Data Preview 0: Definition and planning.

William O'Mullane

2021-08-07

1 Introduction

Table 1 shows the milestones for the Rubin Observatory, many of which concern, or relate to, data previews. Table 4 shows the already achieved milestones. 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 Data Production and System Performance

Milestone	Jira ID	Rubin ID	Due Date	Level	Status	Team
PanDA based workflow system in place	PREOPS-154	L3-MW-0050	2021-03-31	3	To Do	Science Users Middleware
Gen3 butler and pipeline task ready for DP0 production	PREOPS-156	L3-MW-0070	2021-06-10	3	In Progress	Science Users Middleware
Engage with the community to support shared-risk simulated data distribution to community for science with DP0	PREOPS-150	L2-SP-0020	2021-06-30	2	In Progress	Community Engagement
Science Platform ready on for DP0.2	PREOPS-157	L3-PR-0040	2021-06-30	3	To Do	Science Platform and Reliability Engineering
PanDA based workflow system with tooling (e.g. restart) added.	PREOPS-155	L3-MW-0060	2021-06-30	3	In Progress	Science Users Middleware
Evaluate Batch Production System	PREOPS-153	L3-MW-0040	2021-07-31	3	To Do	Science Users Middleware
Demonstrate EPO interface with DP0	PREOPS-152	L3-PR-0030	2021-09-30	3	To Do	Science Platform and Reliability Engineering
DP0.2 Early Access: Provide access to reprocessed images and visit level catalogs from the IDF	PREOPS-159	L2-DP-0040	2021-09-30	2	To Do	Science Platform and Reliability Engineering
DP0.2 Reprocessing Start: Begin early DRP-like reprocessing of DP0 simulated image data, at the IDF.	PREOPS-158	L2-DP-0030	2021-09-30	3	To Do	Execution
Plan for how to use IN2P3 in DP0.2	PREOPS-160	L3-EX-0010	2021-09-30	3	To Do	Execution
Deploy early instantiation of service desk providing second-tier technical support for community	PREOPS-147	L3-PR-0020	2021-09-30	3	To Do	Science Platform and Reliability Engineering
Deliver initial Quality Assessment and Assurance (QA) plan for ComCam Data.	PREOPS-293	FY20-0010	2021-10-30	2	To Do	Verification and Validation
Deliver preliminary implementation plan for real-time and daily monitoring	PREOPS-515	L3-SC-0020	2021-12-31	3	To Do	Survey Scheduling
Deliver preliminary list of metrics for real-time and daily monitoring	PREOPS-514	L3-SC-0010	2021-12-31	3	To Do	Survey Scheduling
Deliver preliminary list of metrics for quarterly monitoring	PREOPS-517	L3-SC-0040	2021-12-31	3	To Do	Survey Scheduling
Deliver LSST Data Products Documentation (DP0)	PREOPS-149	L3-CE-0010	2022-03-31	3	In Progress	Community Engagement
L2 - DP0.2 Public release to delegates	PREOPS-483	L2-DP-0051	2022-06-01	2	To Do	Data Production Management
L2 - DP0.2 Data Release: science-ready catalogs from reprocessed DP0 images released from the IDF	PREOPS-484	L2-PF-0052	2022-06-30	2	To Do	System Performance Management
L2 - USDF Initial setup	PREOPS-492	L2-DP-0081	2022-07-31	2	To Do	Infrastructure and Support

Deliver implementation of real-time and daily monitoring system	PREOPS-516	L3-SC-0030	2022-08-31	3	To Do	Survey Scheduling
Deliver implementation of quarterly metric monitoring	PREOPS-518	L3-SC-0050	2022-12-30	None	To Do	Survey Scheduling
L2 - Announce Initial Survey Strategy	PREOPS-490	L2-SP-0060	2022-12-30	2	To Do	System Performance Management

2 Data Preview 0

In LSO-011 we outlined a number of scenarios for early releases of Rubin Observatory data. The purpose of the these releases are not only to prepare the community for LSST data, but also to serve as an early integration test of existing elements of the Data Management systems and to familiarize the community with our access mechanisms.

Two major new developments have occurred since LSO-011 was drafted:

- There have since been delays in construction such that we are now planning on making Data Previews with Rubin Observatory simulated data or on-sky data from other observatories (see Section C.1.1) which would still allow us to meet some of the goals of the early releases.
- We are planning on carrying these activities at the Interim Data Facility, which is is dedicated to Pre-Ops activities infrastructure needs such as serving data and training operations staff. (Commissioning actives will continue at NCSA and in Chile.)

In this document we outline notable elements of DP0, the first of these planned data previews, from the Data Management and Pre-Operations perspective.

Data Preview 0 itself was broken down in two parts: 0.1 (Appendix C.1) servings existing data products, 0.2 (Section 2.1)reprocessing that data and publishing new catalogs.

Since DP0.1 has been released that text has been moved to an appendix (Appendix C.1).

A DP0.3 has been mentioned but no agreement has been made to do this (apart from tha tit must be real data like HSC). No plannign for that will be done until 2022 when we are confident about DP0.2.

2.1 DP0.2 - processing

The Milestone L2-DP-0040 includes re processing on IDF of the data set previously served as part of L2-DP-0020. This requires a workflow system and associated tools to preferably make this quite automated. Demonstrating a portable set of cloud enabled tools based on Butler Gen3 and PanDA would help to allay the main risk of moving to a new Data Facility in operations. As of today, processing based on Butler Gen3 has been limited to a very small scale, and no scalability testing has been performed. For L2-DP-0040 we intend to reprocess DC2 RC6 dataset late in 2021 or early 2022.

2.1.1 Purpose of DP0.2

The purpose of DP0.2 is manifold, in order of priority:

1. generate a fully self-consistent data release for the scientists to publish papers on
2. Is the purpose to follow a formal data release process with backporting and CCB approvals before allowing new software versions to be used but still taking into account that construction is still ongoing and some flexibility is warranted
3. perform mini runs early on to improve the chosen pipeline release
4. Serve as an operations rehearsal for DRP.

2.1.2 Policy committee

There are certain decisions which will need to be made are best handled in a smaller forum than DPLT. This may include:

- Campaign polices
- Version of pipelines to use and patches which are needed
- Version of QA tooling which needs to run (and where/how to run it)
- Other operational considerations

Such decisions will be endorsed by DPLT but advised by a smaller committee more connected to the issues. The members will be the following (or their delegated representative):

- Hsin-Fang Chiang
- Tim Jenness
- Yusra AlSayyad
- Colin Slater

This is basically one representative each from Science Pipelines, V&V, Middleware, and Execution. It also serves as a trial for operations proper.

2.1.3 Science pipelines release

We have milestone L3-AP-0010 for the DP0.2 release which is satisfied by v22.0.1 of the science pipelines. This will be good to evaluate PanDA. For the actual reprocessing, given the timeline, we will make a v23 release when we have the weekly in a state we feel os good for DP0.2. Should that need fixes they will then be incremental patches on v23.

Hence the delivery of v23 will be driven by the need for DP0.2 rather than time based - this is a more operational way to approach the release. It will also require support of this releases version for a period of time. This implies backporting agreed fixes (through RFC to DMCCB). A support period of one year seems reasonable.

2.2 Workflow engine

BNL have been working to demonstrate PanDA with Gen3 for a while. July 2021 is a decision point on using this for DP0.2 RTN-013 provided the goals for this task. DMTN-168 provides guidelines on how to use this system.

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.

There is some risk that Butler over S3 and Postgres might not be at production grade by DP0. We are working hard on that in construction. There is the possibility to run Gen 3 over a filesystem which would not be ideal on the cloud. If Gen3 does not work at all we will have to have a major rethink and build a much simpler butler. Similarly, the workflow system and associated tools may not be mature enough for large-scale production. Scalability in production is also not understood. We may need to limit the size of DP0 and rethink the system.

3 Planning and team(s) fro DP0

Planning epics have been (and are) being created in the PREOPS Jira project. On the dashboard you can see links to the tickets labeled DP0.1 and DP0.2.

We will have regular (every other week for now) DP0 meetings (see `https://confluence.lsstcorp.org/display/LSSTOps/Data+Production+Meetings`).

3.1 Teams

The Operations era org chart is shown in Figure 1.

The main departments involved in DP0 are Data Production and System Performance. With in those departments various people will be involved from the underlying teams but in small numbers. It makes most sense to approach DP0 with a task force approach. This might best be seen as two teams:

- Data production - with a focus on middleware and execution (Section 3.2);
- System Performance - with a focus on quality assurance and community support (Section 3.3).

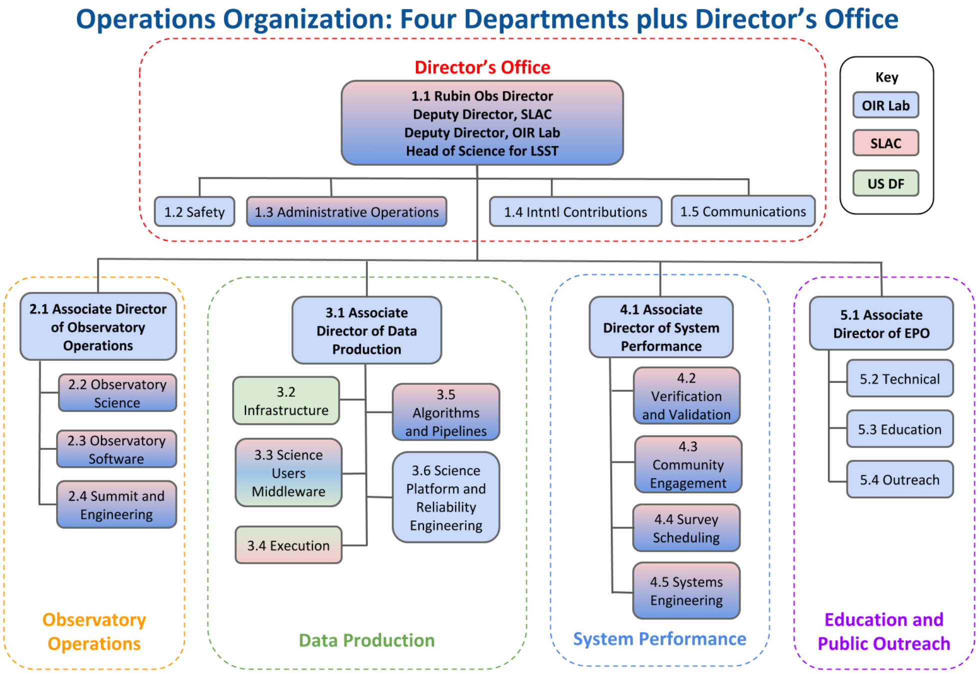

Operations Organization: Four Departments plus Director's Office
Director's Office
1.1 Rubin Obs Director
Deputy Director, SLAC
Deputy Director, OIR Lab
Head of Science for LSST
1.2 Safety
1.3 Administrative Operations
1.4 Intntl Contributions
1.5 Communications
Key
OIR Lab
SLAC
US DF
2.1 Associate Director of Observatory Operations
2.2 Observatory Science
2.3 Observatory Software
2.4 Summit and Engineering
Observatory Operations
3.1 Associate Director of Data Production
3.2 Infrastructure
3.3 Science Users Middleware
3.4 Execution
3.5 Algorithms and Pipelines
3.6 Science Platform and Reliability Engineering
Data Production
4.1 Associate Director of System Performance
4.2 Verification and Validation
4.3 Community Engagement
4.4 Survey Scheduling
4.5 Systems Engineering
System Performance
5.1 Associate Director of EPO
5.2 Technical
5.3 Education
5.4 Outreach
Education and Public Outreach

FIGURE 1: Organization of departments and teams for operations of Rubin Observatory.

As we advance the teams grow and we will transition to the an organization as in Figure 1 with team leads for each team as in Figure 2.

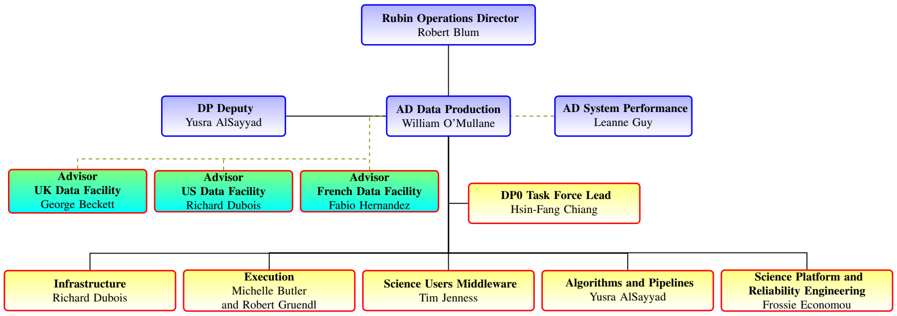


FIGURE 2: Data Production team structure

3.1.1 Task force lead

For DP0 on the IDF a task force approach seems most appropriate given the partial efforts in all teams. Hsin-Fang Chiang shall fulfill this role and coordinate Data Production activities for DP0. Responsibilities of this role include:

- Being point of contact for the IDF provider.
- Setting priorities for all work at the IDF until DP0 is fully complete.
- Evaluate stage-wise operational readiness wrt. to requirements.
- Make all components of the IDF work together (Science Platform, Middleware, Workflow ..)

The task force lead reports directly to the Data Production Associate Director and carries delegated authority for the above responsibilities.

3.2 DP Middleware and Execution

There is preops effort (fractional FTE) available in Execution and Pipelines as well as Middleware teams. The roles etc need some clean up from the ops proposal but the DP Roles are listed in Table 5 though the exact mix of roles is still under discussion.

3.3 SP Quality and Community Support

Note: DP0.1 and DP0.2 Early Access described in this document do not leave time for full-scale quality analysis. The provided data will not be science-ready; system performance milestones are succeeding.

Leanne ..

- How do we intend to do support? Slack? JIRA? CLO?

3.4 Planning

Table 2: Internal timeline

Date	Description	Reference
Jul 2020	Small test datasets identified to help dataset choice	Sec C.1.1
Aug 2020	Decision on DP0.1 dataset	Sec C.1.1
	Software freeze for repo conversion to Gen3 read-only Butler	L3-MW-0030
Dec 2020	Qserv installed and configured on IDF	L3-MW-0010
Jan 2021	Qserv ingestion starts on IDF	
Feb 2021	Qserv scale test	
Feb 2021	TAP service scale test	
Nov 2020	First workflow tools software release	
Jan 2021	Small test datasets available on IDF	
Jan 2021	Batch system configured on IDF	L3-MW-0050
Feb 2021	Test batch processing of the small dataset on IDF	
Apr 2021	Tract size verification run on stack candidate	
Jun 2021	Baseline Software for DP0.2 pipeline stack (v22)	

Table 2 lists internal timeline.

3.4.1 Middleware

There are obvious middleware milestones such as L3-MW-0030 read only Gen3 Butler which are needed from the construction project. There is still installation work needed for the that on Google which includes the need for a Postgress (like) database for the registry. The DAX team are on the hook for this. For DP0.2 we need Butler to handle processing, not just locating files (L3-MW-0070).

3.4.1.1 Qserv should be installed and configured. Though we have some prior art for this we still will need some experimentation to get it correct. Getting DC2 loaded in Qserv is also a DAX activity we will have to do on IDF.

3.4.1.2 Workflow needs to be functioning at scale for DP0.2, ideally we should have basic workflow early on (milestone L3-MW-0050). Then more tooling such as restarting failed jobs (L3-MW-0060).

From the construction side we have BPS as a deliverable which may be useful on IDF also. We shall evaluate BPS as an option later in 2020 (L3-MW-0040). See LDM-636, LDM-633, DMTN-123. BPS translates the quantum graph to DAGMan for execution on HTCondor and submits the jobs. Most work has gone into the graph and execution.

As part of our march toward a potential more DOE oriented Data Facility, BNL will be part of the pre operations team to experiment with PanDA as an environment to monitor and control our processing jobs. This is a slightly parallel effort to construction attempting to take advantage of an existing set of tools for large scale job execution. In an ideal world the quantum graph translation of BPS would feed into a PanDA system to execute (retry etc) our jobs, this is still to be investigated. This may go through CWL.

See also Section 3.4.4.

3.4.2 Science Platform

The science platform and web services need to be deployed. In principle this is reasonable straight forward, an open issue may be configuring of the Portal aspect for the chosen dataset(s).

3.4.3 Pipelines

For DP0.2 we need a Gen3 version of the pipelines to process the dataset. This will have to run at scale for PDR2 or DC2. There may be several runs for quality purposes. Fractional FTE from the Pipelines will provide help in pipeline configuration, data repo preparation, workflow consulting, science verification, data model documenting, troubleshooting, and liaising. **Yusra will provide more info here.**

3.4.4 IN2P3

IN2P3 will contribute in Qserv and pipelines. **Fabio will provide more information here.** They bring experience running Gen3 workflows. The real interest with IN2P3 is to run remote jobs thus emulating the eventual operational DRP runs. This may be difficult to achieve in FY21 but we should make it a milestone for FY22.[1]. A more achievable goal for FY21 would be to duplicate the IDF processing at IN2P3.

Remote execution requires some features in Gen3 to be implemented. We will probably wish to execute jobs with a local registry then merge the results and registries.

[1] Tim, Fabio we should set a date for this

IN2P3 maintains a separate Qserv cluster. The same catalog data will be ingested into the Qserv instance at IDF and the Qserv instance at IN2P3 and the databases will be cross-checked for consistency.

4 Other experiments

Apart from the milestones and planning in Section 3 there are some other activities it may be good to experiment with.

4.1 S3 access to NCSA

Storage remains the cost driver for cloud. We have an S3 interface exposing data a t NCSA, we could attempt some processing on the cloud accessing image data at NCSA.

4.2 Qserv 75% scaling

Qserv scale tests should go to 75% of DR1. This requires a lot of nodes for a short time, we do not need to necessarily keep all those nodes once the test is done. This is an ideal cloud scenario if we have Qserv working in an understood manner on the cloud. DMTN-125 would suggest we can at least do this in principle.

A Dataset choice considerations

For DP0 we are moving ahead with DESC DR6 as the baseline. This section is included for historical context on the decision making.

There were two leading candidates for forming the basis of DP0:

- The Subaru Hyper Suprime-Cam PDR2 dataset, provided permission can be secured from our HSC colleagues. As real (on-sky) data it is likely that users will interact with it in more realistic ways. It is a well understood dataset, and it is regularly re-processed with software that shares a common codebase with the LSST Science Pipelines.

- The simulated precursor to LSST data produced by the Dark Energy Science Collaboration, DESC DC2, provided permission can be secured. This is a very large dataset and putting DC2 catalogs in Qserv would be an excellent demonstration of its abilities.

There is interest from the science collaborations in working with data products from both of these datasets. DC2 was emphasized at the 2019 PCW, and at least one (AGN) has contributed to the simulation inputs since then. A comment at the PCW discussion was that without DC2 in DP0, the science collaborations would not see full frame LSST data until the year before the survey, too late for the needed analysis development.

Data Management is currently in transition between its 2nd and 3rd generation data abstraction layer (aka "Butler"). For DP0 to fulfill its aim as an early deployment/integration exercise, Gen 3 Butler must be used, preferably (stretch goal) using an S3 compliant Object Store as is the intent in production. This has bearing on the choice of dataset.

HSC PDR2 can either be converted from Gen 2 to Gen 3 or (stretch goal but ideally) reprocessed natively with Gen3. A smaller subset may be necessary to avoid production scaling issues. This is the preferred choice in the short term from an engineering point of view.

DC2 is available through Gen2 Butler and as we do not process that data with the Science Pipelines, the only option is conversion to Gen3. Estimates are that this is such a time-consuming process that it cannot be done in time to meet milestone L2-DP-0020. Therefore if DC2 is to be involved in the short term, a significantly smaller subset would have to be selected.

In the case that we do not reprocess the data with updated Science Pipelines, we can serve the data as they are provided to us. For example, in DC2, DESC's codes were used to generate science-ready catalogs, which can be ingested into Qserv without further standardization. Detailed provenance between images in the Butler repo and the Qserv catalogs may not be provided in DP0.1, but improvements will be made in DP0.2 when we reprocess the data.

Questions:

- Which dataset has the broader scientific interest? This question could be answered via a community survey: indeed, the possibility of such a survey was discussed at the 2019 PCW.

- For either dataset if we take a subset to avoid the Gen2-Gen3 conversion issues or production scaling issues, will that reduce the usefulness of the datasets or affect the choice? What would be the smallest data size that is still scientifically interesting?
- Are there HiPS maps available for either of these ?
- Given the delayed construction/commissioning schedule, could we consider including both of these datasets in DP0 over the course of FY21–FY22?

B Data Products in the Butler Repository for DP0.1

Dataset Type	Count
skyMap	1
raw	3652567
icSrc	3651927
calexp	3651777
calexpBackground	3651777
src	3651777
srcMatch	3651777
skyCorr	3650625
deepCoadd	42206
deepCoadd_calexp	42206
deepCoadd_calexp_background	42206
deepCoadd_deblendedFlux	42206
deepCoadd_det	42206
deepCoadd_forced_src	42206
deepCoadd_mcalmax_deblended	7042
deepCoadd_meas	42206
deepCoadd_measMatch	42206
deepCoadd_measMatchFull	42206
deepCoadd_mergeDet	7043
deepCoadd_ngmix_deblended	7005
deepCoadd_nImage	42206
deepCoadd_ref	7043
icSrc_schema	1
src_schema	1
deepCoadd_meas_schema	1
deepCoadd_mergeDet_schema	1
deepCoadd_ngmix_deblended_schema	1
deepCoadd_peak_schema	1
deepCoadd_ref_schema	1
deepCoadd_forced_src_schema	1
deepCoadd_deblendedFlux_schema	1
deepCoadd_deblendedModel_schema	1
deepCoadd_det_schema	3
deepCoadd_mcalmax_deblended_schema	1
camera	1
bias	189
dark	189
flat	1134
cal_ref_cat_2_2	1213
packages	4

TABLE 3: Counts of each dataset type in the Butler Gen3 Registry for DP0.1.

C Historical sections moved for clarity

C.1 Elements of Data Preview 0.1

In this section we discuss the following key topics:

- Dataset choice considerations
- Data products offered
- Services offered
- Audience considerations

C.1.1 Dataset choice considerations

For DP0.1 we are moving ahead with DESC DR6 WFD as the baseline. See Appendix A for historical context on the dataset choice.

We do not have HiPS maps available for DP0.1.

C.1.2 Data Products Offered

We will offer access to images and catalogs, though in more limited ways that will be available in Operations. Images will be stored in read-only Butler Gen3 repo. Catalogs will be stored in Qserv. Source catalogs are not part of the DESC DR6.

For DP0.2 we may provide images and catalogs from different production runs based on the same dataset. For example, in the stretch goal of reprocessing the dataset in Gen 3, catalogs may not be available for Qserv to start ingesting in time. In such a scenario, we may choose to provide existing catalogs from the old run.

From DESC, DC2 catalogs can be obtained with more complete columns extracted from the original FITS files, or a modified schema to roughly match DPDD [LSE-163]. The latter is closer to the eventual data access but the former allows additional scientific analysis. We will provide both in DP0.1. (See milestones L3-MW-0010 and L3-MW-0020 for Qserv loading)

Specifically, in DP0.1, the following tables from DESC will be provided as one database `dp01_dc2_catalogs` in one TAP schema in Rubin's Qserv instance running at the IDF.

- `position`: DESC internal data as ingested in DESC's Qserv instance at IN2P3.
- `reference` (originally named `dpdd_ref` in DESC): DESC internal data as ingested in DESC's Qserv instance at IN2P3.
- `forced_photometry` (originally named `dpdd_forced` in DESC): DESC internal data as ingested in DESC's Qserv instance at IN2P3.
- `object`: DESC internal data v2 at NERSC provided to us in parquet format.
- `truth_match`: DESC internal data v2 at NERSC provided to us in parquet format.

The science data products in the Butler Gen3 repository depend on the availability in the DESC-provided Gen2 data repository and the Gen2-to-3 conversion. In addition to the raw data, the following 4 reruns were obtained from DESC's copy at IN2P3.

- `run2.2i-calexp-v1` (Gen2) `2.2i/runs/DP0.1/calexp/v1` (Gen3)
- `run2.2i-coadd-wfd-dr6-v1-u` (Gen2) `2.2i/runs/DP0.1/coadd/wfd/dr6/v1/u` (Gen3)
- `run2.2i-coadd-wfd-dr6-v1-grizy` (Gen2) `2.2i/runs/DP0.1/coadd/wfd/dr6/v1/grizy` (Gen3)
- `run2.2i-coadd-wfd-dr6-v1` (Gen2) `2.2i/runs/DP0.1/coadd/wfd/dr6/v1` (Gen3)

Data were transferred from IN2P3 to NCSA, and converted into a Gen3 repository at NCSA. The repository provided on IDF is based on the DC2 repo at NCSA on March 16. Newer data ingested into NCSA's repo will not be ported to IDF. Some dataset types, such as warps, will not be provided. See Appendix B for a full list of expected dataset types in DP0.1.

In DP0.2, the exact science data products depend on what pipelines are ready for our reprocessing.

We have ruled out offering bulk download facilities for DP0. The DESC DC2 dataset is public and can be downloaded from https://lsstdesc-portal.nersc.gov/

Questions:

- Are we offering Parquet files? — No in DP0.1; possibly in DP0.2. Currently our SDMified Parquet-generating pipelines are HSC only and Gen2 only. If Parquet files are offered the access will be via the read-only Butler Gen3 repo.

C.1.3 Services Offered

Although DP0 as a milestone described LSO-011 can be fulfilled with simple data distribution, we intend to offer limited Science Platform functionality as part of DP0. This includes:

- Provided the data is stored in Qserv or a Postgres database, catalogue access through TAP
- Access to the Science Platform's notebook-based analysis environment (Nublado); images can be accessed programmatically via the Butler.
- Catalogue access only (no VO image services) via the Portal
- Authentication via Github (new self-service Identity Management system offering Federated Authentication will be offered subsequently to DP 0.1)

Shell access (except through Nublado) will not be offered.

The science platform will be reachable as data.lsst.cloud (”data” is specified by the Product Owner, ”lsst” represents the eventual access to the Legacy Survey of Space and Time, and ”.cloud” represents the GCP-deployed IDF, allowing us to bring up the USDF in parallel under a different TLD such as data.lsst.us.

C.1.4 Audience Considerations

Care should be taken to limit the target audience for the data previews; it is most critical that this is done for DP0.

- We have limited capacity to divert resources to support users.

- We will not have performed scaling tests on the Science Platform services by that point; current Science Platform usage is under 100 users, and any intent to exceed that should be communicated well in advance
- We will not yet have the ability to throttle excessive IDF usage

Authorization will be provided in an all-in basis (users will have the same level of access as project members currently have) since finer access control mechanisms will not be available by DP0; care should be taken in selecting them.

Questions:

- What is the authorization constraints for this data? For example, are DC2 data products only available to DESC science collaboration members? If so, if DC2 is chosen, does only DESC participate in DP0? **No: When agreed, DC2 would be available to all data rights holders.**
- How do we handle access? First come first served? Do we need a sign-up process?

C.2 Completed milestones

Table 4: Milestones for Rubin Observatory Data Production and System Performance

Milestone	Jira ID	Rubin ID	Due Date	Level	Status	Team
Establish initial Key Performance Metrics, as prelude to System Optimization strategy.	PREOPS-294	FY20-0020	2020-09-30	1	Won't Fix	Systems Engineering
Develop a first model for community engagement for DP0.1	PREOPS-151	L3-CE-0020	2021-01-31	3	Done	Community Engagement
IDF DP0-Ready: Complete IDF installation and IDF staff preparations for DP0.	PREOPS-140	L2-DP-0010	2021-01-31	2	Done	Infrastructure and Support
Read only Gen3 butler for DP0 at IDF	PREOPS-143	L3-MW-0030	2021-03-31	3	Done	Science Users Middleware
Science Platform Available on IDF	PREOPS-141	L3-PR-0010	2021-03-31	3	Done	Science Platform and Reliability Engineering
Qserv installation on IDF	PREOPS-142	L3-MW-0010	2021-03-31	3	Done	Science Users Middleware
DP0.1 data loaded into Qserv on IDF	PREOPS-144	L3-MW-0020	2021-04-30	3	Done	Science Users Middleware
DP0.1 QA Access: Provide access to processed images and catalogs from the IDF	PREOPS-146	L2-DP-0020	2021-05-03	2	Done	Science Platform and Reliability Engineering
DP0.1 Data Release: science-ready catalogs released from the IDF	PREOPS-148	L2-SP-0010	2021-06-30	2	Done	Verification and Validation
Pipeline release for DP0.2	PREOPS-145	L3-AP-0010	2021-06-30	3	Done	Algorithms and Pipelines

D References

[DMTN-123], Gower, M., Lim, K.T., 2019, *Batch Production Services Design*, DMTN-123, URL http://DMTN-123.lsst.io

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

[LDM-633], Kowalik, M., Gower, M., Kooper, R., 2019, *Offline Batch Production Services Use Cases*, LDM-633, URL https://ls.st/LDM-633

[LDM-636], Kowalik, M., Gower, M., Kooper, R., 2019, *Batch Production Service Requirements*, LDM-636, URL https://ls.st/LDM-636

[DMTN-125], Lim, K.T., 2019, *Google Cloud Engagement Results*, DMTN-125, URL http://dmtn-125.lsst.io

[RTN-013], O'Mullane, W., Dubois, R., 2020, *Near term workflow for pre-operations with PanDA*, RTN-013, URL http://RTN-013.lsst.io

[LSO-011], O'Mullane, W., Marshall, P., Guy, L., 2019, *OBSOLETE - use RDO-011 Release Scenarios for LSST Data*, LSO-011, URL https://lso-011.lsst.io

[DMTN-168], Padolski, S., Ye, S., 2021, *Running Science Pipelines using PanDA*, DMTN-168, URL https://dmtn-168.lsst.io,
LSST Data Management Technical Note

E Acronyms

Acronym	Description
AGN	active galactic nuclei
AP	Alert Production
BNL	Brookhaven National Laboratory
BPS	Batch Production Service
CCB	Change Control Board

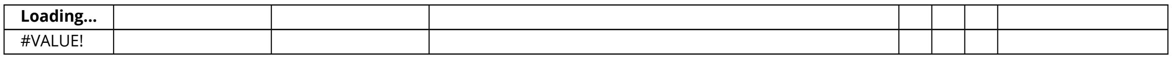
CE	Communications Engagement
CLO	community.lsst.org - use of this acronym is discouraged. The language that should be used in official documents is ”Community Forum” or ”Vera C. Rubin Community Forum”.
CWL	Common Workflow Language
ComCam	The commissioning camera is a single-raft, 9-CCD camera that will be installed in LSST during commissioning, before the final camera is ready.
DAGMan	Directed Acyclic Graph Manager
DAX	Data Access Services
DC2	Data Challenge 2 (DESC)
DESC	Dark Energy Science Collaboration
DMCCB	DM Change Control Board
DMTN	DM Technical Note
DOE	Department of Energy
DP	Data Production
DP0	Data Preview 0
DPDD	Data Product Definition Document
DR1	Data Release 1
DRP	Data Release Production
EPO	Education and Public Outreach
FITS	Flexible Image Transport System
FTE	Full-Time Equivalent
FY20	Financial Year 20
FY21	Financial Year 21
FY22	Financial Year 22
GCP	Google Cloud Platform
HSC	Hyper Suprime-Cam
IDF	Interim Data Facility
IN2P3	Institut National de Physique Nucléaire et de Physique des Particules
L2	Lens 2
L3	Lens 3
LDM	LSST Data Management (Document Handle)
LSE	LSST Systems Engineering (Document Handle)

LSST	Legacy Survey of Space and Time (formerly Large Synoptic Survey Telescope)
NCSA	National Center for Supercomputing Applications
NERSC	National Energy Research Scientific Computing Center
OPS	Operations
PCW	Project Community Workshop
PDR2	Public Data Release 2 (HSC)
PR	Pull Request
PanDA	Production ANd Distributed Analysis system
QA	Quality Assurance
RFC	Request For Comment
RTN	Rubin Technical Note
S3	(Amazon) Simple Storage Service
SC	Science Collaboration
SP	Story Point
TAP	Table Access Protocol
TLD	Top Level Domain
USDF	United States Data Facility
VO	Virtual Observatory
WFD	Wide Fast Deep

F Roles in Data Production FY21

These are the roles and individuals becoming active in FY21. More roles activate later as we approach operations.

Table 5: Team members for Data Production for Rubin Observatory FY21



Loading...							
#VALUE!							