations.

To insure Early Science is enabled, Rubin pre-operations is working with the construction team to provide Data Previews. These are Data Release-like data products made from appropriate commissioning data, both from ComCam and LSSTCam. Due to the relatively short time periods available for commissioning observations (§1.3), these Data Previews will necessarily be limited in their areal and temporal coverage relative to full Data Releases.

Additionally, to enable early alert science, Rubin alert production will include the capability to build "incremental" templates from on-sky imaging as it becomes available during commissioning as well as the first year of operations. Such templates will be built periodically as images accumulate to allow for partial alert generation over an incomplete sky footprint. How extensive these templates are at the start of full survey operations will be influenced on the overall success of commissioning including Science Validation surveys and other significant datasets. These same datasets will provide static sky data release-like products in the Data Previews. The success of Early Science then depends on various scenarios comming out of commissioning as we transition into operations as described next.

1.3 Early Science scenarios

Recent planning on the construction project has led to a reduced amount of on–sky time and SV time. As Rubin construction moves through the challenging phase of System Integration, Test and Commissioning (SIT-COM), on–sky time could get squeezed more. The Operations team is thus planning for various outcomes that might require special attention to producing Early Science opportunities in the first part of regular operations to ensure the community has access to exciting data and data sets while the survey begins its relentless coverage of the sky before DR1.

In all cases, it is assumed Rubin Construction hands Rubin Operations a system that can capture, move, and process science-quality data at the time Operations begins. Planning will then consider three high level options to ensure Early Science . These are initially described in this document, but the document is "living" and we expect the plans to mature in detail over time as we approach full survey operations and the extant SIT-COM program emerges and is executed. At some point, a single option will be adopted and executed:

- *P*lan A: SV is completely successful, move quickly to the LSST and DR1 while providing Early Science data products in Data Previews
- *P*lan B: Early Science Period in the start of full operations (3-6 months) that is different than regular survey operations because on-sky time in SIT-COM is reduced, leading to fewer science ready data before the LSST begins, but the system is otherwise ready
- *P*lan C: further shakedown of operations procedures and data taking is required even though the initial condition above is satisfied and the Rubin System can capture and

produce science quality data (i.e., the system passes the construction completion requirements but the operations team is not yet ready to begin the LSST).

Each option (A, B, or C) will include alert generation of some type, with the major distinction being the relative availability of templates in time, sky position, and filter. Rubin will incrementally generate templates in SIT-COM and year 1 using the best images available and covering as much sky as possible given other needs which must be addressed. Details of the current strategy for alert generation (prompt processing, PP) with incremental templates are given below in section Section 4.

2 Timeline

This timeline provides a list of key dates related to the Early Science program.

2021-11-01: Issue first version of the Rubin Observatory Plans for Early Science document.

2023-01-01: Commissioning Camera (ComCam) on sky

2023-07-01: LSST Science Camera (LSSTCam) on sky

2023-10-01: Start of the science verification surveys

2024-01-15: Current earliest completion date for construction

2024-XX-XX: Current forecast start date for LSST data taking

3 Science Drivers

The various different science drivers outlined in 3 naturally lead to different priorities for template generations, e.g. solar system science prefers templates to be generated in the NES and Milky Way science would to would prefer templates for the galactic plane to optimise alert production in these areas in early operations. Other science will prefer templates in a number of filters to enable .. rather that larger area.

3.1 Time Domain

The Transients and Variable Stars Science Collaboration (tvssc) reviewed the opportunities for Early Science for non time-critical and time-critical science cases in Hambleton et al. (2020) and Street et al. (2020) respectively.

3.2 Solar System

The Solar System Science Collaboration (sssc) reviewed opportunites for Early Science in 2021. LSST is predicted to discover ≈ 6 million solar system planetesimals, providing in total over a billion photometric and astrometric measurements in 6 broad-band filters.

3.3 Static Science

The baseline static science data sets will flow from Science Verification surveys carried out during commissioning.

3.4 Target of Opportunity

Rubin Observatory will be prepared to take advantage of Targets of Opportunties (TOO) in the first year of operations (and hopefully SIT-COM). [**?**] describes potential data processing scenarios for TOO observations in the early operations era.

4 Alert Production in Commissioning and Early Operations

4.1 Processing Overview

The DPDD summarizes the pipelines which will be used during Prompt Processing to produce alerts as well as other prompt data products (§6), including Solar System Processing. In brief, raw images have instrument signatures removed and are photometrically and astrometrically calibrated. When template images for the corresponding region of the sky are available, the template is subtracted from the new processed visit image and sources are detected on the image difference. Alerts are then generated for all DIASources detected at five sigma in the

difference. At the end of the night, DIASources without a history of previous detection are input into Solar System Processing, which attempts to link them with other past detections and identify new Solar System objects.

Both Alert Production and Solar System Processing thus depend on the presence of template images. During steady-state operations, these templates will be constructed during the annual Data Releases and will built from the best available subset of images taken. The input images for DRP-produced templates will accordingly have very good seeing and comprehensive spatial coverage. Coadding multiple images enables artifact rejection [DMTN-080] and reduces noise. All of these template characteristics all help to ensure that image differencing is highly complete and highly pure.

To enable alert production to proceed during commissioning and early operations, it is necessary to accept templates of lower quality. Because we have a smaller set of input images to choose from and uncertain knowledge about future observations, on-the-fly (or incremental) template generation necessarily must balance the trade off of earlier template availibility against template quality and spatial completeness. A template constructed today will enable alerts tomorrow, but that template might produce fewer or lower-quality alerts than one constructed from more data in a week or a month's time. Substantial validation will be required to determine when to build incremental templates to maximize the net throughput of Early Science . Nevertheless our goal is to enable Alert Generation to begin as soon as the data are scientifically useful.

Coadding multiple images is formally required due to the noise-level requirements placed on the DM system. Additionally, the LSST survey is heavily dithered, so without coadding many images onto a common sky plane it is both difficult and inefficient to obtain image differences for a new pointing from past single images. Finally, single-image templates do not permit removal of artifacts, transients, and moving objects from the template, creating additional false positive sources in the resulting differences.

Scientifically it is important that once a template is constructed for a given region of sky, it is used exclusively until it can be updated in the next Data Release. Repeated changes to the template make it extremely difficult to construct usable lightcurves for objects from individual difference image sources: transient objects such as supernovae will be contaminated by changing flux levels from the evolving template, and variable objects such as variable stars and AGN will require repeated corrections for different template flux levels as well.

4.2 Supporting Incremental Template Generation

The Rubin Construction Data Management (DM) Science team (DM-SST), carried out a study of several options for Alert Production in Year 1, reported in DMTN-107 : Options for Alert Production in LSST Operations Year 1. Representatives of the Rubin Project Science Team (PST), DM-SST and Operations reviewed the proposed DM-SST options and converged on a the following strategy for Alerts in year 1:

- Commissioning Data Templates: Build templates, where possible, from all commissioning data before the start of year one, and use them to generate alerts during year one.
- Year One Data Templates: Build templates progressively from data obtained during year one (e.g., on a monthly timescale), and use them to generate alerts during year one, either instead of, or in addition to using commissioning data to build templates.

To handle alert generation outside the template building process attached to the annual DRP, the Construction project initiated a change request to include incremental templates in the DM system workflow. This change has been accepted and is now part of the baselined DM project in constructiom. A summary of the changes is the following:

- LCR-2273: Construct Image Differencing Templates Outside DRP, new requirement 1.4.6 Template Coadds ID: DMS-REQ-0280, The DMS shall periodically create Template Images in each of the u,g,r,i,z,y passbands. Templates may be constructed as part of executing the Data Release Production payload, or by a separate execution of the Template Generation payload. Prior to their availability from Data Releases these coadds shall be created incrementally when sufficient data passing relevant quality criteria is available.
- To enable artifact rejection, templates will be built with at least three images in year one, and five in subsequent years. (Rubin OSS-REQ-0158)
- Once a template is produced for a sky position and filter it will not be replaced until the next Data Release to avoid repeated baseline changes.
- Templates are not necessarily built from the first N images that are collected.

5 Rubin Observatory Commissioning

5.1 Schedule

Outline the current schedule

5.2 Science-driven prioritization of sky templates

By the end of the commissioning period, coadd templates for use in difference imaging will only be available for ≈ 10% of the sky. This leaves open the question of how to prioritize sky areas and bandpasses to optimize the science harvest prior to Data Release 1 (DR1)(§3).

5.3 Template generation during commissioning

5.4 Template verification in commissioning

The LSST SRD places well-defined criteria on the quality of the difference image and the amount of noise that a template can contribute to a difference image. These criteria result in a minimum of three images being needed to construct a template for use in year one. The commissioning period provides an excellent opportunity to investigate how many visits in a given band are sufficient to construct a usable template. Given the desire to maximize the science harvest prior to the Data Release 1 (DR1), relaxing these criteria might be preferable.

5.5 Alert generation during commissioning

Due to the need to verify the instrument characteristics, template quality, and image differencing and Real/Bogus performance, real-time alerts will not be immediately available during the commissioning period. Where the accumulated ComCam data is sufficient for alert generation, we expect to provide alerts at high latency (weeks–months). The goal for these "canned" alerts is to enable alert brokers and science users to understand their characteristics and to help to validate their quality rather than to enable rapid followup and Early Science per se. During LSSTCam commissioning we intend to incrementally generate templates over the maximal sky area supported by the available observations. These templates will be used for Alert Production, with an aim of approaching near-real-time alert distribution to community

brokers by the time of the Science Validation Surveys at the end of LSSTCam commisioning.

6 Early Science Data Products

6.1 Prompt data products

Prompt data products are described in detail in the Data Products Definition Document (DPDD). Alert packets are triggered by difference image source detections and transmitted to community alert brokers[1] and are publicly available. Similarly, daily Solar System Processing identifies new Solar System Objects from difference image sources and reports those publicly to the Minor Planet Center.

Catalog and image products, as well as services for running user-generated processing on the data, are available to Rubin Data Rights holders after 24 hours through the Rubin Science Platform (§6.3). `DIASource`, `DIAObject`, and `SSObject` catalogs are queryable using VO interfaces to the Prompt Products Database.

6.2 Data Release data products

Images and catalogs from the DRP of the commissioning data will be made available to data rights holders via the access mechanisms described in 6.3.

6.3 Access to Early Science data products

Describe how the Early Science data products will be accessed by the community. Describe the community brokers, the science platforms and how the Data Previews play a role in providing access to Early Science data products. Explain that the same data rights policy applies to Early Science data products. Blum & the Rubin Operations Team (RDO-013).

[1] See `https://www.lsst.org/scientists/alert-brokers` for the list of selected brokers.

7 Survey Cadence

In this section we discuss the implications of any Early Science program on the baseline cadence and overall survey strategy. This section should reference the three early science scenarios presented in 1.3 and explain the implications on the survey cadence for each strategy.

8 Communication

Describe here the process by which the community will be consulted and decisions will be made about the early science programme.

Points to address:

- Use of the community platform for engaging the community to provide input,
- Process by which we officially solicit input from the community on preferences for early science, e.g number of filters, vs area vs pointings vs .... Each different science has a different preference.
- Decision making criteria
- Decision making body
- Timeline for making collecting input and making a decision

Rubin community members can open discussions on the topic of early science on the Rubin Community Forum Early Science category

Several science collaborations have already been pro-active in providing input on template generation in year one, 2021, 2020, 2020

A References

et al., M.E.S., 2021, RNAAS

[DMTN-080], AlSayyad, Y., 2018, *Coaddition Artifact Rejection and* `CompareWarp`, DMTN-080, URL `https://dmtn-080.lsst.io`,
LSST Data Management Technical Note, draft version

[RDO-013], Blum, R., the Rubin Operations Team, 2020, *Vera C. Rubin Observatory Data Policy*, RDO-013, URL `https://ls.st/RDO-013`

[DMTN-107], Graham, M.L., Bellm, E.C., Slater, C.T., the DM System Science Team, 2019, *Options for Alert Production in LSST Operations Year 1*, DMTN-107, URL `https://dmtn-107.lsst.io`,
LSST Data Management Technical Note

Hambleton, K., Bianco, F., Clementini, G., et al., 2020, Research Notes of the AAS, 4, 40, URL `https://doi.org/10.3847/2515-5172/ab8129`, doi:10.3847/2515-5172/ab8129

[LSE-163], Jurić, M., et al., 2017, *LSST Data Products Definition Document*, LSE-163, URL `https://ls.st/LSE-163`

Street, R.A., Bianco, F.B., Bonito, R., et al., 2020, Research Notes of the AAS, 4, 41, URL `https://doi.org/10.3847/2515-5172/ab812a`, doi:10.3847/2515-5172/ab812a

B Acronyms

Acronym	Description
AGN	active galactic nuclei
B	Byte (8 bit)
ComCam	The commissioning camera is a single-raft, 9-CCD camera that will be installed in LSST during commissioning, before the final camera is ready.
DM	Data Management
DM-SST	DM System Science Team
DMS	Data Management Subsystem
DMS-REQ	Data Management System Requirements prefix
DMTN	DM Technical Note

DPDD	Data Product Definition Document
DR1	Data Release 1
DRP	Data Release Production
ES	Early Science
LCR	LSST Change Request
LSST	Legacy Survey of Space and Time (formerly Large Synoptic Survey Telescope)
MAF	Metrics Analysis Framework
OSS	Observatory System Specifications; LSE-30
PP	Prompt Processing
PST	Project Science Team
RDO	Rubin Directors Office
RTN	Rubin Technical Note
SIT	System Integration, Test
SRD	LSST Science Requirements; LPM-17
SST	Subsystem Science Team
SV	Science Validation
TOO	Target of Opportunity
VO	Virtual Observatory