

Data Management Monthly Report

January 2017

High-level Summary

The Project Science Team delivered on a coordinated study with JPL that confirmed prior estimates of NEO and PHA discovery. The AP, DRP, SUIT, and DAX Teams continued to develop and enhance the LSST Software Stack. The Networks & Base Site Team contractor Cobra/Telefonica completed phase 2 from the fork to Cerro Pachon and phase 3 from the fork to Cerro Tololo. The entire DM team also continued work toward commissioning, updating System requirements and clarifying Science Quality Assurance through prototyping, developing Verification Plans, and participating in LSST System-level reviews (e.g. Commissioning Review).

Project Status

Selected technical accomplishments:

- The Project Science team
 - Delivered (Jones et al.) the coordinated study with JPL (Chesley et al) on LSST's capabilities as an NEO and PHA discovery machine. Our findings are consistent prior estimates of an expected completeness of 68% for PHAs with $H < 22$ from LSST alone.
- The Architecture Team
 - Supported the Commissioning Review and worked on designs and documents including the overall DM System design, the QA Prototype Concept of Operations, the Data Release Production data flow, and the SuperTask component.
- The AP Team
 - Got back to working on the improved Wcs system.
 - Implemented tests for the jointcal photometry solver.
 - Identified and procured a test dataset to use in the forthcoming `validate_ap`.
- The DRP Team

- Completed development of the new Footprint system, for representing regions of pixels on images corresponding to e.g. astronomical sources. Work is ongoing to update the stack to use it.
 - Agreed a strategy for producing PSF-homogenized coadds as part of regular pipeline processing.
 - Prototyped a Non-negative Matrix Factorization (NMF) deblending algorithm.
- The SUIT Team
 - Added line chart to 2D plotting.
 - Finished time series data search, period finding, and display.
 - PDAC v1 is ready for access with VPN.
- The DAX Team
 - Continued support of integration activities at the PDAC via problem investigations and bug fixes.
 - Completed a synthetic dataset that will be used to gather Qserv KPMs at a level of 20% DR1.
- The Networks & Base Site Team
 - Cobra/Telefonica completed on phase 2 from the fork to Cerro Pachon and phase 3 from the fork to Cerro Tololo. Fiber installation on the AURA property for LSST use is now complete (except for fusion splicing along path in February and termination in Summit Facility when access is possible).
- The NCSA Team
 - Prepared for and participated in the Commissioning Plan Preliminary Design Review.
 - Completed a preliminary resource-loading and scheduling of the 2000-line WBS for the DM-wide replan. Attended the DMLT face-to-face meeting and presented NCSA's portion of the replan.
 - Held a working session with Systems Engineering to integrate our Data Products Production services model with other models of logical and physical systems within the LSST Enterprise Architect framework.
 - Finished provisioning a system to host the Camera Team's DAQ Test Stand.
- The SQuaRE Team
 - Completed SQuaSH integration with the new metrics framework.
 - Resolved problems affecting CI due to security and system changes at the data center. We also reviewed user concerns and formed a plan to address them.
 - Worked with the SuperTask team on architectural issues.
 - Adopted a new binary strategy with the DM System Engineer.
 - Beta-tested a microservice to create repositories from supported templates.
 - Augmented the OSX build node capacity.

Risk Management

The DM Risk Register was reviewed in the monthly process. No new risks were added and no significant changes to existing risk exposure were made.

Detailed Project Progress and Status

LSST Program Office

DM Project Management and Control

Current accomplishments:

The DM Subsystem Lead and the DM Project Manager:

Following the notice given on December 9th, 2016, Mario Juric has officially stepped down as the DM Subsystem Lead effective COB January 13th, 2017. Victor Krabbendam has taken over as the interim DM Project Manager until the arrival of Wil O'Mullane.

Planned activities:

The DM Subsystem Lead and the DM Project Manager will:

DM Science

Current accomplishments:

The DM Project Scientist: The DM Project Science staff has been engaged in completing three studies:

- ***The Large Synoptic Survey Telescope as a Near-Earth Object Discovery Machine***: This study (Jones et al.) shows the LSST would yield a completeness of about 68% for PHAs with $H < 22$, using the current LSST baseline survey strategy, consistent with original estimates. It is consistent with an independent assessment led by JPL and submitted to NASA (Chesley et al.). The manuscript pending before the LSST Publications Board (PUB-39), and will be submitted to Icarus in February.
- ***Optimal Star-Galaxy Separation*** (working title): This study (Slater et al.) will give a theoretical underpinning to the ad-hoc star-galaxy classifier in existing LSST code. It

will allow us to go beyond binary true/false star galaxy separation. So far we have been able to show our algorithm is close to optimal, and are in the process of deriving ROC curves that will allow our end-users to pick a combination of purity/completeness that satisfies their science case. We expect the completion of the study no later than March, followed by publication in a refereed journal.

- ***Simulating LSST Photometric Redshifts*** (working title): This study (Graham et al.) examines how LSST survey parameters might impact the quality of photometric redshifts. We have continued to incorporate feedback and improve the methodology and analysis; the paper will be ready for internal redistribution in February, with submission to a refereed journal to follow. In related work, we prepared the initial report on the optimal photo-z that should be include in Level 2 objects database (<http://ls.st/DM-6367>) including some simulations (<http://ls.st/3oe>). We conclude that the allotted space in the Level 2 database will be sufficient to hold adequate information about the photo-z.

We have also been engaged in organizational activities, including defining the structure of the DM Project Science group, and the R2A2 assignments for the DM Project Scientist and the group staff. Mario Juric has transitioned back to DM Project Scientist role.

Planned activities, in order of priority:

- Develop documents on the Minimal Viable System and DM Science Verification, consistent with the R2A2 matrix developed in January.
- Continue working on studies currently in progress (Graham, Slater).
- On-board Level 1 Project Scientist (Eric Bellm).
- Begin Science Verification tests on the Prototype Data Access Center (new temporary resource, Suberlak).
- Define the internal review schedule for DM, to validate the suitability of post-deplan design as capable of delivering the necessary LSST science.
- Work on completion of changes to LDM-151.

DM System Engineering

Current accomplishments:

All:

- Attended the DMLT meeting January 10-12.

DM System Architect:

- Prepared and presented DM Commissioning activities at the Commissioning Dry Run and Commissioning Review.
- Discussed data flow during the Data Release Production with NCSA staff and prepared a draft document describing it.
- Began drafting a rewrite of LDM-148, the DM System Design document.
- Resolved a variety of coding standards RFCs.

Deputy System Architect:

- Attended a training course on UML and a two day meeting on Supertask.
- Made a detailed analysis of the current status of all the DM RFCs; closing, flagging, adopting or withdrawing where appropriate. This include TCT analysis of relevant tickets.
- Investigated the possibility of a DM risk register based in JIRA.

DM System Interfaces Scientist:

- Completed the initial draft of the 2017 “QA Prototype” system concept of operations and presented it to the DMLT.
- Updated the DM Confluence documentation on DM’s system-level ICDs; prepared a new Docushare collection with all the DM-related ICDs.
<https://docushare.lsstcorp.org/docushare/dsweb/Services/Collection-5201>.
- Discussed the NCSA group’s status w.r.t. “early integration” activities with the other LSST subsystems.

Planned activities:

DM System Architect and Deputy System Architect will:

- Continue the rewrite of LDM-148.
- Assess the current status of the requirements update to LSE-61, including flowing down of new requirements from LSE-29.
- Prepare for JTM (including the System Architect serving on the organizing committee).

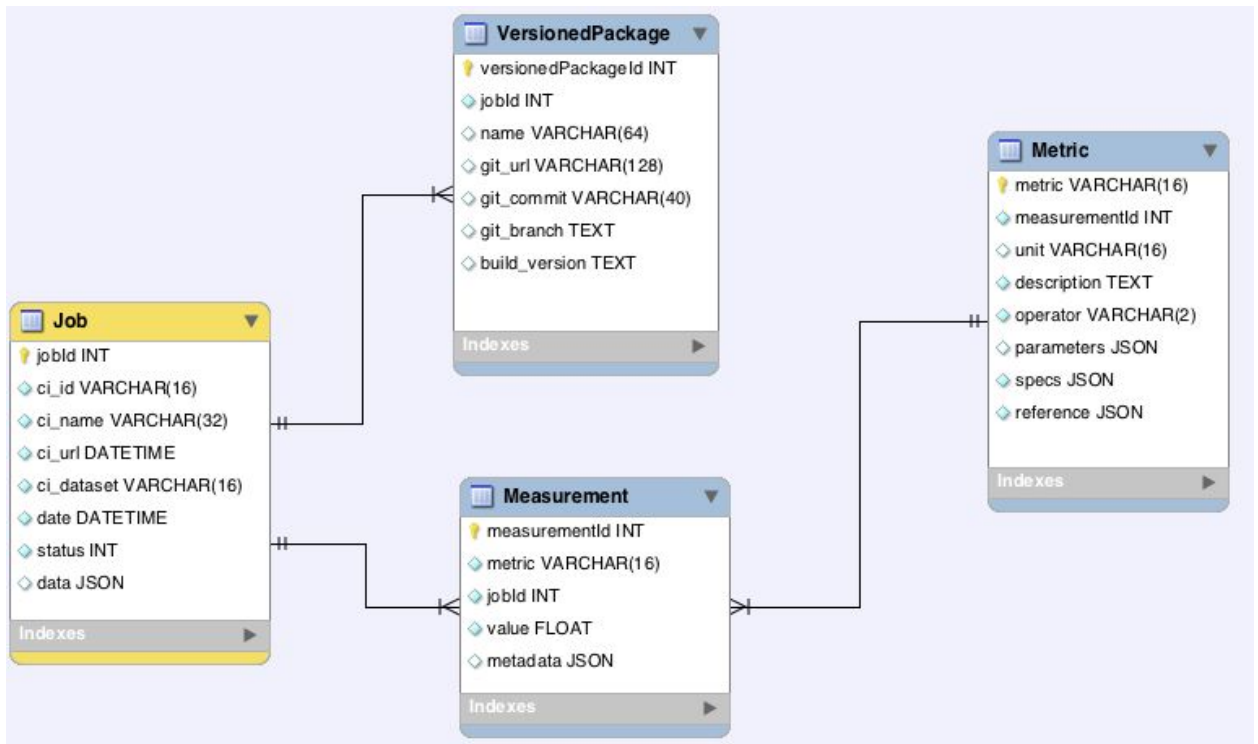
DM System Interfaces Scientist will:

- Create a working group for SuperTask/Butler design decisions.

DM Science Quality and Reliability Engineering (SQuaRE)

Current accomplishments:

- SQuaRE had a working visit from Andy Salnikov (SLAC) and Gregory Dubois Feldman (IPAC/DM) to flesh out aspects of the SuperTask of particular interest to SQuaRE, in particular the interface with the execution environment and as an entry point for documentation. During this meeting we also had a useful discussion with KT Lim and Nate Pease on related Butler functionality. We are looking forward to the Working Group in this area as we have a number of opportunities to work on reference implementations as part of SQuaSH development.
- We worked with the DM System Engineer on the issue of stack binary production. The problem with the current method (conda) is that while popular with users it is a little too fragile, but more importantly it is not used by the developers themselves, leaving SQuaRE as a bottleneck to triaging failures. We set the System Engineer a problem that would allow developers to integrate binaries into their normal workflow, which will both speed up their development (reducing local compilation time) and also allow them to check binary production through the normal development process (by making sure binary production has passed CI). This approach can bring us to the long-standing wish for nightly automated binary production. Tim Jenness recommended that we do this via the eups tarball mechanism and I decided to hold up the 13.0 release so we can switch to the new approach straight away. This was a bad call in retrospect, as there was a bug with that eups functionality that held us up past the point where a release should have been made.
- We identified a security flaw with Jenkins that exposes passwords to the process table. We provided a fix and our upstream patch was merged (see DM-9000 for details)
- We skated a bit ahead of the puck in our decision to store metrics as JSON in the QA database; while native JSON type support is available in databases such as MySQL and Oracle, in MariaDB they are in the current beta (10.2) which is not yet available through RDS. One solution would have been to fall back to MySQL, but this would have created the technical debt of moving to MariaDB again. Instead, we modified SQuaSH to store JSON as blobs in MariaDB 10.1 for now (see DM-8936 etc al). In combination with some database migration issues this delayed the new SQuaSH deployment.
- The major blocker to the SQuaSH release, the schema transition to support the new validation measurement API was completed (see DM-7043). This allows SQuaSH to



display semantically rich plots using our JSON datamodel. The new schema is shown above.

- Last year we introduced an OSX build node capability for Jenkins. This has been failing to keep up due to the longer OSX builds so a second host was added and ESXi configured to prepare it for being added to the Jenkins capacity. This should finally allow us to put OSX builds into general developer usage. A fault with the first host was also identified and fixed.
- Work was done to deal with the effect on SQuaRE services, Jenkins in particular, of the migration from the old lsst-dev to the new one at NCSA.
- A user note ([SQR-015](#)) was written as a guide to writing microservices for api.lsst.codes.
- Following the release of our chatbot infrastructure and our microservice platform api.lsst.codes last month, we are extending sqrbot with the capability of creating and publishing technotes with a single command. This allows technote writers to self-service and reduces demands on the SQuaRE documentation engineer and leverages the existing technote cookiecutter template. A microservice can also be bootstrapped via this method. This gives as an easy way to extend this to other areas where cookiecutter can reduce development friction, eg creation of stub eups packages.
- LTD Keeper's health is now being monitored.
- We are acutely aware of developer dissatisfaction regarding the Jenkins User Interface which is admittedly clunky and confusing. After analysis we have determined that the

most effective way forward is to move to the Jenkins [Blue Ocean](#) interface which is currently in beta. The new interface depends on Jenkins pipelines; while we are now using Jenkins pipelines to construct complex job, the earliest CI jobs, including the critical stack building jobs, are matrix jobs. We will schedule time in S17B to convert these so that we can transition to the superior interface that Blue Ocean provides.

- SQuaRE had planned a sprint for the next SQuaSH interaction to harvest metrics from tasks (so far our usecases have involved afterburners). In order to give the sprint focus, we are prototyping this functionality on metrics from jointcal, and John Parejko has agreed to stand as the product owner for that sprint.
- Produced conda binaries for the Sims group on request.

Planned activities:

- Re-attempt new binary production method.
- Deploy SQuaSH into production.
- Release metric monitoring service.
- Commence work on SQuaSH “jointcal” sprint.

Recruitment update:

Nothing to report.

University of Washington

Current accomplishments:

02C.03.00 - Alert Production Management Engineering and Integration

Krughoff attended the DMLT meeting in Tucson. Rawls attended the AAS meeting. Morrison attended the AAS meeting.

Rawls identified and procured the data necessary for implementing the equivalent validation system for AP that already exists for DRP.

Hiring

No activity.

02C.03.01 - Single Frame Processing Pipelines

Owen started back in on implementing the replacement Wcs system. This included wrapping the new system in pybind11. (DM-9118, DM-9133)

02C.03.02 - Association Pipelines

No advancements in this area.

02C.03.03 - Alert Generation Pipelines

Patterson updated the distribution system to allow for the different pieces: message queue, producer, and consumer, to reside on different hosts. This also allows for more scaling tests to be done (DM-8537, DM-7456).

02C.03.04 - Image Differencing Pipeline

Based on conversations from projects science, Reiss began more testing and verification of his techniques. In particular he is testing whether the two algorithms (modified A&L or ZOGY) converge in the correct way when the template has no noise (DM-8797, DM-8812).

Sullivan continued to work on optimal correction of DCR in the context of template generation. He is now trying a forward modeling technique that is complementary to the systems investigated before.

02C.03.05 - Application Framework for Exposures

Findeisen and Owen continued on with the pybind11 port. This has essentially finished off that effort with some amount of refactoring before merge (DM-8461, DM-8419, DM-8460, DM-8459, DM-8458, DM-8456, DM-8451, DM-8877).

Parejko, Krughoff and Owen also did significant work in the area of bug fixing and technical debt (DM-8827, DM-9014, DM-7532, DM-8825, DM-8978, DM-8708, DM-9041).

02C.03.06 - Moving Object Pipelines

No advancements in this area.

02C.03.07 - Photometric Calibration Pipeline

Parejko implemented tests for the jointcal photometric solutions. This, of course, requires figuring out how to turn photometric calibration on in jointcal (DM-8555, DM-8554, DM-8551, DM-8550).

02C.03.08 - Astrometric Calibration Pipeline

Morrison continued work on comparing a python implementation of the OptimisticB matching algorithm with that implemented in the stack (DM-8645, DM-6824).

Planned activities:

02C.03.00 - Alert Production Management Engineering and Integration

Krughoff and Connolly will attend the DESC meeting at SLAC.

Rawls will continue with implementing the minimum viable alert production system (**validate_ap**).

02C.03.01 - Single Frame Processing Pipelines

Owen will finish off the low level library to support the new Wcs. Findeisen will return to working on geometry objects necessary for completion of the Wcs work.

02C.03.02 - Association Pipelines

No planned work.

02C.03.03 - Alert Generation Pipelines

Patterson will continue scaling the producers and consumers on single node and multi-node distribution system.

02C.03.04 - Image Differencing Pipeline

Reiss will wrap up the recently triggered testing work.

Sullivan will flesh out the forward modeling approach to the DCR problem.

02C.03.05 - Application Framework for Exposures

Bug fixes and technical debt.

02C.03.06 - Moving Object Pipelines

No planned work.

02C.03.07 - Photometric Calibration Pipeline

Parejko will implement the ability to use different reference catalogs for astrometry and photometry and attempt to implement a spatially varying photometric zeropoint model.

02C.03.08 - Astrometric Calibration Pipeline

Morrison will determine whether the current matcher can be used for all our matching problems.

Recruitment update:

N/A

Princeton University and University of California, Davis

Current accomplishments:

02C.04.00 - Data Release Production Management Engineering and Integration

- Travel & meetings:
 - Swinbank & Bosch took part in the DMLT meeting in Tucson during the week of 9 January. Swinbank presented material on the DRP group's development methodology and the status of the DM replan as it pertains to DRP. Bosch presented ideas about the future development of [LDM-151](#), the DM Applications Design Document.
- Data Management replan [[DM-8494](#)]:
 - The current status of the DRP replan was presented at the January DMLT face-to-face.
 - The Calibration Products Production section of [LDM-151](#) has undergone substantial updates in response to suggestions by members of the DMLT. These changes are still undergoing review at time of writing. [[DM-8066](#), [DM-8496](#), [DM-9131](#), [DM-9132](#), [DM-9136](#)]
- Task documentation prototype [[DM-8559](#)]:
 - During this month, initial templates for Task documentation were produced, based on the December discussion on documentation for Science Pipelines. [[DM-8636](#)]
 - Effort is now ongoing to produce documentation for a few example Tasks. This will be shared with Science Pipelines developers for feedback and refinement, ultimately leading to a fully-fledged "how to" manual for Task implementers and documentation authors. [[DM-8717](#), [DM-8723](#)]
- Quality assurance [[DM-8299](#)]:
 - During this month, effort has focused on understanding why there is a discrepancy in the relationship between PSF and CModel magnitudes for the brightest objects. Investigation of this issue has prompted the development of QA techniques which will be incorporated into our standard workflow, as well as uncovering a number of minor bugs and opportunities for configuration improvements. [[DM-8533](#), [DM-9028](#), [DM-9107](#), [DM-9108](#), [DM-9109](#), [DM-9110](#)]

02C.04.01 - Application Framework for Catalogs

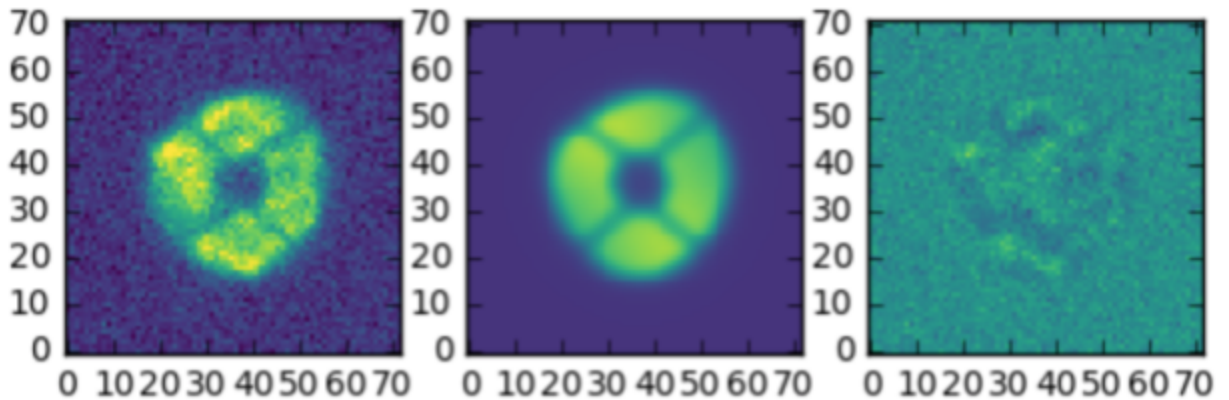
- Conversion from [SWIG](#) to [pybind11](#) [[DM-7717](#)]:
 - The bulk of packages included in the lsst_distrib “umbrella” have now been converted to pybind11, approximately following the expected schedule.
 - A number of stylistic improvements and performance enhancement techniques have been identified during the course of the port. These will be implemented before this work is merged. [[DM-9063](#)]
 - Documentation describing the standards which will be used for future wrapping with pybind11 has been drafted and, at time of writing, is undergoing review. [[RFC-281](#), [DM-8577](#)]
- Implementation of new Footprint system [[DM-3559](#)]:
 - The new Footprint code has been completed. Updated documentation is in progress. [[DM-8107](#), [DM-7177](#)]
 - The new Footprint API is not identical to the old system. Work is currently underway to port existing stack code to the new system. This must be completed before the new Footprints can be fully adopted. [[DM-8108](#)]
- Emergent work and reduction of technical debt [[DM-8136](#)]:
 - The shared-stack on the lsst-dev01 development machine now provides 2017-era weekly builds. This involved resolving a number of build problems. [[DM-8803](#), [DM-9005](#)]

02C.04.02 - Calibration Products Pipeline

- Auxiliary telescope development [[DM-8151](#)]
 - Prototyping of automatic processing of spectra from observations on the CTIO 0.9m telescope has been proceeding successfully. This will inform the design of the pipeline being constructed to reduce data from LSST’s calibration telescope. [[DM-9139](#)]

02C.04.03 - PSF Estimation

- Wavefront measurement and PSF reconstruction [[DM-1909](#)]:
 - It is now possible to detect and extract “donut” images of out-of-focus stars using the LSST stack. The resulting measurements have been used with a [GalSim](#)-based forward modelling approach to measure the Zernike coefficients which are used to characterize the wavefront. Detailed wavefront characterization is still ongoing. [[DM-8644](#), [DM-8565](#)]



An example of fitting a “donut” corresponding to an out-of-focus star in a Hyper Suprime-Cam image. From left to right: the object being measured, the model fit, and the residuals (object / model). Axis units are pixels; the colour scale is arbitrary.

02C.04.04 - Image Coaddition Pipeline

- PSF homogenization [[DM-8289](#)]:
 - A plan for including PSF-homogenized coadds in default LSST pipeline processing has now been adopted. Work on implementing this plan is currently underway. [[RFC-272](#), [DM-8491](#)]
- Warped image comparison [[DM-8290](#)]:
 - A set of PSF-matched coadds have been produced using the tools developed in [[DM-8289](#)]. These have been archived for use in future testing, primarily in the development of the warped image comparison code, but also in deblender development.

02C.04.05 - Object Detection and Deblending

- Deblender development [[DM-8140](#)]:
 - Adjustments to the deblender API were made to make it more convenient to swap in multiple different algorithms for testing. [[DM-8694](#)]
 - A Non-negative Matrix Factorization (NMF) deblender was prototyped, producing promising results in initial tests. Effort over the next sprint will focus on continued development of the NMF deblender, testing its performance relative to the current state-of-the-art. [[DM-8831](#), [DM-8959](#)]
 - A technique to produce simulated galaxy images for use in deblender testing was developed, while a number of known failure modes of the current codebase are being compiled for use in evaluating new techniques. [[DM-8624](#), [DM-8059](#)]

02C.04.06 - Object Characterization Pipeline

- Emergent work and reduction of technical debt [[DM-8306](#)]:
 - [Garmilla proposes](#), in his Princeton PhD thesis developed using Hyper Suprime-Cam data, a technique for star/galaxy separation which is applicable to LSST. We have retrieved Garmilla's prototype code, documented it, and adapted it for use with data produced by the current version of the LSST stack. Full integration with our processing workflow has not yet been undertaken. This code will likely not be run as-is in operations, but it is expected to serve as a useful basis for future classification work. [[DM-9001](#), [DM-9002](#), [DM-9125](#), [DM-9134](#)]
 - New astro- and photometric reference catalogs, based on [Pan-STARRS1 data release 1](#), have been made available for use in LSST processing. [[DM-5989](#)].

Planned activities:

02C.04.00 - Data Release Production Management Engineering and Integration

- Continue to work on the DM replan:
 - Updates to [LDM-151](#), following the discussion at the DMLT meeting in January and as requested by senior management.
 - Assemble a fully integrated plan for DM, in cooperation with Kevin Long (Project Controls Specialist), the DM T/CAMs and the rest of the DMLT.
 - Take advantage of the visit of Guyonnet to Princeton (below) to flesh out planning for Calibration Telescope data reduction.
- Travel:
 - Fisher-Levine and Rykoff will attend a calibration meeting at Harvard during the week of 30 January.
 - Guyonnet will visit Princeton for the week of 6 February.
 - Bosch and Meyers will attend the DESC meeting during the week of 13 February.
 - Meyers will visit Princeton for the week of 20 February.
- Circulate task documentation plans to stakeholders, and act upon feedback.
- Continue tracing down QA issues.

02C.04.01 - Application Framework for Catalogs

- Verify the pybind11 port in complete end-to-end pipeline runs, and resolve remaining technical issues.
- Complete transitioning existing application code to the new Footprint system.

- Cooperate with the Data Access, System Architecture and other groups in clarifying the design of the SuperTask, Butler and other Task Framework systems.

02C.04.02 - Calibration Products Pipeline

- Develop a more detailed model and understanding of producing flat fields using the collimated beam projector.
- Continue reducing CTIO 0.9m spectroscopic data to inform calibration system development.

02C.04.03 - PSF Estimation

- Continue working to reconstruct PSFs based on out-of-focus Hyper Suprime-Cam data.

02C.04.04 - Image Coaddition Pipeline

- Complete the work required to fully and correctly generate PSF-matched coadds.
- Start detailed planning for work on warped image comparison.

02C.04.05 - Object Detection and Deblending

- Develop extensions to existing NMF deblender prototype, and apply them to a variety of challenging test cases. Compare results against the existing state-of-the-art.

02C.04.06 - Object Characterization Pipeline

- No work is planned in this WBS element.

Recruitment update:

- Interviews for the Princeton “Postdoctoral Positions in Software” were carried out in the first week of January. Two candidates have been offered positions working on the LSST Science Pipelines. We expect responses by mid-February.
- Vishal Kasliwal, currently employed at the University of Pennsylvania and working 50% of his time on LSST, will leave the project at the end of February.
- Tim Morton will join the Data Release Production group in Princeton at the beginning of February.

IPAC / California Institute of Technology

Current accomplishments:

02C.05.00

- Continued to work with IPAC IRSA group on collaboration in Firefly development, plan and schedule coordination. IRSA released the first application using JavaScript based Firefly.
- Xiuqin, David, and Gregory attended DMLT face-to-face meeting at Tucson, AZ.
- Continue to work on SUIT requirements and resource estimates to work packages.
- Lijun took 6 days vacation; Xiuqin took 4 days vacation.

02C.05.01 Basic Archive Access Tools

- PDAC:
 - Light curve data search, period finding, and display.
 - Light curve specific default values, image display in expanded mode.
 - Worked with DAX team in API refinement.
 - Ready for access with VPN. Basci Please see introduction here:
<https://confluence.lsstcorp.org/display/DM/Guide+to+PDAC+version+1>
- Java code refactoring and adding unit tests.
- Enable group downloading of images.

02C.05.02 Data Analysis and Visualization Tools

- Attended the weekly meeting with the UIUC camera visualization team.
 - Improved build process and resolved issues related to Tomcat8.
- New chart function: line chart, light curve viewer.
- Bug fixes in reverse axis for magnitude plotting, image display with multi-extension FITS.

02C.05.03 Alert/Notification Toolkit

- No new work done.

02C.05.05 User workspace

- No new work done.

Planned activities: (for Feb., 2017)

02C.05.00

- Continue to work with IPAC IRSA group on collaboration in Firefly development, plan and schedule coordination.
- Finalize the work packages for SUIT.

02C.05.01 Basic Archive Access Tools

- Continue working on SUIT design and design document.
- Junit tests for Java classes.
- PDAC development:
 - finish the connection between object search and light curve plotting for the object in PDAC v1.
 - Enable “all sky” search capability.

02C.05.02 Data Analysis and Visualization Tools

- Stop camera visualization support.
- Bug fixes and necessary changes in JavaScript API.
- Using stack task do to a forced photometry and Firefly to display the result, using Python to control the flow.
- New functions: asymmetric error bars.

02C.05.03 Alert/Notification Toolkit

- No work planned.

02C.05.05 User workspace

- No work planned.

Recruitment update:

- No activities.

SLAC / Stanford University

Current accomplishments:

02C.06.00 Science Data Archive and Application Services Management Engineering and Integration

- Mueller prepared for and attended DMLT meeting in Tucson.
- Mueller continued work to wrap up DM replan activities.

- Mueller and Hanushevsky continued review/improvement of sizing-model inputs.
- Team continued improving robustness of DAX services and otherwise supported SUI integration efforts in the PDAC.
- Gaponenko completed generation of 20% DR1 test dataset.
- Thukral worked to revise/update query selection for DR1 KPMs.

02C.06.01.01 Catalogs, Alerts and Metadata

- Salnikov ran L1/AP db prototype on SSD/NVMe machine at IN2P3, and produced some comparative performance plots. Performance is very similar SATA vs. NVMe. [DM-8966]
- Salnikov experimented with an improved index strategy for DiaObject. [DM-8965]

02C.06.01.02 Image and File Archive

- Lo and Gates verified the imgserv can properly retrieve images by image-id. [DM-8234]
- Lo began work extending imgserv API to handle data from multiple surveys in preparation for loading WISE data in the PDAC.

02C.06.02.01 Data Access Client Framework

- Pease continued work on Butler storage and format refactor.
- Salnikov continued work on SuperTask MVP.

02C.06.02.02 Web Services

- Van Klaveren continued work to revive metaserv service in the PDAC.

02C.06.02.03 Query Services

- Gates addressed column-alias query parse issues. [DM-8226]
- Gates addressed a scan scheduling issue that was causing load-spikes on the IN2P3 cluster. [DM-8983]
- Jammes added improved container deployment script templating. [DM-2580]
- Salnikov fixed a db connection leak in the wmgr service implementation. [DM-8990]
- Jammes loaded a 35TB test dataset into OpenStack environment at IN2P3, and connected it to the Qserv OpenStack provisioning scripts. [DM-8196]
- Mueller updated MariaDB eups components to aid in tracking down a query optimization issue. [DM-9098]

02C.06.02.04 Image Services

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

02C.06.02.05 Catalog Services

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

Planned activities:

02C.06.00 Science Data Archive and Application Services Management Engineering and Integration

- Mueller and Hanushevsky to continue review/improvement of sizing-model inputs.
- Team to continue improving robustness of DAX services and otherwise support SUIT integration efforts in the PDAC as necessary.
- Gaponenko to provide “repaired” SDSS Stripe 82 dataset in PDAC (coverage for holes in overlap area between IN2P3 and NCSA processing).
- Gaponenko to load WISE catalog data and stage WISE images in PDAC.
- Gaponenko to complete load of 20% DR1 test dataset.
- Gates and Thukral to begin taking 20% DR1 KPMs.

02C.06.01.01 Catalogs, Alerts and Metadata

- Salnikov to try parallel load experiment with L1/AP database prototype.

02C.06.01.02 Image and File Archive

- Lo to provide calexp cutout support in imgserv.
- Lo to finish extending imgserv API to handle data from multiple surveys in preparation for accessing WISE data in the PDAC.
- Lo and Van Klaveren to update documentation of existing web service APIs, and also document future directions for these APIs.

02C.06.02.01 Data Access Client Framework

- Pease to continue work on Butler storage and format refactor.
- Salnikov to continue work on SuperTask MVP.

02C.06.02.02 Web Services

- Van Klaveren to reload and deploy metaserv in the PDAC.

02C.06.02.03 Query Services

- Gates to address various Qserv bugs uncovered by ongoing PDAC efforts.

02C.06.02.04 Image Services

- No work is scheduled for this WBS element this month.

02C.06.02.05 Catalog Services

- No work is scheduled for this WBS element this month.

Recruitment update:

No recruitment activity this month. One full-time developer req still open.

NCSA / University of Illinois

Current accomplishments:

02C.07.00 Processing Control and Site Infrastructure Management, Engineering and Integration

Planning and Management

These activities involve the creation of project plans for the construction phase for the DM-wide replan, and planning for the commissioning and operations phases. NCSA leads planning for operations of the Data Products Production (DPP) department.

For the DM-wide replanning effort, we completed preliminary costing of all aspects of construction, service deployment, and ongoing service operations, and assigned these tasks functionally to construction, commissioning, and early operations. We tied schedule of service deployment to commissioning, early operations, and other project-level milestones. We presented on the status of NCSA's portion of the DM replan at the DMLT meeting, including potential opportunities for value engineering and scope control. During the DMLT we reviewed the WBS service-oriented framework with the incoming DM project manager.

For commissioning, we continued working with the commissioning planning team, including traveling to SLAC for the Commissioning Plan Preliminary Design Review. We also incorporated portions of the DM-wide replan that pertain to operating services for commissioning into the commissioning schedule and cost.

For operations planning, we continued to refine the DPP plan and costing based on feedback from the operations planning team. We met with representatives from the Fermi National Accelerator Laboratory to discuss areas of mutual technical interest and potential DOE staffing areas within DPP department.

DM-8517 Replanning, commissioning, and operations planning

Service Management and Monitoring

The Service Manager develops and maintains the service catalog, including service-level targets, service-level agreements, and principal technical and managerial liaisons. This includes monitoring and managing availability, capacity, and IT continuity of services, and other operational matters, as well as forming a definitive opinion about the satisfaction of each customer for the services provided. The service manager supports deployment of services to operation and service transitions.

We continued refining the prototype Service Catalog schema, with consideration of IT security risks, controls, and mitigation actions. We continued developing the DPP services model within the LSST Enterprise Architect (EA) framework, including a trip to Tucson for a working session with Systems Engineering to integrate our model with other models of logical and physical systems in EA, such as those developed within the Chilean ITC tiger team. We also began drafting service-level agreements (SLAs) for NCSA-provided services.

DM-8515 Service Management Definition and Refinement

Hardware acquisition and provisioning planning

These activities involve high-level planning and design of physical systems related to computation, storage, networking, administration, and IT security on which DPP services are running or will be deployed. This includes consideration of near-term construction phase needs, as well as preparation for commissioning and operational needs.

We continued work to design and implement the data backbone hardware, meeting with various vendors to develop an operationally robust set of technologies. We also began updating LDM-144 (Site Specific Infrastructure Estimation Model) to be consistent with planned architecture and provide improved comparison of alternative costing model designs.

DM-8511 Degenerate Backbone Design and Implementation

DM-8506 Costing Model Redesign

02C.07.01 Processing Control

Batch Production Services

These activities support the construction of services that enable the production of data products in a batch environment (e.g., Level 2 data products). The batch production services will execute processing campaigns on computing resources to produce the desired LSST data products, where

campaigns are defined as sets of pipelines (ordered ensembles of computational steps), inputs they are being run against, and methods handling their outputs.

In January work continued to build up the Batch Service façade with the ability to orchestrate and run more complex pipelines. We also refined the use of the SLURM workload manager coupled with the HTCondor batch system and Pegasus workflow management software for handling job submission and execution. Work also continues in preparing a production-scale batch processing service for processing HSC data when the data become available at the end of this February.

DM-8332 Run selected science pipeline using Pegasus

DM-8333 Run selected existing science codes with initial Batch Production Service façade

Level 1 Services

These activities support the construction of services that will support Observatory Operations and enable the production, transport, and archiving of Level 1 data, including the nightly stream of images and events collected and processed in near real time.

We continued to prepare for the OCS-TCS-DM-CCS-DAQ Early Pathfinder Integration Activity #3 being coordinated by LSST Systems Engineering. We continued integrating the existing prompt processing system orchestration code with the newly developed front end elements for Level 1 services, and continued work on the OCS system, designing a logging/monitoring API as a first step in performance evaluation. We began working with Telescope & Site subsystem team members to understand the origin and contents of all OCS metadata that will be available for building image headers and that will be stored in the EFD.

DM-8575 Deploy Mock for OCS with monitoring info

DM-8574 Confirm and Integrate OCS headers

DM-8488 Prep for Integration Activity #3

DM-8320 Exercise DAQ-to-distributor portion of L1 prompt processing system

Common Middleware and Other Tasks

These activities consist of developing and maintaining general processing control software components, including common software used in production processing and by DM developers, as well as emergent unplanned support tasks.

Work in January involved improvements and fixes to common middleware code including code in the logging and ctrl_* orchestration packages.

DM-8518 Emergent Middleware Work (S17a)

02C.07.02 Infrastructure Services

Data Backbone

These activities support the construction of services which archive, catalog, and distribute data to compute resources and data access endpoints across all sites—ingestion, file management, database hosting, etc.

We continued to work with SLAC, the EFD architect, and other parties to detail the operational requirements of hosting 60+ identified databases across the NCSA and Chilean sites, as well as provenance and metadata catalogs that are required for implementing file ingestion, distribution, and retention services within the Data Backbone. We also continued collecting detailed requirements and use cases of the Data Backbone internal file services and service endpoints. These requirements will feed into a functional design which identifies specific software elements to be constructed.

DM-7645 Prototype bookkeeping and interfaces for the data backbone

DM-8331 Create design document for file portion of data backbone

02C.07.03 Environment and Tools

QA Environment

These activities provide environments and tools to support general DM developer activities.

We continued work to provide access to datasets and a Jupyter hub instance on the NCSA Nebula cluster to support upcoming QA activities.

DM-8330 Instantiate VM with Jupyter Hub and GPFS access

02C.07.04 Site Infrastructure

Hardware Acquisition, Deployment and Provisioning

These activities consist of the acquisition, provisioning, configuration, maintenance, and decommissioning of physical capabilities and associated systems which support LSST services running at NCSA.

The move from NFS to GPFS continued with further migration of users and data to the new storage systems. Configuring of a system to host the DAQ test stand was completed, including network configuration, file systems, security evaluation, and providing credentials for Camera system staff access. We also continued work updating the hardware cost model and system data management policies.

DM-8509 DAQ Test Stand API Node Deployment

DM-8503 Migrate Users from NFS

DM-8502 Disaster Recovery for Science Datasets

DM-8508 Finalizing / Publishing “Data Management” Policies

Service Management for LSST development resources, Emergent and Miscellaneous Work

These activities consist of supporting and communicating with users of the current services provided by NCSA to LSST DM developers.

Work for January included addressing user- and developer-driven requests and issues as the development systems continue to mature, improving performance of the verification and compute nodes, upgrading Nebula instances, working with Kubernetes containers, and related system needs.

DM-8519 Service Management & Emergent Work (S17a)

Planned activities:

02C.07.00 Processing Control and Site Infrastructure Management, Engineering and Integration

Planning and Management

For the DM-wide replanning effort, we will continue development of the WBS as design requirements and concerns are addressed, refine the first-pass estimates to a detailed analysis of resource needs, and integrate the plan with the commissioning plan, overall DM plan, and other project milestones and business goals. We will work to create a more robust model of data retention, disaster recovery, and capacity management requirements, and we will work with Systems Engineering to execute change requests submitted to the LSST CCB representing value engineering opportunities. We will begin to migrate the new LDM-230, which is a collection of concepts of operations for DPP services, into the LSST documentation framework. We will continue to refine the commissioning plan within the commissioning planning team. For operations planning, we will continue to refine the DPP plan and costing as requested within the operations planning team. In February we will also prepare a detailed plan for the second half of the S17 cycle, which will begin in March.

DM-8517 Replanning, commissioning, and operations planning
DM-7632 Develop planning packages from DM replanning deliverables

Service Management and Monitoring

We will continue refining the Service Catalog schema, developing the structure of Service Level Agreements for NCSA-provided services based on ITIL/ITSM standards. Development of the DPP services model within the LSST Enterprise Architect framework is ongoing. Work on service monitoring will continue with configuring and testing performance thresholds, developing graphical representations of collected monitoring data, and evaluating alternate technologies for collecting and displaying the data.

DM-8516 Enterprise Monitoring Integration (1)
DM-8515 Service Management Definition and Refinement

Hardware acquisition and provisioning planning

We will continue work to design and implement the data backbone hardware, working with vendors and documenting the design. We will begin planning for hardware deployment in Chile that will occur later this year on which we will deploy project-wide network-based security services, bringing together elements of computing hardware, networking, storage, security, and system administration.

DM-8511 Degenerate Backbone Design and Implementation
DM-8506 Costing Model Redesign
DM-8512 Planning Deployment in Chile

02C.07.01 Processing Control

Batch Production Services

In FY17 we plan to stand up the façade of Batch Production Services. In February we will continue to build up this service façade with the ability to orchestrate and run more complex pipelines in the planned workload management framework using Pegasus and HTCondor, adding orchestration elements as they become available. This includes ongoing work to detail the interface between the production system and supertask. Work also continues in preparing a production-scale batch processing service for processing HSC data in the as-is HSC framework, expected to become available late February.

DM-8332 Run selected science pipeline using Pegasus
DM-8333 Run selected existing science codes with initial Batch Production Service façade

Level 1 Services

We will take part in the OCS-TCS-DM-CCS-DAQ Early Pathfinder Integration Activity #3 being coordinated by LSST Systems Engineering. We will continue integrating the existing prompt processing system orchestration code with the newly developed front end elements for Level 1 services, and continue work on the OCS system, prototyping logging/monitoring as a first step in performance evaluation. We will also continue working with Telescope & Site developers to identify sources and contents of the EFD and OCS metadata available to build headers. Under consideration is a design for acquiring header information from OCS in “catch-up” mode, in which archiving of camera data occurs asynchronously from data acquisition by the camera.

DM-8575 Deploy Mock for OCS with monitoring info

DM-8574 Confirm and Integrate OCS headers

DM-8488 Prep for Integration Activity #3

DM-8320 Exercise DAQ-to-distributor portion of L1 prompt processing system

Common Middleware and Other Tasks

Work for February is anticipated to include changes to common library functions necessary to support Batch Production and Level 1 Services, and addressing further user- and developer-driven issues as they arise.

DM-8518 Emergent Middleware Work (S17a)

02C.07.02 Infrastructure Services

Data Backbone

We will continue to work with SLAC and the DM architect to gather requirements about the operational model of hosting the 60+ identified databases, including estimated query loads, capacity, and other administrative concerns. We will also continue to understand the interface of the Data Backbone with responsibilities of the offline batch processing system related to file movement to and from compute sites, file management external to the custodial file store, and movement of data between storage tiers (e.g., staging from tape to disk) supporting production, as well as serving data in the DAC and through the bulk export service.

DM-7645 Prototype bookkeeping and interfaces for the data backbone

DM-8331 Create design document for file portion of data backbone

02C.07.03 Environment and Tools

QA Environment

Work to provide a Jupyter hub instance and access to datasets to support upcoming QA activities will continue.

DM-8330 Instantiate VM with Jupyter Hub and GPFS access.

02C.07.04 Site Infrastructure

Hardware Acquisition, Deployment and Provisioning

Planned work includes finalizing the migration from NFS to GPFS file systems, beginning migration of services, and planning decommissioning of the NFS file system. We will also publish the as-is data management policies (currently under review). Work on implementing disaster recovery procedures for datasets will be completed.

DM-8503 Migrate Users from NFS

DM-8504 Migrate Services form NFS

DM-8505 Decommissioning NFS

DM-8508 Finalizing / Publishing “Data Management” Policies

DM-8502 Implementation of disaster recovery for /datasets (Phase 2)

Service Management for LSST development resources, Emergent and Miscellaneous Work

Work for February is anticipated to include addressing further user- and developer-driven requests and issues with migration from NFS to GPFS, and related system needs.

DM-8519 Service Management & Emergent Work (S17a)

Recruitment update:

Corey Eichelberger was hired as part of NCSA’s Network Engineering and Research division; he will devote a portion of his time to supporting networking for LSST systems being constructed and operating at NCSA.

AURA

This section covers accomplishments and planned activities in WBS 02C.08 International Communications and Base Site.

Current accomplishments:

02C.08.00 International Communications and Base Site Management Engineering and Integration

02C.08.01 Base Center

- Summit Base ITC Tiger Team: Continued work on the Summit Base ITC Design document that will be proposed to the LSST CCB to become a baselined LSE document. <https://confluence.lsstcorp.org/display/SYSENG/LSST+Summit+-+Base+ITC+Design>
- Summit and Base Networking and Computing: A new transformer has been installed to add capacity in the NOAO Computer room in La Serena in order to install the DWDM equipment and Reuna routers during the interim to the new Data Center completion. The summit LAN RFP was issued to 6 companies in Chile. A selection committee was nominated, a score sheet produced, and we are working with IT Services to establish a drupal site for the procurement material.. Date for bid reception is 28 February, 2017. Purchased Fusion splicer, fiber termination kit and Optical Time Delay Reflectometer (OTDR) for Summit Network installation and test.

02C.08.02 Chilean Data Access Center

- No activity this month.

02C.08.03 Long-Haul Networks

02C.08.03.01 Chile National WAN

- Summits - AURA Gatehouse Network: Cobra/Telefonica completed phase 2 from the fork to Cerro Pachon and phase 3 from the fork to Cerro Tololo. Fiber installation on the AURA property for LSST use is now complete (except for fusion splicing along path and termination in the Summit Facility).



- **DWDM Equipment:** Owing to possible legal action against Raylex (the Chilean Coriant distributor) we backtracked on our original potential contract signing and moved to contract directly with Coriant to purchase the AURA/LSST equipment as FOB from Miami. The new price list was sent but need to clarify some configurations in the BOM. To speed the REUNA part of the purchase the procedure will be: sign the Amendment

between AURA-REUNA so REUNA can proceed to acquire from CORIANT/RAYLEX the DWDM equipment plus the installation services of the whole.

- Santiago-La Serena: Fiber is installed. Waiting for the formal delivery report from Telefonica, which is to include the total length, attenuation and PMD parameters.
- La Serena – AURA Gatehouse: Cobra/Telefonica completed fiber installation from the south side of Highway D41 to Gatehouse. We are still awaiting Ministry approval for the last 400m section along the D41 road north side and the crossing over the highway to the south side, which we expect to complete in February. Work started on the trenches and ducts that will be required to install the local loops for the new fibers from Telefonica into the AURA compound in La Serena.

02C.08.03.02 International Chile - US WAN

- 100 Gbps Managed Ring: No planned activity in this period.
- Management and Coordination Contract: The [AmLight SAACC Spring meeting](#) took place in January 10-11, 2017 at the Kovens Conference Center at the Florida International University, Biscayne Bay Campus (3000 Northeast 151st Street, North Miami, FL 33181). 41 attendees participated (23 in person and 18 remotely). The meeting gathered participants from several university, organizations and research institutions from USA, Latin America and Europe:
 - National Center for Supercomputing Applications (NSCA)
 - Center for Internet Augmented Research and Assessment (CIARA) at Florida International University (FIU)
 - Cerro Chajnantor Atacama Telescope (CCAT)
 - Brazilian National Research and Educational Network (Rede Nacional de Ensino e Pesquisa -RNP)
 - Brazilian e-science/astronomy virtual institute LINEA
 - Pathway of the Americas (AMPATH)
 - European Southern Observatory (ESO)
 - Florida LambdaRail – Florida's Research and Education Network
 - National Radio Astronomy Observatory (NRAO)
 - Internet2
 - National University Network Chile (Red Universitaria Nacional -REUNA)
 - Large Synoptic Survey Telescope (LSST)
 - Cerro Tololo Inter-American Observatory (CTIO), National Optical Astronomy Observatory (NOAO)
 - Ellalink Cabos Submarinos
 - Academic Network at São Paulo (ANSP), Brazil
 - Information Science Institute (ISI) at University of Southern California (USC)

- Cornell University
- Federico Santa María Technical University (UTFSM), Chile
- Vanderbilt University
- National Science Foundation (NSF)
- Latin American Advanced Networks Cooperation (Cooperación Latino Americana de Redes Avanzadas-RedCLARA)
- School of Integrated Science and Humanity (SISH) at Florida International University
- Energy Sciences Network (ESnet)
- The LSST Network Engineering day (Jan 11th) was also held in Miami to discuss the end-to-end network connectivity to support the Large Synoptic Survey Telescope (LSST) project. Because all LSST participant network operators have ongoing network-engineering efforts in place to support the LSST project, the main goal of this meeting was to define the next steps to interconnect all efforts towards a manageable, cost-effective, secure and scalable end-to-end network infrastructure. A report of the meeting results has been drafted and is under review by the participants. The meeting had the following goals which were all accomplished:
 - Introducing the network engineers involved from each network operator;
 - Detailing the physical and logical network configurations, interconnections and activations planned for the 2017-2018 time frame;
 - Discussing the status and next steps of the LSST End-to-End Test Plan;
 - Presenting all ongoing efforts to support the LSST network for the 2019-2020 time frame;
 - Align network deployments and dates with the LSST project plan.
- Spectrum Contract: No planned activity in this period.
- US National WAN: No planned activity in this period.

Planned activities:

02C.08.01 Base Center

- Summit Base ITC Tiger Team: Continue work to complete draft of Summit - Base ITC Design Document by the completion of the Joint Technical Meeting in March.
- Summit and Base Networking and Computing: Complete RFP question period for Summit Network and Computing.

02C.08.03 Long-Haul Networks

02C.08.03.01 Chile National WAN

- Summits - AURA Gatehouse Network: Complete the fiber installation with fusion splicing.
- DWDM Equipment: Purchase AURA/LSST DWDM equipment FOB from Coriant. Expected delivery now March and installation end of April for AURA/LSST equipment and end of June for REUNA equipment. REUNA to work on the deployment plan including the measurements to validate Telefonica's parameters using the DWDM.
- Santiago-La Serena: Waiting for the formal delivery report from Telefonica, which has to include the total length and attenuation parameters and is now expected in February.
- La Serena - AURA Gatehouse: Complete the highway 41 crossing and installation to the AURA Gatehouse. Continue conduit work to existing data center in AURA recinto.

02C.08.03.02 International Chile - US WAN

- 100 Gbps Managed Ring: There are no planned activities for this period.
- Management and Coordination Contract: Complete review and publish report of LSST Network Engineering Day.
- Spectrum Contract: There are no planned activities for this period.
- US National WAN: There are no planned activities for this period.

Recruitment update:

- Advertisements for the new hires Network Engineer, IT Technician and Systems Administrator were sent out early January with closing date of 15 February. Established selection committee and worked with HR to set up sharepoint site for selection committee to see applications, upload scoresheets, etc. Developed candidate scoresheet. Now expect hire date of May 31, 2017 for first two and June 30, 2017 for last one.