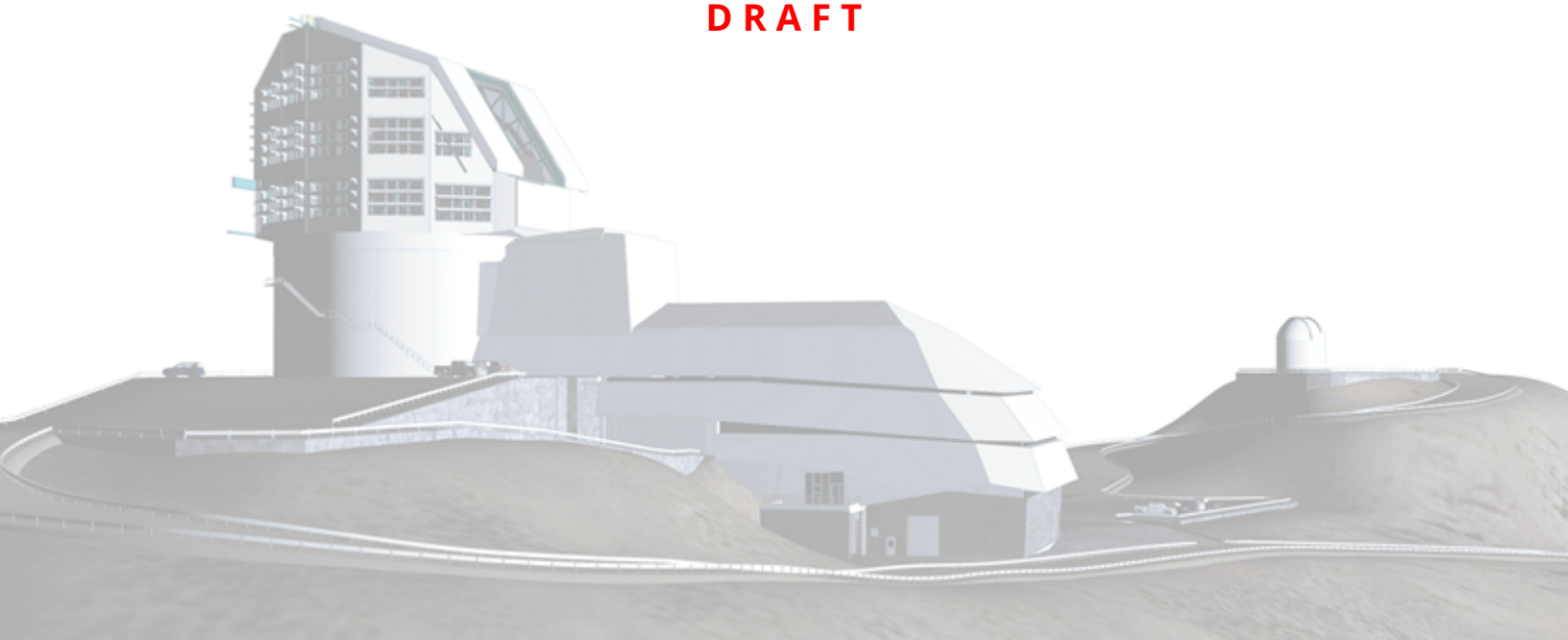

Vera C. Rubin Observatory
Rubin Observatory Operations

Management and Execution plan for Data Management Operations.

William O'Mullane

RTN-046

Latest Revision: 2024-05-17

DRAFT

Abstract

This is the management plan for operations of Data Management - this includes software products and data products.

Change Record

Version	Date	Description	Owner name
0.1	2023-01-10	Draft.	William O'Mullane

Document source location: `https://github.com/lsst/rtn-046`

Contents

Management and Execution plan for Data Management Operations.

1 Introduction

1.1 Purpose

This document defines the mission, goals and objectives, organization and responsibilities of Vera C. Rubin Observatory Data Management Operations.

1.2 Mission Statement

Maintain, improve and operate a suite of Vera C. Rubin data management services to produce and serve to the community high-quality data products from the Legacy Survey of Space and Time.

1.3 Goals and Objectives

These are similar to our construction goals outlined in LDM-294. Rubin Data Management Operations will:

- Produce the data products as outlined in LSE-61
- Maintain and improve data production mechanisms.
- Maintain and improve data access mechanisms.
- Maintain and improve data abstraction mechanisms.
- Assess current and operations-era technologies for use in providing engineered solutions for Vera C. Rubin Observatory .
- Maintain appropriate cybersecurity measures throughout Vera C. Rubin Observatory and especially on external facing services.
- Document the operational procedures associated with using and maintaining DM capabilities.
- Evaluate, select, recruit, hire/contract and direct permanent staff, contract, and in-kind resources in Rubin and from partner organizations participating in DM initiatives.

The goals in selecting and, where necessary, developing Rubin software solutions are:

- We prefer to acquire and configure existing, off-the-shelf, solutions. Where no satisfactory off-the-shelf solutions are available, we develop the software and hardware systems necessary to meet our objectives. This extends into maintenance where we will continue to probe choices and may replace custom systems with off-the-shelf solutions where appropriate.
- The software architecture is actively managed at the subsystem level. A well engineered and cleanly designed codebase is less buggy, more maintainable, and makes developers who work on it more productive. We continue to follow and maintain the developer guide[1].
- Other than when prohibited by licensing, security, or other similar considerations, all newly developed source code, and in particular that pertaining to scientific algorithms, is public. Our primary goals in publicizing the code are to simplify reproducibility of LSST data products and to provide insight into algorithms used. Achieving these goals requires that the software must be properly documented.
- Background decision material on choices made will be documented in technical notes with ”DMTN”, ”RTN” or similar series handles. (see `lsst.io`)

2 Architecture

The construction era DM architecture is defined in LDM-148.

As stated in the introduction our operational goals now include production of the data products. In broad terms we may think if two prongs in data management: Data Production and Data Serving. This is depicted in Figure 3.

We also now have three operational data facilities for data release production and a Cloud Facility on Google for the science users. This is all depicted in Figure 1.

Details about the build up the data facilities is given in RTN-021.

3 Functionality based teams and organisation

While Figure 2 Shows the reporting structure Figure 3 puts this more in a operations concept. We consider the main functions to be Data Production and Data Serving,

[1] `developer.lsst.io`

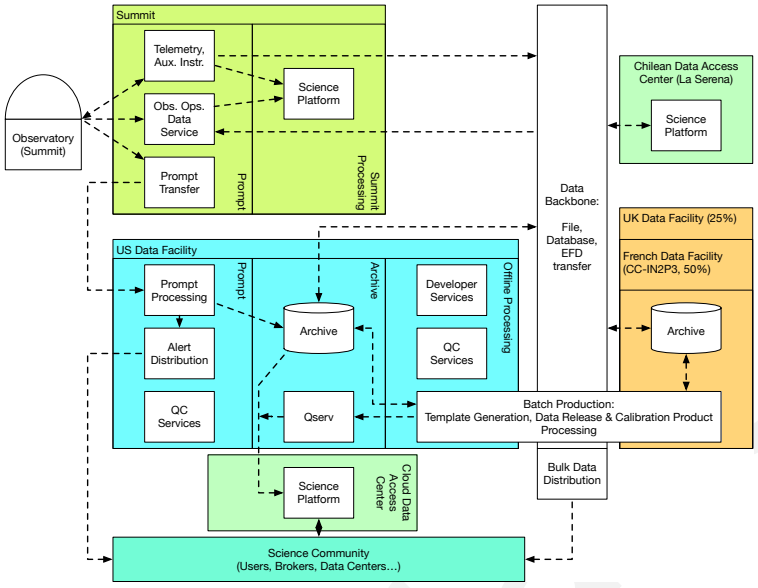

FIGURE 1: Simplified operations architecture for Data Management.

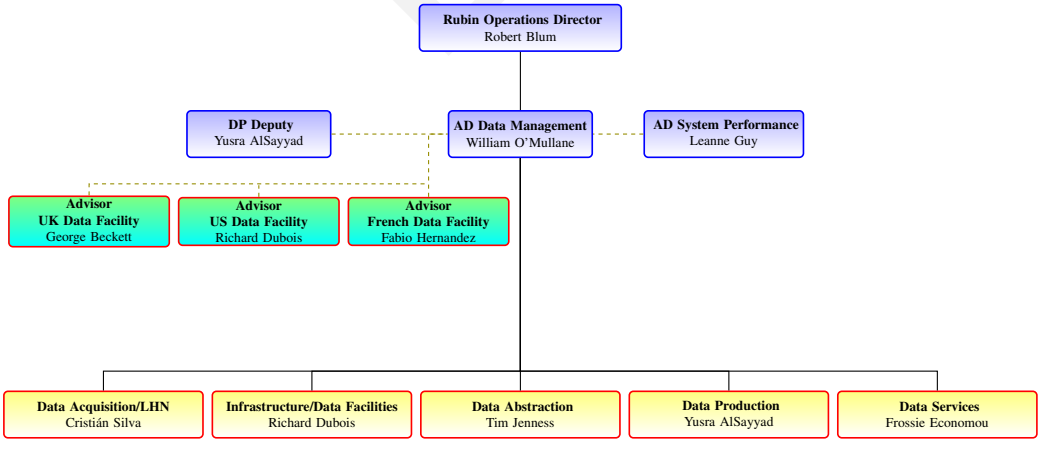

FIGURE 2: Reporting lines in Data Management Operations.

these are supported by the data abstraction team and the data facilities.

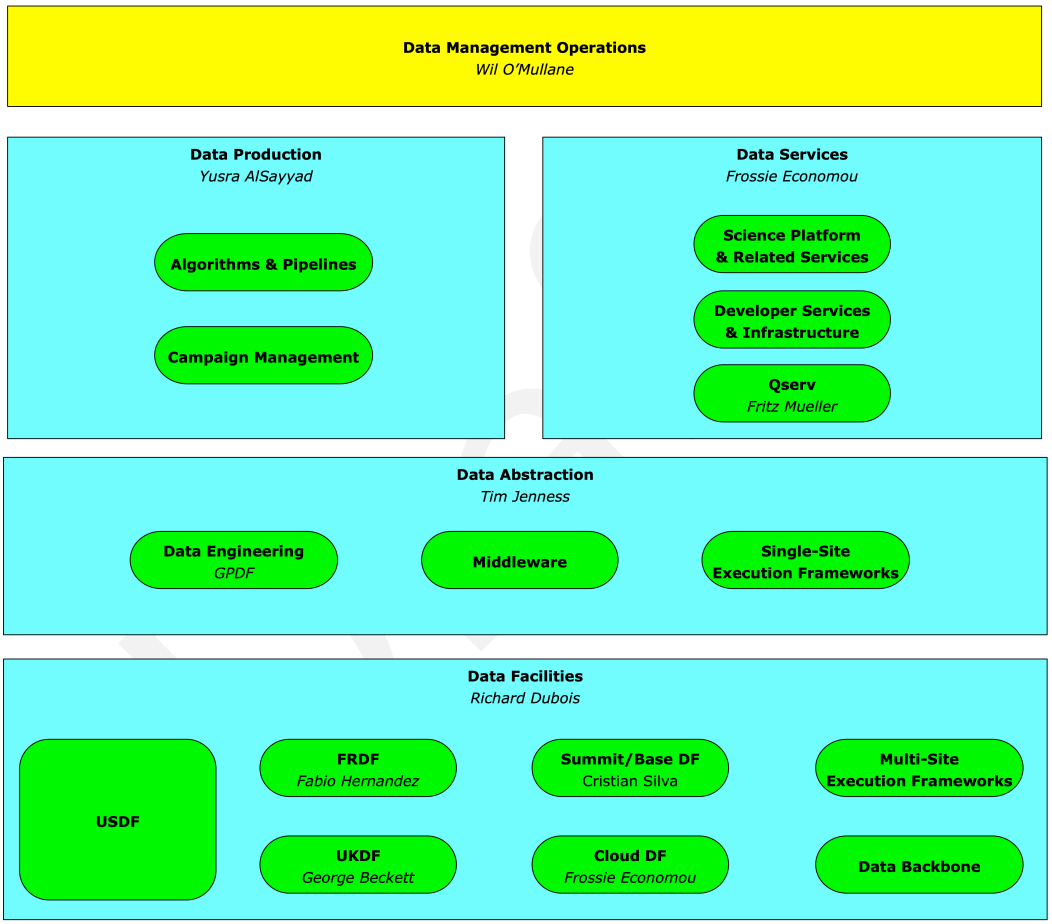

FIGURE 3: Functions in operations of Rubin Data Management.

3.1 Support model

DM strives to have automated and reliable systems. We are not sized for twenty four seven support but some of us are available on call as a best efforts basis. The assumption is after the initial set up calls would be infrequent (order less than once a month) for problems. Appropriate numbers are lodged in 1Password and available to summit observers. We will also set up some Squadcast alerts for specific things.

3.1.1 Alerts

Alerts must be kept running and hence we do need weekend on call support for to ensure we do not have consecutive days without alerts.

3.1.2 Potential Outline

If we do on call on best effort basis we should agree on some ground rules. Institutionally it may be difficult to formalize these . These might be:

- Should an individual get called out of hours they would not need to be in any meetings or workplace the following morning.
- Any hour worked out of hours may be taken as 1.5 hours leave at another point in time.
- If requested specific mobile phones for on call should be provided (may be hard given our distributed nature).

3.2 Data services

All services associated with data serving are in this group. As depicted in Figure 3 this includes:

- The Science Platform
- Developer Services Infrastructure
- Qserv advanced Database
- The Engineering Facilities Database

A more complete list of items under may be found in the section 7

3.3 Data Acquisition and Long Haul Networks

After the camera team close the shutter and they read out an image and put header on it. From this point on Data Management take that image process, store and transmit it to the USA..

The hardware and software infrastructure for doing this and for running all telescope control software is part of the Summit facility. The Chile DevOps team delivers and maintains the base and summit facilities (see subsection 4.5).

The team provides networking and machines plus a kubernetes layer ready for deployment of services from data management as well as telescope and site software.

On the summit there are also some bare metal machines which are setup usually with puppet.

This team also controls the more classic IT support for desktops and laptop connections etc.

3.4 Making Data

All services associated with data making are in this group. As depicted in Figure 3 this includes:

- The Science Pipelines code
- Execution of science pipelines to produce data products
  - Alert production
  - Data Release Prodcution

A more complete list of items under may be found in the section 7

3.5 Data Abstraction

Underpinning Data Making and Data Serving is out abstraction of data and services. This includes middleware such as butler and batch production systems etc. But also Prompt Processing execution and Data engineering. It is crucial for our system portability to maintain the abstraction layer.

A comprehensive list is given in section 7.

Some of these require a little more discussion here.

3.5.1 Data Engineering

- Support the metadata translation infrastructure (astro_metadata_translator) and monitor correctness of FITS headers.
- Advise on file formats and file metadata for all systems writing files that are to be archived.
- Support the Felis system for specifying schemas.
- Define the data models for everything in the consolidated database. (“global data model” schema?)
- Write and support code that populates the consolidated database (for example, code that analyzes the EFD and creates the “exposure” and “visit” tables).
- Gregory Dubois-Felsmann is product owner (“Data Scientist”)
- Staffing: tiny in construction. 0.5 FTE in ops + fractional GPDF.

3.5.2 Middleware Assumptions

As can be seen in the product list there are a lot of elements to Middleware. A few assumptions are made.

- Assumes butler/Rucio integration is entirely handled by the infrastructure team.
- Assumes database administration is done by the infrastructure team.
- Staffing: 0.5 Andy S; 0.5 new NateP; NateL 0.25; Matthias 0.5; MichelleG 0.5; Kowalik 0.25; at least 0.25 PanDA ongoing person for ctrl_bps_panda.

4 Data facilities and access centers

Hardware underpins all of our operations. This is arranged in three data facilities in US, UK and France as outlined below. We also have two on project Data Access Centers to provide services to the scientific users.

The plan for building up the data facilities is in RTN-021. A more complete list of items under may be found in the section 7

4.1 USDF

The USDF will be the main archive of Rubin data. It performs the daily processing of data including alert generation. It performs 25% of th DRP processing. There is a full description in DMTN-189. User batch will run at the USDF [DMTN-223].

4.2 FrDF

The French Data facility will hold a copy of the Raw data. The FRDF will run 50% of the DRP processing.

4.3 UKDF

The FRDF will run 25% of the DRP processing.

4.4 US DAC

The USDAC is hosted on Google Cloud. Most image data remains at USDF but some catalogs and possibly coadds will be kept on Google. All User files spaces and the RSP will be on google [DMTN-209].

4.5 Chile facilities

There are facilities at the base and the summit in Chile.

4.5.1 Summit Facilities

We maintain a data center on the summit of cerro Pachon. A large part of this is a kubernetes cluster which runs the control system components (CSCs) and the science platform to allow the team access to the images coming of the cameras. There are many other machines for individual control - mostly within the data center though some are physically close to the devices they control. We are responsible for all fibre cabling and networking within the facility to allow interconnection of the equipment.

4.5.2 Base Facilities

Within the NOIRLab base data center we maintain a full DAQ and ancillary machines for testing the control system. This is detailed in .. ITTN-00 The base facility is also where our DWDM connects to the summit and to the LHN . Finally some virtual machines run here for services and backups of services in the US such as email.

4.5.3 Chile DAC

The Chilean Data Access Center will be built after operations commences. Some discussions are still pending on its exact shape see LDM-572. This will be hosted in the base data center.

5 Project Controls

DM operations tracks work and milestones in Jira following RTN-005.

Risk management folows the Operations risk management plan [RDO-71].

Security is covered in O'Mullane et al. (RTN-030).

Disaster recovery is civered in various documents:

- ITTN-058 covers Disaster Recovery for Infrastructure Support Devices
- ITTN-057 covers Disaster Recovery for Computing
- ITTN-056 covers Disaster Recovery for the Network

5.1 Work Breakdown Structure

Table Table 1 gives the WBS structure for DM in operaitons.

Table 1: WBS elements for Rubin Observatory Data Management Operations

#REF!		

6 Roles and Responabilities

Table 2: Management roles for Rubin Observatory Data Management Operations

WBS	Role Title	Role Description
3.1a	Associate Director for Data Management Operations	The AD of the Data Management Operations Department is one of the principal leaders of the Rubin Observatory operations phase. This position requires a Ph.D. level astronomer with extensive astronomical survey and science management experience, and reports directly to the Rubin Observatory Director. The primary responsibilities of this position include the management of the Data Management Operations Department (including budget and planning for this WBS element), participation in the leadership of Rubin Observatory survey, and coordination with other Rubin Observatory Departments. The AD for Data Management Operations also has overall responsibility and authority for safely running the Rubin Observatory Data Facilities (DF) including the generation of prompt data products (alerts) and the annual data release processing. This person will supervise a technical staff that will be responsible for all aspects of data processing, preparation of data products, archiving, and operation of the Chilean, French and UK DACs/DFs, as well as the US DF. They will be responsible for coordinating with project level Contract Management and Supplier Management when dealing with issues of business impact, and accountable for ensuring a disaster recovery plan is effective and able to be invoked. The AD of Data Management Operations is also responsible for supervising the data flow from the Recinto to the USDF.

3.1b.1	Data Management Operations Deputy Manager	Assist the AD of Data Management Operations in all aspects of the ongoing project. The deputy manager will be enabled to act on behalf of the AD in any situation.
3.1b.2	Data Management Operations Deputy Manager	Assist the AD of Data Management Operations in all aspects of the ongoing project. The deputy manager will be enabled to act on behalf of the AD in any situation.
3.1c.2	Data Management Operations Advisor - US DF	Each Rubin Observatory Data Facility (US, French and UK) has an advisory role to the AD for Data Management Operations in terms of execution across the data facilities. The US DF Advisor role is held by the US DF Lead; this person is responsible for Data Facilities oversight, analgous to the other oversight areas in the department.
3.1c.3	Data Management Operations Advisor - Fr DF	Each Rubin Observatory Data Facility (US, French and UK) has an advisory role to the AD for Data Management Operations in terms of execution across the data facilities.
3.1c.4	Data Management Operations Advisor - UK DF	Each Rubin Observatory Data Facility (US, French and UK) has an advisory role to the AD for Data Management Operations in terms of execution across the data facilities.
3.1d.1	Data Abstraction Oversight	Oversees the Pipeline Middleware and Data Engineering Teams, functionally supervising the leaders of those teams and reporting on their activity to the AD for RDM.
3.1d.2	Data Production Oversight	Oversees the Algorithms and Pipelines and Campaign Management Teams, functionally supervising the leaders of those teams and reporting on their activity to the AD for RDM.
3.1d.3	Data Services Oversight	Oversees the SQuaRE and Advanced Databases Teams, functionally supervising the leaders of those teams and reporting on their activity to the AD for RDM.
3.1d.4	Data Acquisition Oversight	Oversees the SQuaRE and Advanced Databases Teams, functionally supervising the leaders of those teams and reporting on their activity to the AD for RDM.

6.1 Data Serving Roles

Table 3: SQuARE roles for Rubin Observatory Data Management Operations

WBS	Role Title	Role Description
3.10a	Technical Lead/Manager	Responsible for technical leadership and management of the Service Quality and Reliability Engineering Team. This includes running stand ups and looking after budgets and staff issues as well as making technical calls where decisions are needed. The technical lead is responsible overall for the architecture of the team's services and for ensuring it is fit for purpose for the observatory's evolving needs. They also represent the team's work both inside the organization and to the astronomical and computing technical community.
3.10b.1	Senior Full Stack and Documentation Engineer - NOIRLab	A Senior Full-Stack Engineer is a DevOps Engineer who is versed in architecture and implementation of both backend and frontend architectures. They are able to synthesize requirements, propose system architectures and independently implement services. They have advanced skills in DevOps engineering, including Continuous Deployment and Infrastructure As Code, and must also be able to improve the infrastructure and debug problems which can span hardware, network and operating system all the way to the end user delivered service. A Documentation Engineer is well versed in architecting effective ways to support documentation oriented workflows such as documentation continuous delivery systems and integration with communication platforms. They have advanced skills in service implementation and deployment, familiarity with documentation infrastructure libraries such as Sphinx, and best practices in software documentation, including code-level, package-level and user guides.

3.10b.2	Security Architect / Senior DevOps Engineer - NOIRLab	A Security Architect is a software engineer with advanced expertise in designing and building software services with particular regard to security concerns such as authentication & authorisation, hardening, auditability, penetration testing and dependency management. A Senior Devops Engineer has advanced skills in DevOps engineering, including Software Engineering, Continuous Deployment and Infrastructure As Code, and must also be able to improve the infrastructure and debug problems which can span hardware, network and operating system all the way to the end user delivered service.
3.10c.1	Front End Engineer / Senior Front End Engineer - NOIRLab	Front-end engineers with strong javascript skills, web services architecture, user interface design and astronomical search and visualisation to focus on the Science Platform portal. At least one of these FTEs needs to be at IPAC at a senior engineer level to reflect our current technical investment in Firefly and coordinate effort appropriately in conjunction with other IPAC Firefly development. An appropriately skilled FTE may be found elsewhere or may be added to an IPAC subcontract if available
3.10c.2	Front End Engineer / Senior Front End Engineer - IPAC	Front-end engineers with strong javascript skills, web services architecture, user interface design and astronomical search and visualisation to focus on the Science Platform portal. At least one of these FTEs needs to be at IPAC at a senior engineer level to reflect our current technical investment in Firefly and coordinate effort appropriately in conjunction with other IPAC Firefly development. An appropriately skilled FTE may be found elsewhere or may be added to an IPAC subcontract if available

3.10d.1	Science Platform & Production Services Engineer	A Science Platform and Production Services Engineer is responsible for identifying and resolving issues with production services, including user-facing services such as the Science Platform as well as facility services (such as deployments of the Science Platform at the telescope summit). They are also responsible for evolving services on the basis of rapidly changing user demand, for example the JupyterLab-based component of the Science Platform. They have strong software engineering skills including coding, testing and service deployment and can perform technical work in an independent manner.
3.10d.2	Services Architect / Senior DevOps Engineer	A Services Architect is a software engineer with advanced expertise in designing and building software services with particular regard to scalability, performance, deployment infrastructure and interface design. They are able to synthesize requirements, propose system architectures and independently implement services. A Senior Devops Engineer has advanced skills in DevOps engineering, including Software Engineering, Continuous Deployment and Infrastructure As Code, and must also be able to improve the infrastructure and debug problems which can span hardware, network and operating system all the way to the end user delivered service.

3.10e.1	Senior Scientific Programmer / Data Exploration Specialist	A Senior Scientific Programmer has deep domain expertise and is adept in capturing and implementing data analysis solutions. Their scientific background allows them to anticipate user data exploration needs and other requirements and to effectively communicate them to other engineers on their team. A Data Exploration Specialist communicates data exploration needs to the engineers, documents and develops tools, demonstrates how to achieve scientific goals with the tools provided. This would explicitly include technical consultations to the Community Science Team and EPO scientists. Their scientific background allows them to translate the technical needs of the users of services (external and internal) into an engineering request, as well as functioning as liaison to other interrelated systems (in particular Middleware and Infrastructure) to which Science Platform services have a dependence.
3.10e.2	Senior Devops Engineer / Data Engineer - NOIRLab	A Senior Devops Engineer has advanced skills in DevOps engineering, including Software Engineering, Continuous Deployment and Infrastructure As Code, and must also be able to improve the infrastructure and debug problems which can span hardware, network and operating system all the way to the end user delivered service. A Data Engineer is a software engineer with the ability to model and abstract data flow and representation, as well as design and implement appropriate systems for the effective curation and visualisation of particular kinds of data (eg. time series). A Data Exploration Specialist communicates data exploration needs to the engineers, documents and develops tools, demonstrates how to achieve scientific goals with the tools provided. This would explicitly include technical consultations to the Community Science Team and EPO scientists. Their scientific background allows them to translate the technical needs of the users of services (external and internal) into an engineering request, as well as functioning as liaison to other interrelated systems (in particular Middleware and Infrastructure) to which Science Platform services have a dependence.

3.10e.3	Senior Devops Engineer / Data Engineer - SLAC	A Senior Devops Engineer has advanced skills in DevOps engineering, including Software Engineering, Continuous Deployment and Infrastructure As Code, and must also be able to improve the infrastructure and debug problems which can span hardware, network and operating system all the way to the end user delivered service. A Data Engineer is a software engineer with the ability to model and abstract data flow and representation, as well as design and implement appropriate systems for the effective curation and visualisation of particular kinds of data (eg. time series). A Data Exploration Specialist communicates data exploration needs to the engineers, documents and develops tools, demonstrates how to achieve scientific goals with the tools provided. This would explicitly include technical consultations to the Community Science Team and EPO scientists. Their scientific background allows them to translate the technical needs of the users of services (external and internal) into an engineering request, as well as functioning as liaison to other interrelated systems (in particular Middleware and Infrastructure) to which Science Platform services have a dependence.

Table 4: Qserv roles for Rubin Observatory Data Management Operations

WBS	Role Title	Role Description
3.11a	Qserv Lead	To be written!
3.11b	Database Engineer (Qserv) - SLAC	Develops, maintains, and implements the science Databases e.g. QSERV database, data butler, Prompt Products Database. May also work on other middleware as needed.

3.11c	Dev/Ops Software Engineer - IN2P3	Develops, maintains, and implements DF software, including: QSERV database, data butler, DAX, Alert Filtering Service, orchestration software, workflow software, data backbone software, integration testing framework, authentication services, pipeline construction tools, operational fabric codes, logging, messaging, monitoring and health and status software, hosting environment for Rubin Observatory Data Space, Data Space batching services, and bulk export to other sites.

6.2 Data Production Roles

Table 5: Data Production roles for Rubin Observatory Data Management Operations

WBS	Role Title	Role Description
3.6a	Pipeline Middleware Lead	Organizes the software maintenance effort and assigns work in a way that provides for continuity of maintenance for all Rubin Observatory maintained software. Is primarily responsible for further defining and enforcing software engineering rules related to maintenance, including maintenance of documentation, correct security practices, testing, and other aspects of delivery of a complete change set. Ensures that software tasks are consistent with authorized changes. Carries share of maintenance load. Participates in reviews.
3.6b.1	Pipeline Middleware Engineer - NOIRLab	Develops, maintains, and implements pipeline and workflow software, including: Data Butler, orchestration software, workflow/-workload software, integration testing framework, pipeline construction tools, and pipeline infrastructure libraries.
3.6b.2	Pipeline Middleware Engineer - SLAC	Develops, maintains, and implements pipeline and workflow software, including: Data Butler, orchestration software, workflow/-workload software, integration testing framework, pipeline construction tools, and pipeline infrastructure libraries.

Table 6: Algorithm and Pipeline roles for Rubin Observatory Data Management Operations

#REF!		

Table 7: Campaign Management roles for Rubin Observatory Data Management Operations

WBS	Role Title	Role Description
3.9a	Campaign Management Lead	Lead the Campaign Management Team. Coordinate scheduling and training. Perform day to day technical operation of data processing campaigns, including Alert Production, Data Release Production, Calibration, and other campaigns. Diagnose and address processing deficiencies, bottlenecks, and failure modes. Coordinate with the V&V team in testing and implementing the campaign strategies and finalizing the campaign configurations. Coordinate with the Workflow/load Management Engineer and the US DF Infrastructure Group in troubleshooting infrastructure failures. This may require 7 days per week coverage.
3.9b.2	Processing Scientist - SLAC	Perform day to day technical operation of data processing campaigns, including Alert Production, Data Release Production, Calibration, and other campaigns. Diagnose and address processing deficiencies, bottlenecks, and failure modes. Coordinate with the V&V team in testing and implementing the campaign strategies and finalizing the campaign configurations. Coordinate with the Workflow/load Management Engineer and the Infrastructure Group in troubleshooting infrastructure failures. This may require 7 days per week coverage.

3.9b.1	Processing Scientist - NOIRLab/UW	Perform day to day technical operation of data processing campaigns, including Alert Production, Data Release Production, Calibration, and other campaigns. Diagnose and address processing deficiencies, bottlenecks, and failure modes. Coordinate with the V&V team in testing and implementing the campaign strategies and finalizing the campaign configurations. Coordinate with the Workflow/load Management Engineer and the Infrastructure Group in troubleshooting infrastructure failures. This may require 7 days per week coverage.
3.9b.5	Processing Scientist - NOIRLab/NCSCA	Perform day to day technical operation of data processing campaigns, including Alert Production, Data Release Production, Calibration, and other campaigns. Diagnose and address processing deficiencies, bottlenecks, and failure modes. Coordinate with the V&V team in testing and implementing the campaign strategies and finalizing the campaign configurations. Coordinate with the Workflow/load Management Engineer and the Infrastructure Group in troubleshooting infrastructure failures. This may require 7 days per week coverage.
3.9b.3	Processing Scientist - IN2P3	Perform day to day technical operation of data processing campaigns, including Alert Production, Data Release Production, Calibration, and other campaigns. Diagnose and address processing deficiencies, bottlenecks, and failure modes. Coordinate with the V&V team in testing and implementing the campaign strategies and finalizing the campaign configurations. Coordinate with the Workflow/load Management Engineer and the Infrastructure Group in troubleshooting infrastructure failures. This may require 7 days per week coverage.

3.9b.4	Processing Scientist - UK	Perform day to day technical operation of data processing campaigns, including Alert Production, Data Release Production, Calibration, and other campaigns. Diagnose and address processing deficiencies, bottlenecks, and failure modes. Coordinate with the V&V team in testing and implementing the campaign strategies and finalizing the campaign configurations. Coordinate with the Workflow/load Management Engineer and the Infrastructure Group in troubleshooting infrastructure failures. This may require 7 days per week coverage.
3.9c.2	Campaign Management Developer - SLAC	To be written!

6.3 Data Abstraction Roles

Table 8: Data Engineering roles for Rubin Observatory Data Management Operations

WBS	Role Title	Role Description
3.7a	Data Scientist	Hold project history on data models etc and guide VO developments
3.7b.2	Data Engineer - SLAC	Develops, maintains, and implements pipeline and workflow software, including: Data Butler, orchestration software, workflow/-workload software, integration testing framework, pipeline construction tools, and pipeline infrastructure libraries.
3.7c.1	Build Engineer	Maintain and improve the builds for the various Data Management Operations software products. This includes fixing the continuous integration builds and improving the packaging system (currently aiming for Conda but could potentially move to another system over ten years). With hundreds of packages this is not a trivial task. This role would also support the observatory software builds which could potentially remain different to the Data Management Operations builds.

3.7c.2	Build Engineer	Maintain and improve the builds for the various Data Management Operations software products. This includes fixing the continuous integration builds and improving the packaging system (currently aiming for Conda but could potentially move to another system over ten years). With hundreds of packages this is not a trivial task. This role would also support the observatory software builds which could potentially remain different to the Data Management Operations builds.

Table 9: Pipeline Middleware roles for Rubin Observatory Data Management Operations

WBS	Role Title	Role Description
3.6a	Pipeline Middleware Lead	Organizes the software maintenance effort and assigns work in a way that provides for continuity of maintenance for all Rubin Observatory maintained software. Is primarily responsible for further defining and enforcing software engineering rules related to maintenance, including maintenance of documentation, correct security practices, testing, and other aspects of delivery of a complete change set. Ensures that software tasks are consistent with authorized changes. Carries share of maintenance load. Participates in reviews.
3.6b.1	Pipeline Middleware Engineer - NOIRLab	Develops, maintains, and implements pipeline and workflow software, including: Data Butler, orchestration software, workflow/-workload software, integration testing framework, pipeline construction tools, and pipeline infrastructure libraries.
3.6b.2	Pipeline Middleware Engineer - SLAC	Develops, maintains, and implements pipeline and workflow software, including: Data Butler, orchestration software, workflow/-workload software, integration testing framework, pipeline construction tools, and pipeline infrastructure libraries.
3.6b.3	Pipeline Middleware Engineer - NOIRLab/NCSA	Develops, maintains, and implements pipeline and workflow software, including: Data Butler, orchestration software, workflow/-workload software, integration testing framework, pipeline construction tools, and pipeline infrastructure libraries.

3.6b.4	Pipeline Middleware Engineer - NOIR-Lab/Princeton	Develops, maintains, and implements pipeline and workflow software, including: Data Butler, orchestration software, workflow/-workload software, integration testing framework, pipeline construction tools, and pipeline infrastructure libraries.

6.4 Data Facility Roles

Table 10: Chile Summit and Base Facility roles for Rubin Observatory Data Management Operations

WBS	Role Title	Role Description
3.5a	Chilean DevOps Technical Lead/Manager	Responsible for technical leadership and management of the Chilean DevOps Support Team, including planning and overseeing the installation, operation and maintenance of all Chilean computing hardware. This includes all summit hardware, networks and cabling. The summit systems are run as infrastructure through service deployment; the Chilean devops team must maintain a deployment ready layer (foreman, puppet, kubernetes) for software teams to work with. This also includes maintaining the camera test stands in Chile and in Tucson in a similar as that for the summit. Interface with NOIRLab COS-IT to coordinate delivered IT services such as helpdesk and network support. Work with the I&S Team's DF/DACs Technical Coordinator to build the Chilean DAC into the Rubin DAC Network.
3.5b.1	Chilean DevOps Engineer	Responsible for the installation, operation and maintenance of all Chilean and test stand (Tucson) computing hardware. This includes guidance on technical decisions for technology both in terms of hardware and software. This may require negotiation/persuasion of COS/IT to support desired Rubin technology in the future and open mindedness to offered solutions from NOIRlab. Training and inclusion of less experienced staff is an essential part of this role.

3.5b.2	Surge Chilean DevOps Engineer Cover	Responsible for the installation, operation and maintenance of all Chilean and test stand (Tucson) computing hardware. This includes guidance on technical decisions for technology both in terms of hardware and software. This may require negotiation/persuasion of COS/IT to support desired Rubin technology in the future and open mindedness to offered solutions from NOIRlab. Training and inclusion of less experienced staff is an essential part of this role.

Table 11: USDF roles for Rubin Observatory Data Management Operations

WBS	Role Title	Role Description
3.2a	US Data Facility Lead	Provides management and oversight for the US Data Facility. Provides for evolution of service architecture and supporting ITC. Responsible for design and evolution of the US DF and its interactions with the French and UK Data Facilities, and the Chilean and other DACs.
3.2b.2	US DF Technical Lead - SLAC	Responsible for leading the Data, Compute and IT Security team involved in providing foundational services for file-based data, data resident in database engines, and facility-wide services including AAA and operational network security infrastructure.
3.2b.1	US DF Technical Lead - NOIRLab	Responsible for leading the Data, Compute and IT Security team involved in providing foundational services for file-based data, data resident in database engines, and facility-wide services including AAA and operational network security infrastructure.
3.2c	DF/IDACs Technical Coordinator	Coordinate technical needs among the collection of Data Facilities and IDACS. Respond to issues that come up day to day in connections, transfers etc.

3.2d	Data Movement Engineer	Maintain Rucio system which will be involved in the tracking and moving of data between multiple sites,including hosting environment for Rubin Observatory Data Space, Data Space batching services, and bulk export to other sites. This is in close conjunction with the Storage Engineers in the Data Facilities. Rucio is an open source HEP product which we have adopted on Rubin Observatory.
3.2e.3	Data Wrangler - IN2P3	The data wrangler ensures that data (science raw data, calibration data, data products, etc.) is replicated at IN2P3 and data products resulting from the local processing performed at IN2P3 are replicated to the USDF They also ensure that the data archived at IN2P3 and needed for the annual processing are recalled from tape on time for the image processing tasks to be performed. They ensure that the tools and systems used for replicating data at IN2P3 are operational. This role needs tight coordination with the team at the USDF that is responsible for data distribution.
3.2e.4	Data Wrangler - UK	The UK data wrangler ensures that data (science raw data, calibration data, data products, etc.) is replicated and data products resulting from the local processing performed in the UK are replicated to the USDF They also ensure that the data archived in the UK and needed for the annual processing are recalled from tape on time for the image processing tasks to be performed. They ensure that the tools and systems used for replicating data in the UK are operational. This role needs tight coordination with the team at the USDF that is responsible for data distribution.

3.2f.2	Database Administrator - SLAC	Administers QSERV, reformatted EFD, and other databases at the US Data Facility and Chilean Base Center. Makes sure the databases are backed up, properly interfaced to the authentication infrastructure. Interfaces with the Security Administrator to ensure access by only authorized users (confidentiality, integrity, and availability). Additional database responsibilities for supporting misc workflows and processes within the Data Facilities. The Database Administrator is also responsible for ensuring data transfers to disaster recovery stores are implemented and functioning according to policy. This assumes significant value engineering is realized by collecting miscellaneous databases into enterprise-level database, with potential additional savings by reusing database infrastructure for other database needs elsewhere in the project. Additional .5-1 FTEs otherwise.
3.2f.1	Database Administrator - NOIRLab	Administers QSERV, reformatted EFD, and other databases at the US Data Facility and Chilean Base Center. Makes sure the databases are backed up, properly interfaced to the authentication infrastructure. Interfaces with the Security Administrator to ensure access by only authorized users (confidentiality, integrity, and availability). Additional database responsibilities for supporting misc workflows and processes within the Data Facilities. The Database Administrator is also responsible for ensuring data transfers to disaster recovery stores are implemented and functioning according to policy. This assumes significant value engineering is realized by collecting miscellaneous databases into enterprise-level database, with potential additional savings by reusing database infrastructure for other database needs elsewhere in the project. Additional .5-1 FTEs otherwise.

3.2g.3	Catalog Manager - IN2P3	The catalog manager ensures the day-to-day operations of the astronomical catalog database at IN2P3. This includes ingesting new data and removing and archiving obsolete catalogs. They also interact with the data wrangler to ensure that the catalog data produced at other sites are imported and ingested into the IN2P3 catalog and that the catalog data produced at IN2P3 is ingested into the local catalog and replicated to other sites. They also ensure that the software releases for the catalog database are compatible with those releases used at other sites operating a catalog database, in particular, the USDF.
3.2g.4	Catalog Manager - UK	The catalog manager ensures the day-to-day operations of the astronomical catalog database at IN2P3. This includes ingesting new data and removing and archiving obsolete catalogs. They also interact with the data wrangler to ensure that the catalog data produced at other sites are imported and ingested into the IN2P3 catalog and that the catalog data produced at IN2P3 is ingested into the local catalog and replicated to other sites. They also ensure that the software releases for the catalog database are compatible with those releases used at other sites operating a catalog database, in particular, the USDF.
3.2h	US DF Infrastructure Group Leader	Lead US DF Software Developer, coordinating activities of the US DF Infrastructure Group as well as participating in group's technical activity.
3.2i	US DF Network & Core Services Engineer - SLAC	Provides network hardware and operational functionality used from a site's border router to Rubin Observatory end equipment. Collaborates with the security engineer and also IT services related to dynamic reallocation of US DF enclaves to support these functions with network features. Supplies higher-level network services as needed at each site, such as DNS, NTP, domain name registrations, netflows, and support for security. Provides Rubin S3DF enclave core services, including login services, name services, and environment maintenance.

3.2j	US DF Software Developer - SLAC	General and broad computing skills, responding to the myriad of issues/needs that will come up - including via user support tickets. Provides workload management (eg kubernetes) of the clusters. Other skills to be drawn from python coding, web application development and deployment, continuous integration tooling (eg Jenkins), database expertise, code and I/O optimization. Some of the work involves maintaining and improving the builds for the various Data Management Operations software products. This includes fixing the continuous integration builds and improving the packaging system (currently aiming for Conda but could potentially move to another system over ten years). This part of the role would also support the observatory software builds which could potentially remain different to the Data Management Operations builds.
3.2j.1	Surge US DF Software Developer - SLAC	General and broad computing skills, responding to the myriad of issues/needs that will come up - including via user support tickets. Provides workload management (eg kubernetes) of the clusters. Other skills to be drawn from python coding, web application development and deployment, continuous integration tooling (eg Jenkins), database expertise, code and I/O optimization. Some of the work involves maintaining and improving the builds for the various Data Management Operations software products. This includes fixing the continuous integration builds and improving the packaging system (currently aiming for Conda but could potentially move to another system over ten years). This part of the role would also support the observatory software builds which could potentially remain different to the Data Management Operations builds.

3.2j.2	US DF Software Developer - AURA/NCSA	General and broad computing skills, responding to the myriad of issues/needs that will come up - including via user support tickets. Provides workload management (eg kubernetes) of the clusters. Other skills to be drawn from python coding, web application development and deployment, continuous integration tooling (eg Jenkins), database expertise, code and I/O optimization. Some of the work involves maintaining and improving the builds for the various Data Management Operations software products. This includes fixing the continuous integration builds and improving the packaging system (currently aiming for Conda but could potentially move to another system over ten years). This part of the role would also support the observatory software builds which could potentially remain different to the Data Management Operations builds.
3.2k	Workflow/Load Management Engineer	Maintain smooth operation and configuration of workflow/workload tools for pipeline operation.
3.2l.3	Image Handler - IN2P3	The image handler ensures the image processing stages assigned to IN2P3 are performed on time. This person also ensures that the processing (e.g. software releases, configuration files, etc.) is compatible with what is agreed upon with the other processing sites, in particular with the USDF. They also ensure that IN2P3's image processing infrastructure (batch processing, workflow management system, etc.) is operational and correctly configured for LSST needs. They also ensure the day-to-day operations of the annual image processing campaign. This role needs tight coordination with the team in charge of image processing at the USDF.

3.2l.4	Image Handler - UK	The image handler ensures the image processing stages assigned to the UK are performed on time. This person also ensures that the processing (e.g. software releases, configuration files, etc.) is compatible with what is agreed upon with the other processing sites, in particular with the USDF. They also ensure that the UK's image processing infrastructure (batch processing, workflow management system, etc.) is operational and correctly configured for LSST needs. They also ensure the day-to-day operations of the annual image processing campaign. This role needs tight coordination with the team in charge of image processing at the USDF.
3.2m.2	US DF Network & Core Services Engineer - NOIRLab	Provides network hardware and operational functionality used from a site's border router to Rubin Observatory end equipment. Collaborates with the security engineer and also IT services related to dynamic reallocation of US DF enclaves to support these functions with network features. Supplies higher-level network services as needed at each site, such as DNS, NTP, domain name registrations, netflows, and support for security. Provides Rubin S3DF enclave core services, including login services, name services, and environment maintenance.
3.2n	US DF Technician	Provides technical support for the Rubin US DF hardware at SLAC. Responsible for physical installation tasks: server racking, cabling and power. Lifecycle tasks include maintenance, troubleshooting and repair. Responsibilities may extend to monitoring and updating firmware.
3.2o	Wide Area Network Technical Manager	Responsible for providing coordination amongst and managing relationships with the four independent WAN operators. Acts as the interface for services provided to the US DF in the context of the WAN. Responsible for managing the risk associated with each WAN operator, including developing mitigation strategies and proposed project responses to credible risks. Leads the Joint Wide Area Network Working Group. Well connected to DOE ESNet.

3.2p	Wide Area Network Architect	Familiar with WAN implementation technologies generally available in the networks supporting the Rubin Observatory. Familiar with technology roadmap of the ESNet WAN provider. Synthesizes and evolves network techniques and provisioning supporting the Rubin Observatory mission, as network technology evolves. Drawn from staff of WAN groups but explicitly supported by and work in the context of the Rubin Observatory.

Table 12: FRDF roles for Rubin Observatory Data Management Operations

WBS	Role Title	Role Description

Table 13: UK roles for Rubin Observatory Data Management Operations

WBS	Role Title	Role Description

6.5 Support model

DM strives to have automated and reliable systems. We are not sized for twenty four seven support but some of us are available on call as a best efforts basis. The assumption is after the initial set up calls would be infrequent (order less than once a month) for problems. Appropriate numbers are lodged in 1Password and available to summit observers. We will also set up some Squadcast alerts for specific things.

6.5.1 Alerts

Alerts must be kept running and hence we do need weekend on call support for to ensure we do not have consecutive days without alerts.

6.5.2 Potential Outline

If we do on call on best effort basis we should agree on some ground rules. Institutionally it may be difficult to formalize these . These might be:

- Should an individual get called out of hours they would not need to be in any meetings or workplace the following morning.
- Any hour worked out of hours may be taken as 1.5 hours leave at another point in time.
- If requested specific mobile phones for on call should be provided (may be hard given our distributed nature).

7 Products

Product	Manager	Owner	Notes
DM Ops	Wil O'Mullane		Data Management (Ops)
Data Abstraction	Tim Jenness		Data Abstraction
Build Engineering	Tim Jenness		Build Engineering
Data Eng	Gregory Dubois Felsmann		Data Engineering
Felis	TBD		Felis
Metadata	TBD		Metadata
Pipe Middleware	Tim Jenness		Pipelline Middleware
Sis Exec	TBD		Single-site Exec
OCPS	KT Lim ?		Observatory Controlled Processing System
Prompt f/ w	TBD		Prompt forwarder
s3daemon	TBD		s3 object transfer daemon
BPS	TBD		BPS
Butler	Tim Jenness	Y (Jlm B)	Butler
Control Interface	TBD		Control Interface
ctrl_bps	TBD		ctrl_bps
ctrl_mpexec	TBD		ctrl_mpexec
user batch envelope	TBD		user batch envelope
Pipeline interfaces	TBD		Pipeline interfaces
pex_config	TBD		pex_config
pipe_base	TBD		pipe_base
EUPS	TBD		EUPS and EUPS.lsst.code
Data Acquisition LHN	Cristian Silva		Data Acquisition and Lonf Haul Networks
Chile DevOps	Cristian Silva		Chile DevOps
Data Facilities	Richard Dubois		Data Facilities
Data Curation	TBD		Data Curation
Data Backbone	TBD		Data Backbone
Backups	TBD		Backups
Bulk Download	TBD		Bulk Download
Consolidated DB	TBD		Consolidated DB
Butler repos	TBD		Butler repos
Rucio	TBD		Rucio
OODS	Steve Pietrowicz		Observstory Operations Data System
Infrastructures	TBD		Infrastructures
CDAC	Frossie Economou		Chile Data Access Center
FrDF	Fabio Hernandez		French Data Facility

SDF	Cristián Silva		Summit and Base Data Facility
UKDF	George Beckett		UK Data Facility
USDF	TBD		US Data Facility
Multi-site & User Exec	TBD		Multi-site & User Exec
PanDA	TBD		PanDA
User Batch	TBD		User Batch
UWS	KT lim		Universal Worker Service
Data Production	Yusra AlSayyad		Data Production
Campaign Management	?	Y (N/ A)	Campaign Management
Algorithms & Pipelines	Yusra AlSayyad	? (Jim Bosch)	Algorithms & Pipelines:In ops our construction POs (JimB+EricB) become our group leads, so PO prob not necessary.
Data Services	Frossie Economou		Data Services
Complex. DB	Fritz Mueller	Y (Colin)	Complex Database Support:Should this be complex Databases or somethign ? Qserv under
Big Databases	Fritz Mueller		Big Databases
PromptDV	TBD		Prompt Products DB
Qserv	Fritz Mueller		Qserv
User Databases	TBD		User Databases
SQuaRE	TBD		SQuaRE
Doc Services	TBD		Doc Services
Documentation standards	TBD		Documentation standards
LtD	TBD		LtD
Templating	TBD		Templating
Phalanx	TBD		Phalanx
Authorisation	TBD		Authorisation
Reliability Engineering	TBD		Reliability Engineering
Secrets	TBD		Secrets
Planned Obs.	Willam O'Mullane		Planned Observation Publication
RSP	TBD		Rubin Science Platform
APIs		Y (GPDF)	APIs:IVOA and non-VO Apis
data.lsst.cloud	TBD		data.lsst.cloud
Authentication			Authentication:and security engineering
Notebook		Y (KSK)	Notebook
Portal		Y (GPDF)	RSP Portal
Square One	TBD		Square One
User Support			User Support:clo service and helpdesk
Sasquatch	TBD		Sasquatch
EFD	TBD		EFD
Metrics	TBD		Metrics
Telemetry Gateway	TBD		Telemetry Gateway

A References

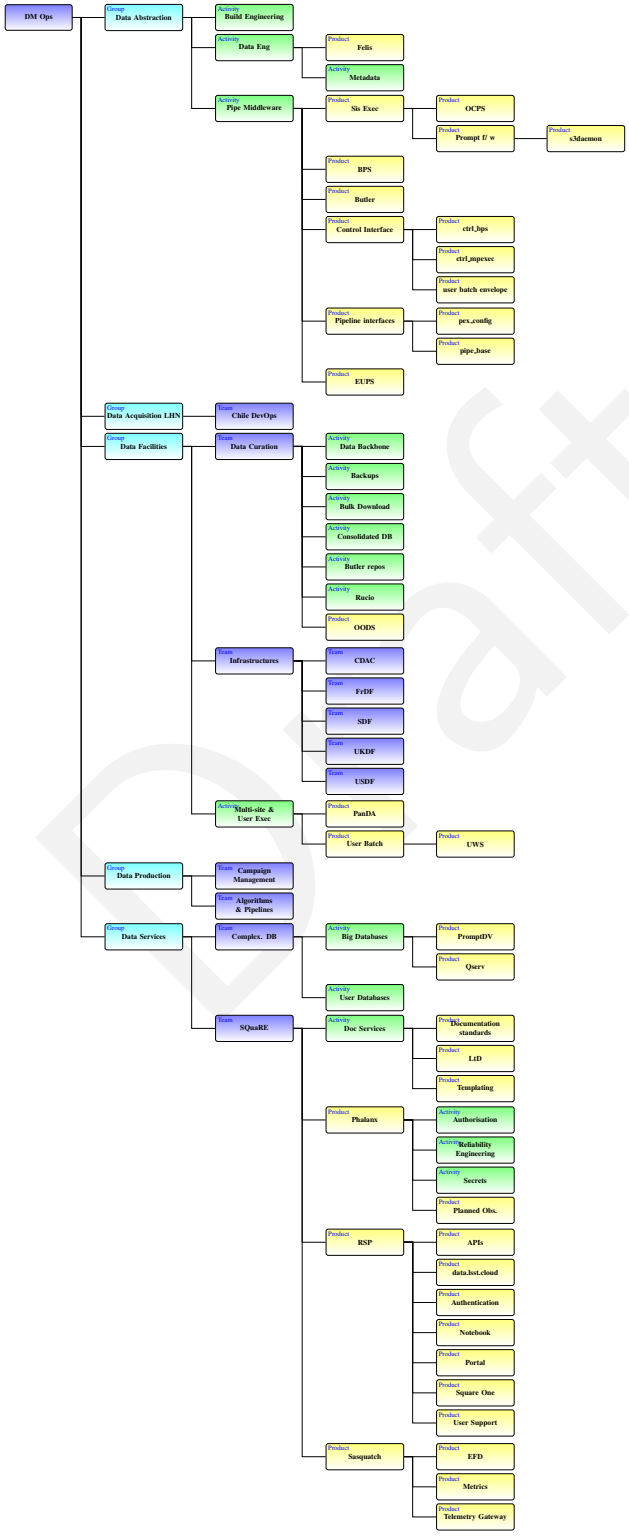

Figure 4: DM operations product tree

[RTN-021], Dubois, R., O'Mullane, W., 2022, Data Facilities Transition Plan, URL https://rtn-021.lsst.io/,
Vera C. Rubin Observatory Technical Note RTN-021

[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

[DMTN-189], Lim, K.T., 2021, Data Facility Specifications, URL https://dmtn-189.lsst.io/,
Vera C. Rubin Observatory Data Management Technical Note DMTN-189

[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

[LDM-572], O'Mullane, W., 2021, Chilean Data Access Center, URL https://ldm-572.lsst.io/,
Vera C. Rubin Observatory Data Management Controlled Document LDM-572

[DMTN-223], O'Mullane, W., 2023, User batch - possibilities and plans., URL https://dmtn-223.lsst.io/,
Vera C. Rubin Observatory Data Management Technical Note DMTN-223

[DMTN-209], O'Mullane, W., Economou, F., Huang, F., et al., 2021, Rubin Science Platform on Google: the story so far., URL https://dmtn-209.lsst.io/,
Vera C. Rubin Observatory Data Management Technical Note DMTN-209

[RTN-030], O'Mullane, W., Allbery, R., Dubois, R., Lim, K., 2023, Rubin Data and Information Security Plan, URL https://rtn-030.lsst.io/,
Vera C. Rubin Observatory Technical Note RTN-030

[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

[RTN-005], O'Mullane, W., Bauer, A., Blum, R., et al., 2024, Rubin Operations Work Management and Budget Planning, URL https://rtn-005.lsst.io/,
Vera C. Rubin Observatory Technical Note RTN-005

[RDO-71], Roberts, A., Blum, R., Claver, C., et al., 2024, Rubin Observatory Risk and Opportunity Management Plan, URL https://rdo-71.lsst.io/, Vera C. Rubin Observatory RDO-71

[ITTN-057], Silva, C., 2022, Disaster Recovery - Computing, URL https://ittn-057.lsst.io/, Vera C. Rubin Observatory ITTN-057

[ITTN-058], Silva, C., 2022, Disaster Recovery - Infrastructure Support Devices, URL https://ittn-058.lsst.io/, Vera C. Rubin Observatory ITTN-058

[ITTN-056], Silva, C., 2023, Disaster Recovery - Network, URL https://ittn-056.lsst.io/, Vera C. Rubin Observatory ITTN-056

B Acronyms

Acronym	Description
B	Byte (8 bit)
BPS	Batch Production Service
CDF	Cumulative Distribution Function
DAC	Data Access Center
DAQ	Data Acquisition System
DB	DataBase
DF	Data Facility
DM	Data Management
DMTN	DM Technical Note
DP	Data Production
DRP	Data Release Production
DWDM	Dense Wave Division Multiplex
EFD	Engineering and Facility Database
EUPS	Extended Unix Product System
FITS	Flexible Image Transport System
FRDF	French Data Facility
FTE	Full-Time Equivalent
FrDF	French Data Facility

IT	Information Technology
ITTN	IT Technote
IVOA	International Virtual-Observatory Alliance
LDM	LSST Data Management (Document Handle)
LHN	long haul network
LSE	LSST Systems Engineering (Document Handle)
LSST	Legacy Survey of Space and Time (formerly Large Synoptic Survey Telescope)
NOIRLab	NSF's National Optical-Infrared Astronomy Research Laboratory; `https://noirlab.edu`
OCPS	OCS Controlled Pipeline System
OODS	Observatory Operations Data Service
OPS	Operations
PO	Program Operations
PanDA	Production ANd Distributed Analysis system
RSP	Rubin Science Platform
RTN	Rubin Technical Note
SQuaRE	Science Quality and Reliability Engineering
TBD	To Be Defined (Determined)
UK	United Kingdom
UKDF	United Kingdom Data Facility
US	United States
USDF	United States Data Facility
UWS	Universal Worker Service (IVOA standard)
VO	Virtual Observatory
VRO	(not to be used)Vera C. Rubin Observatory
WBS	Work Breakdown Structure
bps	bit(s) per second