Rubin Observatory

Data Preview 0: Definition and planning.

William O'Mullane

2020-05-29

1 Introduction

Table 1 shows the FY21 milestones for the Vera C. Rubin Observatory, many of which concern, or relate to, data previews. Section 2 defines what Data Preview 0 is about and covers possible risks and mitigations to that definition. Section 3 Sets out the planning for achieving DP0.

Table 1: Milestones for Rubin Observatory FY21

Milestone	Label	Year	Q	Type	Team
Read only Gen3 butler for DP0 at IDF	DP-IN-01	FY21	Q1	Code Release	Science Users Middleware
IDF DP0-Ready: Complete IDF installation and IDF staff preparations for DP0.		FY21	Q1		Infrastructure and Support
Submit FY20 POP Annual Progress Report to NSF (via OIR Lab POPPR)		FY21	Q1		Rubin Observatory Directorate
Open up the Help Desk for LSST Operations staff to use.		FY21	Q1		Community Engagement
Establish new media presence on at least one new channel for Rubin operations.		FY21	Q1		Outreach
Produce FY22 POP input for DOE budget briefing	DP-SP-01	FY21	Q2	Reporting	Rubin Observatory Directorate
DP0.1 Early Access: Provide access to processed images and visit level catalogs from the IDF		FY21	Q2		Science Platform and Reliability Engineering
Announce Initial Survey Strategy		FY21	Q2		Survey Scheduling
Deliver Q1 Report to NSF on POP21 status (via OIR Lab Q report)		FY21	Q2		Rubin Observatory Directorate
USDF Decision: obtain confirmed location of US Data Facility		FY21	Q3		Rubin Observatory Directorate
Deliver Q2 Report to NSF on POP21 status (via OIR Lab Q report)		FY21	Q3		Rubin Observatory Directorate
Identify Observatory Operations Team Leads (Observatory Software and Summit and Engineering Operation) or launch external searches.		FY21	Q3		Observatory Operations Management
Submit FY23 DOE FWP(s)	DP-SP-02	FY21	Q3	Reporting	Rubin Observatory Directorate
Gen3 butler backed by S3 for processing DP0		FY21	Q3		Science Users Middleware
DP0.2 Reprocessing Start: Begin early DRP-like re-processing of DP0 simulated image data, at the IDF.	DP-SP-M-01	FY21	Q3	Event	Execution
Begin operations of IDF to support shared-risk simulated data distribution to community		FY21	Q3		Community Engagement
Demonstrate EPO interface with DP0		FY21	Q3		Science Platform and Reliability Engineering
Stand up Users Committee so that it is active for DP0.1 feedback from Science Collaborations.	DP-PM-01	FY21	Q3	Process Definition	Community Engagement
DP0.1 Data Release: science-ready catalogs released from the IDF		FY21	Q3		Verification and Validation
USDF Transition Plan: work with selected USDF team to plan start-up of USDF.	DP-SP-03	FY21	Q3	Process Definition	Data Production Management

Draft

Deliver Q3 Report to NSF on POP21 status (via OIR Lab Q report)	DP-SP-04	FY21	Q4	Reporting	Rubin Observatory Directorate
		FY21	Q4		Science Platform and Reliability Engineering
		FY21	Q4		Science Platform and Reliability Engineering
		FY21	Q4		Rubin Observatory Directorate
		FY21	Q4		Rubin Observatory Directorate
		FY21	Q4		Education
Establish Communications Strategy for Operations	DP-MW-M-01	FY21	Q4	Process Definition	EPO Management
Establish Communications Strategy for Operations	DP-MW-M-02	FY21	Q4	Process Definition	Communications

2 Data Preview 0

The first ideas about initial releases of Rubin Observatorydata were presented in LSO-011. There have since been delays in construction such that we have considered making Data Previews with simulated data or other non Rubin Observatorydata (see Section **??**).

We now propose to do this via the Intermediate Data Facility. This facility would only be for pre operation activities e.g. serving data and training operations staff. Commissioning actives would continue at NCSA and Chile.

Here we discuss DP0 in more detail.

2.1 Interface

More ambitiously than LSO-011 we would like to allow access to DP0 via the Science Platform.

2.1.1 Images

Images would be accessible via a read only Gen3 butler repo.

2.1.2 Catalogs

Catalogs would be served via the VO TAP interface preferably backed by Qserv.

Draft

2.2 Dataset(s)

For DP0 we will use existing catalogs and products, images should be in a Gen3 Object Store backed repository (with S3 interface).

From a usability perspective the best data set to use would be HSC PDR2. Though public use of this should be cleared with the HSC colleagues. This data set is well understood and semi regularly processed with our stack. It is real data which is potentially more interesting for exploration than simulated data. The current repos are Gen2 so would need conversion meaning some subset may have to be used.

The main simulated data set would be DESC DC2. We need permission from the DESC colleagues. This is a large dataset and therefore interesting. It is currently only in Gen2 butler. There is a conversion script, however it runs serially and would take prohibitively long to convert to Gen3. We could use a subset which would be fine. Putting DC2 catalogs in Qserv would be an excellent demonstration of its abilities.

2.3 Risks and mitigation

The biggest schedule risk is not getting an interim data facility in place in time. This would delay the entire schedule and there is not much mitigation.

In the long run costs may be higher than expected in a cloud based IDF. This will be due to storage. An mitigation to this would be to store data on our own systems (NCSA or Chile) and expose it through S3. NCSA already have this in place and we should consider testing this for lesser used data sets.

3 Plannignn and team(s) fro DP0

A References

Draft

References

[LSO-011], William O'Mullane, L.G., Phil Marshall, 2019, *Release Scenarios for LSST Data*, LSO-011, URL https://lso-011.lsst.io

B Acronyms

Draft