



Vera C. Rubin Observatory
Rubin Observatory Operations

Data Preview 0.2 and Operations rehearsal for DRP.

William O'Mullane, Yusra Alsayyad, Hsin-Fang Chiang,
Frossie Economou, Fritz Mueller, Tim Jenness

RTN-041

Latest Revision: 2022-09-02



Abstract

DM delivered software to operations to perform processing of the DESC DC2 data as well as enhancements to the portal and Qserv for interaction with the results. The release of this was called Data Preview 0.2 and the production of the data products and publication of them were carried out in an operational manner. This provides valuable insights for operational data releases.

Change Record

Version	Date	Description	Owner name
1	2022-08-02	Unreleased.	William O'Mullane

Document source location: <https://github.com/lst/rtn-041>

Contents

1 Introduction	1
2 Management and communication	1
2.1 Oversight	1
2.2 Coordination	2
2.3 Work Management	2
2.4 Change Control	3
3 DP0.2 processing on Google	4
4 Data Product Quality Assurance	4
5 Science Platform, user front end for DP0.2	5
6 Community engagement	5
A References	5
B Acronyms	5

Data Preview 0.2 and Operations rehearsal for DRP.

1 Introduction

Between December 2021 and May 2022 the DESC DC2 (LSST Dark Energy Science Collaboration (LSST DESC) et al., 2021) was reprocessed with Rubin Science pipelines V23 [DMTR-351].¹ Between May and June the catalogs were ingested to Qserv, tutorials and documentation were updated and the Data Preview 0.2 data release was made on time at the end of June. A number of procedures were developed and practiced to achieve this. Planning for DP0.1 and DP0.2 are in RTN-001. We shall discuss the process in the following sections:

- Management and communication is discussed in Section 2
- An overview of the processing is given in Section 3
- Quality assurance and feedback to processing is discussed in Section 4
- Community engagement, tutorials and documentation are discussed in Section 6

2 Management and communication

Here we cover the management structures in place for DP0.2 this includes the groups and meetings like the change control for the pipeline version.

2.1 Oversight

The Data Production Leadership Team (DPLT) consisted of representatives from all the teams involved in the data preview process as well as the data facilities. The DPLT met fortnightly to discuss any issues and minutes were recorded on Confluence.

The membership of the DPLT was:

- Wil O'Mullane

¹https://pipelines.lsst.io/v/v23_0_2/index.html

- Bob Blum
- Leanne Guy
- Frossie Economou
- Yusra Alsayyad
- Hsin-Fang Chiang
- Michelle Butler (NCSA)
- Richard Dubois (USDF)
- George Beckett (UKDF)
- Fabio Hernandez (FrDF)

2.2 Coordination

During the data preview 0.2 process there was a regular coordination meeting every two weeks (out of phase with the DPLT meeting) with minutes recorded on Confluence. This meeting was attended by all the people directly involved in the data preview process: management, the processing infrastructure team, the science platform team, the execution team, the pipelines team, the verification and validation team, and the community engagement team.

This allowed the different teams to report on status and bring up any issues that needed to be addressed and made everyone aware of progress. Data Preview 0.1 had been released during this period and this allowed us to also include feedback from the Community Engagement Team as they interacted with the existing delegates and prepared updated tutorials for DP0.2.

Data Facility representatives were present at this meeting even though all the processing was being done on the Interim Data Facility at Google (O'Mullane et al., 2021).

2.3 Work Management

We used Jira to track work related to the Data Preview. Epics and milestones were created in the PREOPS Jira Project. Story tickets were then attached to each epic but in order to properly integrate the work into existing Data Management processes, any tickets that would result in

code changes in pipelines software or middleware packages were created in the Data Management Jira project.

The status of the epics and how they related to the relevant milestones was monitored as part of the weekly coordination or DPLT meetings.

2.4 Change Control

Data Preview 0.2 used v23.0.x of the LSST Science Pipelines Software and that was derived from a weekly release from September 2021 (w. 2022.40). We decided to group processing into distinct “steps” that allowed updates to the software used in later stages of processing to be worked on whilst earlier steps were executing.

We continued to want to use the v23 release for all data processing and that required that we had a process to determine which patches would need to be back-ported to the release branch as needed before each step could begin.

The Data Management Change Control Board (DMCCB) and DPLT delegated authority to a new Campaign Management Board that had the following membership:

- Yusra AlSayyad, representing the pipelines team.
- Leanne Guy and Colin Slater, representing the verification and validation team.
- Hsin-Fang Chiang, representing the execution team.
- Tim Jenness, representing the data processing architecture team.

The Board met weekly on Tuesday at 8:30am Pacific Time and also had a Slack channel to discuss any issues that would come up between meetings. Minutes for the meetings were recorded on Confluence.

The process for deciding on a back-port is as follows:

1. A request is made that a ticket should be applied to the release branch by applying a `backport-v23` tag to the Jira ticket.

2. The board would then discuss the relative merits of the back-port and if approved a backport-approved label would be added.
3. The work on the back-port would then be scheduled by the relevant T/CAM following instructions in the developer guide.²
4. Once the code is on the v23.0.x branch a backport-done label would be applied.

A Jira query was constructed to find all the tickets and track their porting status. There were 64 tickets with back-port requests and 61 of those were approved and applied to the release branch. For the three that were not approved, one is for a clean-up to the database schema that was discovered after we had finalized the processing; another was for an improvement to the graph-building efficiency but would have involved a very difficult back-port because there had been a package reorganization since the release branch had been created; and the final ticket was an improvement to the matched catalog filtering that was determined to be non-critical.

Once all the necessary back-porting has been completed for a specific step, the release manager would be instructed to start the process of creating a new patch release of the Science Pipelines. During DP0.2 we made two additional formal releases of the version 23 software: v23.0.1 and v23.0.2. This allowed us to state which release was used for each step, although we ensured that changes in later patch releases would not affect the processing from steps that were already completed using older patch releases.

3 DP0.2 processing on Google

Description of the workflow and the Jira system - refer to RTN-039

Discuss some of the issues which came up and how they were resolved.

4 Data Product Quality Assurance

Description on the checks run, perhaps a few plots . Anything to do differently next time ?

²<https://developer.lsst.io/work/backports.html>

5 Science Platform, user front end for DP0.2

DP0.1 upgrade image services any new issues but mainly deployment and control, issue handling

Qserv Load and new metadata model

6 Community engagement

Organisation and execution of the tutorials, docs, assemblies ..

A References

[DMTR-351], Carlin, J., 2022, *Characterization Metric Report: Science Pipelines Version 23.0.0*, DMTR-351, URL <https://dmtr-351.lsst.io/>,
Vera C. Rubin Observatory Data Management Test Report

LSST Dark Energy Science Collaboration (LSST DESC), Abolfathi, B., Alonso, D., et al., 2021, *ApJS*, 253, 31 (arXiv:2010.05926), doi:10.3847/1538-4365/abd62c, ADS Link

[RTN-001], O'Mullane, W., 2021, *Data Preview 0: Definition and planning.*, RTN-001, URL <https://rtn-001.lsst.io/>,
Vera C. Rubin Observatory Technical Note

O'Mullane, W., Economou, F., Huang, F., et al., 2021, arXiv e-prints, arXiv:2111.15030 (arXiv:2111.15030), ADS Link

[RTN-039], Yanny, B., 2022, *Compute Resource Usage of DP0.2 production run*, RTN-039, URL <https://rtn-039.lsst.io/>,
Vera C. Rubin Observatory Technical Note

B Acronyms

Acronym	Description
CAM	CAMera
DC2	Data Challenge 2 (DESC)
DESC	Dark Energy Science Collaboration
DM	Data Management
DMCCB	DM Change Control Board
DMTR	DM Test Report
DPO	Data Preview 0
DPLT	DP Leadership Team
DRP	Data Release Production
FrDF	French Data Facility
LSST	Legacy Survey of Space and Time (formerly Large Synoptic Survey Telescope)
NCSA	National Center for Supercomputing Applications
OPS	Operations
RTN	Rubin Technical Note
T/CAM	Technical/Control (or Cost) Account Manager
UKDF	United Kingdom Data Facility
USDF	United States Data Facility