Vera C. Rubin Observatory
Data Management Monthly Report
May 2020

High-level Summary

Community Interactions, Meetings and Workshops

The DM Leadership Team meeting was held May 12 through 14. This was a virtual meeting, but nevertheless proved very productive. Discussions on a variety of key topics, including middleware development; time-series interfaces; plans for the Data Facility; rebaselining and schedule; and the design of the prompt processing system.

In addition, several members of the team participated in the [IVOA 2020 Virtual Interoperability Meeting](#), May 4–8. This meeting included important discussion about standardized representation for world coordinate system (WCS) information and the development of science platforms, as well as a number of other topics relevant to Rubin development.

Preparations are being made for June 15–17 Pre-Verification Review of Networks and our second Operations Rehearsal, June 2–4.

Technical Progress

Various documents were released or updated:

- [DMTN-148](#), was made available in draft form. This document defines the overall architecture and strategy for management of calibration products in the construction, commissioning, and operational eras.
- [DMTN-150](#), which describes the Google Cloud Proof of Concept, was released. Work began on data transfer, Data Release Production execution, and Alert Production execution in the cloud.
- [LDM-732](#), the Rubin Observatory Network Verification Baseline, was published.
- [SQR-039](#) was published. This document proposes improvements to the Science Platform Authentication and Authorization service. These are of immediate relevance to the deployment of the Intermediate Data Facility.

- [LCR-2313](), to baseline updates to [LSE-78]() — the Rubin Observatory Network Design — was submitted.

The DM team achieved a significant milestone this month with the generation of the first alert packets by the Alert Production system. This represents a major new capability for the DM system. Currently, these alerts are not being available by the real time Alert Distribution system which will be used in the operational era. However, sample alerts based on precursor data will be made available to the community for bulk download, in particular with a view to assisting potential authors of community brokers.

A number of other key improvements were made to the Science Pipelines. Particular highlights include implementing the first version of "sky sources" — corresponding to measurements made on empty patches of sky — in single frame processing. This provides an important QA tool. The team also provided driver scripts for processing with simulated ("fake") sources inserted at large scales.

Members of the DM System Science Team, in collaboration with members of the project's Commissioning Team having been making rapid progress on the new scalable system for calculating performance metrics based on the results delivered by those pipelines. This represents a key capability for both ongoing QA of our processing and ultimately for formal verification and validation of the as-delivered system.

The Qserv build process was updated to take advantage of Conda (refer to last month's report for a discussion of the new Conda environment as applied to the Science Pipelines), and improvements and updates forced by upstream changes were made to the Conda environment. The new Qserv parallel data ingest tooling was heavily exercised with both reprocessed Hyper Suprime-Cam and DESC Data Challenge 2 datasets, and a simulated catalog data creation tool is now being integrated to enable the next regime of Qserv scale testing.

New monitoring software for tracking images as they are transferred to the Data Facility at NCSA was deployed. This includes all images from all cameras (both the Auxiliary Telescope and ComCam from the Summit Facility or Base Data Center, and from test environments at NCSA and SLAC). A new software environment, including the ComCam Header Service, Archiver Service, Forwarder, and Observatory Operations Data Service, was deployed at the Base Data Center.

Risk Management

The DM Risk Register was reviewed in the monthly process. No new risks were added and

no significant changes were made to the existing exposure.

Milestone Summary

Milestones Completed

Alert Production

DM-AP-18: Alert packet generation
Due 2020-05-29; completed 2020-05-14
This work was delivered on activity [DM-24324](#).

Milestones Delayed

Data Release Production

DM-DRP-8: Calibration product generation for the Auxiliary Telescope
Due 2020-01-21
The DM team is currently analyzing initial observations performed with the Auxiliary Telescope, including the generation of calibration products. This milestone will be completed when that system achieves a relatively mature capability.

DM-DRP-12: Background estimation over the full visit
Due 2020-05-29
Background estimation improvements are underway on activity [DM-25270](#) during summer 2020, although addressing the full focal plane is not currently seen as the scientifically highest priority.

DM-DRP-13: PSF estimation over the full visit
Due 2020-05-29
Pending the integration of [Piff](#), which is scheduled for summer 2020 on activity [DM-25319](#).

DM-DRP-29: Moving point source model fitting now available
Due 2018-11-30
This milestone was delayed due to staff shortages; it will be reprioritized and addressed during the current calendar year.

DM-DRP-37: Artifact rejection and background matching during coadd construction
Due 2018-11-30
Artifact rejection steps are complete and well tested. Background matching is regarded as a scientifically lower priority, and will be addressed during calendar year 2020.

Science User Interface and Tools

DM-SUIT-5: Search and display processed HSC data

Due 2019-02-28

This is still awaiting the HSC data to be in ObsCore, which is underway (WISE data is already available in ObsCore).

DM-SUIT-10: SUIT deployment procedure

Due 2019-05-31

Still awaiting formal documentation in the shape of [DMTN-136](#).

Science Data Archive & Application Services

DLP-802: Alert Production Database (APDB) Design

Due 2016-11-30

The combination of aggressive performance and science requirements for this database have made a working design at scale more difficult than was originally anticipated. A functional prototype at reduced scale has been produced, and was evolved with feedback from the AP team. Work in search of a full-scale solution continues in F20 with scheduled experiments involving different back-end database technologies.

DM-DAX-5: Database ingest in support of HSC reprocessing (i.e., large catalog ingest)

Due 2019-02-28

Blocking milestone DM-DRP-11 (Support for database ingestion of results) has been completed, and work is now underway. HSC reprocessing runs are now being regularly ingested into the Qserv database using the new tooling. Final work now underway to fully automate this process. Completion anticipated in F20.

LSST Data Facility

DM-NCSA-11: Verified acquisition of raw and crosstalk-corrected exposures at raft scale, incl. correct metadata

Due 2019-07-29

Following LCR-1923, we no longer expect to acquire crosstalk-corrected exposures, and the scope of this milestone is modified accordingly. The milestone will be met after ComCam is available at the Base Data Center and is capable of providing raft-scale raw images, currently expected in June 2020.

DM-NCSA-20: ComCam Archiving Service

Due 2019-09-25

Software to provide this service was deployed to the Base Data Center in May. We expect it to achieve initial operational capability June, at which point this milestone will be completed.

International Communications & Base Site

IT-725-M: Base Data Center Network Complete
Due 2019-07-29
This milestone is still pending the acceptance of the Base Facility (including the Data Center) by COS. The baseline of network and computing infrastructure is complete and only documentation is pending to be completed. The expected completion of documentation is June 2020. The critical path is not impacted.

Detailed Project Progress

1.02C.01: System Management

Current accomplishments

The DM Project Manager chaired the [DM Leadership Team Virtual Face-to-Face Meeting](), May 12–14, and participated in the AURA Management Committee for Rubin meeting on May 27 & 28.

The DM Project Manager also focused on preparations for operations, and, in particular on the Intermediate Data Facility (IDF).

Planned activities

The DM Project Manager will work further on the IDF and start preparing for the Joint Status Review this summer.

1.02C.02.01: Data Management Science

Current accomplishments

The DM Subsystem Scientist:

- Participated in the [DM Leadership Team Virtual Face-to-Face Meeting]().
- Contributed to a memo for NSF on the scientific impact of image delay with the Project Science Team.
- Opened discussions with the DESC on algorithms for building coadds.

- Reviewed DM construction paper outlines with the DM System Science Team, resolving open questions on paper scope and content.
- Produced a plan for follow-up of the Rubin Algorithms Workshop.

Other DM System Science Team (DM-SST) activities include:

- Eric Bellm and Gregory Dubois-Felsmann presented use cases and recommendations for interfaces for querying time series data to the DM Leadership Team.
- Jeff Carlin & Simon Krughoff presented progress on a new scalable framework for calculating performance metrics. Initial results look promising; this study will continue through June.
- Hsin-Fang Chiang completed reprocessing Hyper Suprime-Cam Public Data Release 2 (HSC-PDR2) for test and QA purposes.
- Hsin-Fang Chiang demonstrated a first prototype of a Qserv ingest workflow based on precursor data from Hyper Suprime-Cam.

Planned activities

The DM Subsystem Scientist will:

- Give a presentation at the LSSTC board meeting on the status of DM.
- Give a presentation at the LSST Solar System Readiness (Virtual) Sprint 3 on Year 1 Template Generation and Year 1 Operations.
- Issue an update on the due date for the Rubin Observatory Call for Full Proposals for Community Alert Brokers and the status of sample alerts in consultation with the SAC[1].
- Work on a concrete plan for the evolution of the Data Products Definition Document (DPDD; LSE-163) with Deputy Subsystem Scientist Colin Slater.

Other DM Science Team activities will include:

- Jeff Carlin & Simon Krughoff will work on scaling tests and the implementation of additional performance metrics in the new metric calculation framework
- Jeff Carlin will run acceptance testing of the 20.0.0 release of the science pipelines.
- Melissa Graham will work on an interim plan for community support, bridging the construction and operations project.

---

[1] Science Advisory Committee

- Melissa Graham will work in proposal for updates for Host Galaxy Association for DIAObjects
- Colin Slater will present progress on stellar crowded fields processing

1.02C.02.02: DM System Architecture

Current accomplishments

Architecture

The Architecture team participated in the DM Leadership Team virtual face-to-face meeting, which included the transfer of product ownership of most Prompt Services to Robert Gruendl.

The Google Cloud proof of concept scope document, [DMTN-150](#), was released, and work began on data transfer, Data Release Production execution, and Alert Production execution in the cloud. A prototype harness for simulating image data transfer was written, and Google Transfer Services were ruled out as a data transfer mechanism. Miami was identified as an appropriate interconnect point between the Rubin Observatory network and Google Cloud. A sizing model for cloud deployment of an Interim Data Facility was developed. The draft product tree, [DMTN-140](#), was updated including feedback from management.

The Conda environment was updated to accommodate new Science Pipelines dependencies, emergent incompatibilities, and the Simulations code. Upstream modifications by the conda-forge project forced a rebuild of the environment. The Qserv build process was updated to take advantage of Conda. The Jenkins system was further modified to handle patch releases built using older configurations of the Jenkins system itself, but ultimately changing the entire configuration, including Docker container names and the system/compiler used, without affecting concurrent Jenkins usage proved to be difficult to engineer and the patch release was abandoned. Future patch releases that do not require the use of past build system versions should be possible, and the Architecture team began development of a long-term roadmap to re-engineer the release process. Small bugs were fixed in the lsstsw build system and the newinstall.sh Science Pipelines installation script, which are now maintained by the Architecture Team.

The logging of timing information in Science Pipelines tasks was made more controllable, the severity of log messages about reading non-standard FITS file headers was decreased, and all instances of the Science Pipelines writing non-standard lowercase FITS headers

were fixed. The obsolete obs_lsstSim package was removed from continuous integration; it has been replaced by appropriate code in obs_lsst.

Middleware

- Continued to add new command line utilities for Butler repository manipulation. These included 2 to 3 conversion, import of exported data, visit definition, ingest of raw data. We have also continued to iterate on the unification of the command line options based on user feedback.
- We continued to focus on registry stability and performance and continued to test with the PostgreSQL and Oracle database management systems.
- We demonstrated that we can run the ci_hsc integration test using PostgreSQL and [Amazon S3](#) cloud storage.
- Pipeline configurations and software versions are now stored in Butler repositories to record run configurations.

Planned activities

Architecture

Architecture will represent Data Management at the Rubin Observatory Network Pre-Verification Review.

Input into the design of the DM-relevant components of the LOVE system for observatory operations will be provided.

Initial data transfer tests from Chile to Google Cloud storage using the public Internet and the Python interface will be conducted. The interconnect between the Rubin Observatory network and Google Cloud will be ordered. Small-scale testing of Data Release Production and Alert Production using Google Compute Engine will be conducted.

Deployment of the Science Pipelines on CentOS 8 will be enabled in the installation tooling and continuous integration system, although CentOS 7 remains the official baseline. The ability to test the Alert Production against in-progress ticket branches of the ap_verify package will be enabled.

Improvements will be made to the Verification Control Document generation process, including an expanded summary table.

Middleware

- Begin implementation of registry version migration infrastructure.

- Add support for compressed FITS files, including the addition of a generic facility supporting write parameters for formatters and selection of pre-defined write recipes.
- Continue to add new Butler command line tools.
- Continue to investigate registry performance.

1.02C.03: Alert Production

Current accomplishments

02C.03.00 – Management and Leadership

- The AP leadership team organized, prepared for, and participated in the [DM Leadership Team (virtual) Face-to-Face meeting](#) which took place from 12 to 14 May. In addition to participating in general discussions,
    - Eric Bellm, together with Gregory Dubois-Felsmann, led a discussion on [plans for IVOA and Python interfaces to time series data](#).
    - John Swinbank led a discussion on [the management of calibration products](#).
- The team undertook "rolling wave" planning for the Fall 2020 development cycle (which starts in June).

02C.03.01 – Single Frame Processing

- No work was completed in this WBS element.

02C.03.02 – Catalog Association for Alert Production

- No work was completed in this WBS element.

02C.03.03 – Alert Distribution System

- Alert serialization and distribution [[DM-24345](#), [DM-25143](#)]:
    - The Alert Production pipeline now serializes alert data into schema-compliant Avro packets, suitable for use in the alert distribution system. This is a major step forward for the Alert Production system: it represents the first time at which it actually produces alerts! [[DM-24324](#)]
    - The alert_packet package, which provides convenience libraries for working with Avro-format alert packets, has been added to the main Science Pipelines distribution. [[DM-24500](#)]
    - Major progress has been made on the Alert Stream Simulator. As described in [DMTN-149](#), this is the tool which will be provided to (potential) community

broker authors to provide them with a simulated version of the future Rubin alert stream. The bulk of development has been completed this month; we anticipate release to the community during June. [DM-24642, DM-24644, DM-24649, DM-24653, DM-24654, DM-24681, DM-24718, DM-24729, DM-24941]

○ This completes work on activity DM-24345; work on the alert distribution system will continue on DM-25143.

02C.03.04 – Alert Generation Pipeline

● Work is ongoing to apply the lessons learned in the re-engineering of the decorrelation afterburner (reported last month) to our implementation of the Zackay, Ofek and Gal-Yam (2016) image differencing algorithm. Results will be reported in June.

02C.03.05 – Tools for Science Pipelines

● Emergent work [DM-24339]:
  ○ A number of bugs were fixed and feature requests serviced across the codebase. Highlights included:
    ■ The lsst.log package has been updated to follow the current Science Pipelines documentation standards. [DM-15813]
    ■ A substantial amount of deprecated and unmaintained functionality has been removed in preparation for the upcoming release of version 20.0.0 of the Science Pipelines. [DM-20753, DM-21711, DM-22192, DM-23343, DM-24565, DM-24747]
    ■ Minor errors in the implementation of the improved decorrelation afterburner (as described last month) have been resolved. [DM-24708]
    ■ A race condition in building the daf_persistence package was resolved. [DM-24768]
    ■ Minor improvements were made to the "linting" system used to check DM code for syntax errors and standards violations. [DM-24768, DM-24906]
  ○ This completes work on activity DM-24339; future emergent work will be addressed on DM-25139.
● Pipeline infrastructure updates [DM-24341]:
  ○ The AP team continued to convert existing code to the new "Generation 3" middleware. In particular, this month:
    ■ The ap_verify datasets — which provide conveniently packaged

verification datasets for testing the AP system — have been updated to work in the Generation 3 system. [DM-24260]

- Conventions around the naming of filters in the Alert Production system were resolved. This made it possible to use the source association system in the Generation 3 system. [DM-25040]
  - This completes work on activity DM-24341; future pipeline infrastructure work will be addressed on DM-25145.

02C.03.06 – Moving Objects Processing System (MOPS)

The team is focused on preparation for the Solar System Science Collaboration Readiness Sprint which will take place in June.

02C.03.07 – Transform Fitting on Stacks of Images

- Implementation of proper motion and parallax in Jointcal [DM-24343]:
  - The test and verification data for the Jointcal system has been updated based on the latest Gaia and PanSTARRS reference catalogs. This required substantial reworking of the Jointcal testing system, as well as bulk data reprocessing. However, it resulted in a substantially enhanced set of test cases, which are now suitable for verifying upcoming work on proper motion and parallaxes. [DM-17597, DM-24976]
  - This completes activity DM-24343; work on proper motion and parallax support in Jointcal will continue on DM-25141.

02C.03.08 – Integration

- QA tooling and pipeline performance monitoring [DM-24344]:
  - An extensive comparison of the results achieved by differencing against templates which are generated by the standard CompareWarp algorithm and those which have been corrected for differential chromatic refraction (DCR) on simulations of crowded fields has been completed. Full results are available in this Jupyter notebook. In brief, we conclude that the DCR mitigation algorithm will struggle to operate effectively in crowded fields. [DM-23674]
  - An analysis of the changes to AP verification results due to the integration of the upgraded decorrelation afterburner (reported last month) was completed. Changes to the verification results were shown to be in-line with expectations. Full results are available in this Jupyter notebook. [DM-24730]
  - This completes activity DM-24344; QA analysis of alert production pipeline

outputs will continue on [DM-25146](#).

Planned activities

02C.03.00 – Management and Leadership

- Transition to the Fall 2020 development cycle.

02C.03.01 – Single Frame Processing

- No work is scheduled in this WBS element.

02C.03.02 – Catalog Association for Alert Production

- No work is scheduled in this WBS element.

02C.03.03 – Alert Distribution System

- Enhance alert packets to include cut-out images and WCS and PSF information.
- Provide Alert Stream Simulator and example alert data to the community.

02C.03.04 – Alert Generation Pipeline

- Resolve remaining issues in our [ZOGY](#) image subtraction code.
- Produce a technical note describing the current state of, and future plans for, the Rubin image differencing code.

02C.03.05 – Tools for Science Pipelines

- Complete the transition to Generation 3 middleware, by delivering an operational AP pipeline in the new system.

02C.03.06 – Moving Objects Processing System (MOPS)

- Participate in the Solar System Science Collaboration Readiness Sprint.

02C.03.07 – Transform Fitting on Stacks of Images

- Incorporate proper motions and parallaxes read from reference catalogs into Jointcal processing.

02C.03.08 – Integration

- Host the first in a series of meetings of the AP team with the Commissioning and DM Science Verification teams to discuss QA metrics within the Alert Production

system.

Staffing update

- John Swinbank, AP team T/CAM and Deputy Data Management Project Manager, will be leaving the project in October of this year. We are in the process of putting in place arrangements for his replacement.

1.02C.04: Data Release Production

Current accomplishments

02C.04.00 – Management and Leadership

- The AP leadership team prepared for, and participated in the DM Leadership Team (virtual) Face-to-Face meeting which took place from 12 to 14 May.
- The team undertook "rolling wave" planning for the Fall 2020 development cycle (which starts in June).

02C.04.01 – Software Primitives

- "Generation 3" middleware development [DM-23737]:
  - Throughout this month, members of the DRP team continued to participate in development of the "Generation 3" middleware (task execution framework and "Butler" data access abstraction) and to convert existing algorithmic code to the new framework. This work is carried out under the overall coordination of the Architecture team. [DM-24435, DM-24614, DM-24664, DM-24698, DM-24734, DM-24780, DM-24892, DM-24940, DM-25000, DM-25014, DM-25080]
- Emergent work [DM-23740]:
  - A number of bugs were fixed and feature requests serviced across the codebase. Highlights included:
    - Improvements were made to the code that writes Parquet format source tables such that it can now extract calibration and WCS information from appropriate exposures if required. [DM-24379]
    - A "driver" script for running single-frame processing while inserting simulated ("fake") sources was produced. This enables us to use our existing infrastructure for handling simulated sources as part of large-scale test runs. [DM-24553]
    - An initial version of DMTN-148 was published. This major work

defines the overall architecture and strategy for management of calibration products in the construction, commissioning, and operational eras. This first version will serve as the basis for discussion across the subsystem over the following months. [DM-24707]

- A total of 11 tickets, totaling 30 story points, were completed. [DM-6981, DM-24857, DM-24864, DM-24908, DM-24946, DM-25033, DM-25112, DM-25149]
- This completes work on activity DM-23740; future emergent work will be addressed on DM-25271:.

02C.04.02 – Calibration Products

- Detector characterization and instrument signature removal [DM-23745]:
  - The set of bad amplifiers being tracked for the LSSTCam has been updated and verified. [DM-23762]
  - Defects are now stored in "normalized" form (that is, with overlapping and redundant definitions merged). [DM-24556]
  - It is now possible to configure sigma-clipping in the photon transfer curve task. [DM-24756]

02C.04.03 – Image Characterization

- PSF characterization [DM-23736]:
  - A major product to resolve a discrepancy between the measured size of stars and the derived model of the PSF on co-added images has been successfully completed and the issue resolved. [DM-24691]
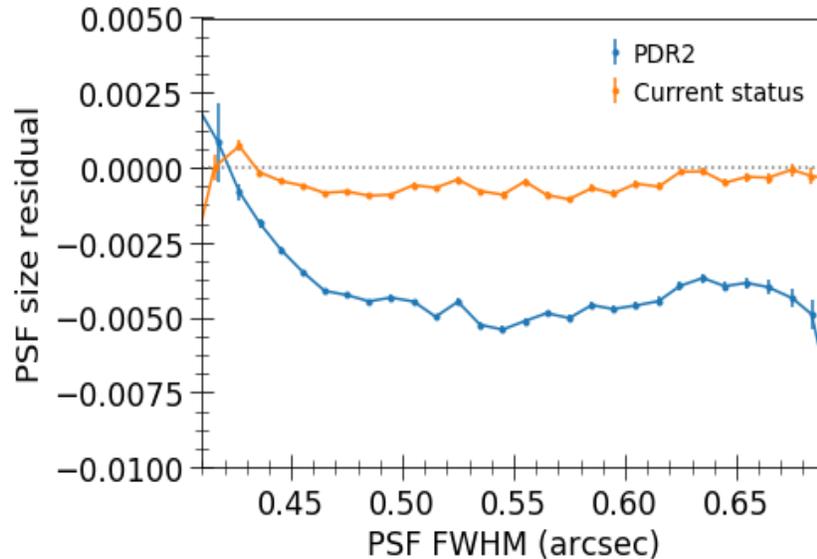
Residuals characterizing the difference between the PSF

- ○ The DRP team plans to implement the Gaussian aperture and Photometry ("GaaP") algorithm (Kuijken, 2008). This algorithm requires accurate fluxes calculated within elliptical Gaussian apertures. An investigation was performed to establish whether this would require the use of sinc coefficients; the conclusion was that it is not. Full details are available in this Jupyter notebook. [DM-24691]
  - ○ This completes work on activity DM-23736. Work to implement the GaaP algorithm will continue on DM-25320, while improvements to our PSF estimation algorithms will proceed on DM-25319.
- ● Background estimation [DM-23739]:
  - ○ "Sky sources" — corresponding to measurements performed on empty patches of sky — are now available in single frame processing pipeline outputs. These are essential for a variety of QA use cases, and are described in detail in this community.lsst.org post. [23078, DM-24782]
  - ○ This completes work on activity DM-23739; improvements to background estimation will continue on DM-25270.
- ● Commissioning FGCM (Burke et al, 2018) [DM-23742]:
  - ○ Various improvements were made to I/O in the FGCM system. [DM-24935, DM-25020]
  - ○ This completes work on activity DM-23742; development and integration of FGCM continues on DM-25271.

02C.04.04 – Coaddition

- No work was completed in this WBS element.

02C.04.05 – Detection and Deblending

- Deblender development [DM-23738]:
  - The SCARLET deblender system can now perform convolutions in real space. This has performance benefits, and makes it possible to take account of PSF variation across the field. [DM-23788]
  - SCARLET can now directly compute gradients, rather than relying on the autograd library to do it. This brings further performance improvements. [DM-24443, DM-24915]
  - An investigation was performed into migrating SCARLET from the autograd library to JAX for performance reasons. We concluded that the JAX library is not currently an adequate replacement. We may re-evaluate this position in future, depending on the state of JAX development. [DM-24609]
  - Following reports of poor results from SCARLET on crowded fields (in particular, the field of global cluster NGC 6569), some initial investigations have been carried out by the DRP team. Although some improvements have been made (see this Jupyter notebook), we expect to undertake a more comprehensive analysis in future on DM-24736. [DM-24713]
  - Incorporating upstream updates to the SCARLET project have provided performance improvements. [DM-24862]

02C.04.06 – Characterization and Measurement

- Galaxy photometry metrics [DM-23743]:
  - The MultiProFit galaxy fitting tool now includes a component in its fit to take account of the background. [DM-23399]
  - A comparison between CModel (the current default galaxy fitting system in the Science Pipelines), MultiProFit, and ngmix was carried out on simulated data. The results are presented in this Jupyter notebook. [DM-23805]
  - This completes work on activity DM-23743; galaxy photometry development will continue on DM-25274.

02C.04.07 – Maintenance, Quality and Documentation

- QA and analysis of DRP pipelines [DM-23741]:
  - The pipe_analysis QA tool has been updated to show distributions and

metrics of the aperture and PSF fluxes measured from so-called "sky objects" (that is, measurements made at positions where no astronomical object has been detected). [DM-16595]

- The pipe_analysis tool has been updated to remove calls to the Data Butler from low-level functions. This is a necessary prerequisite for running pipe_analysis with the Generation 3 middleware. [DM-22255]
- The pipe_analysis tool has also been updated to avoid calling deprecated functionality functions which will no longer be available in future Science Pipelines releases. [DM-24705]
- This completes work on activity DM-23741; further QA on DRP pipelines will be carried out on DM-25272.

Planned activities

02C.04.00 – Management and Leadership

- Transition to the Fall 2020 development cycle.

02C.04.01 – Software Primitives

- Continue Generation 3 middleware development.

02C.04.02 – Calibration Products

- Convert existing calibration products to the DMTN-148 model.
- Continue reduction of, and pipeline development for, Auxiliary Telescope data.

02C.04.03 – Image Characterization

- Add background offset tracking to fgcmcal.

02C.04.04 – Coaddition

- Complete and merge code for finding and masking streaks as part of coadd assemble.

02C.04.05 – Detection and Deblending

- Improve SCARLET initialization to improve convergence speed and solution accuracy.

02C.04.06 – Characterization and Measurement

- Calculate [Rowe statistics](#) as part of regular QA processing.
- Implement priors in [MultiProFit](#).

02C.04.07 – Maintenance, Quality and Documentation

- Augment the pipe_analysis tooling.
- Improvements to simulated source insertion.

Staffing update

- No changes to report.

1.02C.05: Science User Interface & Tools

This WBS supports the activities of the Science Platform Scientist Gregory Dubois-Felsman, which are reported under the Architecture (1.02C.02.02) and DM Science (1.02C.02.01) teams. An additional fractional FTE is used to support bug fixes on Firefly which are not reported here in detail.

1.02C.06: Science Data Archive & Application Services

Current accomplishments

02C.06.00 Management & Leadership

- Mueller and Pease prepared for, and participated in the [DM Leadership Team (virtual) Face-to-Face meeting](#) which took place from 12 to 14 May.
- The team undertook "rolling wave" planning for the Fall 2020 development cycle (which starts in June).

02C.06.01.01 Catalogs, Alerts and Metadata

- Salnikov evaluated the Scylla real-time database as an alternative to Cassandra for use as a base technology in the Alert Production Database prototype. [[DM-24692](#)].

02C.06.01.02 Image and File Archive

- No work was carried out in this WBS element this month.

02C.06.02.01 Data Access Client Framework

- Management of work for this WBS element has been transferred to ARCH, and corresponding progress for this and future months will be reported in that section (02C.02.02) of this report.

02C.06.02.02 Web Services

- Management of work and staffing for this WBS element have been transferred to SQuaRE, and corresponding progress for this and future months will be reported in that section (02C.10) of this report.

02C.06.02.03 Query Services

- Gaponenko extended the Replication/Ingest framework and REST API with algorithms and tools for indexing catalogs [DM-24684].
- Gaponenko created a way to grant privileges when publishing newly ingested catalogs [DM-24732].
- Gaponenko improved error reporting in the REST services of the Replication System's master controller [DM-24788].
- Jammes migrated Qserv documentation to LSST-the-docs [DM-23256].
- Mueller worked in support of adapting Qserv build and continuous integration to the new Conda-based stack environment, and continued work on Qserv build and runtime container refactorization.
- Gates continued work on integrating catalog data synthesizer with new Qserv ingest to generate dataset for 50%-DR1 Qserv scale tests.

02C.06.02.04 Image Services

- Lo continued work on ObsCore image metadata extraction.

02C.06.02.05 Catalog Services

- No work was carried out in this WBS element this month.

02C.06.03 Task Framework

- Salnikov created a CollectionsManager implementation that uses an integer primary key [DM-24594].
- Salnikov fixed an issue related to deserializing a QuantumGraph [DM-25016].
- Salnikov made pipetask's graphviz dot files quote component dataset type names [DM-24796].
- Salnikov made gen3 pipetask store per-run information (configs, software versions)

in butler repo [DM-24797].

- Salnikov designed and implemented a way to store and run a quantum graph [DM-25113].

Planned activities

02C.06.00 Management & Leadership

- Transition to the Fall 2020 development cycle.

02C.06.01.01 Catalogs, Alerts and Metadata

- Salnikov and Hanushevsky to conclude and write up APDB Cassandra/Scylla performance characterization work.

02C.06.01.02 Image and File Archive

- No work is planned for this WBS element in the upcoming month

02C.06.02.03 Query Services

- Gates to continue work on integrating data synthesizer with new Qserv ingest to generate dataset for 50%-DR1 scale tests.
- Gaponenko, Gates, and Hanushevsky to continue Qserv profiling, tuning, and performance optimization.
- Mueller to continue rework of Qserv build procedures to decrease build time, decrease container sizes, and simplify dependency management.
- Jammes to continue oversight of Qserv deployment on new hardware cluster at CC-IN2P3 and prepare the cluster for data-loading.

02C.06.02.04 Image Services

- Lo to continue work on ObsCore image metadata extraction, and deploy finalized HSC ObsCore metadata into LSP ObsTap instance at the LDF.

02C.06.02.05 Catalog Services

- No work is planned for this WBS element in the upcoming month.

02C.06.03 Task Framework

- Salnikov and Pease to continue with enhancements/fixes to the task framework in support of ongoing Gen3 middleware development.

Staffing update

- No changes to report.

1.02C.07: LSST Data Facility

Current accomplishments

02C.07.05 LSST Data Facility Management, Service Architecture, and Project Controls

The management team at NCSA:

- Continued initial planning for Operations Rehearsal #2 ([LDM-643](#)).
- Continued regular steering meetings for internal oversight over all technical areas of the LSST Data Facility.
- Continued working toward a working NCSA test stand for all Configurable SAL Components (CSCs) involved in the LATISS and ComCam environments. We extended the network for 141.142.139.* for the teststand network to give more IP addresses in the test stand for the CSCs.
- Continued participation with System Science Team, LDF Infrastructure meetings, Data Management Leadership Team, Commissioning meetings, ComCam meetings, AuxTel meetings, IT support committee, Telescope & Site meeting, T/CAM meetings, middleware meeting, CCB, SST, LSST Science Platform meetings and operations proposal writing team.

02C.07.06 LDF Production Services

- Continued monthly reprocessing of each the RC2 and DC2 datasets to support software stack testing and pipeline code development.
- The condor flock systems were put back into place, and condor testing with Butler/G3 continues.
- All nodes for the [Kubernetes](#) -stable and -int clusters were put into production.

02C.07.07 Data and Compute Services

- We continuously improve the Kubernetes cluster infrastructure.
- The Engineering and Facility Database (EFD) environment was installed into the test Kubernetes cluster, and then — after approval — it was installed also into the production side of the Kubernetes cluster. The data could have data flow immediately from the EFD at the summit once that system is turned back on. As the

BDC begins the comcam data image generation, more data will flow to the EFD in the Kubernetes environment.

02C.07.08 LDF Service Software

- For Prompt Service software we continued incremental development and improvements for the Archiver, Forwarder, Observatory Operations Data Service (OODS) and Header Service software to support LATISS in the test stand awaiting the mountain to power back on.
- For Prompt service software, we continued with major development strides for ComCam. The header service, Archiver, Forwarder and OODS all were able to view and work with the 9-CCD image (raft scale).
- For Batch Production Service development, we continue to integrate the HTCondor workload management system, as well as responding to emerging changes in the Gen3 middleware code base.
- For the Data Backbone new data transfer mechanism was tested and installed for comcam and Latis in the teststand.
- We continued supporting data transfer from test stands at SLAC, monitoring file transfer and verifying data ingestion.
- We continued work adding features to the new Disaster Recovery service, aimed at hardening and improvements to operationally support all LSST data.

02C.07.09 ITC and Facilities

- We continued to work with CSCs on the NCSA test stand and it's new releases.
- We continued responding to incidents and requests concerning as-is development systems and services, as well as addressing emerging hardware and system needs, and reviewing vulnerability reports and performing security reviews for system changes.

Planned activities

02C.07.05 LSST Data Facility Management, Service Architecture, and Project Controls

The management team at NCSA plans to:

- Continue regular steering meetings for internal oversight over all technical areas of the LSST Data Facility.
- Continue developing and documenting test plans for Data Facility components and services.

- Complete and upload to the project management control system planning activities for the next development cycle.
- Continue regular meetings with CC-IN2P3 to coordinate near-term work and discuss operations.
- Continue participation with the Summit/Base Tiger Team, Science System Team, LDF Infrastructure meetings, Data Management Leadership Team, IT North/South Team, Data Management Change Control Board, ITS committee, Comcam/ATS meetings, and Commissioning Team.

02C.07.06 LDF Production Services

- We will continue periodic reprocessing of datasets in support of stack testing and pipeline development.
- We will begin data ingest from the ComCam environment from BDC, continue with LATISS observing and continue ingest from other SLAC data, enabling Science Platform (LSP) viewing.
- We will continue to develop the comcam software for raft scale images.
- Continue evolving the LDF services. The rome node will be brought into production as a development mode.

02C.07.07 Data and Compute Services

- We will continue work implementing the current phase of file transfer, service endpoint, and file ingestion into the Data Backbone for early commissioning data generated by the LATISS instrument on the summit, the ComCam test stand, and test stands at SLAC.
- We will continue working with developers on improvements to the Kubernetes service as well as Kafka installations to support development for the EFD and alert distribution and bring real data to the NCSA Kubernetes systems from the Summit Facility for the EFD.

02C.07.08 LDF Service Software

- We will continue work on Prompt Service software with new libraries and Xml needed, the LATISS needs as they arise, as well as upcoming support for ComCam testing and integration on the test stand in Tucson.
- We will continue modifications to Forwarder, Archiver, OODS, and Header Service components to support the new DAQ hardware and software on the NCSA Test Stand.
- We will continue to work with Gen3 middleware, using emergent versions of the

middleware software to test pipeline execution using Gen3 components with workflow management and database systems, as well as responding to changing needs for database support.

- For the Data Backbone, we will continue facilitating and monitoring automated file transfer and ingestion in support of test instruments at SLAC and at BDC as well as continue to prototype elements needed for future versions of the data backbone.

02C.07.09 ITC and Facilities

- Activities for April, will include configuration for systems to support ComCam at the base data center, as well as systems that will support LATISS on the summit.
- GPFS will be monitored from the BDC for the NFS mounts to the commissioning cluster.
- We will continue responding to incidents and requests concerning as-is development systems and services, as well as addressing emerging hardware and system needs.

Staffing update

- No changes to report.

1.02C.08: International Communications & Base Site

Current accomplishments

02C.08.01 – Infrastructure

- Setup ComCam hosts with dependencies needed to run the archiving software and attached network interfaces directly to the LHN for the Ops. Rehearsal.

- Created a Slack channel to communicate and troubleshoot with all the Network and Infrastructure engineers working in the long haul network

- Integrated IPA in Foreman to allow users to run privileged commands con ComCam hosts.

04C.12.05 – Local Network

- Enabled routed traffic to be sent to the long haul network. The commission cluster

and a few hosts of the temporary deployment of ComCam at the base can now send traffic to the LHN without the need of a physical direct connection to the Rubin Border Router

- Improved the logging reporting of network devices by sending them to a centralized logging deployment

02C.08.03 – Long-Haul Networks

- Prepared networks for June 1–5 Operations Rehearsal 2 using ComCam on bench
- Prepared for June 15–17 Pre-Verification Review of networks
  - Submitted LCR for updates to LSE-78 Rubin Observatory Network Design.
  - Updated LDM-732 Rubin Observatory Network Verification Baseline
  - Populated review website
- Note: The Dense Wave Division Multiplex equipment in Cerro Pachón was shut down on March 20th as part of suspension of construction site activity due to Covid-19.

02C.08.03.01 – Chile National WAN

- REUNA/Vera Rubin Telecommunications workshop held to transfer information regarding DWDM technology and the infrastructure that is currently implemented in both Long Haul and the Summit and Base.
- Summit - AURA Gatehouse:
  - The link remained shut down since the March 20 shutdown.
- Dense Wave Division Multiplex (DWDM) Equipment:
  - The DWDM system over the Long haul path remained operational and working properly until the March 20 shutdown, and remains operational from La Serena to Santiago. Testing of the DWDM protection between La Serena and Cerro Pachon is still pending to be executed. Due to Coronavirus this activity will be scheduled once the Summit is accessible.
- Santiago-La Serena:
  - Two major cuts were suffered on the fiber in this section this month. Due to the service provider exceeding the contract time for the repair, the associated fines were issued. Fail-over to the secondary occurred with no loss of data.
- La Serena–AURA Gatehouse:
  - The link remained shut down since the March 20 shutdown.

02C.08.03.02 – International Chile–US WAN

- *100 Gbps Managed Ring:*
    - The links remain operational and working properly.
- *Management and Coordination Contract:*
    - The LSST NET meeting was held on May 14, 2020. The meeting had the following topics:
        - NET Action Items Status
        - Continental US Networks (CONUS) update
        - Chilean/International Networks update
        - Transatlantic Networks (NCSA - CC-IN2P) update
        - LSST Operations Network Service-Level Agreements (SLAs) and Virtual Network Operations Center (VNOC)
        - LSST Network Verification and Validation
        - Network Design and End-to-End Test Plan and QoS planning
    - Conferences and Workshops
        - Prepared report from South American Astronomy Coordination Committee (SAACC) Virtual Meeting - April 21-22, 2020
- *Miami - Boca Raton - Atlanta:*
    - The links remain operational and working properly.
- *São Paulo – Miami Spectrum:*
    - The links remain operational and working properly.
- *São Paulo – Santiago Spectrum:*
    - CenturyLink and Silica are provisioning the 100G wave from Porto Alegre to Santiago. The spectrum from Santiago to Porto Alegre/Brazil is ready for limited service, with full installation by Q3/Q4 2020.
    - Due to an issue with the power company installation of the fiber from Porto Alegre to São Paulo, RNP is working with carriers and ISPs to swap fibers in the segment. The goal is to have a solution by the end of 2020.
- *US National WAN:*
    - The 2 x 20 Gbps link remained operational.
- *Transatlantic WAN:*
    - GEANT's 100G connection to Lyon, France is in progress and will be available by 3QFY20.

Planned activities

02C.08.01 – Infrastructure

- Complete deployment of CCS software.

04C.12.05 – Local Network

- Complete integration of the NOIRLab Wireless with Rubin's network.

02C.08.03 – Long-Haul Networks

- Conduct network tests during June 1–5 Operations Rehearsal 2 using ComCam on bench (now delayed due to ComCAM/Operations Rehearsal 2 delay)
- Conduct June 15–17 Pre-Verification Review of networks

02C.08.03.01 – Chile National WAN

- Summits - AURA Gatehouse Network:
    - Enable transfer of LATISS traffic over this segment pending resumption of summit activity.
    - Conduct network data transfer tests La Serena - NCSA
- La Serena - AURA Gatehouse:
    - Enable transfer of LATISS traffic over this segment pending resumption of summit activity.
    - Conduct network data transfer tests La Serena - NCSA
- Santiago-La Serena:
    - Enable transfer of LATISS traffic over this segment pending resumption of summit activity.
    - Conduct network data transfer tests La Serena - NCSA
- DWDM:
    - Restart equipment pending resumption of summit activity.

02C.08.03.02 – International Chile–US WAN

- *São Paulo – Miami Spectrum:*
    - Enable transfer of LATISS traffic over this segment pending resumption of summit activity.
    - Conduct network data transfer tests La Serena - NCSA
- *São Paulo – Santiago Spectrum:*
    - Monitor work on all phases

- *US National WAN:*
  - Enable transfer of LATISS traffic over this segment pending resumption of summit activity.
  - Conduct network data transfer tests La Serena - NCSA

Staffing update

- No changes to report.

1.02C.10: Science Quality and Reliability Engineering

Current accomplishments

Highlights include:

- We published a document ([SQR-039](#)) proposing improvements to the Science Platform Authentication and Authorization service and incorporated comments we received in response. These changes will prepare this critical service for operation at the Intermediate Data Facility and at the future US Data Facility site.
- We have developed the capability to automatically stand up a "bleeding edge" environment for the Science Platform. This allows us to continuously integrate our deployment infrastructure, as well as to easily bring up ad-hoc environments to test compatibility ahead of future upgrades.
- We have made additional improvements to our deployment infrastructure to help us better manage the different configurations numerous instances of the Science Platform we currently support (we have 8 partial or full deployments on 7 different clusters at the current time)
- We developed a plan for replacing the Confluent Kafka replication connector we have been using for the telemetry stream with an open-source alternative as we could not obtain affordable pricing. See [DM-24774](#) for details.
- We investigated a bug where the ArgoCD login command does not respect port forwarding, fixed it and contributed it upstream.
- We developed "Ook", a metadata framework required to index LSST The Docs documents using [Algolia](#).
- We improved the front end of our LSST the Docs search prototype by refining the style design. This brought us to the point where we could demo this long-wished for capability at the DMLT virtual face-to-face meeting.
- We improved logging in Nublado so as to better trace individual user interactions

when many users are interacting with the system simultaneously.
- We deployed some science platform elements in the NCSA test-stand
- Last month we described switching session information for Nublado from SQLite to PostgreSQL. we have now rolled out this change on all deployments (Base Data Center, Tucson test stand etc.)
- We imported the DESC DC2 schema into the TAP schema database
- Notable external interactions:
    - Christine Banek participated in the virtual IVOA meeting.

Planned activities

- Publish an analysis of security concerns for the Science Platform
- Start documentation infrastructure migration to Github Actions
- Create Faust records from Avro schemas and apply a compute method to data within defined time windows
- Investigate an issue with the notebook workflow prototype where it seems to run as the wrong user
- Add JS9 extensions to nublado
- Refactor the testing infrastructure to improve understanding of failures
- Release the EFD example notebooks into the production branch

Staffing update

- No changes to report.