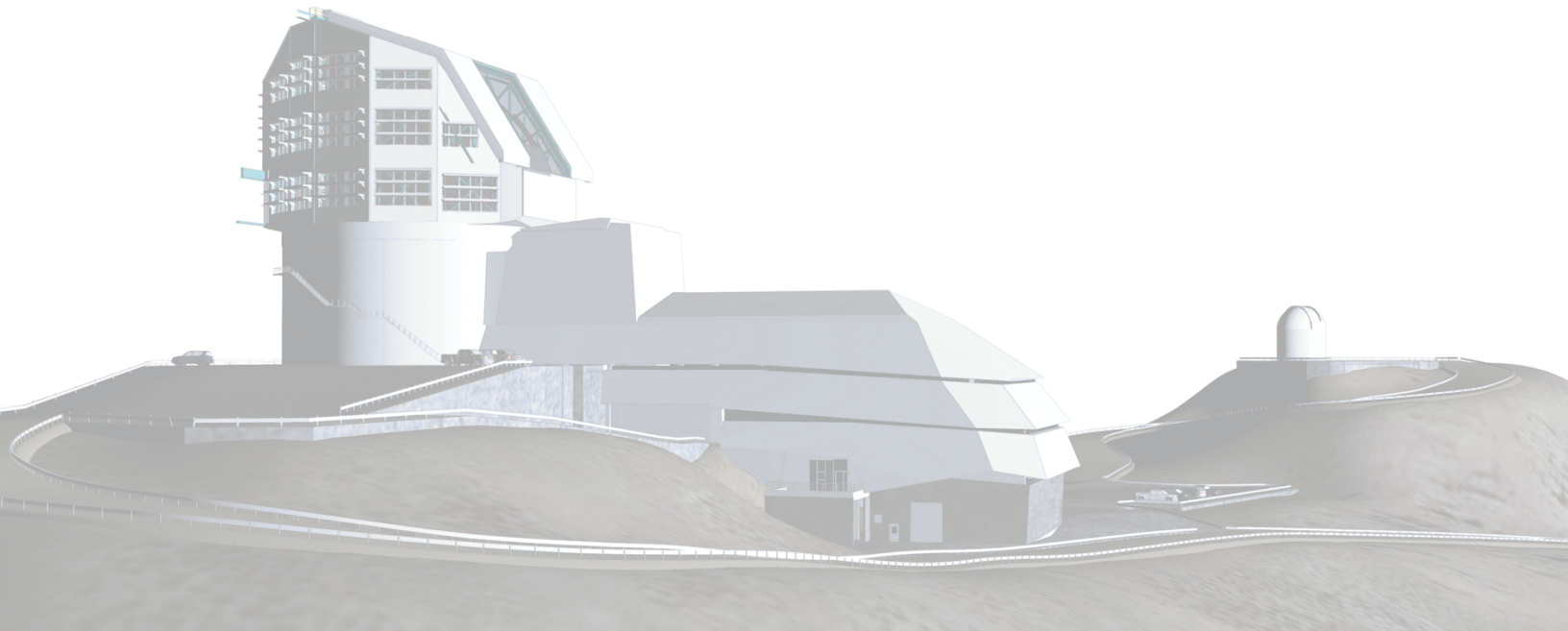
Release Scenarios for Rubin

LSST Commissioning and Survey Data

Phil Marshall, Yusra Alsayyad, Eric Bellm, Leanne Guy, Wil O'Mullane, Zeljko Ivezic, Chuck Claver, Bob Blum, Amanda Bauer, Lauren Corlies, Frossie Economou

RDO-011

Latest Revision Date: February 10, 2023



Change Record

Release	Date	Description	Owner Name
0.1	2021-04-29	Document layout and initial content	Phil Marshall
0.1.1	2021-05-20	Handle switched from RDP-25 to RDO-11	Phil Marshall
0.2	2021-11-05	Schedule updated and initial text finalized.	Phil Marshall
0.3	2022-04-15	Update based on rebaselined (SFR2) construction Project schedule	Phil Marshall
0.4	2022-08-07	Updated dates to connect directly to the Project “celebratory” milestones. Include introduction of Early Science scenarios.	Phil Marshall
1.0	2023-02-10	Streamlined and brought up to date with the v3 Early Science Plan RTN-011	Phil Marshall

Table of Contents

Release Scenarios for Rubin LSST Commissioning and Survey Data

Summary

Rubin Operations is processing and releasing simulated and commissioning data during the pre-operations period in order to develop operating capability and support the LSST science community as they develop their analyses. This short document aims to answer the following questions:

- What is the high level framework for Rubin data previews and releases?
- What variations to the baseline data release scenario might we expect?

The plans for which LSST science pipeline data products are provided in each data preview and data release are described elsewhere, most notably the Early Science Plan, RTN-011.

Reference Documents

“Data Products Definition Document” (DPDD, LSE-163)

Data Preview and Release Planning Tool (DPRPT)

“Rubin Observatory Plans for an Early Science Program” (ESP, RTN-011)

Release Scenarios for Rubin LSST Commissioning and Survey Data

1 Introduction

During survey operations, Vera C. Rubin Observatory will function as a factory of data products from its Legacy Survey of Space and Time (LSST), producing periodic data releases of reduced images and catalogs, a stream of community brokered transient alerts, and a steady accumulation of difference images and catalogs to enable time domain analysis. In order to be ready to process, serve and support these survey data products, and to support the LSST science community in its preparations to analyze those products, the Rubin operations team will issue a series of data releases in advance of construction completion and the start of the LSST. These “Data Previews” will contain simulated LSST data, and commissioning data taken by the Simonyi Survey Telescope, and present challenges of increasing dataset volume, complexity and velocity.

This document outlines the overall framework for data previews and releases derived in conjunction with the Rubin Observatory operations plan. It does not contain a release schedule or plan, but rather it sketches a *baseline release scenario* that the Rubin Operations team is targeting, including both the Data Previews in the pre-operations period, and the data release activity during survey operations, and some discussion of *possible variations* to that baseline. Which of these scenarios plays out in practice will depend on the availability of both the data and the resources to support it.

All data releases are seen as operations or pre-operations activities: the Rubin MREFC Construction project, including commissioning, did not foresee serving data to the

community. During the commissioning and operations period, the plan to release data to the LSST science community will have to be adjusted according to the success of the data processing (the quality of the data products), resources available for preparing and supporting a release, and the need for incremental releases (i.e. will a new intermediate release be a significant improvement on the previous release).

1.1 LSST Data Products

The definitive guide to the LSST data products is the Data Products Definition Document (DPDD, http://ls.st/LSE-163). In order to define the goals of each data preview, we define various broad groupings of the DPDD products. In the early data previews, only a subset of these groupings will be produced; as the pre-operations phase proceeds, more groupings will be added, until the complete set is produced at DR1.

Table 1: Groupings of LSST data products.

LSST DPDD Data Products	Comments
DRP Processed Visit Images and Visit Catalogs	"Visit Catalogs" includes the "Source" table, and CcdVisit, Visit, and VisitMetadata metadata tables. Sources are primarily for diagnostic and calibration purposes.
DRP Coadded Images	
DRP Object and ForcedSource Catalogs	Object characterization measurements plus visit image PSF flux lightcurves.
DRP Difference Images and DIASources	DIA re-run as part of DRP, to refine transient/variable object (and host) characterization.
DRP ForcedSource Catalogs including DIA outputs	Objects derived from both CoaddSources and DIAObjects, measurements on both visit images and difference images.
PP Processed Visit Images	Source catalogs are not accessible (neither with butler, nor in the PPDB).
PP Difference Images	Subject to template availability.

PP Catalogs (DIASources, DIAObjects, DIAForcedSources)	Available in PPDB.
PP SSP Catalogs	SSSource, SSObject, MPCORB. Prereqs: Difference Images, templates. MPCORB is updated (by the MPC) daily, and our hosted version updated with it.
DRP SSP Catalogs	DRP SSP uses only LSST data (no MPCORB info) to understand selection effects. This is only really valuable once we have a significant amount of LSST data, so DR1 at the earliest.

1.2 Access to LSST Data

Data release in this document concerns the delivery of data to LSST Data Rights holders (US, Chilean, and named international data rights holders). LSST data has a 2-year proprietary period. Public release of LSST data after its proprietary period will be described elsewhere.

2 Baseline Data Release Scenario

Three Data Previews are planned, DP0, DP1 and DP2. DP0 is based on simulated data, DP1 and DP2 on commissioning data. DP0 is divided into three phases, each supporting different operations capability development and science analyses. The DP data releases should be spaced roughly equally in time and clearly advertized to the LSST science community via the Early Science Plan.

The DP progression leads up to the first two survey data releases, currently named DR1 and DR2. In the baseline scenario, DR1 will contain data products generated from approximately the first 6 months of the LSST survey. Six months of processing time have been allocated to perform a final test of the processing pipelines, generate the data products, validate their scientific usefulness, document them (and any of their limitations), deliver them to consumers (including Rubin Education and Public Outreach), and prepare the Data Access

Center systems to handle them. The sum of these time periods means that the expected release date for DR1 is one year after the start of the LSST survey. DR2 will contain data products from a whole first year, and is expected to take approximately 12 months to produce following completion of the Year 1 observations.

Beyond DR1 and DR2, we envision approximately annual data releases, with exact release dates announced some months in advance. Since for each annual release it will take approximately 1 year to finalize the pipeline, carry out the data release processing, and then verify and validate the data products, the data release that contains the images and catalogs derived from the Year N observations will be released about 2 years after the start of Year N. Roughly speaking it takes one year to survey the sky, then 1 year to process and release the data.[1] Table 7 shows one possible scenario, in which annual data releases are issued on the same date each year. In practice, the annual data release dates are likely to vary from year to year, in response to changing circumstances and science needs.

Table 2: Nominal survey data release dates for the baseline scenario, assuming a nominal LSST survey start in January 2025.

Data Release	Latest Year of Observations	Nominal Date Range of Observations			Nominal Release Date	Nominal Public Availability
DR1	First 6 months	January 2025	through	July 2025	January 2026	February 2028
DR2	Year 1	January 2025	through	January 2026	January 2027	February 2029
DR3	Year 2	January 2025	through	January 2027	February 2028	February 2030
DR4	Year 3	January 2025	through	January 2028	January 2029	February 2031
DR5	Year 4	January 2025	through	January 2029	January 2030	February 2032

[1] The timing of the data release processing means that DR2, which will contain the data taken during Year 1, will be released at the end of Year 2. This may prove confusing, having DR2 released during Year 2 but not containing any of the Year 2 data. Rubin is considering indexing the survey data releases from zero instead, such that DR6 contains all the Year 6 data (and will be released at the end of survey Year 7).

DR6	Year 5	January 2025	through	January 2030	January 2031	February 2033
DR7	Year 6	January 2025	through	January 2031	February 2032	February 2034
DR8	Year 7	January 2025	through	January 2032	January 2033	February 2035
DR9	Year 8	January 2025	through	January 2033	January 2034	February 2036
DR10	Year 9	January 2025	through	January 2034	January 2035	February 2037
DR11	Year 10	January 2025	through	January 2035	February 2036	February 2038

While the table gives nominal release dates in order to illustrate the annual pattern, date *ranges* are our preferred way to communicate forecast data preview and release dates, since they provide a way to indicate our uncertainty on when these releases will be possible. That uncertainty is currently dominated by the construction Project completion schedule, but in the future will be driven by operational considerations to do with the observing schedule, summit maintenance, and image processing. Rubin Operations will maintain and publish a set of key milestones, with associated date ranges, and then narrow the uncertainty ranges as the release dates approach. Some uncertainty will remain throughout survey operations, with release date forecasts refined and announced as data processing progresses.

3 Variations on the Baseline Data Release Scenario

In this section, which is under continuous development, we explore some possible alternatives to the baseline release scenario outlined above.

RTN-011 “Rubin Observatory Plans for Early Science” contains detailed discussion of the possible contents of DP2, DR1, and the alert stream during those two eras. This discussion allows for an alternative commissioning scenario in which the Science Validation Surveys are carried out partly by the operations team, before the LSST begins. In this scenario there is no

change to the contents of DP2, just an extension of the timeline for DP2, DR1, and subsequent data releases.

Tech note RTN-003 “Guidelines for Rubin Independent Data Access Centers” mentions an “Object Lite” data product containing a subset of the columns of the full object table (see Table 3 in RTN-003 for a rough estimate of its size). This data product is expected to be attractive for distribution to “Lite” IDACs, and may also be attractive for broader distribution within the Rubin community. Definition of the contents of an Object Lite data product has not yet been determined, but is expected to flow from detailed science and technical use cases currently under discussion. Once defined, inclusion of Object Lite may be considered in the data preview and first data release plan.

While the LSST has long been envisioned as providing annual data releases to enable the science community to keep up with the survey and exploit it in a timely manner. However, we can imagine scenarios where, in response to changing circumstances and the discoveries made in and experience gained from the survey, the schedule of data releases needs to be adjusted. This section will be used to sketch out some possible scenarios and how they would be navigated.