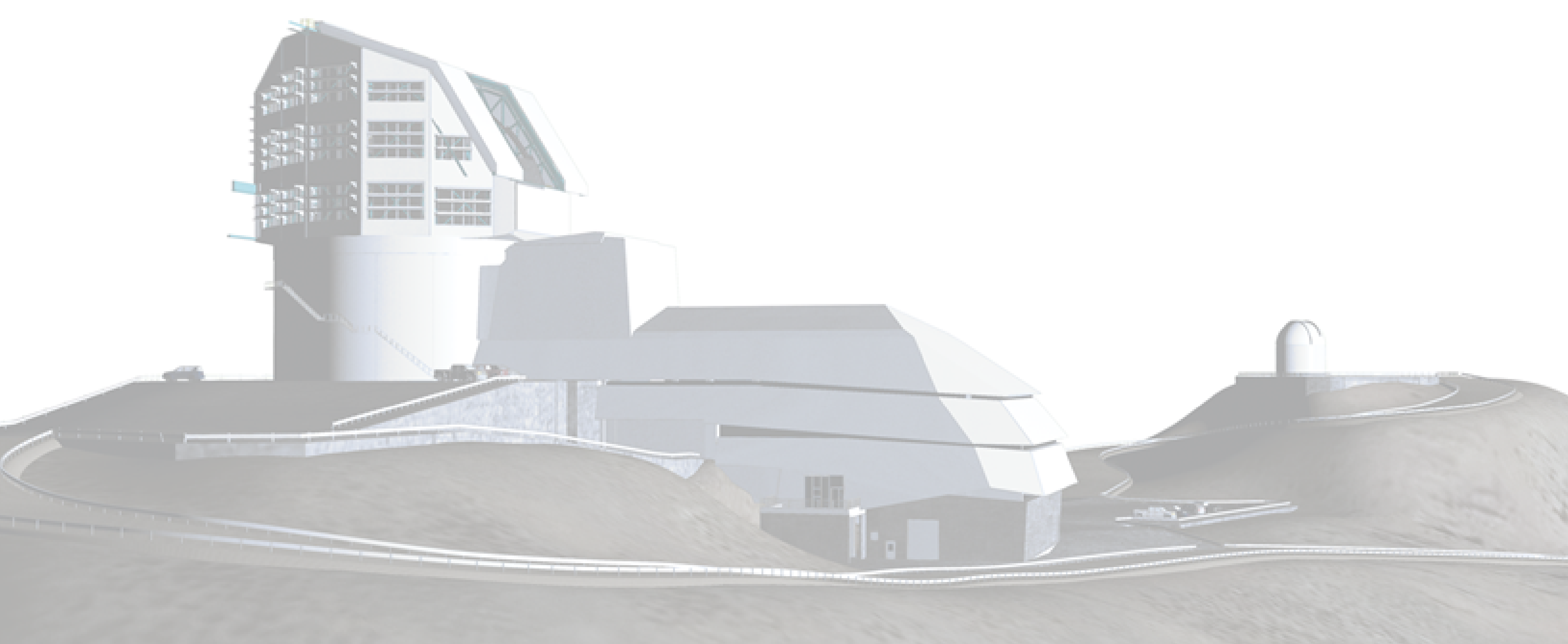
## Vera C. Rubin Observatory
## Rubin Observatory Operations

# Statement of Work for the Rubin Observatory US Data Facility

**William O'Mullane and Robert Blum**

**RTN-080**

**Latest Revision: 2020-05-20**



## Abstract

This statement of work describes the scope of work for the US Data Facility for which the Department of Energy (DOE) will issue a Funding Opportunity Announcement (FOA) in 2020 to support Rubin Observatory Operations. Hardware and staffing levels needed to successfully support the US Data Facility within the Rubin Observatory Operations organization are described in detail.

# Change Record

| Version | Date | Description | Owner name |
|---|---|---|---|
| 0.8 | 2024-04-08 | Put on lsst.io as RTN-080, no changes | WOM |
| 0.7 | 2020-05-20 | Draft for Agencies | WOM, RDB |
| 0.6 | 2020-05-13 | Further edits | RDB |
| 0.5 | 2020-05-12 | Appendix B, subsystem requirements, added | WOM |
| 0.4 | 2020-05-10 | Many Edits, new figures (1, 3, 4) | RDB |
| 0.3 | 2020-05-05 | Break to multiple files add WP0 | WOM |
| 0.2 | 2020-05-03 | Minor Edits | RDB |
| 0.1 | 2020-04-23 | First draft | WOM |

*Document source location:* `https://github.com/lsst/rtn-080`

# Contents

# Statement of Work for the Rubin Observatory US Data Facility

## 1 Introduction

In this statement of work (SOW) we detail the needs of the Vera C. Rubin Observatory United States Data Facility (USDF). The USDF is the main data processing, archive, and access center for Rubin Observatory. The USDF is integrated within the Data Production Department of Rubin Observatory Operations.

DOE will select a USDF awardee through a Funding Opportunity Announcement FOA. This will be a open process with independent review. Proposals to run the USDF for Rubin will respond to the scope of work and requirements detailed in this document and associated references.

It is expected that the selected organization resulting from the FOA process will provide all the North American computing for the Vera C. Rubin Observatory Legacy Survey of Space and Time (formerly Large Synoptic Survey Telescope) (LSST).

The FOA process will take some time to complete, so an interim data facility will be put in place to enable continued progress in pre-operations planning and activity related to the Rubin operations data facilities. Some transition between this interim facility and the USDF will be required during the remaining period before construction completes, a period that now runs in parallel with Rubin pre-operations.

This document summarizes important considerations for the USDF in the context of the FOA.

In particular, it

- describes how the Data Facility is integrated within the Rubin Operations structure (including its integration with the Data Production Department) and with the scientific community (subsection 3.1);
- details key requirements and constraints on the Data Facility (section 2);
- discusses studies which have been carried out to date on executing LSST processing in cloud environments (section 5);

- specifies other data centers that the USDF must be able to work with (2.3);
- specifies a start up ramp needed before start of operations subsection 2.1.

## 1.1 Context: Rubin Observatory Data Management and the US Data Facility

The Data Management Subsystem (DMS) will be used to receive, process, and serve to the community data collected by Rubin over the course of system commissioning, pre-operations, its ten-year mission, and final processing after data collection is complete (roughly now until 2035). It combines a range of both hardware and software, including — for example — long-haul networks; systems for ingesting data from the telescope; compute clusters for processing that data; scientific pipelines and algorithms; and databases and interfaces which will be used to publish the resulting data products to the scientific community. The requirements on the DMS are enumerated in LSE-61; DMS architecture is described in LDM-148; the data products which DMS will produce and distribute are detailed in LSE-163.

The DMS is being developed by the Rubin Observatory Construction project's Data Management (DM) team. The DM team consists of around 100 individuals, organized into functional teams that align broadly with their institutional affiliation; details of its aims, organization, and management are presented in LDM-294. The Rubin Data Facility team *in construction* is one of these constituent teams within DM, based at National Center for Supercomputing Applications (NCSA), and charged with:

- developing the middleware systems which will collect data from Camera and Observatory systems, archive them, and make them available for processing;
- developing the systems which will execute and manage scientific data processing during the operational era;
- supporting the activities of other teams within the DM Subsystem, and across the Rubin Observatory Construction project, by providing them with compute facilities, data storage, etc., as required to build and commission the Rubin Observatory system.

The overall design of the system which is currently under construction by the Data Facility team is described in LDM-129.

During *operations*, the Data Facility forms one of the key elements within the Data Production Department. It will operate and maintain the systems which were developed during construction to produce and provide to the community Rubin Observatory scientific data products.

It is clear from the above that the Data Facility, in both construction and operations, is responsible for providing both hardware resources and software development (and maintenance). Although one of the design principles of the DMS is that, where appropriate, DM elements should be portable between facilities (a topic to which we will return in section 5), we are mindful that moving the facility may have an impact on the ongoing development effort in construction. We are considering not just compute resources, which may be easily sourced from elsewhere, but also expertise, experience, and ongoing software development effort.

## 1.2 Scope

This document describes the scope of activity to be executed at the USDF by the awardee. LDM-129 outlines the design era services foreseen for the data facility. This should be taken as informative rather than prescriptive. LDM-148 describes the full Data Management Design and may be useful to understand the systems which will be deployed in the data facility.

## 1.3 Motivation

It is assumed this document will form part of the FOA document package.

## 1.4 Risk Factors

Hardware pricing and how it is modeled over ten years is difficult to predict and assumptions on out year costs and replacement frequency vary between institutions based on their experience and scale of operation. There is thus a significant risk of underestimating the cost of compute and storage.

Rubin Observatory staff are working with DOE and NSF to understand security concerns with the data path for Rubin between Cerro Pachón in Chile and the USDF as well as data content security concerns. At this point, no significant concerns (with potential significant associated cost) have been identified. The FOA will specify any security requirements placed on the USDF.

## 1.5 Rubin Observatory Documents


[1] **[LDM-141]**, Becla, J., Lim, K.T., 2013, Data Management Storage Sizing and I/O Model, URL `https//ls.st/LDM-141`,
Vera C. Rubin Observatory LDM-141

[2] **[DMTN-020]**, Becla, J., Economou, F., Gelman, M., et al., 2018, Data Management Project Management Guide, URL `https://dmtn-020.lsst.io/`,
Vera C. Rubin Observatory Data Management Technical Note DMTN-020

[3] **[LDM-612]**, Bellm, E., Blum, R., Graham, M., et al., 2020, Plans and Policies for LSST Alert Distribution, URL `https://ldm-612.lsst.io/`,
Vera C. Rubin Observatory Data Management Controlled Document LDM-612

[4] **[LDO-31-OBS-RDO-018]**, Blum, R., et al., 2020, OBSOLETE NOW RDO-018 - LSST Operations Proposal , URL `https//ls.st/LDO-31-OBS-RDO-018`,
Vera C. Rubin Observatory LDO-31-OBS-RDO-018

[5] **[DMTN-104]**, DMLT, 2020, Data Management Detailed Product Tree, URL `https://dmtn-104.lsst.io/`,
Vera C. Rubin Observatory Data Management Technical Note DMTN-104

[6] **[LSE-81]**, Dubois-Felsmann, G., 2013, LSST Science and Project Sizing Inputs, URL `https//ls.st/LSE-81`,
Vera C. Rubin Observatory LSE-81

[7] **[DMTN-136]**, Dubois-Felsmann, G., 2021, LSST Science Platform Portal Aspect Design and Maintenance Manual, URL `https://dmtn-136.lsst.io/`,
Vera C. Rubin Observatory Data Management Technical Note DMTN-136

[8] **[LSE-61]**, Dubois-Felsmann, G., Jenness, T., 2019, Data Management System (DMS) Requirements, URL `https://lse-61.lsst.io/`,
Vera C. Rubin Observatory LSE-61

[9] **[DMTN-123]**, Gower, M., Lim, K.T., 2019, Batch Production Services Design, URL `https://dmtn-123.lsst.io/`,
Vera C. Rubin Observatory Data Management Technical Note DMTN-123

[10] **[DMTN-087]**, Juric, M., Eggl, S., Moeyens, J., Jones, L., 2020, Proposed Modifications to Solar System Processing and Data Products, URL https://dmtn-087.lsst.io/, Vera C. Rubin Observatory Data Management Technical Note DMTN-087

[11] **[LSE-163]**, Jurić, M., Axelrod, T., Becker, A., et al., 2023, Data Products Definition Document, URL https://lse-163.lsst.io/, Vera C. Rubin Observatory LSE-163

[12] **[LSE-78]**, Lambert, R., Kantor, J., Huffer, M., et al., 2017, LSST Observatory Network Design, URL https//ls.st/LSE-78, Vera C. Rubin Observatory LSE-78

[13] **[DMTN-125]**, Lim, K.T., 2019, Google Cloud Engagement Results, URL https://dmtn-125.lsst.io/, Vera C. Rubin Observatory Data Management Technical Note DMTN-125

[14] **[LDM-152]**, Lim, K.T., Dubois-Felsmann, G., Johnson, M., Juric, M., Petravick, D., 2019, Data Management Middleware Design, URL https://ldm-152.lsst.io/, Vera C. Rubin Observatory Data Management Controlled Document LDM-152

[15] **[DMTN-114]**, Lim, K.T., Guy, L., Chiang, H.F., 2019, LSST + Amazon Web Services Proof of Concept, URL https://dmtn-114.lsst.io/, Vera C. Rubin Observatory Data Management Technical Note DMTN-114

[16] **[LDM-148]**, Lim, K.T., Bosch, J., Dubois-Felsmann, G., et al., 2020, Data Management System Design, URL https://ldm-148.lsst.io/, Vera C. Rubin Observatory Data Management Controlled Document LDM-148

[17] **[Agreement-51]**, LSST, 2015, Memorandum of Agreement regarding collaboration in the scientific exploitation of data acquired with LSST by specified Principal Investigators and scientists at IN2P3, URL https//ls.st/Agreement-51, Vera C. Rubin Observatory Agreement-51

[18] **[LDM-156]**, Myers, J., Jones, L., Axelrod, T., 2013, Moving Object Pipeline System Design, URL https//ls.st/LDM-156, Vera C. Rubin Observatory LDM-156

[19] **[LDM-294]**, O'Mullane, W., Swinbank, J., Juric, M., Guy, L., DMLT, 2023, Data Management Organization and Management, URL https://ldm-294.lsst.io/, Vera C. Rubin Observatory Data Management Controlled Document LDM-294

[20] **[LDM-129]**, Petravick, D., Johnson, M., Butler, M., 2018, LSST Data Facility Logical Information Technology and Communications Design, URL `https//ls.st/LDM-129`,
Vera C. Rubin Observatory LDM-129

[21] **[DMTN-018]**, Salnikov, A., 2016, Re-visiting L1 Database Design, URL `https://dmtn-018.lsst.io/`,
Vera C. Rubin Observatory Data Management Technical Note DMTN-018

[22] **[DMTN-113]**, Salnikov, A., 2019, Performance of RDBMS-based PPDB implementation, URL `https://dmtn-113.lsst.io/`,
Vera C. Rubin Observatory Data Management Technical Note DMTN-113

[23] **[SQR-006]**, Sick, J., 2016, The LSST the Docs Platform for Continuous Documentation Delivery, URL `https://sqr-006.lsst.io/`,
Vera C. Rubin Observatory SQuaRE Technical Note SQR-006

[24] **[LDM-151]**, Swinbank, J., Axelrod, T., Becker, A., et al., 2020, Data Management Science Pipelines Design, URL `https://ldm-151.lsst.io/`,
Vera C. Rubin Observatory Data Management Controlled Document LDM-151

[25] **[DMTN-091]**, Wood-Vasey, M., Bellm, E., Bosch, J., et al., 2024, Test Datasets for Scientific Performance Monitoring, URL `https://dmtn-091.lsst.io/`,
Vera C. Rubin Observatory Data Management Technical Note DMTN-091

# 2 Work to be performed

## 2.1 WP-00: USDF Startup

The USDF is expected to start operations in late FY23 or early FY24. There will need to be overlap with construction and interim operations facilities in FY23 and possibly sooner to enable a smooth and successful transition to the USDF.

Developer support is one aspect of this transition – there need to be several machines similar to the target machines for checking code and testing. Such transition activities will involve close conversations between the USDF awardee and Rubin staff; however, a first estimate would be to provide similar machines to those defined in `https://developer.lsst.io/services/lsst-dev.html`.

In the following, requirements are called out in **bold face** with integer indices for each type of requirement. The first set listed below are precursor requirements that establish the minimum capability of the USDF before operations of the full LSST begins.

**INIT-020** At least one year before LSST Operations Year 1 (LOY1) the USDF shall have developer support setup. This should include services similar to the LDF services listed on `https://developer.lsst.io/services`.

**INIT-040** At least one year before LOY1 the USDF shall have storage resources available for development and prototypical Data Access services. At this time, we estimate that 5 PB of object, 9PB of normal and 16 PB of tape storage shall be available (See also Table 1.)

Table 1: This table outlines the estimated needs pre LOY1

| Year | 2022 |
|---|---|
| Instantaneous cores (DRP) Total | 4,673 |
| Instantaneous cores (Alerts) | 1188 |
| Qserv nodes (US DAC/ Staff) | 14 |
| **Total owned nodes** | **567** |
| Fast Storage (TB) | 50 |
| Normal Storage (TB) | 9241 |
| Latent Storage (TB) | 4966 |
| High Latency (TB) | 16733 |

For Table 1 the storage descriptions are given in Table 3 and the machines are described in Table 2.

Table 2: Machine types used in sizing model.

| Type of machine | Cores | Memory(GB) | Eff cores/ node |
|---|---|---|---|
| Xeon | 32 | 192 | 27 |
| Qserv | 12 | 128 | 12 |
| current compute node | 24 | 128 | 24 |

The available memory per core should average at least 6 GB, with 10 GB for Qserv cores. The actual number of cores and thus core-hours may need to be adjusted based on performance of the selected node type.

Table 3: Storage types used in sizing model.

| Storage type |
|---|
| fast – NVMe (50GB/ s each) / TB |
| normal - SATA GPFS file systems/ TB |
| latency – slower but on disk |
| high latency – very slow – on tape |

## 2.2 Networking

The distributed nature of the LSST facilities necessitates a complex network design, which is described in LSE-78. This includes local networking at the Summit Facility, at the Base Site, and at the Data Facility, as well as long-haul networks connecting Chile and North America.

**INIT-060** At least six months before LOY1 the USDF shall have arranged 100Gbs, path redundant, network capacity to Energy Sciences Network (ESNet) or equivalent to connect to the Rubin Observatory facility in Chile.

**INIT-080** The USDF shall ensure the capacity to IN2P3 is not impeded on the US side. IN2P3 are responsible for the transatlantic transfer of data.

**INIT-100** The USDF shall ensure low latency on the links to Chile to enable alert processing within 60s. The latency due to the network shall be 3s or better - the USDF must ensure their contribution to latency is within this envelope.

**INIT-120** The facility shall reserve an allocation of 10 Gb/s for alert stream transfer from the USDF to community brokers; see LDM-612.

**INIT-140** The USDF shall provide high-performance (100 Gb/s) network paths between cluster compute nodes. We expect that detailed tuning of the cluster networking system may be necessary for optimal performance, and would require support for this from the Data Facility host.

**INIT-160** The USDF shall provide specific network overlays, such as Weave, if required for the Kubernetes cluster (see Req. INFR-280).

**INIT-180** The USDF shall allow Rubin to partition the infrastructure network from the Internet so that internal services and administrative interfaces are not generally accessible.

**INIT-200** The USDF shall enable Rubin to define Internet ingresses, including a mechanism for engineers to access the internal network and administrative interfaces, via a bastion host, VPN, or other similar mechanism.

## 2.3 WP-01: USDF Infrastructure

The main deliverable is a computing infrastructure capable of supporting the Rubin Observatory Data Production operations. The organizational structure of Rubin Observatory is presented in detail in the Operations Plan document [LDO-31-OBS-RDO-018].

Table 4 gives the estimated high level storage needs for operations. Table 5 gives the estimated high level compute needs for ops. More details are provided in Appendix A and a graphical representation of storage needs is given in Figure 1.

These are of course estimates, algorithms may be less (or more) efficient and sizes may prove to be slightly off. The USDF should monitor the actual situation with Rubin Observatory and decide how to address any situation which arises including raising it to the funding agencies if appropriate. Furthermore, the technical requirements in these documents capture key platforms and services that will be delivered by the Construction Project to Operations. They are not an exhaustive list, and as big data science is a rapidly evolving field, may evolve or be superseded during Operations.

There is an existing agreement with IN2P3 [Agreement-51] to provision and execute 50% of the total Data Release Production; the USDF awardee will need to work with IN2P3 to enable this production. It is incumbent upon the USDFto develop and deploy systems for effectively managing split-site data processing.

Another facility may yet provide a further 25% of processing power which would also have to be integrated in the model. If that happens, then the corresponding scope would be removed from the USDF. For purposes of the FOA, it can be assumed the USDF will do 50% of the Data Release Production.

The price of disk and tape media purchases over time have a profound effect on the USDF budget over 10 years. The FOA proposal should explicitly state the assumed pricing factors over the life of the survey for annual hardware purchases. It is assumed that the cost of media will decrease over time, and the annual decrements should be called out in the proposal (net of annual consumer inflation).

Table 4: On floor LDF storage estimates during Operations

| LDF Storage (on the floor) | unit | LOY1 | LOY2 | LOY3 | LOY4 | LOY5 | LOY6 | LOY7 | LOY8 | LOY9 | LOY10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| APDB | TB | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Qserv Czar/ Object | TB | 182 | 347 | 562 | 643 | 711 | 774 | 835 | 894 | 951 | 1006 |
| **Total Fast** | **TB** | **206** | **371** | **586** | **667** | **735** | **798** | **859** | **918** | **974** | **1029** |
| Normal | TB | 38983 | 67976 | 99538 | 131025 | 162321 | 193727 | 225288 | 256991 | 288788 | 320731 |
| Qserv Storage | TB | 4094 | 9257 | 17275 | 24144 | 31277 | 38734 | 46555 | 54716 | 63206 | 72017 |
| LSSTCam Raw Images | TB | 6982 | 11798 | 16614 | 21430 | 26246 | 31062 | 35878 | 40694 | 45510 | 50326 |
| LSSTCam Output Images | TB | 3933 | 10676 | 16857 | 23599 | 30342 | 37084 | 43827 | 50570 | 57312 | 64055 |
| LSSTCam Output Coadd Images | TB | 8636 | 16364 | 23182 | 23182 | 23182 | 23182 | 23182 | 23182 | 23182 | 23182 |
| LSSTCam Output Parquet | TB | 9302 | 25248 | 47839 | 71758 | 95678 | 119597 | 143516 | 167436 | 191355 | 215275 |
| Object Store | TB | 28854 | 64086 | 104491 | 139969 | 175447 | 210925 | 246403 | 281881 | 317359 | 352837 |
| LSSTCam Raw Images | TB | 6982 | 11798 | 16614 | 21430 | 26246 | 31062 | 35878 | 40694 | 45510 | 50326 |
| All Data Products/ Backup | TB | 47626 | 106967 | 194226 | 309242 | 451681 | 621564 | 818915 | 1043758 | 1296109 | 1575992 |
| All Object Store-Only Products | TB | 8636 | 16364 | 24091 | 31818 | 39545 | 47273 | 55000 | 62727 | 70455 | 78182 |
| Tape | TB | 63245 | 135129 | 234931 | 362490 | 517473 | 699899 | 909793 | 1147179 | 1412074 | 1704500 |

Table 5: Compute needs during Operations

| **Data Release Production** | **units** | **LOY1** | **LOY2** | **LOY3** | **LOY4** | **LOY5** | **LOY6** | **LOY7** | **LOY8** | **LOY9** | **LOY10** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| LSSTCam visit input size | TB | 1911 | 3822 | 5733 | 7644 | 9556 | 11467 | 13378 | 15289 | 17200 | 19111 |
| DRP compute | core-hours | 4.5E+07 | 8.2E+07 | 1.2E+08 | 1.6E+08 | 2.0E+08 | 2.5E+08 | 2.9E+08 | 3.3E+08 | 3.7E+08 | 4.1E+08 |
| Alert Production | units | LOY1 | LOY2 | LOY3 | LOY4 | LOY5 | LOY6 | LOY7 | LOY8 | LOY9 | LOY10 |
| AP cores | cores | 1,188 | 1,188 | 1,188 | 1,188 | 1,188 | 1,188 | 1,188 | 1,188 | 1,188 | 1,188 |
| US DAC | units | LOY1 | LOY2 | LOY3 | LOY4 | LOY5 | LOY6 | LOY7 | LOY8 | LOY9 | LOY10 |
| LSP cores | cores | 517 | 933 | 1,399 | 1,866 | 2,332 | 2,798 | 3,265 | 3,731 | 4,198 | 4,664 |
| Qserv data per node | TB/ node | 43 | 43 | 86 | 86 | 86 | 86 | 173 | 173 | 173 | 173 |
| Qserv nodes | nodes | 95 | 216 | 309 | 348 | 364 | 451 | 436 | 408 | 367 | 418 |
| LSP cores/ science user | cores/ user | 0.1 | 0.2 | 0.2 | 0.2 | 0.3 | 0.4 | 0.4 | 0.5 | 0.6 | 0.6 |
| Chilean DAC | units | LOY1 | LOY2 | LOY3 | LOY4 | LOY5 | LOY6 | LOY7 | LOY8 | LOY9 | LOY10 |
| LSP cores | cores | 103 | 187 | 280 | 373 | 466 | 560 | 653 | 746 | 840 | 933 |
| Qserv nodes | nodes | 95 | 216 | 309 | 348 | 364 | 451 | 436 | 408 | 367 | 418 |
| Staff LSP | units | LOY1 | LOY2 | LOY3 | LOY4 | LOY5 | LOY6 | LOY7 | LOY8 | LOY9 | LOY10 |
| LSP cores | cores | 52 | 93 | 140 | 187 | 233 | 280 | 326 | 373 | 420 | 466 |

**INFR-020** The cost model for work package 2.3 shall include the nonlabor costs to purchase computing hardware meeting the operation needs of the Rubin Observatory US Data Facility as summarized in Table 4, Table 5, and Figure 2.

**INFR-040** The cost model for work package 2.3 shall include the labor costs to manage and operate the computing for the Rubin Observatory US Data Facility.

The infrastructure team will oversee and manage the Data Facilities' performance and strategy. The USDF infrastructure team will instantiate and operate a combination of hardware and software as services, including problem management, incident management, request response, and installing and validating changes in conjunction with the IT change control process. This includes maintenance of configuration information at the service level, e.g., an application map showing the reliance of the service on all software and ITC, being aware of security configurations and other operational matters, and handling both network security and authorization infrastructure, and operational security associated with network-based security.

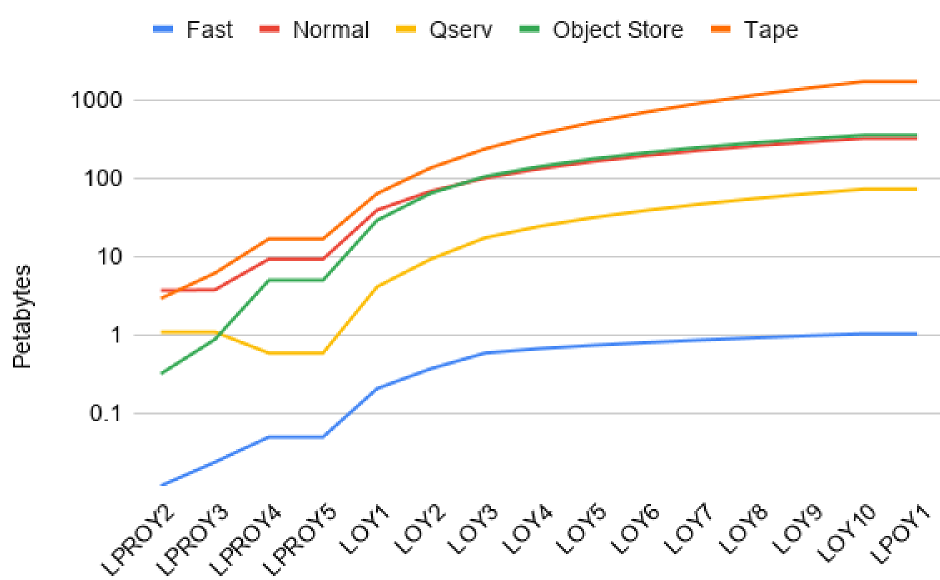



Figure 1: Evolution of **cumulative** storage needs for Rubin Observatory. Details are given in the appendix. This plot shows the total storage needed at the USDF of different types. Storage at other data facilities, including the Chilean data access center, are not included here. The USDF is responsible for the last two years of pre operations (LPROY4-5) which are transition to USDF, the survey years (LOY1-10), and the post operations years (only the first year of two is shown).

**INFR-060** The awardee shall integrate staff in the Rubin infrastructure team from Institut National de Physique Nucléaire et de Physique des Particules (IN2P3). It is assumed that any additional data facility involved in Data Release Production (DRP) will also integrate staff into this team.

**INFR-080** The USDF shall have a high availability architecture including: automated monitoring of services, including alerts to Rubin engineers about degraded service, redundancy of service, the capability to upgrade infrastructure in a rolling fashion to minimize outages, and (preferably) automated ejection of mis-behaving infrastructure elements from the resource pool.

**INFR-100** The facility shall have no more than a total of three days of unplanned downtime per year.

**INFR-120** The facility shall schedule planned downtimes only after consultation with Rubin and shall provide a service level agreement for the facility with ticket turnaround times, etc.

**INFR-140** The facility shall provide access to logs or a logging service (e.g., Kibana or Splunk) to any infrastructure services that may be affecting Rubin systems. This potentially means access to Kubernetes or storage host logs.

## 2.4 Security

The provider shall be responsible for the security of the infrastructure and keep that infrastructure patched and configured according to security best practices, including regular security testing and remediation of any high-severity findings.

Rubin Observatory expects to have thousands of users in America and beyond. Facility security polices should not prevent direct internet access to our public-facing services from registered and unregistered (as appropriate) users (e.g., by mandating VPN-only access see Req. INFR-240).

Rubin already has a data base of users and uses CI-Login for federated Authentication, facility security policies should allow Rubin to or someone we contracted with, to manage user authentication and to our services. (See also Req. INFR-220.

Rubin requires the right to reject any security policy that would require us to permit decryption of TLS traffic to the infrastructure provider or third parties e.g., to network filtering appliances.

Facility security policies must allow authorized administrators from Rubin Observatory to investigate errors and debug technical issues on kubernetes nodes or other service hosts. (See also Req. INFR-200 Req. INFR-140.)

**INFR-160** Rubin has many SSL certificates and has many domains registered facilities should allow Rubin to continue managing SSL certificates, using our own domain registrar and deploying public-facing services under our own domains (e.g., lsst.io) and with an external DNS service (e.g., Amazon's Route53)

**INFR-180** The provider shall be responsible for the security of the infrastructure and keep that infrastructure patched and configured according to security best practices, including regular security testing and remediation of any high-severity findings.

**INFR-200** Administrative access to the infrastructure shall require two-factor authentication.

## 2.5 Science Platform

**INFR-220** The facility shall support federated login for the LSST Science Platform (now Rubin Science Platform) (LSP) such as Continuous Integration (CI)-Logon.

**INFR-240** The facility shall support access to the LSP services from unrestricted IP address - i.e. not requiring a Virtual Private Network (VPN).
Some services may not require authentication.

**INFR-260** The facility shall support and demonstrate knowledge of the data production deployment mechanisms e.g. Helm, ArgoCD, Kubernetes and Puppet.

Any facility should consider that the Science Platform is a continuously deployed system that exposes shell access and ad-hoc capabilities to users and the data center resources as a platform to developers. So it is a poor fit to more "buttoned-down" data center models. Rolling upgrades should also be standard to allow for less downtime. There is significant redundancy in the system to allow for this.

### 2.5.1 Kubernetes

Specifically concerning K8S there are several requirements to be considered.

**INFR-280** The facility shall provide Managed Kubernetes, including all necessary administrative access to create/destroy/administer clusters and debug pod and storage problems, no more than one minor version behind current (e.g. if current is 1.18, 1.17 is required). See for example DMTN-136

**INFR-300** The facility shall provide self serve tools for machine and cluster management. e.g. K8S admin.

**INFR-320** USDF self serve tools shall include command-line access to any managed services through Unix/Linux systems. Command-line access (via an API, for example) to be available to engineers in addition to any web-console access.

**INFR-340** The facility shall provide ability for Rubin engineers to solely or jointly manage ingress services to the Kubernetes cluster(s)

**INFR-360** The facility shall provide the ability for Rubin services to utilize Kubernetes Dynamic Volume Provisioning.

**INFR-380** The facility shall enable storage to be exposed as a POSIX filesystem to our services permitting exclusive file locks as well as lock reservations (e.g. NFSv4 or NFS v.3 with the ability to specify/configure lock daemon behavior).

**INFR-400** The facility shall allow Rubin services to control UID/GID of users in the POSIX filesystem (see Req. INFR-380).

**INFR-420** The facility shall allow select services pods (not users) to access storage with escalated privileges

See also subsection 2.2.

## 2.6 Alerts

As a part of regular operations, the project will scan all images taken by the LSST Camera for transient and variable sources, announcing results to the scientific community within 60 s of the data being taken through an alert stream provided to a set of preselected community alert brokers. LDM-612 provides an overview of the Rubin Observatory alert system.

The external "community brokers" will receive the full alert stream generated by Rubin Observatory, and bear the primary responsibility of redistributing relevant alerts to science users. Rubin will generate up to 10 million alerts per night, with the average alert packet size being 82 KiloByte (KB) (see Req. INIT-120).

The current system architecture locates all scientific processing pipelines, including those used to identify transients and variables (the "Alert Generation Pipeline"), at the USDF. Each exposure corresponds to around 8.2 GB of raw data, which must be shipped with extremely low latency over the dedicated Long Haul Network (LHN) to the USDF to enable the Alert Generation Pipeline to execute and deliver the results to the community alert brokers within the allocated time window (see Req. INIT-100).

Increased bandwidth allocation from the Data Facility would provide an opportunity to increase the number of community brokers supported.

**INFR-440** The USDF shall host an specific dedicated cluster for prompt processing as defined in LDM-151. See also subsection B.1.

One could discuss this as a service level rather than dedicated resources.

**INFR-460** The USDF shall support Kafka[1] for alert distribution.

## 2.7 Solar system object processing

During the 24 hours following the completion of an observing night, the Solar System Processing Pipeline will be executed to carry out real-time identification of objects within our solar system. This procedure relies on knowledge of all previously detected solar system ob-

[1] https://kafka.apache.org/

jects. The Minor Planet Center (MPC) maintains such a catalog; Rubin will ingest the latest version of this catalog every evening and orbits will be computed using all available data, not only Rubin observations. Resultant new identifications and associations will be transmitted publicly in the event alert stream; candidate discoveries are sent to the MPC for inclusion in the next night's catalog.

**INFR-480** The facility shall interface with the MPC both to ingest updated catalogs, and to transmit Rubin Observatory candidate discoveries, as part of the regular daily operations cycle.

Full details of the Solar System Processing Pipeline can be found in LDM-156 and DMTN-087.

## 2.8 Batch Computing

The computing model in Table 5 assumes two major processes along side LSP usage: the alerts (see Req. INFR-440 and DRP.

**INFR-500** The USDF shall provide a batch processing system that integrates with the Rubin Observatory Middleware to run the release processing.

DMTN-123 describes the batch processing in detail - the baseline for this will be Condor and Postgres (for the registry). The middleware is ready to use an Object Store back end which will greatly facilitate job distribution (see also section 5).

**INFR-520** The USDF shall reserve about 10% of processing for user jobs - this may be provided by the same batch system in place for Req. INFR-500. Or they could use a simpler system.

**INFR-540** The USDF shall provide monitoring tools for the batch processing to allow tracing of problems, restarting of jobs, etc.

## 2.9 Databases

The USDF will need to host several databases (see also DMTN-104:

- Managed Consolidated Database - general-purpose relational database management that supports other services. It includes metadata and provenance, but it does not include the large catalog science data products that are generated as files and loaded into the Qserv parallel distributed database. This should be Postgres.
- LSP Database - mainly for user data and meta information about the system. This is currently Postgres
- Engineering and Facility Database (EFD) (Cache) - Engineering data is stored in an Influx database. A copy may need to be hosted near the Science Platform and some high level (averaged) values may need to be stored in a a relational system such as the consolidated database.
- Alert Production DataBase (APDB) - Performance critical internal database used to support Alert Production; will not support end- user queries. See also DMTN-018.

## 2.10 Bulk Download

Rubin Observatory USDF will support transfer of the full dataset, or large subsets, to data centers in the US and other countries, subject to future agreements. The USDF may also support bulk downloads to scientific collaborations which wish to perform additional systematic processing (e.g., shift-and-stack image analysis to search for outer Solar System objects).

**INFR-560** The USDF shall provide a bulk download service, with concomitant implications for external bandwidth from Rubin storage to the public and research Internets, as well as for the provision of a storage management layer that facilitates reliable incremental export (e.g., A scientific data management system developed at European Organization for Nuclear Research (CERN); `https://rucio.cern.ch` (Rucio)).

Some or all costs for bulk downloads may be borne by the users.

## 2.11 Other Services

There are other services outlined in construction to run at the USDF.

**INFR-580** Prospective Awardees shall enumerate the construction requirements listed in Appendix B derived from DMTN-104 and ensure they are covered. LDM-129 provides a set of services covering these which could be considered (or alternatives suggested). The other services are shown in Figure 2 and enumerated in DMTN-104.

The important point here is to cover the requirements not necessarily the services list.

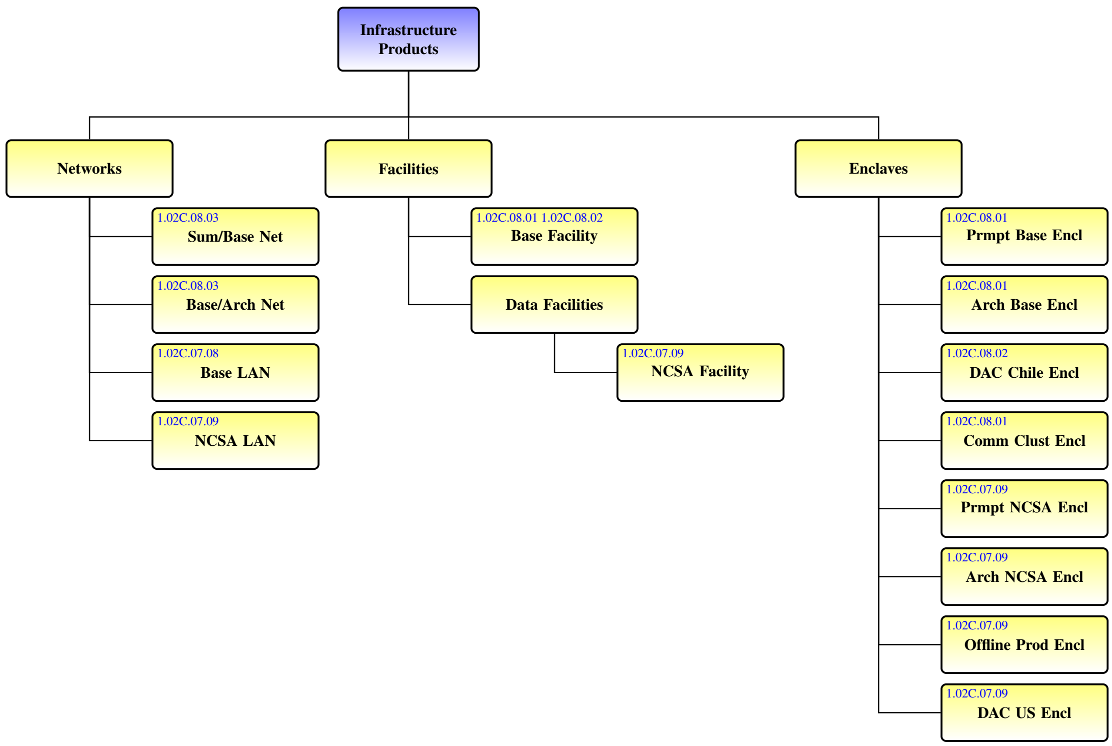



Figure 2: Subset of the product tree from DMTN-104 pertaining to Infrastructure

### 2.11.1 Data transfer and preservation

Within the services mentioned in Req. INFR-580 there are a set of services concerning data transfer, preservation and tracking. These bear particular scrutiny.

**INFR-600** Prospective awardees shall transfer data from Chile. This data shall be archived, tracked and made available to Rubin processing systems. See specifically **??**

Butler currently keeps track of files in Postgres and can be used with filesystems or S3. S3 seems more scalable long term. See also Req. MDLW-080

**INFR-620** Image data shall be ingested into the Rubin butler. Files shall be accessible via an S3 compliant object store interface.

**INFR-640** The facility should provide a Postgres like database service for the Butler metadata. This service shall allow Rubin to select extensions like PGSphere. It shall by performant for databases up to 100 TB.

**INFR-660** Any data services provided (filesystems, object stores, databases) shall be regularly backed up for disaster recovery, unless otherwise specified (e.g., /scratch).

These services are currently known as Data backbone (DBB) - these could be installed/ported.

## 2.12 WP-02: Rubin middleware

The Rubin middleware team acts as a center of general software development expertise to the Data Production department. Specific duties are:

- Maintain and evolve middleware which provides hardware and I/O abstractions for us in science and user codes.
- Maintain and evolve the Qserv distributed database.
- Coordinate software build and release activities across the Data Production department.
- Conduct investigations into underlying technological and/or infrastructural changes (for example, potential migration of some or all services to commodity cloud infrastructure).

**MDLW-020** The facility shall support the middleware team preferably by integrating staff in the team which is under direct Rubin management.

**MDLW-040** Within the middleware team the facility shall assist with integration of the Gen3 butler and other middleware.

**MDLW-060** Within the middleware team the facility shall assist with integration of the task framework with local processing framework such as HTCondor. See also [LDM-152] and [DMTN-123]. As noted in subsection 2.3 processing will also take place in IN2P3.
Workflow is an area potentially requiring significant some work to deploy in a new operations USDF. This could be outsourced from the USDF - other groups are expert in workflow and could be useful.

**MDLW-080** The butler requires a meta data database for which a Postgres installation is required. See also subsection 2.9.

### 2.13 WP-03: Rubin Execution

The Rubin execution team requires some level of support at the data facility. The Execution team will run the pipelines which generate prompt and data release products for the community, as well as calibration, environmental, quality and metadata products for the data production and system performance departments. There may be a need to run services to satisfy specific use cases as yet unidentified. Execution staff at the USDF will integrate reusable services, data layer, software, services provided by MoU, and Information Technology Center (ITC) to produce functioning services.

**EXEC-020** The awardee shall integrate staff in the Rubin execution team which already has members from SLAC National Accelerator Laboratory (SLAC) and IN2P3. It is assumed any additional data facility involved in DRP will also integrate staff in this team.

## 3 Project Management

The Rubin Operations plan details the operational organization. An abbreviated description of the Rubin System is given in the document "Vera C. Rubin Observatory System and Organization Description" attached to the FOA.

The USDF is part of the Data Production department. The department is built on teams as depicted in Figure 3. Most USDF staff will be in the Infrastructure team, with several in middleware and execution. These teams will have leads that report to the Associate Director for Data Production. The USDF Infrastructure lead will be a staff member of the USDF. The other DP team leads are based at SLAC or NOIRLab.

It is essential that the Data Facility, whatever its host institution, integrates with this Operations structure; see the attached description for more details. In particular, we emphasize that although Rubin staff are spread across multiple groups and multiple institutions, they are expected to collaborate as members of a single functional organization, working together across institutional boundaries to achieve the best outcomes for the project. This organization can be described as a mini-matrix in Data Production. See the attached description document, section 4, Figure 8. The USDF will have an administrative structure of its own with a point of contact that will provide advice and input to the AD for Data Production along with similar individuals from other data centers (see also Req. MNGT-020 below).

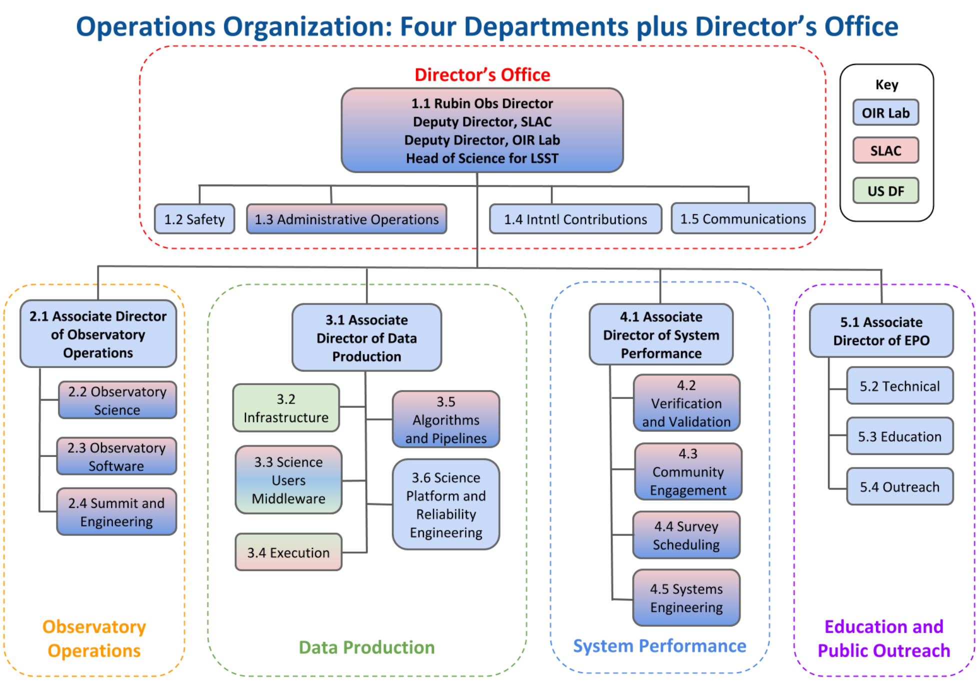



Figure 3: The Vera Rubin Observatory Organization Chart. Shaded boxes indicate shared staff across operations partners. Staff at affiliate institutions are included within their associated operations partner (i.e. not separately).

## 3.1 Management

**MNGT-020** There shall be a the single point of contact for all managerial aspects of the work.

**MNGT-040** This work shall be carried out within the Rubin Observatory management structure under the Data Production department.

**MNGT-060** The awardee shall inform Rubin Observatory management of planned changes in the availability of staff in support of the work.

**MNGT-080** The awardee shall conform to Rubin Observatory management practices including use of tracking and reporting tools adopted by the observatory. DMTN-020 is an example of construction era practices.

**MNGT-100** The awardee will support Rubin management in all oversight committee meetings, joint agency reviews, and other management body activities as deemed necessary or desired by AURA and SLAC.

## 3.2 Performance

**PERF-020** The awardee shall be responsible to Rubin Observatory management organizations AURA and SLAC for performance of the activities and capabilities detailed in this SOW through a negotiated set of performance metrics.

**PERF-040** The awardee's staff will receive annual performance input from Rubin management to be included in the awardee annual performance assessment process.

**PERF-060** Rubin management will have the authority to request that awardee staff who under perform be replaced.

## 3.3 Reporting

**REPT-020** The awardee shall provide a regular progress report on the status of activities, support provided, status of anomaly investigations, etc. This report shall cover all WPs described in this statement of and shall include contributions from sub-contractors as appropriate. This may be integrated in a more general Rubin Observatory report. DOE may place other, independent, reporting requirements on the awardee.

**REPT-040** The frequency of reporting shall be monthly, quarterly, and annual, as appropriate.

**REPT-060** Monthly reporting shall include SLA metrics to be agreed with Rubin Observatory such as cumulative downtime, issue turn around time etc.

## 3.4 Communications

Regardless of funding streams, Rubin Observatory Operations should function as one project. While we recognize that this is not always easy for staff already embedded in another institution, it is important that we share the same tools to avoid silos and to communicate effectively about our work and to our communities. The three primary platforms for communication at Rubin are the JIRA ticketing system, the Slack chat system, and the web forum, community.lsst.org. Technotes are produced via LSST the Docs [SQR-006] for Data Production. Official documentation is placed in Docushare and we also heavily use confluence. We would expect that Data Facility team members would engage with all of these.

An important part of both LSST Construction and Operations will be writing and sharing code. All software written for LSST is open source and publicly available, and is developed following the workflows and engineering standards described in the LSST Developer guide at `developer.lsst.io`. Code developed at institutions must be developed and made available under the same conditions.

**COMS-020** USDF staff shall use the Rubin Observatory communication tools such as JIRA, Slack and LSST the Docs.

**COMS-040** Any code developed for the Rubin Observatory Project shall be developed in the project repository (currently github) and shall carry the project open source license (currently GNU Public License (GPL)).

### 3.5 Meetings and Travel

**TRVL-020** The awardee shall support teleconferences and travel in support of the above work packages as required and agreed by both parties. At a minimum this would include weekly video calls and in person meetings two times per year.

**TRVL-040** The awardee shall participate in the NET group and attend the monthly telecons and meetings once or twice a year as needed.[2]

## 4 Deliverables

**DLVR-020** Any code deliverables shall adhere to the standards and guidelines of Rubin Observatory as on `developer.lsst.io`

## 5 Data Management System portability and cloud computing

Although the project baseline has physical compute facilities in Chile and USDF, with split-site processing at IN2P3, we have designed the system to be flexible with regard to the environment within which it is deployed. It is also possible the computing systems at the USDF and Chilean Access Centers will not satisfy the entire demand for near-the-data processing and peak load; computing models that allow externally-funded resources to be easily and efficiently used are desirable. Public clouds, by handling the accounting and resource management for multi-tenancy, provide an interesting solution for this, provided that potentially-large data storage costs can be mitigated. Any USDF partner shall be open to continuing investigations and partnerships of this type in collaboration with Rubin Observatory.

In this vein, we have recently undertaken studies to investigate the possibility of performing LSST data processing on cloud computing platforms provided by both Google and Amazon. On the Google cloud platform, we demonstrated that several of the major components of the Data Management System could be run effectively. In particular,

- we deployed the Qserv database system, demonstrating that it could achieve 80% of the

[2] `https://confluence.lsstcorp.org/pages/viewpage.action?pageId=20284335`

performance we achieved in-house on physical hardware;

- we demonstrated data transfer adequate for Prompt Processing, within the limits of the current LHN networks available for testing;.
- we deployed and tested the Prompt Products Database;
- we deployed an instance of the LSP.

Note that the LSP in particular is engineered around the Kubernetes provisioning system (K8S) system (subsubsection 2.5.1), and therefore deploys extremely smoothly to the Google Cloud. The results of this study are described in detail in DMTN-125.

The Google study did not investigate the single largest compute load that the LSST will face: Data Release Processing. We are now addressing this on Amazon Web Services / Elastic Compute Cloud, as described in DMTN-114. This work is ongoing at time of writing; initial results are positive.

We believe that these studies demonstrate the flexibility of the DM System to a variety of deployment environments, and particularly illustrate the value — and importance to the project of — K8S.

# A Sizing details

The following simplified sizing may be used to give the input sizes for a cost model. The storage sizes are given in Table 4 while the compute is given in Table 8 and Table 9. The cumulative storage requirements are also shown in Figure 1. The cumulative processing required is shown below in Figure 4.

Some useful inputs are provided in Table 6.

Table 6: Various inputs for deriving costs - 2019 represents current holdings.

| Year | 2019 | 2020 | 2021 | 2022 | 2023 |
|---|---|---|---|---|---|
| Core-hours Needed Total (DRP) | | 4.41E+06 | 4.41E+06 | 1.12E+07 | 4.53E+07 |
| Annual Increase | | 4.41E+06 | 0.00E+00 | 6.81E+06 | 3.40E+07 |
| Time to Process days | | 100.0 | 100.0 | 100.0 | 200 |
| Time to Process hours | | 2,400 | 2,400 | 2,400 | 4,800 |
| Instantaneous cores (DRP) Total | | 1,836 | 1,836 | 4,673 | 9,430 |
| Instantaneous cores (DRP) Annual increase | 1152 | 1,836 | 0 | 2,837 | 7,093 |

| | | | | | |
|---|---|---|---|---|---|
| Instantaneous cores (Alerts) | | 0 | 0 | 1188 | 1188 |
| Cores (Alerts) Annual increase | | 0 | 0 | 1188 | 0 |
| Instantaneous cores (US DAC/ Staff) | 540 | 540 | 540 | 141 | 568 |
| Cores (US DAC/ Staff) Annual increase | | 0 | 0 | 0 | 428 |
| Instantaneous cores (Chilean DAC) | | 0 | 0 | 26 | 103 |
| Cores (Chilean DAC) Annual increase | | 0 | 0 | 26 | 78 |
| Qserv nodes (US DAC/ Staff) | | | | 14 | 95 |
| Qserv nodes (US DAC/ Staff) Annual Increase | | | | 14 | 81 |
| Qserv nodes (Chilean DAC) | | | | 14 | 95 |
| Qserv nodes (Chilean DAC) Annual Increase | | | | 14 | 81 |
| **Total Cores Annual Increase** | | **1,836** | **0** | **4,051** | **7,599** |
| Fast Storage (TB) | | 12 | 24 | 50 | 206 |
| Annual Increase (Fast) | | 12 | 12 | 26 | 156 |
| Normal Storage (TB) | 3000 | 3680 | 3748 | 9241 | 38983 |
| Annual Increase (Normal) | | 680 | 68 | 5494 | 29742 |
| Latent Storage (TB) | | 319 | 876 | 4966 | 28854 |
| Annual Increase (Latent) | | 319 | 557 | 4090 | 23888 |
| High Latency (TB) | | 2910 | 6128 | 16733 | 63245 |
| Annual Increase (High Latency) | | 2910 | 3218 | 10605 | 46512 |
| Chilean DAC Fast Storage (TB) | | | | | 156 |
| Annual Increase (Fast Chilean DAC) | | | | | 156 |
| Chilean DAC Latent Storage (TB) | | | | | 28854 |
| Annual Increase (Latent Chilean DAC) | | | | | 28854 |
| Annual price decrease CPU | | 10% | | | |
| Annual price decrease Storage | | 5% | | | |
| Annual price decrease Qserv | | 8% | | | |

## A.1 Processing Plan

This model assumes the following processing:

- Precursor data (HSC RC2 and a similarly-sized DESC DC2 subset) is reprocessed each month during the Construction period using the Data Release Production (DRP).
- A large precursor reprocessing of HSC PDR2 (or equivalent) is completed twice a year. Products from one of these reprocessings will be released as Data Preview 0 (DP0) by the operations team. This will not be done at the USDF.
- One or more of these processings will be devoted to ComCam and LSSTCam science data during Commissioning. Some processing at the USDF might occur in the transition in late FY22. If not then, certainly in FY23 in advance of full operations. ComCam data will be released as DP1. LSSTCam commissioning data will be released as DP2 soon *after* the start of full operations in FY24.
- Alert Production (AP) processing happens continuously as LSSTCam science images are obtained. AP hardware is purchased in FY23 to support this.

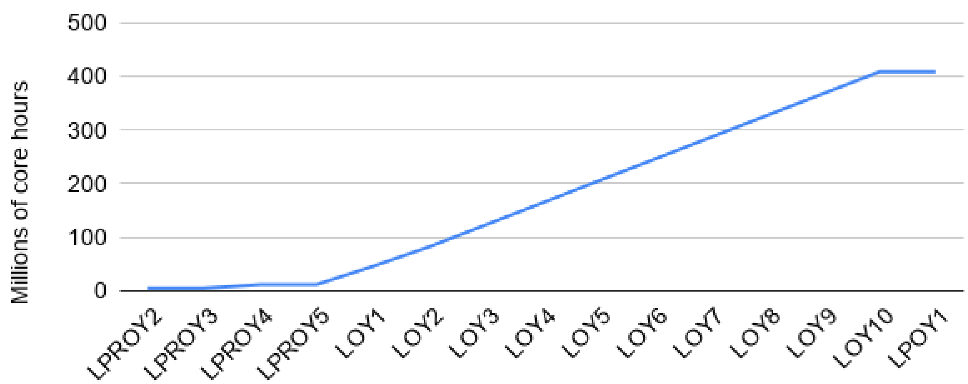



Figure 4: Evolution of **cumulative** computing needs for Rubin Observatory. Details are given below. Compute at other data facilities, is not included here. The USDF is responsible for the last two years of pre operations (LPROY4-5) which are transition to USDF, the survey years (LOY1-10), and the post operations years (only the first year of two is shown).

- DR1 processing begins after the first 6 months of the survey; the hardware for this can be part of the purchase during FY23.
- Annual DRP execution starts at the beginning of LSST Operations Year 2 with the processing for DR2. The hardware for each year's processing must be purchased and ready for use at the beginning of the year, so it is allocated in the tables to the prior fiscal year, when the images for that processing were taken.

Some storage for raw data needs to be in place at the beginning of the fiscal year, but it can be ramped up over the course of the year. As a simplification it is allocated to the fiscal year in which it will be used.

## A.2 Operations Storage Model

### A.2.1 Overview

Values are computed for the amount of storage expected to be "on the floor" at the beginning of each fiscal. Key scientific and algorithmic assumptions made include:

- All significant intermediates and data products generated by Data Release Production processing need to be kept on filesystem disk until the DRP is complete. Some scratch space is provided to hold small, temporary intermediates. If some intermediates could be removed during DRP when it is known they will no longer be needed, some space savings could be realized.

- HSC RC2 processing is representative of the outputs that DRP will generate (see e.g. PDR2[3] The coadd storage is doubled to account for an additional ”good-seeing” coadd along with the existing ”deep” coadd.
- Processed visit images (PVIs) and catalogs in Parquet format start on ”normal” filesystem disk but then move to object storage at the completion of the DRP, with lossy compression of the PVIs at that time. This is in accordance with RFC-325, although the relevant LCR has not yet been approved. Object storage is expected to be cheaper and more scalable for read-only data products; filesystem storage is used for data that is being generated or modified.
- Raw images and coadd images are only temporarily stored on filesystem disk and are then rapidly moved to object storage, where they are retained.
- Intermediates like warped images for coaddition are not survey data products and do not need to be kept beyond the end of the DRP and subsequent QA.

All data is backed up to tape permanently, including annual snapshots of filesystems. Any incremental backups are assumed to be reusable or otherwise purged and hence not significant.

#### A.2.2 Parameters

The numbers of science users are estimates, using ”Stack Club” users and Commissioning users for FY20 and 2021, followed by US science users in FY22 and FY23 for Data Preview data. The bulk of US science users are not expected to arrive until after Data Release 1 at the beginning of LOY2.

Storage per science user is estimated based on today’s usage at NCSA, scaled up as users become more active, and approaching the number given in LSE-81 as Operations begins. Note that it is expected that there will be a wide distribution of usage by user, with some using almost none and some using much more than their proportional share.

[3] https://hsc-release.mtk.nao.ac.jp/doc/index.php/sample-page/pdr2/

The LSSTCam image size is uncompressed and includes overscan, 4 bytes of raw data per pixel, and both science and corner rafts (guide and focus sensors).

The raw image compression factor was measured on simulated LSST images. The lossy image compression factor for processed visit images is the ratio between the lossy-compressed file size (estimated at 1/6 of uncompressed) and the lossless-compressed file size (estimated at 66% of uncompressed). Note that PVIs do not compress losslessly as well as raw images due to their floating point planes.

The number of observing nights per year and the number of visits per night are maximal estimates. Two images per visit is still the baseline and a possibility that must be accounted for. The number of calibration images per day was derived from the calibration plan.

Two complete all-sky coadds are assumed, one for ”good seeing” and one deep.

As stated above, the number of LSSTCam science images is scaled by 2/12 for FY23 given the length of science validation time. The number of test images, taken on test stands, is estimated as a ramp up to the full science cadence. The numbers of engineering (unprocessed) and calibration images are estimated as ramping-down fractions of the number of science and test images, with calibration images ending at the number per day given previously.

Sizes of rows in various data product tables are taken from LDM-141, which was in turn derived from the Data Product Definition Document (DPDD).

Qserv replicates its data for fault tolerance; a typical replication factor is selected here.

### A.2.3 Data Product Sizing

Images and the results of processing them are the dominant factor controlling the storage sizing which is outlined in Table 7. Precursor survey and LSSTCam images are the largest; ComCam, at less than 5% of the size of LSSTCam and with little on-sky science time is negligible, as is LATISS, which is less than 1% of the size of LSSTCam, though it has considerable on-sky time.

The sizing of the Alert Production Database (APDB) is based on experiments in [22] which found that 57,000 visits took 4.5 TB including indexes. A simple linear scaling to a full year's visits was performed, with half that purchased in 2020 for large (but not full) scale testing.

HyperSuprime-Cam (HSC) RC2 is a relatively small dataset used for monthly processing tests, but it is highly representative of the currently-known DRP work and so is used as the basis for scaling. The size of the input images was taken from [25]; the size of the outputs (image and Parquet/other non-image files) was measured from the latest execution. A similar size dataset based on DESC DC2 is assumed to be being used for an additional monthly processing test. Note that this is a very small subset of the full DESC DC2, which is expected to cover 300 square degrees to 10-year LSST depth (approximately 1000 epochs per point on the sky). The full DESC DC2 is not currently scheduled to be reprocessed by the construction team. Instead, twice-a-year processings of the full HSC SSP PDR2 dataset (including PDR1) are assumed to occur. The size of this dataset was measured on disk; it is 2,564,358 CCD images, each at 18.2 MB (approximately three times the size of PDR1 alone). The Operations team plans to host DESC DC2 as part of DP0 and may do some reprocessing of DESC DC2 for training and readiness purposes. But this will not be done at the USDF.

Output sizes are assumed to scale linearly with input size, and by the same factor for each instrument, except for coadds which scale by the sky area processed. While the Object catalog ought to be proportional to sky area as well, its size is expected to be dominated by Source and ForcedSource, so we conservatively make them all proportional to input size (visits) for the precursor data where we do not have object count estimates. For LSSTCam, we use the catalog row estimates to derive Qserv table sizes, but the Parquet file sizes are scaled based on HSC, as they may differ from the Qserv schema.

Scratch space is set at 10% of the output image storage for LSSTCam processing; it is assumed to be already present for precursor processing.

Qserv Czar fast (SSD) storage is assumed to be used for the primary Object table; additional space for the so-called ”secondary index” mapping object identifiers to spatial chunks is negligible in comparison.

The main Qserv database storage is based on the Parquet file sizing for precursor data and on the estimated numbers of Objects, Sources, and ForcedSources for LSSTCam data.

Note that no space is explicitly reserved for Qserv query result storage.

An additional 20% disk and tape storage is added to account for all other needs.

Table 7: Dataset sizes used to calculate storage needs during Operations

| Dataset Sizing | unit | LOY1 | LOY2 | LOY3 | LOY4 | LOY5 | LOY6 | LOY7 | LOY8 | LOY9 | LOY10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| LSSTCam Area | deg2 | 17000 | 17000 | 17000 | 17000 | 17000 | 17000 | 17000 | 17000 | 17000 | 17000 |
| APDB | TB | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 |
| Object store datasets: | | | | | | | | | | | |
| Incremental LSSTCam Raw Images | TB | 4816 | 4816 | 4816 | 4816 | 4816 | 4816 | 4816 | 4816 | 4816 | 4816 |
| LSSTCam Output Coadd Images | TB | 7727 | 7727 | 7727 | 7727 | 7727 | 7727 | 7727 | 7727 | 7727 | 7727 |
| Normal disk datasets: | | | | | | | | | | | |
| LSSTCam Output Images | TB | 13485 | 26970 | 40456 | 53941 | 67426 | 80911 | 94397 | 107882 | 121367 | 134852 |
| LSSTCam Output Parquet | TB | 7973 | 15946 | 23919 | 31893 | 39866 | 47839 | 55812 | 63785 | 71758 | 79731 |
| Scratch | TB | 1349 | 2697 | 4046 | 5394 | 6743 | 8091 | 9440 | 10788 | 12137 | 13485 |
| Qserv Czar/ Object | TB | 156 | 190 | 215 | 238 | 258 | 279 | 298 | 318 | 335 | 353 |
| Qserv Database | TB | 3510 | 5748 | 8018 | 10378 | 12881 | 15475 | 18199 | 21042 | 23965 | 27010 |
| Science User Home | TB | 2000 | 3000 | 4200 | 5250 | 6000 | 6750 | 7500 | 8250 | 9000 | 9750 |
| Other/ Misc | TB | 8208 | 13424 | 18684 | 23932 | 29148 | 34382 | 39642 | 44926 | 50226 | 55550 |

### A.2.4 Storage Sizing

Finally, storage is allocated to specific types as shown in Table 4. Fast storage (SSD) is used for the APDB and Qserv Czar, which accumulates data from year to year until Data Releases are retired. Normal storage is used for the datasets labeled as such, including output images (initially), output catalogs, and scratch. Local Qserv storage is used for Qserv catalogs. It is assumed that precursor data will be removed from Qserv once LSST data is available, but the LSST data accumulates from year to year.

Raw images (lossless-compressed) are written immediately to object storage, as are Parquet-format catalogs. PVIs are lossy-compressed and placed in object storage. The complete set of raw images is available, whereas the catalogs from only the last two Data Releases and the one in preparation are kept, and the PVIs from only the last Data Release and the one in preparation are online.

All data products and new raw images for each Data Release are copied to tape, but scratch space and the Qserv-schema catalogs are not.

Note that no replication is assumed in the object store.

## A.3 Compute Model

### A.3.1 Overview

This simplified computing model (Table 5) divides computation into three classes: Data Release Production (DRP), Alert Production, and Rubin Science Platform (for Rubin staff internal use). Calibration Products Production is assumed to be negligible. The number of cores for Alert Production does not change with time.

Scaling compute needs based on an execution of the nascent DRP pipeline on HSC PDR1 data and nightly executions of the nascent `ap_pipe` pipeline on HiTS2015 data is appropriate, but the fact that several steps are still missing from these pipelines must be taken into account.

Elapsed times are measured on existing hardware and converted into core-hours on a nominal CPU (Intel Xeon E5-2680v3 at 2.50 GHz). For example, if a pipeline running on precursor data took an average of one hour on a 32-core nominal CPU, 32 core-hours would be used as its compute requirement. This estimation methodology incorporates all I/O, memory bandwidth, cache miss, and other overheads into the core-hour measurement, simplifying calculations. Note that the nominal CPU does not evolve with time; if future CPUs do more work per core, the actual core-hours may be less than estimated here.

Scaling to other CPUs of the same architecture is based on the ratios of nominal GHz clock rates and core counts. For different architectures (e.g. Rome), the scaling is based on the ratio of industry-reported achievable FLOPS for the two architectures.

Key scientific and algorithmic assumptions are:

- DRP compute time is proportional to the input data size (or, equivalently, the number of visits). While certain tasks are undoubtedly proportional to sky area or number of Objects, overall the pipeline elapsed times are a better fit to the number of visits. Some of this may be because the Object density increases as the number of visits to the same sky patch increases.
- HSC PDR1 processing is generally representative of the final DRP, with an allocation for future additional steps as described below.
- Qserv node counts should remain proportional to the size of data loaded into the database in order to maintain sufficient disk bandwidth and query processing capability, but the

proportionality constant changes with time as new generations of system bus with greater bandwidth become available.

- The US DAC LSP is sized at 10% of the DRP compute budget in core-hours, readjusted to be spread over an entire year. The Rubin staff LSP is sized at 10% of the US DAC.

The DAC and staff LSP instances are sized based on the assumed percentages of DRP compute.

The amount of Qserv data that can be handled by a node is assumed to grow with time, doubling every four years (PCI Express has gone from 1.0 GB/sec to 16 GB/sec between 2003 and 2019). The number of Qserv nodes is calculated by dividing each Data Release's storage by the storage-per-node figure for its year; older nodes are assumed to be retired.

#### A.3.2 Parameters

The key parameters in Table 5 are described below.

HSC PDR1 was executed on the NCSA verification cluster, which uses the nominal CPU. The Alert Production executes on Kubernetes nodes, which are a bit slower; to be conservative, this is neglected.

A 2018 run of DRP on HSC PDR1 data is described at `https://confluence.lsstcorp.org/x/WpBiB`. The input data size is measured; note that the input data files are lossless-compressed. Most jobs (but not most of the time) could run on relatively small-memory machines with 24 cores and 5 GB RAM per core. The largest and longest-running jobs, however, required up to 4 times as much memory, using half or a quarter of the cores. To be conservative, we assume that half the cores were used for the large-memory jobs. The percentage of DRP core-hours that will need to execute on large-memory nodes is estimated.

Since the HSC PDR1 processing did not include several steps from the Science Pipelines Design document [LDM-151] such as image differencing and full multi-epoch characterization, the core-hours used are scaled up to the expected pipeline consumption. Note that these algorithmic adjustments are multiplicative.

The SQuaSH system reports the execution time of `ap_pipe` in seconds per CCD. A mean was taken over all processed CCDs, and it was assumed that each CCD is processed on a single core. These CCDs are from DECam, which is half the size of an LSST CCD, so the total time is doubled. A factor is added to account for additional steps like differential chromatic refraction compensation and false positive detection that are not well-represented in the current pipeline. Multiplying by the number of LSSTCam science CCDs gives the total number of core-hours per visit.

The amount of Qserv data that can be handled by one node is estimated based on the amount of disk that can be scanned in 12 hours at an aggregate rate of 1 GB per second. (Since the Qserv data replicas are not all anticipated to be accessed at the same rate, this is a conservative estimate.)

### A.3.3 Data Release Production

The number of nominal core-hours per TB of input data is multiplied by the precursor (HSC RC2 and DESC DC2 subset for 12 months and HSC PDR2 twice a year) and LSSTCam input data sizes (with lossless compression) to determine the total number of core-hours needed in each year. This is shown in Table 8. Approximately one-third of these core-hours need to be provided by small-memory (4-5 GB/core) machines; the other two-thirds need to come from large-memory (8-20 GB/core) machines.

Table 8: Compute needs for DRP and AP

| **Data Release Production** | **units** | **FY20** | **FY21** | **FY22** | **FY23** | **Notes** |
|---|---|---|---|---|---|---|
| Precursor input size | TB | 206 | 206 | 206 | 206 | |
| LSSTCam visit input size | TB | | | 319 | 1911 | raw images / images/ visit, lossless-compressed |
| Precursor compute | core-hours | 4.4E+06 | 4.4E+06 | 4.4E+06 | 4.4E+06 | |
| LSSTCam compute | core-hours | | | 6.8E+06 | 4.1E+07 | |
| **Total DRP compute** | **core-hours** | **4.4E+06** | **4.4E+06** | **1.1E+07** | **4.5E+07** | |
| Alert Production | units | FY20 | FY21 | FY22 | FY23 | Notes |
| AP cores | cores | | | 1,188 | 1,188 | minimum necessary to keep up |

### A.3.4 Alert Production

The core-hours per visit are divided by the minimum visit length (30 sec plus 1 sec shutter motion plus 2 sec readout) to give the minimum number of cores needed to keep up with image taking. This is shown in Table 8. These cores are expected to be provided over multiple ”strings” of nodes. Note that the current AP design is not readily able to take advantage of more than one core per CCD.

### A.3.5 LSST Science Platform

LSST Science Platform needs for US DAC science users are derived as 10% of the DRP core-hour requirement and are shown in Table 9. The LSP core-hours are assumed to be spread over a year, giving the total number of nominal cores needed in the DAC. Peak loads are expected to be handled by ”borrowing” elastically from the DRP compute pool.

As a reasonableness check, the number of cores per science user is computed, but it must be noted that an oversubscription factor needs to be taken into account since not all users are expected to be simultaneously active.

Similar computations for the Chilean DAC (at 20% of the US DAC) and the LSST staff LSP (at 10% of the US DAC) are also in Table 9.

The number of Qserv nodes needed is computed from the storage devoted to it and the storage per node number. Note that staff use of Qserv is taken into account by loading the Data Release products into an internal-only Qserv instance and then making that instance part of the DAC at Data Release, so the compute sizing is part of the US DAC.

Table 9: Compute needs for the Science Platform instances

| **US DAC** | **units** | **FY20** | **FY21** | **FY22** | **FY23** | **Notes** |
|---|---|---|---|---|---|---|
| LSP cores | cores | | | 128 | 517 | 10% of DRP, over a year |
| Qserv nodes | nodes | | | 14 | 95 | |
| LSP cores/ science user | cores/ user | | | 0.03 | 0.10 | includes oversubscription |
| Chilean DAC | units | FY20 | FY21 | FY22 | FY23 | Notes |
| LSP cores | cores | | | 26 | 103 | 20% of US DAC |
| Qserv nodes | nodes | | | 14 | 95 | |
| Staff LSP | units | FY20 | FY21 | FY22 | FY23 | Notes |
| LSP cores | cores | | | 13 | 52 | 10% of US DAC |

# B Construction requirements relevant to the USDF

In this appendix, Data Management Subsystem Requirements (DMSR) for construction are shown associated with the different aspects of the Data Facility and it’s scope of operations. The DMSR are captured in the Rubin document LSE-61. The USDF operator will be responsible for making sure the USDF supports these requirements in collaboration with the Rubin Operations team (specifically the Data Production Department).

## B.1 Prompt processing

The USDF will need to run certain software in near real time namely (further detailed in DMTN-104:

- Alert Distribution
- Prompt Processing Ingest
- Offline Quality Control
- Prompt Quality Control
- Prompt Processing
- APDB

This is referred to as the prompt enclave in DMTN-104. This means supporting the DMSR requirements below.

| **Related Requirements** | |
|---|---|
| CA-DM-CON-ICD-0019 | Camera engineering image data archiving |
| DMS-REQ-0002 | Transient Alert Distribution |
| DMS-REQ-0004 | Nightly Data Accessible Within Specified Time |
| DMS-REQ-0008 | Pipeline Availability |
| DMS-REQ-0096 | Generate Data Quality Report Within Specified Time |
| DMS-REQ-0098 | Generate DMS Performance Report Within Specified Time |
| DMS-REQ-0100 | Generate Calibration Report Within Specified Time |
| DMS-REQ-0102 | Provide Engineering & Facility Database Archive |
| DMS-REQ-0131 | Calibration Images Available Within Specified Time |
| DMS-REQ-0161 | Optimization of Cost, Reliability and Availability in Order |
| DMS-REQ-0162 | Pipeline Throughput |
| DMS-REQ-0165 | Infrastructure Sizing for "catching up" |
| DMS-REQ-0166 | Incorporate Fault-Tolerance |
| DMS-REQ-0167 | Incorporate Autonomics |
| DMS-REQ-0284 | Level-1 Production Completeness |
| DMS-REQ-0314 | Compute Platform Heterogeneity |
| DMS-REQ-0318 | Data Management Unscheduled Downtime |

| | |
|---|---|
| EP-DM-CON-ICD-0023 | Nightly DM Transfer of Processed Visit Images (PVI)-Based Images to EPO |

## B.2 Batch System

The USDF will need to run certain software in batch mode (further detailed in DMTN-104:

- Batch Production
- Offline Quality Control
- Bulk Distribution

This is referred to as the offline production enclave in DMTN-104. This means supporting the DMSR requirements below.

| Related Requirements | |
|---|---|
| DM-TS-CON-ICD-0003 | Wavefront image archive access |
| DMS-REQ-0004 | Nightly Data Accessible Within Specified Time |
| DMS-REQ-0008 | Pipeline Availability |
| DMS-REQ-0131 | Calibration Images Available Within Specified Time |
| DMS-REQ-0161 | Optimization of Cost, Reliability and Availability in Order |
| DMS-REQ-0162 | Pipeline Throughput |
| DMS-REQ-0163 | Re-processing Capacity |
| DMS-REQ-0166 | Incorporate Fault-Tolerance |
| DMS-REQ-0167 | Incorporate Autonomics |
| DMS-REQ-0284 | Level-1 Production Completeness |
| DMS-REQ-0289 | Calibration Production Processing |
| DMS-REQ-0314 | Compute Platform Heterogeneity |
| DMS-REQ-0318 | Data Management Unscheduled Downtime |
| DMS-REQ-0320 | Processing of Data From Special Programs |
| DMS-REQ-0334 | Persisting Data Products |
| DMS-REQ-0341 | Providing a Precovery Service |
| EP-DM-CON-ICD-0037 | EPO Compute Cluster |

## B.3 US Data Access Center

The USDF will need to host the US Data Access Center (DAC) comprising the elements below (further detailed in DMTN-104 r.:

- LSP Nublado
- LSP Portal
- WebDAV API
- Simple Image Access (International Virtual-Observatory Alliance (IVOA) standard) (SIA) API
- Server-side Operations for Data Access (IVOA standard) (SODA) API
- Table Access Protocol (IVOA standard) (TAP) API
- LSP Database

This is referred to as the DAC US Enclave in DMTN-104. This means supporting the DMSR requirements below.

| **Related Requirements** | |
|---|---|
| DMS-REQ-0004 | Nightly Data Accessible Within Specified Time |
| DMS-REQ-0077 | Maintain Archive Publicly Accessible |
| DMS-REQ-0089 | Solar System Objects Available Within Specified Time |
| DMS-REQ-0094 | Keep Historical Alert Archive |
| DMS-REQ-0102 | Provide Engineering & Facility Database Archive |
| DMS-REQ-0119 | DAC resource allocation for Level 3 processing |
| DMS-REQ-0131 | Calibration Images Available Within Specified Time |
| DMS-REQ-0161 | Optimization of Cost, Reliability and Availability in Order |
| DMS-REQ-0162 | Pipeline Throughput |
| DMS-REQ-0166 | Incorporate Fault-Tolerance |
| DMS-REQ-0167 | Incorporate Autonomics |
| DMS-REQ-0193 | Data Access Centers |
| DMS-REQ-0194 | Data Access Center Simultaneous Connections |
| DMS-REQ-0196 | Data Access Center Geographical Distribution |

| | |
|---|---|
| DMS-REQ-0284 | Level-1 Production Completeness |
| DMS-REQ-0287 | DIASource Precovery |
| DMS-REQ-0310 | Un-Archived Data Product Cache |
| DMS-REQ-0311 | Regenerate Un-archived Data Products |
| DMS-REQ-0312 | Level 1 Data Product Access |
| DMS-REQ-0313 | Level 1 & 2 Catalog Access |
| DMS-REQ-0314 | Compute Platform Heterogeneity |
| DMS-REQ-0318 | Data Management Unscheduled Downtime |
| DMS-REQ-0322 | Special Programs Database |
| DMS-REQ-0334 | Persisting Data Products |
| DMS-REQ-0336 | b Regenerating Data Products from Previous Data Releases |
| DMS-REQ-0341 | Providing a Precovery Service |
| DMS-REQ-0344 | Constraints on Level 1 Special Program Products Generation |
| DMS-REQ-0363 | Access to Previous Data Releases |
| DMS-REQ-0364 | Data Access Services |
| DMS-REQ-0366 | Subsets Support |
| DMS-REQ-0367 | Access Services Performance |
| DMS-REQ-0368 | Implementation Provisions |
| DMS-REQ-0370 | Older Release Behavior |
| EP-DM-CON-ICD-0001 | US DAC Provides EPO Interface |
| EP-DM-CON-ICD-0002 | EPO is an Authorized Science User |
| EP-DM-CON-ICD-0034 | Citizen Science Data |
| OCS-DM-COM-ICD-0029 | Archive Latency |

## B.4 Data transfer and preservation

The USDF needs to look after Rubin data in terms of (further detailed in DMTN-104):

- Ingest/ Metadata Management
- Lifetime Management
- Transport/ Replication/ Backup
- Storage

This is referred to as DataBackbone services in DMTN-104. This means supporting the DMSR requirements below.

| Uses: | Used in: |
|---|---|
| **Related Requirements** | |
| DMS-REQ-0008 | Pipeline Availability |
| DMS-REQ-0068 | Raw Science Image Metadata |
| DMS-REQ-0074 | Difference Exposure Attributes |
| DMS-REQ-0077 | Maintain Archive Publicly Accessible |
| DMS-REQ-0089 | Solar System Objects Available Within Specified Time |
| DMS-REQ-0094 | Keep Historical Alert Archive |
| DMS-REQ-0102 | Provide Engineering & Facility Database Archive |
| DMS-REQ-0120 | Level 3 Data Product Self Consistency |
| DMS-REQ-0122 | Access to catalogs for external Level 3 processing |
| DMS-REQ-0123 | Access to input catalogs for DAC-based Level 3 processing |
| DMS-REQ-0126 | Access to images for external Level 3 processing |
| DMS-REQ-0127 | Access to input images for DAC-based Level 3 processing |
| DMS-REQ-0130 | Calibration Data Products |
| DMS-REQ-0131 | Calibration Images Available Within Specified Time |
| DMS-REQ-0132 | Calibration Image Provenance |
| DMS-REQ-0161 | Optimization of Cost, Reliability and Availability in Order |
| DMS-REQ-0162 | Pipeline Throughput |
| DMS-REQ-0163 | Re-processing Capacity |
| DMS-REQ-0164 | Temporary Storage for Communications Links |
| DMS-REQ-0165 | Infrastructure Sizing for ”catching up” |
| DMS-REQ-0166 | Incorporate Fault-Tolerance |
| DMS-REQ-0167 | Incorporate Autonomics |
| DMS-REQ-0176 | Base Facility Infrastructure |
| DMS-REQ-0185 | Archive Center |
| DMS-REQ-0186 | Archive Center Disaster Recovery |
| DMS-REQ-0197 | No Limit on Data Access Centers |
| DMS-REQ-0266 | Exposure Catalog |
| DMS-REQ-0269 | DIASource Catalog |
| DMS-REQ-0271 | DIAObject Catalog |
| DMS-REQ-0273 | SSObject Catalog |
| DMS-REQ-0287 | DIASource Precovery |
| DMS-REQ-0291 | Query Repeatability |

| | |
|---|---|
| DMS-REQ-0292 | Uniqueness of IDs Across Data Releases |
| DMS-REQ-0293 | Selection of Datasets |
| DMS-REQ-0299 | Data Product Ingest |
| DMS-REQ-0309 | Raw Data Archiving Reliability |
| DMS-REQ-0310 | Un-Archived Data Product Cache |
| DMS-REQ-0313 | Level 1 & 2 Catalog Access |
| DMS-REQ-0314 | Compute Platform Heterogeneity |
| DMS-REQ-0317 | DIAForcedSource Catalog |
| DMS-REQ-0318 | Data Management Unscheduled Downtime |
| DMS-REQ-0322 | Special Programs Database |
| DMS-REQ-0334 | Persisting Data Products |
| DMS-REQ-0338 | Targeted Coadds |
| DMS-REQ-0339 | Tracking Characterization Changes Between Data Releases |
| DMS-REQ-0346 | Data Availability |
| DMS-REQ-0363 | Access to Previous Data Releases |
| DMS-REQ-0364 | Data Access Services |
| DMS-REQ-0365 | Operations Subsets |
| DMS-REQ-0366 | Subsets Support |
| DMS-REQ-0369 | Evolution |
| DMS-REQ-0370 | Older Release Behavior |
| DMS-REQ-0372 | a Archiving Camera Test Data |
| DMS-REQ-0386 | a Archive Processing Provenance |
| DMS-REQ-0387 | b Serve Archived Provenance |
| DMS-REQ-0388 | Provide Re-Run Tools |
| OCS-DM-COM-ICD-0047 | Image Archived Event |

# C Glossary and Acronyms

# D Glossary

**Alert Production DataBase** A dedicated, internal database system used to support LSST Alert Production. Does not support end-user access..

**APDB** Alert Production DataBase.

**calibration** The process of translating signals produced by a measuring instrument such as a telescope and camera into physical units such as flux, which are used for scientific analysis. Calibration removes most of the contributions to the signal from environmental and instrumental factors, such that only the astronomical component remains.

**Camera** The LSST subsystem responsible for the 3.2-gigapixel LSST camera, which will take more than 800 panoramic images of the sky every night. SLAC leads a consortium of Department of Energy laboratories to design and build the camera sensors, optics, electronics, cryostat, filters and filter exchange mechanism, and camera control system.

**Center** An entity managed by AURA that is responsible for execution of a federally funded project.

**CERN** European Organization for Nuclear Research.

**CI** Continuous Integration.

**Construction** The period during which LSST observatory facilities, components, hardware, and software are built, tested, integrated, and commissioned. Construction follows design and development and precedes operations. The LSST construction phase is funded through the NSF MREFC account.

**DAC** Data Access Center.

**Data Access Center** Part of the LSST Data Management System, the US and Chilean DACs will provide authorized access to the released LSST data products, software such as the Science Platform, and computational resources for data analysis. The US DAC also includes a service for distributing bulk data on daily and annual (Data Release) timescales to partner institutions, collaborations, and LSST Education and Public Outreach (EPO)..

**data collection** A data collection in the second-generation (Gen2) Butler (referred to as a data repository in earlier generations) consists of hierarchically organized data files, an inventory or registry of the contents (i.e., metadata from the data files) stored in an sqlite3 file, and a Mapper file that specifies to the LSST Stack software the camera model to apply when accessing the data in the data repository.

**Data Management** The LSST Subsystem responsible for the Data Management System (DMS), which will capture, store, catalog, and serve the LSST dataset to the scientific community and public. The DM team is responsible for the DMS architecture, applications, middleware, infrastructure, algorithms, and Observatory Network Design. DM is a distributed team working at LSST and partner institutions, with the DM Subsystem Manager located at LSST headquarters in Tucson.

**Data Management Subsystem** The Data Management Subsystem is one of the four subsystems which constitute the LSST Construction Project. The Data Management Subsystem is responsible for developing and delivering the LSST Data Management System to the LSST Operations Project.

**Data Management System** The computing infrastructure, middleware, and applications that process, store, and enable information extraction from the LSST dataset; the DMS will process peta-scale data volume, convert raw images into a faithful representation of the universe, and archive the results in a useful form. The infrastructure layer consists of the computing, storage, networking hardware, and system software. The middleware layer handles distributed processing, data access, user interface, and system operations services. The applications layer includes the data pipelines and the science data archives' products and services.

**Data Release Processing** Deprecated term; see Data Release Production.

**Data Release Production** An episode of (re)processing all of the accumulated LSST images, during which all output DR data products are generated. These episodes are planned to occur annually during the LSST survey, and the processing will be executed at the Archive Center. This includes Difference Imaging Analysis, generating deep Coadd Images, Source detection and association, creating Object and Solar System Object catalogs, and related metadata.

**Department of Energy** cabinet department of the United States federal government; the DOE has assumed technical and financial responsibility for providing the LSST camera. The DOE's responsibilities are executed by a collaboration led by SLAC National Accelerator Laboratory.

**DM** Data Management.

**DMS** Data Management Subsystem.

**Document** Any object (in any application supported by DocuShare or design archives such as PDMWorks or GIT) that supports project management or records milestones and deliverables of the LSST Project.

**DOE** Department of Energy.

**DPDD** Data Product Definition Document.

**DRP** Data Release Production.

**EFD** Engineering and Facility Database.

**ESNet** Energy Sciences Network.

**FOA** Funding Opportunity Announcement.

**GPL** GNU Public License.

**IN2P3** Institut National de Physique Nucléaire et de Physique des Particules.

**ITC** Information Technology Center.

**IVOA** International Virtual-Observatory Alliance.

**JIRA** issue tracking product (not an acronym but a truncation of Gojira the Japanese name for Godzilla).

**K8S** Kubernetes provisioning system.

**KB** KiloByte.

**LHN** Long Haul Network.

**LOY1** LSST Operations Year 1.

**LSP** LSST Science Platform (now Rubin Science Platform).

**LSST** Legacy Survey of Space and Time (formerly Large Synoptic Survey Telescope).

**metadata** General term for data about data, e.g., attributes of astronomical objects (e.g. images, sources, astroObjects, etc.) that are characteristics of the objects themselves, and facilitate the organization, preservation, and query of data sets. (E.g., a FITS header contains metadata).

**middleware** Software that acts as a bridge between other systems or software usually a database or network. Specifically in the Data Management System this refers to Butler for data access and Workflow management for distributed processing..

**MPC** Minor Planet Center.

**NCSA** National Center for Supercomputing Applications.

**Operations** The 10-year period following construction and commissioning during which the LSST Observatory conducts its survey.

**Opportunity** The degree of exposure to an event that might happen to the benefit of a program, project, or other activity. It is described by a combination of the probability that the opportunity event will occur and the consequence of the extent of gain from the occurrence, or impact. There are two levels of opportunities. At the macro level, a project itself is the manifestation of the pursuit of an opportunity. At the element level, tactical opportunities exist, whereby certain events, if realized, provide a cost or schedule savings to the project or increase technical performance.

**Prompt Processing** The data processing which occurs at the Archive Center based on the stream of images coming from the telescope. This includes both Alert Production, which scans the image stream to identify and send alerts on transient and variable sources, and Solar System Processing, which identifies and characterizes objects in our solar system. It also includes specialized rapid calibration and Commissioning processing. Prompt Processing generates the Prompt Data Products..

**Qserv** LSST's distributed parallel database. This database system is used for collecting, storing, and serving LSST Data Release Catalogs and Project metadata, and is part of the Software Stack.

**Rucio** A scientific data management system developed at CERN; `https://rucio.cern.ch`.

**Science Platform** A set of integrated web applications and services deployed at the LSST Data Access Centers (DACs) through which the scientific community will access, visualize, and perform next-to-the-data analysis of the LSST data products.

**SIA** Simple Image Access (IVOA standard).

**SLAC** SLAC National Accelerator Laboratory.

**SLAC National Accelerator Laboratory** A national laboratory funded by the US Department of Energy (DOE); SLAC leads a consortium of DOE laboratories that has assumed responsibility for providing the LSST camera. Although the Camera project manages its own schedule and budget, including contingency, the Camera team's schedule and requirements are integrated with the larger Project. The camera effort is accountable to the LSSTPO..

**SODA** Server-side Operations for Data Access (IVOA standard).

**Subsystem** A set of elements comprising a system within the larger LSST system that is responsible for a key technical deliverable of the project.

**TAP** Table Access Protocol (IVOA standard).

**US** United States.

**USDF** US Data Facility.

**VPN** Virtual Private Network.