

# Vera C. Rubin Observatory

# Data Management Monthly Report

## August 2020

## High-level Summary

### Community Interactions, Meetings and Workshops

Most members of the DM subsystem participated in the [Rubin Observatory Project and Community Workshop](#), August 10–14. DM team members participated in the Scientific Organizing Committee for the meeting, and were involved in organizing and chairing several key sessions, including:

- [An Introduction to the Rubin Observatory: Systems, Jargon, and Acronyms](#)
- [Rubin Algorithms Workshop Follow-up: Answers to Community Questions](#)
- [Community Alert Brokers](#)
- [Low Surface Brightness Science and Robust Sky Estimation](#)

In addition, the DM team organized the now-traditional PCW [board gaming session!](#)

DM was represented at the [DotDotAstro](#) online conference on August 17, 19, and 26 by Meredith Rawls (University of Washington).

At the end of the month, members of the DM team were involved in the annual [Joint Status Review](#). This started on August 31, and continues into September.

### Technical Progress

Many documents were updated or published in August in preparation for the Joint Status Review. Notable highlights include:

- LDM-148, the DM System Architecture
- LDM-294, the DM Management Plan
- LDM-503, the DM Test Plan
- LDM-564, the DM Release Schedule
- [DMTN-157](#), documenting initial results of the Google Cloud proof-of-concept
- [SQR-048](#), providing security guidelines for Kubernetes clusters in use by DM

In addition, the documentation portal at [www.lsst.io](http://www.lsst.io) now has a tag-driven user guide listing in preparation for exposing the service more to external users.

Substantial progress was made towards adopting the new “Generation 3” data access and task execution middleware this month, as we deployed a complete “end-to-end” Alert Production pipeline using this new system. In addition, Hyper Suprime-Cam RC2 — one of our core precursor test and verification datasets — is now available in the Generation 3 system, with the data registry held in a PostgreSQL database at NCSA. Pipeline processing of this HSC RC2 data is now possible in the Generation 3 framework.

Several key refinements were made to DM’s data access and database systems during August. First, the WISE and Gaia Data Release 2 datasets were ingested and made available in a new Qserv deployment at NCSA. Secondly, the prototype Alert Production pipeline was upgraded to use a database based on PostgreSQL (rather than SQLite, as had been used previously); this provides much improved scalability and enables large scale testing. Finally, data aggregated (or otherwise processed in-stream) from [Apache Kafka](#) (such as the Engineering Facilities Database telemetry stream) can now be stored to [S3](#) in [Apache Parquet](#) format.

A number of useful improvements were also made to our science algorithms. In particular, DM’s implementation of the [ZOGY](#) image differencing algorithm has been extended to support a spatially-varying PSF, and our task to inject simulated sources into data release processing can now accept arbitrary input images.

The generation of 21-raft images by the Camera subsystem at SLAC had major impacts across the whole Rubin team and beyond this month. DM’s involvement included transferring the data to NCSA, ingesting it into standard data repositories, and making it available for viewing and manipulation via Jupyter notebooks in the Science Platform.

The DM team is preparing for some major personnel changes in the upcoming weeks. Effective Oct. 1, Jeff Kantor will reduce his effort to 25% Rubin Obs. Cristian Silva, based in La Serena, will take over networks and co-chair the NET team. On a similar timescale, Deputy DM Project Manager and Science Pipelines Manager John will transition out of the project. His leadership role in the Science Pipelines group will be taken over by the new Science Pipelines Manager Yusra AlSayyad and Alert Production T/CAM Ian Sullivan.

## 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

No milestones were completed during this month.

## Milestones Delayed

### Level 1 milestones

DLP-526: Archive Center Integration Complete

Due 2020-04-22

*This milestone is obsolete following recent funding agency guidance about the location of the US Data Facility. [LCR-2453](#) fixes the governing requirements and [LCR-2446](#) will remove the milestone.*

### Level 2 milestones

LDM-503-10b: Large Scale CCOB Data Access

Due 2020-04-22

*This testing is complete, and the test plan executed. The test report is being written and reviewed and will be reported on in the September monthly report.*

LDM-503-11a: ComCam Ops Readiness

Due 2020-07-15

*This testing is complete and the test plan executed. The test report is being written and reviewed and will be reported on in the September monthly report.*

LDM-503-11b: Test report: Pipelines Release Fall 2019

Due 2020-07-15

*The Pipelines release was successfully completed and test cases were executed in late August. Review of the accompanying test report ([DMTR-192](#)) was ongoing at the end of the month; we expect it to be approved in early September.*

## Alert Production

DM-AP-10: Advanced single frame measurement pipeline for Alert Production

Due 2020-06-01

*This milestone is pending final integration of the upgrades to astrometry which were prototyped in Summer 2019. However, this is not currently blocking further scientific progress with the AP pipeline, so it has been temporarily deferred.*

## Data Release Production

DM-DRP-21: Integrated image characterization pipeline for Data Release Production.

Due 2020-07-15

*Currently, the photometric calibration uses measurements with the initial per-CCD background. This milestone will be complete after adding a task between full-focal plane sky estimation and coaddition that remeasures sources with the full-focal plane background model. Development on this new task will begin after the migration to Generation 3 middleware. Eventually, this extra iteration will be augmented with a higher-fidelity PSF model as well.*

DM-DRP-22: Template generation integrated

Due 2020-06-01

*Currently, a single set of coadded templates are generated per data repository for image differencing. The capability to generate and store multiple sets of templates based on subsets of visits or seeing conditions is blocked pending the availability of Generation 3 middleware.*

DM-DRP-24: Physically motivated PSF model

Due 2020-06-01

*Modeling out of focus PSFs with separate contributions from the atmosphere and telescope system has been demonstrated. Modeling code to be integrated with the stack expected S21. Inferred PSFs expected F21.*

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-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 and is continuing into Fall on activity DM-26785.*

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. The background matching implementation was demonstrated on SDSS data, but we will claim it complete after demonstrating on HSC which has a more complex instrumental background.*

## Science User Interface and Tools

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

Due 2019-02-28

*This is still awaiting the HSC image data to be available via IVOA services, including a) making HSC image metadata available in ObsCore format, which is underway, and b) the provision of an ObsTAP service, which has been prototyped by the DAX group. (ObsCore-format image metadata from AllWISE has already been used to test the prototype service.)*

## Science Data Archive & Application Services

DLP-802: Alert Production Database (APDB) Design

Due 2016-11-30

*A scalable prototype based on the Cassandra NoSQL database system has been developed and tested (see [DMTN-156](#)). Further scale testing is necessary. Work ongoing with completion expected in S21.*

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

Due 2019-02-28

*HSC reprocessing runs are now being ingested into the Qserv database system as they are produced. Final work now underway to fully automate this process. Completion anticipated in F20.*

## LSST Data Facility

DM-NCSA-8: Alert Filtering Service receives alert streams

Due 2020-05-29

*This milestone is currently delayed awaiting alert streams. It is anticipated that this work will be completed in December of 2020.*

DM-NCSA-15: Batch Processing Service for offline spectrograph data processing

Due 2020-05-29

*This milestone is currently delayed awaiting automatic BPS for when files arrive at LDF. This milestone can be run by hand today specifically with previously identified data and pipeline tasks, but in a "service" automatic way as data comes from the summit is work being done in*

July/August/September.

## 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 AURA Center Operation Services. The baseline of network and computing infrastructure is complete, documentation is also completed, several technotes have been released covering physical and logical infrastructure. The critical path is not impacted.*

## Detailed Project Progress

### 1.02C.01: System Management

#### Current accomplishments

The DM Project Manager led the subsystem through the Project and Community Workshop and the Joint Status Review.

#### Planned activities

The DM Project Manager will finish the Joint Status Review and chair a [virtual face-to-face meeting of the DM Leadership Team](#).

### 1.02C.02.01: Data Management Science

#### Current accomplishments

The DM Subsystem Scientist:

- Prepared for and participated in the [Project & Community Workshop 2020](#), co-chairing a session on [Rubin Algorithms Workshop Follow-up](#) with Deputy DM Subsystem Scientist Colin Slater and a session on [Community Preparations for Early Science with Rubin](#) with Interim Director of Operations Bob Blum.
- Prepared for and participated in the [Joint Status Review 2020](#).

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

- Melissa Graham chaired, and Jeff Carlin was a member of, the [Project & Community Workshop 2020](#) Scientific Organizing Committee.
- Eric Bellm and Melissa Graham co-chaired a PCW session on [Community Alert Brokers](#).
- Most of the DM-SST attended and participated in the [Project & Community Workshop 2020](#).
- Gregory Dubois-Felsmann gave a presentation at an AAS hosted workshop on the topic of issuing persistent references to data.

## Planned activities

The DM Subsystem Scientist will:

- Participate in the Data Management [virtual face-to-face meeting](#).
- Give an invited presentation as part of the Duke University seminar series.

Other DM System Science Team (DM-SST) Activities will include:

- Issuing a technical note detailing the interim model for community support.
- Issuing a first version of the roadmap for serving photometric redshifts in data release catalogs.
- Jeff Carlin and Simon Krughoff will give a status update on development of the new framework for computing metrics.

## 1.02C.02.02: DM System Architecture

### Current accomplishments

#### Architecture

The Architecture Team attended the virtual Project and Community Workshop to better understand science needs. Proposed responses to the Network Review were assessed.

Members of the team wrote up the initial results of the Google Cloud proof of concept engagement ([DMTN-157](#)).

In the area of developer services and utility code, we switched to a much more efficient mechanism for specifying an exact conda environment for the Science Pipelines and updated several packages in that environment. The build system was improved by using dependency hashes rather than sequential build numbers, allowing parallelism. Several

problems were fixed (sims use of snco-smo, excessive environment differences, excessive environment rebuilding). Details of the conda environment are now automatically captured for provenance purposes. Versions are now required to be specified for deprecations.

In the area of release, verification, and document generation, the team generated [DMTR-261](#) and documents for the Joint Status Review; improved the inclusion of issue lists and images in test reports; and developed templates for verification element baseline documents LDM-752 and LDM-753. An initial prototype of new release automation was completed, encouraging continued development.

## Middleware

The Middleware Team:

- Improved handling of timespans in registry.
- Integrated WebDAV datastore (submitted by IN2P3).
- Performance improvements for registry queries.
- Continued development of click-based pipetask command.
- Released the new pipelines\_check package as a simple test that gen3 infrastructure is working correctly to execute a pipeline.
- Made datastore more resilient to tasks being killed during writes.
- Added missing ObsCore metadata to exposure registry.
- Demonstrated composite disassembly in ci\_hsc\_gen3 integration test.

## Planned activities

### Architecture

The Architecture Team will support the Joint Status Review and attend the DM Leadership Team Face-to-Face virtual meeting.

Final results are expected from the Google Cloud proof of concept.

The new lsstinstall installation tool will be released via conda-forge. Automatic generation of specialized containers for the Rubin Science Platform for use with Observatory control will be fixed. Further improvements in conda usage will result from a transition to a CentOS 7-based conda environment using a conda metapackage.

## Middleware

The Middleware Team will:

- Migrate obs\_cfht to Gen3.
- Work on calibration collections.
- Continue improving registry performance and schema stability.

## 1.02C.03: Alert Production

### Current accomplishments

#### 02C.03.00 – Management and Leadership

- The AP team participated in the successful [Project & Community Workshop 2020](#) during the week of August 10. Of particular note was the [Community Alert Brokers](#) session, co-organized by Eric Bellm (AP Science Lead) and Melissa Graham (DM Science Analyst). At this session, the team fielded questions from the community on the alert distribution system and the community broker selection process, and broker teams discussed broker features and community interactions.
- AP leadership participated in the [Joint Status Review 2020](#) during the week of August 31.
- Meredith Rawls attended the [DotDotAstro](#) online-only meeting on August 17, 19, and 26.
- A [new section](#) has been added to the [Science Pipelines documentation](#) describing the major user-facing changes to data products which were introduced in the last release. This is expected to become a regular part of the notes accompanying each future Science Pipelines release. [\[DM-26263\]](#)

#### 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

- No work was completed in this WBS element.

#### 02C.03.04 – Alert Generation Pipeline

- Image differencing algorithm development [\[DM-25144\]](#)
  - The implementation of the [ZOGY](#) image differencing algorithm has been extended to support a spatially-varying PSF, and made more robust against

- invalid data in the images. [[DM-25115](#), [DM-26182](#)]
- AP Pipeline QA [[DM-25146](#)]
    - Regular monthly reprocessing of DECam HiTS data was completed. No substantial changes relative to the previous month were flagged for detailed investigation. [[DM-25910](#), [DM-26380](#)]
    - New timing measurements of the Difference Image Analysis tasks were added to the SQuaSH instrumentation. This enables us to track the execution time of the various stages of the AP system, an essential capability both to monitor for performance regressions and to evaluate progress towards hitting latency targets for the Rubin alert production system. [[DM-26040](#)]
  - AP performance in crowded fields [[DM-25147](#)]
    - New data from two years of DECam g- and i-band observations covering the galactic bulge was added to our curated collection at NCSA. This data will be used to evaluate the performance of the AP pipeline in crowded fields. [[DM-25105](#)]

#### 02C.03.05 – Tools for Science Pipelines

- Emergent work [[DM-25139](#)]:
  - It is now possible to execute the Alert Production pipeline using a database based on PostgreSQL, rather than SQLite (as has been used to date). PostgreSQL can scale to much greater data volumes and transaction rates than SQLite; this new capability therefore enables us to test the alert production system at much larger scales than previously. Note that even PostgreSQL is not expected to scale to the level required in the operational era: the team continues to collaborate with the Data Access team (02C.06) to explore next-generation database technologies for this purpose (see, e.g., [DMTN-156](#)). [[DM-26051](#)]
  - A number of minor bugs were addressed and feature requests serviced across the codebase. [[DM-26007](#), [DM-26187](#), [DM-26265](#), [DM-26419](#)]
- Alert Production Pipeline infrastructure [[DM-25145](#)]:
  - The complete, “end-to-end” AP pipeline can now be executed using Data Management’s upgraded “Generation 3” middleware. This has been verified using the ap\_verify test system, which is automatically and regularly executed using DM’s Jenkins continuous integration system. This represents a critical milestone in both the development of the AP system and in the adoption of Generation 3 middleware across the subsystem, and is a necessary precursor for future AP system development. [[DM-21919](#)]

- New Hyper Suprime-Cam datasets have been added to our continuous integration test platform using the Generation 3 Butler. [[DM-24262](#)]
- An issue whereby the ap\_verify tool could not be executed using the same workspace more than once has been resolved. [[DM-26237](#)]

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

- The team prepared and submitted abstracts for the AAS Division of Planetary Sciences meeting which will take place in October of this year. These abstracts will cover the linking algorithm adopted for Rubin (HeliLinC2) and the analysis tooling that is being developed to demonstrate its performance. [[DM-26365](#), [DM-26366](#)]
- The team is preparing for a “hack week” in conjunction with the [Minor Planet Center](#) (MPC), scheduled to take place next month. This week will focus on developing data interchange systems between Rubin pipelines and the MPC.
- Work is ongoing on a prototype system for associating newly-observed DIASources (ie, source detections on difference images) with known solar system objects. We expect this to be available for testing next month, although work will remain to integrate it with the general purpose Association Pipeline.

#### 02C.03.07 – Transform Fitting on Stacks of Images

- No work was completed in this WBS element.

#### 02C.03.08 – Integration

- No work was completed in this WBS element.

### Planned activities

#### 02C.03.00 – Management and Leadership

- AP leadership will participate in the [DM Leadership Team virtual Face-to-Face meeting](#), September 15-17.
- John Swinbank and Ian Sullivan will prepare for handover of T/CAM responsibilities for the AP team.

#### 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

- No work is scheduled in this WBS element.

#### 02C.03.04 – Alert Generation Pipeline

- Draft technote detailing mathematical understanding of image differencing with regards to ZOGY.
- Continue development of regular test and QA procedures based on Hyper Suprime-Cam data.
- Resolve issues discovered in Single Frame Processing of crowded fields with DECam data.
- Investigate issues with coaddition of crowded fields.
- Create a Task for generating fake sources for AP processing.

#### 02C.03.05 – Tools for Science Pipelines

- Redesign definition of a filter to support conversion to Gen 3 middleware.

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

- Continue to develop tests and analysis tooling.
- Create test data and interfaces for “hack week” with members of the Minor Planet Center.

#### 02C.03.07 – Transform Fitting on Stacks of Images

- Apply proper motions and parallaxes to matched sources in Jointcal.

#### 02C.03.08 – Integration

- No work is scheduled in this WBS element.

### Staffing update

- John Swinbank, AP team T/CAM and DM Deputy Project Manager, will leave the project in October 2020. Ian Sullivan (University of Washington) will replace him as AP team T/CAM, while Yusra AlSayyad (Princeton University) will assume overall management of Science Pipelines development.

- Spencer Nelson will be on paternity leave through the end of calendar year 2020.

## 1.02C.04: Data Release Production

### Current accomplishments

#### 02C.04.00 – Management and Leadership

- The team chaired sessions and presented at the [Project and Community Workshop](#).
- Leadership prepared for and participated in the [Joint Status Review](#).

#### 02C.04.01 – Software Primitives

- “Generation 3” middleware development [[DM-23737](#)]:
  - Significant speedup of quantum graph generation by adding a system for pre-fetching, caching, and in-memory joins of dimension information [[DM-24938](#), [DM-26430](#)]
  - This completes activity on [DM-25268](#); Work toward gen3 feature parity will continue on epic [DM-26786](#)
- Emergent Work [[DM-25271](#)]:
  - Remove a redundant function [[DM-25768](#)]
  - Add warning to statisticsStack’s mean if you supply weights, but don’t ask for a weighted mean [[DM-25253](#)]

#### 02C.04.02 – Calibration Products

- Convert existing calibration products to the [DMTN-148](#) model for Gen 3 migration [[DM-26452](#), [DM-26414](#), [DM-26316](#), [DM-26217](#)]. This completes activity [DM-25321](#).
- Continue reduction of, and pipeline development for, Auxiliary Telescope data.
  - Astrometric solutions can now be fit [[DM-24592](#)].
  - This completes activity [DM-25275](#); work will continue on epic [DM-26800](#).
- Continue detector characterization
  - Providing a physically motivated correction to PTC gain measurements. This work also will allow faster brighter-fatter kernel production [[DM-26453](#), [DM-25954](#), [DM-26352](#), [DM-26166](#), [DM-25711](#)]
  - This completes activity [DM-25276](#); work will continue on [DM-26805](#).

#### 02C.04.03 – Image Characterization

- Initial exploration on alternative astrometry algorithms [[DM-25277](#)] was completed with the report attached to [DM-26640](#). Work will continue on epic [DM-](#)

[26792](#).

- The epic to model and subtract scattered halos of bright stars [[DM-25278](#)] is completed with a Task in progress to stack a library of bright star stamps; Activity will continue on [DM-26793](#).
- The epic to integrate PIFF, [DM-25319](#), is completed with a preliminary prototype implementation of a PiffPsf that has not yet been demonstrated or merged to lsst\_distrib. Work will continue on epic [DM-26785](#).
- FGCM development [[DM-25273](#)]
  - Fix fgcmcal zero-point offset due to background offset [[DM-26085](#)]
  - This completes work on activity [DM-25273](#); Converting fgcmcal to gen3 will continue on [DM-26791](#)

#### 02C.04.04 – Coaddition

- No work was completed in this WBS element.

#### 02C.04.05 – Detection and Deblending

- Deblender development toward making Scarlet more robust [[DM-25269](#)]
  - Create regression tests for third party package Scarlet to protect Rubin from accidental upstream regressions [[DM-25060](#)]
  - This completes work on activity [DM-25269](#); development based on tract-scale testing will continue on [DM-26787](#)

#### 02C.04.06 – Characterization and Measurement

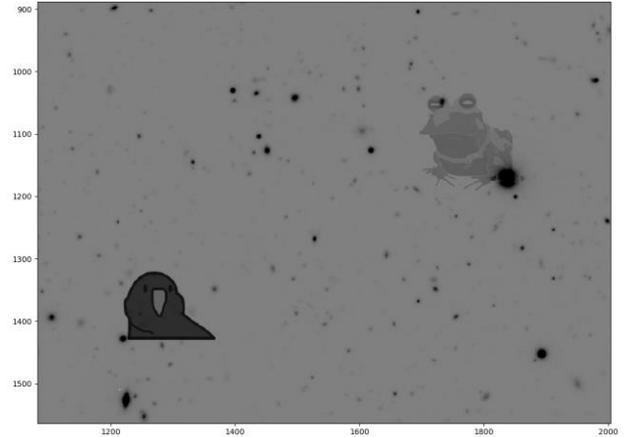
- Complete activity to add weak lensing “Rowe Statistics” to regular DRP processing [[DM-25320](#)]
- Continue Implementing Gaussian Aperture and PSF Photometry (GAaP) for measuring galaxy colors on [DM-26794](#)
- Complete activity to experiment with improvements to galaxy fitting during F20a on [DM-25274](#). Conclusions include that size priors only offer modest improvement to galaxy photometry.

#### 02C.04.07 – Maintenance, Quality and Documentation

- Added a “detect\_isPrimary” flag to Source Tables to provide users a clear way to filter duplicate sources, analogous to the same flag in the Object Table [[DM-25782](#),

[DM-26438](#)]

- Augment DRP's QA tooling in the pipe\_analysis package and investigate of DRP pipeline performance [[DM-25272](#)]:
  - Enhancements to tasks for injecting simulated sources into precursor data:
    - API updates and bug fixes [[DM-26418](#), [DM-25662](#)]
    - Add functionality to inject arbitrary images into real precursor data [[DM-25747](#)] (*Right: Figure credit Sophie Reed. Images that are easily identifiable as simulated are shown injected into an HSC patch.*)



## Planned activities

### 02C.04.00 – Management and Leadership

- Participate in the DM Leadership Team virtual Face-to-Face meeting.

### 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, including conversion to use “Generation 3” middleware.
- Compare data management’s calibration products with those produced by the camera team’s eotest.

### 02C.04.03 – Image Characterization

- Continue integration of PIFF.
- Continue conversion of FGCM Task to Generation 3 middleware. Make FGCM the default photometric calibration algorithm for the HSC precursor data set.

- Investigate effects of various astrometric models.

#### 02C.04.04 – Coaddition

- No work is planned on this WBS

#### 02C.04.05 – Detection and Deblending

- Begin validating Scarlet at tract scale (multiple square degrees)

#### 02C.04.06 – Characterization and Measurement

- Continue Implementing Gaussian Aperture and PSF Photometry (GAaP) for measuring galaxy colors

#### 02C.04.07 – Maintenance, Quality and Documentation

- Generate completeness plots for measurements on the coadds using simulated sources.
- Begin porting DRP plots and metrics used for internal testing to Generation 3 middleware.

### Staffing update

- Following the departure of John Swinbank in October 2020, Yusra AlSayyad of the DRP team will take over as Science Pipelines team manager.

## 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

- All team members attended the [Project & Community Workshop 2020](#) meetings.
- Mueller prepared for and participated in the [Joint Status Review](#).

- Mueller updated budgets and statements of work for the upcoming annual contract renewal.
- Mueller and Gaponenko had vacations this month.

#### 02C.06.01.01 Catalogs, Alerts and Metadata

- Salnikov published [DMTN-156](#), with results from Cassandra APDB tests [[DM-25641](#)].

#### 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 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.06.02.03 Query Services

- Gaponenko fixed issues in the Qserv Replication services which occurred on receipt of spurious packets [[DM-26486](#), [DM-26521](#)].
- Gaponenko improved batched mode of Qserv table contribution ingest [[DM-26530](#)].
- Gaponenko deployed the production version of Qserv to the new cluster at NCSA [[DM-25949](#)].
- Gaponenko fixed a master replication controller crash in the Kubernetes environment [[DM-25986](#)].
- Gaponenko fixed a possible race condition in the Qserv worker ingest service [[DM-26034](#)].
- Gates created a prototype system for distributed creation of test data [[DM-23716](#), [DM-26203](#), [DM-26409](#)].
- Jammes loaded DESC DC2 data into the Qserv cluster at CC-IN2P3 [[DM-24587](#)].
- Jammes enabled core file creation and gdb debugging with qserv-operator [[DM-26293](#)].
- Jammes merged a externally submitted pull request regarding Kubernetes network policies [[DM-26294](#)].

- Jammes upgraded the Qserv image inside the Qserv-operator [[DM-26426](#)].
- Mueller fixed the Qserv qhttp ajax unit test [[DM-26264](#)].
- Salnikov designed and implemented a new scale testing harness for the upcoming KPM50 tests [[DM-26100](#)].

#### 02C.06.02.04 Image Services

- Lo continued to refine ObsCore image metadata extraction tooling, in preparation for deployment of HSC ObsCore metadata at the LDF.

#### 02C.06.02.05 Catalog Services

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

#### 02C.06.03 Task Framework

- Salnikov improved handling of crashes in the PipeTask framework [[DM-25641](#)].
- Salnikov improved diagnostics on multiprocessing timeouts in the PipeTask framework [[DM-26230](#)].
- Pease implemented various improvements to the Butler Gen 3 command line tools [[DM-25101](#), [DM-25627](#), [DM-26388](#), [DM-26415](#)].

### Planned activities

#### 02C.06.00 Management & Leadership

- Mueller and Pease will participate in the [DM Leadership Team virtual Face-to-Face meeting](#), September 15-17.

#### 02C.06.01.01 Catalogs, Alerts and Metadata

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

#### 02C.06.01.02 Image and File Archive

- No work is planned for the 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.

#### 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:

- Participated in the Rubin Observatory Joint Status Review.
- Continued regular steering meetings for internal oversight over all technical areas of the LSST Data Facility.
- Contracts for FY20 completed into the project, and budgets for FY20 and FY21 complete into Uofl. With the winding down of construction, and operations funding is moving to IDF to support the number of people being supported by the project at NCSA is on a downward trend.
- 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 on a monthly cadence for each including running fakes which is a separate run.
- We have included Gen2 conversion to Gen3 butler on Postgres database.
- Preparing for RC3 as we can.
- Preparing to run RC2 with Gen3 environment on condor.
- Every monday a small HSC processing is run to check the stack against PostgreSQL and other needs of Butler/G3
- The testing of Condor Flock systems with Butler/G3 continues.

#### 02C.07.07 Data and Compute Services

- GPFS native on all K8 nodes. Stability and performance increased all around.
- Planning for large maintenance in September including new systems (Login nodes, and devl nodes to replace dev nodes, and submit nodes for condor, new release of Slurm installed)

#### 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 and ComCam in the test stand awaiting the mountain to power back on.
- Gen2 and Gen3 ingestion in the Observatory Operations Data Service at the Base Data Center for ComCam.
- 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). These are in the test stand at NCSA and at the BDC for AuxTel. The 2 milestones were achieved for producing 9raft images with proper headers and the AuxTel archiving service at the summit.
- 21 raft images were generated and sent for SLAC to NCSA and viewed by notebook on the Rubin Science Platform. This also means that the file was sent, ingested into Butler/G2 with a repo, and then the files were placed into file systems that can be viewed by the RSP systems.
- For Batch Production Services, 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, a new data transfer mechanism was tested and installed for ComCam and LATISS after testing on the test stand at the BDC. Changing the software from the SCP transfer protocol. This environment was also applied to all the other 6 installations of data transfer for data to NCSA. BOT and CCS from SLAC, NTS ComCam, NTS AuxTel, BDC AuxTel and summit AuxTel.
- We continued supporting data transfer from test stands at SLAC, monitoring file transfer and verifying data ingestion. NCSA continues with ingestion development now with Butler/G3.
- 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 work with Configurable SAL Components (CSCs) on the NCSA test stand and its 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 along with the September DMLT.
- Continue developing and documenting test plans for Data Facility components and services.
- Complete and upload to P6 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 continue to maintain four separate testing environments: NTS (NCSA Test Stand)

AuxTel, NTS ComCam, BDC (Base Data Center) AuxTel, and Summit AuxTel (currently disabled due to the Covid shutdown).

- We will continue periodic reprocessing of datasets in support of stack testing and pipeline development. We are moving forward with additional testing of Butler/G3 with postgres and the weekly stack builds.
- We will continue to develop the AuxTel software for raft scale images and then on to 21 rafts.
- Continue evolving the LDF services.

#### 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 have Butler/G2 and Butler/G3 ingests for the OODS and ingestion at the LDF ready when needed.
- 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 K8 systems from the mountain 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..
- 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 (NTS).
- 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 September, will include configuration for systems to support ComCam

at the base data center, as well as systems that will support LATISS on the summit if turned back on.

- We will investigate Prompt service software to be able to execute small jobs on the commissioning cluster and at NCSA and then expand from there in a service architecture. .
- We will continue responding to incidents and requests concerning as-is development systems and services, as well as addressing emerging hardware and system needs.

## Recruitment update

- none to report

## 1.02C.08: International Communications & Base Site

### 02C.08.01 – Infrastructure

- Provided Phishing training for all Rubin staff. Participation reached 84% and had a positive approval of the selected training modules. A report of the exercise is in the works.
- Several long-standing fixes have been applied to the deployment platform. This includes fixes to the Puppet platform, IPA, DNS, etc. and removal of multiple leftovers in the network.
- Created a layer 3 diagram of all components of the base datacenter. The diagram includes IP addresses, port speeds, vlans, etc.

### 02C.08.03 – Long-Haul Networks

- Continued La Serena - NCSA network tests, adapted due to delay of implementation of 40G upgrade of La Serena - Santiago Secondary link.
- 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.
- Jeff Kantor announced that effective Oct. 1, 2020 he will be reducing his effort to

25% Rubin Obs and 25% to other projects at NOIRLab. Cristian Silva [csilva@lsst.org](mailto:csilva@lsst.org), the Assistant Director of IT for Rubin Observatory, and based in La Serena, will be the co-chair of the NET team. Budget authority will be transferred to Cristian Silva. Cristian will handle annual amendments to contracts beginning in FY22. Jeff will be briefing Cristian and Bob Blum on remaining contract issues and start the process of identifying a successor co-chair for the SAACC after the next committee meeting. Monthly reports are to go to Cristian and cc Jeff. Jeff will remain available for information on Rubin. Eventually the Rubin Observatory IT organization will transition to NOIRLab, along with Gemini and NOAO. Chris Morrison will be head of NOIRLab.

#### 02C.08.03.01 – Chile National WAN

- 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:
  - We have been delayed in upgrading the secondary from 4 G to 40 G by the COVID 19 shutdown in La Serena. This will soon begin to impact network testing and verification and we recommend this upgrade be classified as essential work and scheduled for completion as soon as possible after reopening.
- 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 Vera Rubin Observatory NET monthly meeting was held on August 20, 2020.
  - Conferences and Workshops
    - Jeronimo Bezerra, Italo Valcy, Vasilka Chergarova, Daniela Toranzo,

and Renata Frez, participated in a virtual Kytos summit on August 13, 2020. The one-day workshop included current development of the Kytos SDN platform, improvements, new functionalities, and strategy for future development.

- Julio Ibarra and Vasilka Chergarova attended the virtual CI/CS Workshop on Tuesday, August 18 and Thursday, August 20, 2020. The workshop was focused on the cyberinfrastructure and cybersecurity needs of National Science Foundation research facilities and other research projects.
- *Miami - Boca Raton - Atlanta:*
  - The links remain operational and working properly. The resilience of the Miami (MI1) to Boca Raton (MI3) path was increased with the addition of a new fiber route provided by Crown Castle. The fiber route is 134 km long and it was connected to the Ciena 6500 together with the new Ciena OPS module (Optical Protection Switch). As it is a point-to-point system, each side (MI3/Boca Raton and MI1/Miami) needs a Ciena 6500 chassis plus an OPS module to light up the backup fiber route and protect the backhaul between Miami and Boca. The maintenance took place on August 10, 2020. After the activation, three tests were performed to evaluate the convergence time from the primary fiber route deployed in 2019 to the backup fiber route. Total convergence time (or outage time) is less than 100ms.
- *São Paulo - Miami Spectrum:*
  - The links remain operational and working properly.
- *São Paulo - Santiago Spectrum:*
  - Silica has installed the Vera Rubin 100G wave from Buenos Aires to Santiago. Century Link is provisioning the 100G wave from Porto Alegre to Buenos Aires with ETD of September 2020. Century Link will deliver a lit 100G channel for immediate use.
  - From Porto Alegre to São Paulo, RNP is working with carriers and ISPs to swap fibers in the segment from Porto Alegre to São Paulo. The goal is to have a solution by the end of 2020. It was discovered that RNP has employed Huawei equipment on this segment and we are exploring alternative paths/approaches that do not traverse any Huawei equipment.
- *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 the end of FY20.

## Planned activities

### 02C.08.03 – Long-Haul Networks

- Continue network tests with synthetic data (perfSonarI, iPerf) of data transfers between Chile and NCSA, and Interim Data Facility as soon as known.

#### 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

- IT Engineer position published, it will remain open until September 10th
- Luis Corral left Rubin Obs. on August 31st.

## 1.02C.10: Science Quality and Reliability Engineering

### Current accomplishments

- We now can store data aggregated from [Apache Kafka](#) (such as the Engineering Facilities Database telemetry stream) to [S3](#) in [Apache Parquet](#) format (*see figure*)
- We added a user guide listing to [lsst.io](#) in preparation for exposing the service more to external users.
- We address a number of issues with the logging service. However this service was provided as a temporary solution with the plan to eventually use our infrastructure provider's logging service, so we would like to find a way to phase it out.
- We investigated using private container registries in parallel to [Docker Hub](#), which was an unplanned requirement driven by Telescope & Site use of Nublado in combination with their non-open-source, licensed packages. There were a number of challenges uncovered so the work will be implemented next month.
- We evaluated [COmanage](#) as a solution to identify management requirements for the Science Platform ([SQR-045](#))
- We published security guidelines for our Kubernetes clusters ([SQR-048](#))
- We Improved JupyterLab 2.0 compatibility for Nublado, our Science Platform notebook service
- We published a document ([SQR-047](#)) as a commentary to Nublado requirements
- We made some changes to our documentation platform, LtD, to bring it more inline in how we now package and continuously integrate services.

## Plotting the aggregated stream

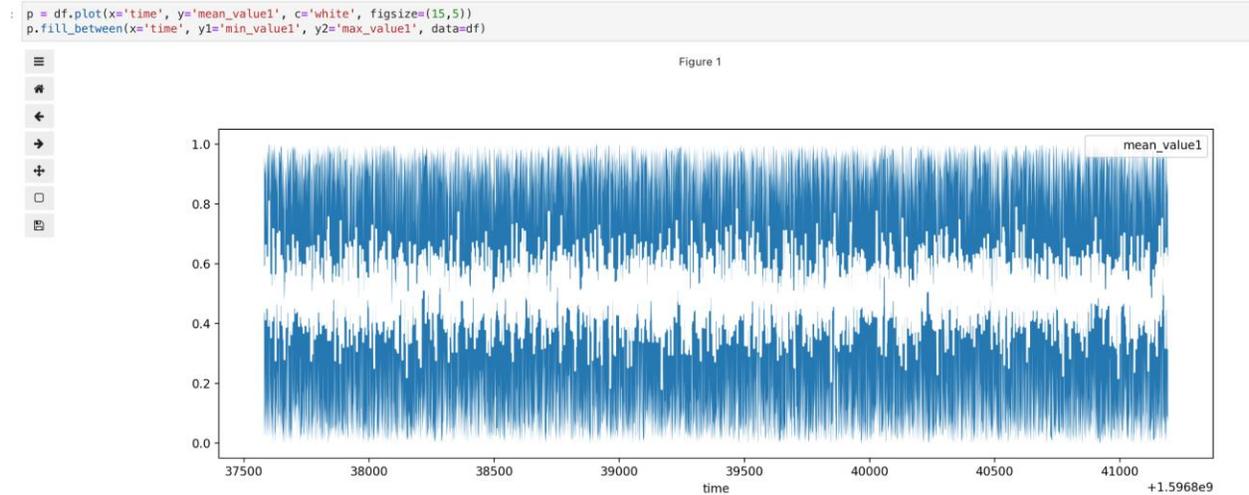


Figure: extract from a notebook running on our Science Platform demonstrating how to access the parquet repository of aggregated telemetry streams. The blue lines demonstrate a 10Hz stream and the white line plots the mean in 1Hz aggregated windows.

## Planned activities

- Deploy Nublado with JupyterLab 2.0
- Add the SQuaSH applications to lsp\_deploy
- Create a Helm chart for the SQuaSH API
- Convert the Science Platform to use letsencrypt certificates
- Change Nublado to be able to pull from private container registries

## Staffing update

- No changes to report.