

Large Synoptic Survey Telescope Data Management Monthly Report

July 2017

High-level Summary

July 24-26 saw the first Camera Data Acquisition (DAQ) hardware, on loan from SLAC, installed at NCSA. The SLAC Camera Team visited and successfully installed the DAQ system and connected it to the NCSA-provided host machine. For further details of this event, see <https://www.lsst.org/news/camera-daq-installation-ncsa>. Testing of the DAQ system will continue next month.

While visiting NCSA, the team held a workshop on early subsystem integration. This has resulted in a refined plan for addressing this vital part of the construction plan.

DM put together a good review data package and presented the DM status successfully to the NSF/DOE review panel hosted by NCSA during the week of July 24. We await the report but the general mood at the review and initial feedback was very positive. DM will move forward on the momentum of this review.

A number of important documents were issued over this period. This includes a variety of technical notes, including both a description of DM's interaction with the Engineering & Facilities Database (EFD; DMTN-050) and the approach being taken by DM to providing photometric redshifts (DMTN-049), a series of test reports, and the Summit Computer Room requirements document. In addition, the Summit to Base ITC Design Document (LSE-309) was baselined by the LSST CCB.

Activities within the Alert Production group focused on the development of a prototype end-to-end system which will be used as the basis for future development and early verification of the AP system. This has included identifying appropriate test data and developing the necessary test harnesses.

The SynPipe synthetic object injection code was adapted to run within the framework of the DM stack. A new "robust" coaddition algorithm which includes advanced facilities for identifying and rejecting artefacts was demonstrated. DAX services are now compatible with Python 3 and the metadata service is more aligned with the IVOA VOResource standard.

The Prototype Data Access Center (PDAC) now incorporates group information from the NCSA CILogin system. This is an important step toward providing an integrated user

experience.

The Summit LAN procurement was awarded to Dimension Data and the initial order was placed. The AURA/LSST Dense Wavelength Division Multiplexing (DWDM) equipment arrived in Chile and will be installed in September.

Following hardware relocation and upgrades to the automated management jobs, we are once again providing continuous integration & test services on the macOS platform.

Risk Management

The DM Risk Register was reviewed in the monthly process. Attention was paid to the opportunities this time with one new opportunity identified (DM-325: New or different technology provides saving in hardware/effort). No new risks were added and no significant changes to existing risk exposure were made.

Detailed Project Progress

1.02C.01: System Management

Current accomplishments

The DM Project Manager prepared for and led DM through the NSF/DOE External review. This was the main activity of the month. The week before the review July 17-20, also in NCSA, the first Data Acquisition system was installed by SLAC and operated by DM. This was an important milestone for DM and as such the DM Project Manager attended.

Planned activities

The DM Project Manager will prepare for the Joint Status Review. This will focus on drafting responses to the DM Review recommendations. He will also prepare for and attend the Project and Community workshop.

1.02C.02.01: Data Management Science

Current accomplishments

We began simulations of improvements to LSST photometric redshifts when infrared data from e.g., Euclid or WFIRST, are included (Graham). Final comments on DMTN-049 were integrated and the technical note was issued ("LSST DRP (Level 2) Catalog Photometric Redshifts"; Graham et al.). Melissa Graham also prepared and delivered a presentation on the LSST Data Products to the TVS SC and continued participation in the LSST Community Science Center working group.

Colin Slater started developing a technical note on Level 2 database bandwidth requirements for various likely usage scenarios, and the implications for supporting Level 3 processing.

We began developing a test suite to quantify the efficiency of LSST codes in crowded fields, to support further development of crowded field processing strategy (Suberlak & Slater).

Juric was largely focused on contributions to the DM review, and coordinating the team.

Planned activities

- Preparations for LSST 2017 sessions
- Further testing of potential level 3 processing scheme modifications.
- Continued work on quantifying the performance expectations and options with respect to crowded field processing.

1.02C.02.02: DM System Architecture

Current accomplishments

The DM Software Architect:

- Prepared for and participated in the NSF/DOE DM Review, including presenting the overall DM architecture.
- Learned about the Camera Data Acquisition (DAQ) system API and discussed future integration testing during the DAQ Test Stand installation at NCSA.
- Discussed and worked out next steps for the Prototype Data Access Center and other LSST Science Platform deployments including Kubernetes and GPFS access.
- Discussed the Engineering and Facilities Database and wrote up a technical note (DMTN-050) with current understanding of how DM will handle it.
- Contributed to Summit-Base Tiger Team output document (LSE-309) and Network Engineering Team understanding of north/south data flows.

The DM Systems Engineer:

- Added LSE-299 to the SysML model.
- Extracted the historical content from LDM-135 and Trac into multiple tech notes, and test reports, thereby allowing LDM-135 to focus on the current database design.
- Prepared for and participated in the NSF/DOE DM review. This included giving a presentation; dealing with document review RFCs; and processing review documents.
- Submitted LCR-1024; tying DPDD into the requirements hierarchy via LSE-30.

The DM System Interfaces Scientist:

- Supported the DAQ hardware installation at NCSA, and facilitated the subsequent DAQ workshop that brought in additional TCS and CCS participation.
- Prepared for and participated in the NSF/DOE DM review, including both a plenary and a breakout presentation on the Science Platform.
- Further developed the “early integration pathfinder activities” plan involving DM, Camera DAQ and CCS, TCS, and OCS.

Planned activities

The DM Software Architect will:

- Update the product tree to add services and rationalize variants; incorporate into the DM subsystem model in MagicDraw.
- Attend LSST2017.
- Advise the DM Butler Working Group.

The DM Systems Engineer will:

- Complete LSE-61 verification matrix.
- Attend LSST2017.
- Modify sconUtils to use pytest for test execution.
- Process LCR-921.
- Start working on the DM Butler Working Group.

The DM System Interfaces Scientist will:

- Work on writing the SuperTask design document, DMTN-055.
 - Provide SuperTask-related input to the Butler Working Group;
- Attend LSST2017;
- Work with LSST Systems Engineering to incorporate the early integration pathfinder plan into PMCS.

1.02C.03: Alert Production

Current accomplishments

02C.03.00 – Alert Production Management Engineering and Integration

- Rawls, Reiss, Findeisen, Bellm and Krughoff worked on implementing the design of the minimum viable system and the framework in which it will run to eventually verify the AP system.

- Bellm identified metrics necessary for the initial verification system (DM-10922).
- Reiss worked on the data side of the AP verification system by defining the packaging for verification data as well as collecting and ingesting the data for the initial verification system (DM-11343, DM-11305, DM-11116, DM-11420, DM-11117, DM-11414)
- Rawls worked through configuration of the system to ensure image difference decorrelation is activated and that the linearity correction is being applied in ISR (DM-10980, DM-10974).
- Findeisen built the stubbed out version of the framework for running the verification system. It is now waiting on Rawls to provide the library the stubs can call to execute the various steps of the pipeline (DM-11118).

02C.03.01 – Single Frame Processing Pipelines

- Owen worked to replace the current Wcs classes with AST backed versions.

02C.03.02 – Association Pipelines

- Morrison started in on the difference image analysis object generation work. Sources are detected on each difference image and are then aggregated into object with lightcurves. Morrison designed and implemented the class to hold DIAObjects and their associated lightcurves (DM-11038, DM-11039). He also started designing the system that will associate new sources with known objects (DM-11283).

02C.03.03 – Alert Generation Pipelines

- Patterson continued on investigating the end-to-end alert distribution system.

02C.03.04 – Image Differencing Pipeline

- Sullivan processed some data containing known quasars to see how the DCR correction system would handle them, specifically if quasars could be separated from stars using a single band. Unfortunately, the data did not contain enough dynamic range in airmass to make the results very conclusive (DM-11036)

02C.03.05 – Application Framework for Exposures

- Parejko improved the output messages for table like objects (DM-6718).
- Findeisen improved verbosity levels for some stack code and updated the C++ documentation guidelines (DM-11237, DM-10220).

02C.03.06 – Moving Object Pipelines

- No advancements in this area.

02C.03.07 – Photometric Calibration Pipeline

- No advancements in this area.

02C.03.08 – Astrometric Calibration Pipeline

- Parejko added the ability to use a reference catalog for constraining the photometric fits (DM-11095).

Planned activities

02C.03.00 – Alert Production Management Engineering and Integration

- All will attend the project and community workshop in Tucson.
- All will continue to work toward a minimum viable system and the framework to verify it.

02C.03.01 – Single Frame Processing Pipelines

- Owen will finish work on the WCS replacement with the AST backed version.

02C.03.02 – Association Pipelines

- Morrison will implement DIAObject association classes including persistence.

02C.03.03 – Alert Generation Pipelines

- Patterson will work toward the minimum viable alert distribution system.

02C.03.04 – Image Differencing Pipeline

- Sullivan will work on the “stackification” of the DCR correction system.

02C.03.05 – Application Framework for Exposures

- Various technical debt.

02C.03.06 – Moving Object Pipelines

- No work planned.

02C.03.07 - Photometric Calibration Pipeline

- Parejko will finish the initial implementation of the spatially variable photometric model in Jointcal.

02C.03.08 – Astrometric Calibration Pipeline

- No work planned.

Recruitment update

- No activity.

1.02C.04: Data Release Production

Current accomplishments

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

- Travel & meetings:
 - Meyers and Rykoff attended the Dark Energy Science Collaboration Meeting at Brookhaven National Laboratory during the week of 10 July.
 - Bosch and Swinbank attended the [NSF/DOE Review of DM](#) in Champaign, Illinois during the week of 24 July.
 - Meyers spent two days during the week of 17 July in Princeton.
- Management & planning [[DM-10816](#)]:
 - Management effort during this month focused on preparation for and participation in the NSF/DOE Review of DM. Members of the DRP team delivered presentations in the plenary and Science Pipelines sessions of the review, as well as being on hand to answer ad-hoc questions and reviewer requests. [[DM-11046](#), [DM-11047](#), [DM-11048](#), [DM-11170](#), [DM-11278](#)]
 - An early draft of [LDM-534](#), the test specification for the LSST Level 2 Software System, was produced. This is expected to evolve substantially over the next several months. [[DM-10963](#), [DM-11160](#)]
 - In August, Swinbank will begin to assume the role of T/CAM for the Alert Production group at the University of Washington. In preparation for that, he has had a series of preparatory discussions with Simon Krughoff, the previous AP T/CAM. In addition, a [transition plan](#) has been developed to guide the handover of responsibilities.
- Forward Global Calibration Method (FGCM) port to LSST [[DM-10584](#)]:
 - The FGCM code for reading and writing data from FITS files has been removed, and, instead, the codebase now uses standard DM conventions for data access. [[DM-10856](#)]
 - A prototype mechanism for representing spatially and wavelength variable filter transmission within the DM stack is under development; a test version is expected to be available next month. [[DM-10860](#)]
 - Work is ongoing to create a command-line task to act as the entry point to FGCM. This is expected to be completed next month. [[DM-11313](#)]
- Emergent work and pipeline support [[DM-10382](#)]:
 - The [SynPipe](#) synthetic object injection system, originally written by Song Huang (UC Santa Cruz) has been adapted to run within the DM stack framework. [[DM-10848](#), [DM-11292](#)]

- It is now possible to use the ModelPsfMatchTask to match any type of PSF to any other type. Previously, this operation was limited to only a subset of the available PSF classes (specifically, SingleGaussianPsf and DoubleGaussianPsf). [[DM-10953](#)]
- A series of minor bugs were fixed and usability enhancements were made across the codebase. [[DM-2795](#), [DM-10901](#), [DM-11166m](#) [DM-11196](#), [DM-11223](#), [DM-11355](#), [DM-11360](#), [DM-11378](#)]

02C.04.01 – Application Framework for Catalogs

- Middleware and framework development [[DM-10586](#)]:
 - As part of the DRP contribution to the SuperTask Working Group, a number of design concepts were prototyped. These will feed into the ultimate working group report. [[DM-10600](#)]
- Emergent work and reduction of technical debt [[DM-10383](#)]:
 - The bitfield used to represent pixel masks that may be overlaid on images (to represent e.g. saturated or interpolated pixels) has been increased to 32 bits, doubling the number of masks that may be used on any given image. In the process, the API for working with masks underwent some refinements. [[DM-7477](#)]
 - The ci_hsc test package, which performs realistic pipeline processing on selected Hyper Suprime-Cam engineering data, has been updated to work with Python 3. This is a crucial step towards making the whole of the DM stack Python 3 compliant. [[DM-7324](#)]
 - Work was ongoing throughout the month to replace the framework used for performing statistical calculations throughout the codebase with a more flexible, usable and faster version. This work is expected to be completed next month. [[DM-10843](#)]
 - A series of minor bugs were fixed and usability enhancements were made across the codebase. [[DM-3705](#), [DM-7084](#), [DM-11010](#), [DM-11163](#), [DM-11236](#), [DM-11269](#)]

02C.04.02 – Calibration Products Pipeline

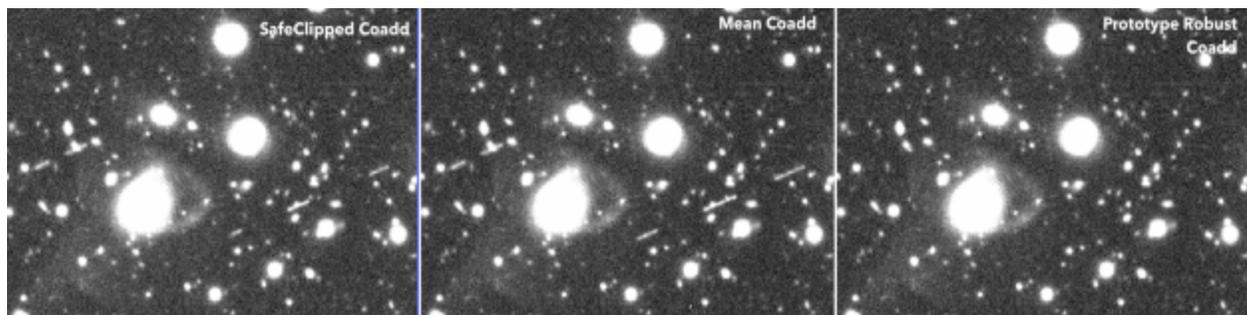
- Auxiliary telescope development [[DM-10581](#)]:
 - No significant progress has been made during this period due to unavailability of key personnel.
- Processing camera test stand data [[DM-10897](#)]:
 - Work has begun on verifying detector characterization using test stand data within the DM stack. Initially, this consists of cross-checking detector acceptance tests produced during sensor testing by the Camera Team to ensure that DM processing produces equivalent results. This work will be ongoing throughout August. [[DM-10895](#), [DM-11348](#)]

02C.04.03 – PSF Estimation

- Wavefront measurement and PSF reconstruction [[DM-10355](#)]:
 - Work continued to rotate the Zernike polynomials measured with respect to focal plane coordinates into an unrotated frame corresponding to the telescope. This work was not completed by the end of the reporting period. [[DM-10850](#)]
 - Work is underway on a technical note (“DMTN”) describing the work carried out to date in this area and presenting suggestions for future enhancements. This is expected to become available during August. [[DM-10355](#)]

02C.04.04 – Image Coaddition Pipeline

- Warped image comparison [[DM-8290](#)]:
 - An initial prototype of an algorithm for identifying “junk” (including asteroids, ghosts, cosmic rays, satellites, etc) and rejecting it during coaddition to form clean coadds for detection has been implemented. This achieves excellent results by tracing the source of the imperfections back to individual input exposures, rather than performing a per-pixel statistical operation. Work is now ongoing to deploy this algorithm within the context of the DM software stack. [[DM-10005](#)]



A demonstration of the results of the new “robust” coadd construction code compared to a simple mean (center) or the previous state-of-the-art “safe clipping”, at left. Note in particular the removal of linear features — asteroid trails — in the center-left and center-right of the image.

02C.04.05 – Object Detection and Deblending

- Deblender development [[DM-10616](#)]:
 - An example of the Generalized Linear Mixed Model (GLMM) algorithm developed for the new (Non-negative Matrix Factorization, or NMF) deblender has been prepared as a Jupyter notebook. This forms the basis of an upcoming publication describing the algorithm. [[DM-10827](#)]
 - Updates were made to the NMF deblender to adapt to API changes in underlying packages. [[DM-11139](#), [DM-11404](#)]
 - An investigation was completed into the circumstances in which the NMF deblender could fail to converge, and the major causes of bad divergence

were resolved. In order to make this possible, extra tracking of history was added to the algorithm for use in debugging. [[DM-11169](#), [DM-11229](#), [DM-11230](#)]

02C.04.06 – Object Characterization Pipeline

- Experiments in shear measurement on coadds [[DM-10579](#)]:
 - Work to create the interface between the Bayesian Fourier Domain (BFD) and Metcalibration codebases, which will be evaluated in this work package, was performed. This includes writing the code to generate simulated coadds and to determine the approach used to propagate uncertainties to those simulated coadds. [[DM-10820](#), [DM-10821](#), [DM-10822](#)]
- Improved galaxy flux measurement algorithms [[DM-1111](#), [DM-10580](#)]:
 - Given the record of a particular source measurement, it is now possible to reconstruct exactly the image on which that measurement was made. This procedure is somewhat involved: measurement involves masking out other sources on the image with random noise, so it is necessary to regenerate exactly the same noise field as previously used. [[DM-10359](#)]
 - Given a CModel fit, it is now possible to reconstruct the data that went into the fit. This enables us to re-run the code with varying configurations, investigating possibilities for improving the fit. [[DM-10000](#)]

Planned activities

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

- Prepare for the September 2017 LSST Status Review.
- Complete work on the initial FGCM command line task and spatial/wavelength variability in filters.
- Most members of the team will attend the LSST2017 Project & Community Workshop in the week of 14 August.
- Adopt SynPipe as an official LSST-provided package.

02C.04.01 – Application Framework for Catalogs

- Complete the implementation of the new statistics system.
- Finish the DRP contribution to the SuperTask Working Group, including writing documentation.
- Participate in the Butler Working Group.

02C.04.02 – Calibration Products Pipeline

- Demonstrate key detector diagnostics implemented within the DM stack.

02C.04.03 – PSF Estimation

- The rotation of Zernike polynomials from focal plane to telescope coordinate frames will be completed.
- The technical note describing the PSF fitting work during the S17 and F17 cycles will be made available.

02C.04.04 – Image Coaddition Pipeline

- Provide the new robust coaddition code as a pipeline task and validate its performance on real science data at scale.

02C.04.05 – Object Detection and Deblending

- Continue algorithmic refinement and large-scale testing of the NMF deblender.

02C.04.06 – Object Characterization Pipeline

- Continue with efforts to establish the potential performance of galaxy shear measurement on coadds.
- Work on galaxy modeling will continue, but at a reduced rate due to staff paternity leave.

Recruitment update

- Swinbank will start at the University of Washington on 2017-10-15, but will continue to act as T/CAM for the DRP group.
- We are currently recruiting a project manager at 50% FTE to assist Swinbank in the T/CAM role when the latter moves to UW. Review of applications to begin 2017-09-01.
- Also advertising for a scientific software developer. Review of applications to begin 2017-10-01.

1.02C.05: Science User Interface & Tools

02C.05.00 – Science User Interface & Tools Management Engineering and Integration

- Continued to work with IPAC IRSA group on collaboration in Firefly development, plan and schedule coordination. Continued to work on SUIT requirements in the text of Science Platform.
- David C., Gregory, and Xiuqin attended and presented at the NSF/DOE review at NCSA.

02C.05.01 – Basic Archive Access Tools

- File upload function expansion, user can now upload files of FITS, VOTable, IPAC

table, CSV, and TSV formats. (DM-7225)

- PDAC development:
 - Build SUIT portal docker images in gradle build script (DM-11122)
 - Setup PDAC to make use of the newly added group info provided by CILogon (DM-11308)
 - Allow user to put long running searches in background and provide monitor dialog. (DM-10345)
 - Better error messages (DM-11013)
 - Time series page tile improvement (DM-11094)
 - Bug fix: Time series image display ignored the cutout size (DM-11299)
- Manual UI testing

02C.05.02 – Data Analysis and Visualization Tools

- New functions
 - Heatmap (density) plot in multiple data trace architecture (DM-10832)
 - Create a Firefly grid view/layout entry point for more flexible layout of data display (DM-10948)
- Firefly_client
 - Python binding for the grid layout (DM-11263)
 - Expose the channel and connid from WebSocketClient to enable Python API access to Jupyter widget. (DM-10838)

02C.05.03 – Alert/Notification Toolkit

- No new work done.

02C.05.05 – User Workspace

- No new work done.

Planned activities

02C.05.00 – Science User Interface & Tools Management Engineering and Integration

- Continue to work with IPAC IRSA group on collaboration in Firefly development, plan and schedule coordination.
- All team members will attend the LSST all hands meeting at Tucson.
- Gregory will attend JupyterCon at NYC.
- Several people will take vacation for August total solar eclipse.

02C.05.01 – Basic Archive Access Tools

- Various bug fixes reported by testers
- Work on security and authorization related issues.
- PDAC development:

- Deployment of a PDAC-like app in Isst-dev or Nebula (DM-11546, DM-11547)
- Work with metaserv v1 API (DM-8215)

02C.05.02 – Data Analysis and Visualization Tools

- New functions
 - Make Firefly to handle the Plotly chart type not explicitly support by Firefly (DM-11349)
- Python documents and examples

02C.05.03 – Alert/Notification Toolkit

- No work planned

02C.05.05 – User workspace

- No work planned

Recruitment update

- No activities.

1.02C.06: Science Data Archive & Application Services

Current accomplishments

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

- Team continued improving robustness of DAX services and otherwise supported SUIT integration efforts in the PDAC
- Mueller completed document and presentation prep and attended 2017 DOE/NSF DM Review in Champaign/Urbana
- Gaponenko, Gates, Jammes, and Pease had vacations this month

02C.06.01.01 Catalogs, Alerts and Metadata

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

02C.06.01.02 Image and File Archive

- Van Klaveren implemented an improvement to allow for more descriptive error messages to be returned from imgserv [DM-11370]
- Lo continued work on new version of imgserv which supports IVOA SODA protocol for image cutouts

02C.06.02.01 Data Access Client Framework

- Pease addressed some unit-test issues in the Butler [DM-9761, DM-10972]
- Pease updated Python test names in daf_persistence per RFC-215/RFC-229 [DM-10972]
- Pease added support for the “standardize” function to Butler composite datasets [DM-11251]
- Pease addressed an issue that was causing the Butler to write many identical entries into the _parents list of some POSIX registries [DM-11284]
- Pease addressed an issue that prevented the Butler from outputting to a non-empty non-repo folder [DM-11217]
- Salnikov implemented improvements to the SuperTask command-line activator [DM-10839]
- Van Klaveren implemented a fix for an issue with afw.catalog.asAstroPy when multiple columns have the same name [DM-9503]

02C.06.02.02 Web Services

- Van Klaveren completed porting the DAX Web Services to Python 3 [DM-7415, DM-7417, DM-7418, DM-11183]
- Van Klaveren completed initial implementation of metaserv v1 API, including a schema based on VOResource and a command-line tool for updating database catalog information [DM-10862, DM-10863, DM-10864]

02C.06.02.03 Query Services

- Salnikov implemented asynchronous query dispatch (immediate return with job ID) within Qserv [DM-11190]
- Jammes implemented a proof-of-concept container refactorization that packaged XRootD services on Qserv nodes as a micro-service in their own independent container [DM-10562]
- Jammes implemented extensions to the Qserv Travis CI configuration such that containers generated by Travis CI are pushed automatically to docker-hub [DM-11159]
- Gates and Mueller packaged the jemalloc memory allocator in EUPS and cut over to using this allocator in Qserv builds, to facilitate memory allocation profiling while pursuing large-result performance issues [DM-11214, DM-10971]
- Gaponenko continued work on data distribution design and prototyping
- Hanushevsky worked converting czar code to use the new XRootD SSI v2 API for content-addressed messaging between czar and workers

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

- Team to continue improving robustness of DAX services and otherwise support SUIT integration efforts in the PDAC as necessary
- Entire team to attend 2017 LSST Project and Community Workshop in Tucson, AZ
- Gates, Gaponenko, and Pease have vacation scheduled this month

02C.06.01.01 Catalogs, Alerts and Metadata

- No work is planned for this WBS element this month

02C.06.01.02 Image and File Archive

- Lo to complete initial implementation of image service supporting image cutouts via IVOA SODA protocol

02C.06.02.01 Data Access Client Framework

- Pease to implement afw BaseCatalog reader and writer for Butler

02C.06.02.02 Web Services

- Van Klaveren to complete support for reading/writing afw tables from/to SQL databases
- Van Klaveren to begin working on support for asynchronous queries at the web service layer

02C.06.02.03 Query Services

- Hanushevsky to complete converting czar code to use the new XRootD SSI v2 API for content-addressed messaging between czar and workers
- Salnikov to implement result retrieval and status monitoring for Qserv asynchronous queries
- Salnikov to migrate Qserv admin scripts to Python 3
- Jammes to work on refactoring MariaDB into an independent container in Kubernetes-based Qserv deployments
- Gaponenko to continue work on Qserv data distribution/replication framework
- Gates to tie up implementation loose ends with Qserv query cancellation and large-result handling

02C.06.02.04 Image Services

- No work is planned for this WBS element this month

02C.06.02.05 Catalog Services

- No work is planned for this WBS element this month

Recruitment update

- No recruitment activity this month

1.02C.07: LSST Data Facility

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, high-level engineering and design specification, and planning for Data Facility operations during the commissioning and operations phases. NCSA leads planning for operations of the Data Products Production (DPP) department.

We completed the milestone matrix assessment and facility roadmap, which traces from use cases and goals of project- and subsystem-level milestones through the Data Facility service layers and down to physical components. The milestone matrix describes which services, features, and other elements are needed to achieve the needs of the related milestone. The facility roadmap summarizes the status of the Data Facility at six-month and annual intervals, and altogether describes how the Data Facility will evolve during construction, commissioning, and pre-operations. From this roadmap, we defined Data Facility milestones and service bundles, which are collections of services planned for concurrent deployment. We also finalized planning packages for entry into PMCS.

We assessed the hardware provisioning plan, phasing acquisitions according to service bundles and included value engineering options (consolidated database, use of facility commons) and new baselined technologies (e.g., Kubernetes). These were incorporated into PMCS planning packages.

We prepared for and participated in the NSF/DOE Data Management Status and Replan Review, including the Data Facility plenary and breakout presentations, as well as dry runs and related prep.

- DM-10692 Overall Facility Planning and High-Level Engineering
- DM-7632 Develop planning packages from DM replanning deliverables

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 work improving service management processes, including implementation and consolidation of Incident Management and Change Management. To achieve this, we worked with LSST headquarters system administrators to develop a form in Jira, providing a single point-of-contact for users and developers to report issues, make routine requests (e.g., account creation), and request new features. Internally, we prototyped processes for prioritizing, assigning, and communicating progress on tickets, as well as interfacing with other facility ticketing systems (e.g., NCSA-internal Jira), and are able to assess to first-order the progress on ticket resolution.

We also continued work developing the service architecture supporting infrastructure, beginning to populate the framework and testing data exchange between system architecture tools and the service management system.

Emergent work involved remediating system performance issues for qserv developers.

- DM-10693 Service Management Service Operation Workflow Development
- DM-10694 Service Management Framework Implementation
- DM-10695 Service Management - Emergent (F17a)

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 planning for procurement of the full-scale L1 test system, as well as updating plans for expected remaining FY17 hardware expansion as described in the FY17 Annual Acquisition Strategy Document, verifying needs against planned capabilities. FY17 acquisitions addressed include a refresh of the lsst-db system, and planning for an intermediate-sized Oracle database system. We also finalized the logistics for sending equipment to Chile, including tagging, inventory, shipping, insurance, installation support in Chile, and management processes.

- DM-10696 FY17 Procurement Activity #2 Planning
- DM-10697 FY17 Procurement Activity #3 Planning

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.

We continued development of the offline Batch Production System, setting up a non-shared-disk computing scenario and running with the CiHsc production workflow. We continued to prototype use of the DESDM-based production workflow, verifying functionality with the latest LSST stack and beginning an assessment of scaling and ingesting large-scale datasets into an Oracle database. We continued biweekly reprocessing of the HSC RC dataset based on DM needs, to both verify the software stack and to test the development of science algorithms. We also worked toward finalizing an analysis of the initial HSC full reprocessing that took place in May, including summarizing operational metrics such as capacity usage, processing times, problem management (root-cause analysis), etc. Finally, we identified operational requirements and production use cases, mostly learned from DESDM operations, of the production workflow system to support development of the Pegasus-based workflow system.

- DM-10700 Batch Processing System Phase 1b (Pegasus)
- DM-10701 Batch Processing System Phase 1a (DES)
- DM-10702 HSC reprocessing campaigns

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 refining the front-end interfacing elements of the archiving and prompt processing Level 1 services, converting existing python 2 code to python 3, writing specifications for auditing framework code, and optimizing prompt processing device code. We continued work with Camera and Telescope & Site developers to identify sources and contents of the EFD, OCS metadata available to build headers, and contents of headers required by CCS developers and downstream DM processing. The initial implementation for supporting ITL/e2V sensors was completed, and we began implementing python

multiprocessing in the header writer. We also hosted a workshop associated with the installation of the DAQ test stand, and began testing the interface with our code and the provided API.

- DM-10703 L1 Services & Integration Activity Support
- DM-10704 Header Service Design & Implementation

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.

Emergent work for July involved updating the Developer Guide Common Dataset Organization and Policy for RFC-362, and addressing related dataset access needs.

- DM-10705 Emergent Middleware Work (F17a)

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 investigating methods for data movement over a WAN supporting the spectrograph test stand in January 2018, as well as long-term solutions for data distribution in the Data Backbone. We expanded the prototyping system using the Rucio data distribution package by setting up and using a remote test node as a proxy for the target access points at the Base Facility and CC-IN2P3. We continued working toward managed consolidated database services, soliciting metadata and provenance schema information from relevant DM team members, and modeling the complete LSST raw and processed data flow to facilitate design of a database architecture that supports centralized data management.

- DM-10707 File distribution and management framework
- DM-10708 Consolidated database technology evaluation

Other Infrastructure

These activities support the construction of general reusable services and infrastructure components, as well as emergent infrastructure work not included above.

We continued working with Kubernetes, HTCondor, and Docker on a small dedicated computing cluster to prototype on-demand, elastic provisioning and understand resource

management capabilities. A local Docker registry was set up, populated and tested, and additional configuration options were investigated. We also had a meeting with DM developers constructing containerized applications to discuss our technical roadmap and requested near-term support. Our investigations and input collected provide the basis for the FY18 Acquisition Strategy.

- DM-10711 Containerized Application Management Services

02C.07.03 – Environment and Tools

No work is planned in this WBS for the F17 cycle.

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.

In July, the decommissioning and return of the Level 1 Phase 1 Test system was completed. We began receiving components for the Chilean Base instance of network-based security (intrusion detection, log collection, scanning, etc.) and authentication & authorization systems. These systems will be “burned-in” at NCSA before being repackaged, shipped to Chile, and installed in the existing NOAO facility in La Serena. NCSA 3003 lab refresh activities continued, with all physical components in place; with full deployment scheduled for the August maintenance window. We assisted the Camera team with installation of their simulation hardware, which completed the DAQ test stand deployment activity. We also assisted Elastic Services developers by reconfiguring systems in their test cluster to meet evolving needs.

- DM-8510 DAQ Teststand Single Raft Deployment
- DM-10712 Chile Base AA Acquisition & Installation
- DM-10713 Level 1 Test System Provisioning Phase 1
- DM-10715 NCSA 3003 Lab Refresh
- DM-10719 Elastic Services Testbed Provisioning

Service Management for LSST developer 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, as well as general improvements and additions to hardware and system infrastructure.

We continued to work issues with Disaster Recovery (DR) software with the software vendor; multiple fixes were tested and a stable solution was found. Outstanding DR work is expected to be completed in August. Work continued on system monitoring improvement;

basic system-level monitoring was installed on the PDAC, work began looking into more refined monitoring of the Nebula cluster, and an initial look at using the Graylog log monitoring tool was completed. We also integrated the monitoring infrastructure with Puppet configuration management. Ingestion, transformation, and analysis of system-level monitoring information at the service level is on-going as the service-level monitoring framework is developed (see above).

- DM-8502 Disaster Recovery for Science Datasets
- DM-10716 System Monitoring Refinement (F17a)

Planned activities

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

Data Facility staff will attend the LSST2017 conference in Tucson, AZ and make several presentations about current services, planned services and development.

Planning and Management

In August, we will pursue NSF Review follow-up activities, including further integration of L3 milestones with the DM plan. We will make preparations for the NSF/DOE Joint Status Review as needed. We will complete planning of Epic-level activities for the second half of the F17 cycle. We will begin working on articulating requirements documents and defining verification tests consistent with the Data Facility and overall DM roadmap.

We will work on articulating the Data Facility roadmap, milestone matrix, and service dependencies in a Configuration Management Database (CMDB). The goal is to record the service dependencies, components, and processes, which supports overall management of the facility, provides a basis for detailed cycle planning, and supports change control of Data Facility services. We currently plan to use a TOGAF-standard tool, Archimate, to model the CMDB.

We will begin planning FY18 capabilities in preparation for the FY18 Acquisition Strategy document.

We will continue to participate in the DM subsystem science team and DM subsystem engineering team and associated activities.

- DM-10692 Overall Facility Planning and High-Level Engineering

Service Management and Monitoring

We will continue work improving service management processes, including single entry points for incident response and request response. We will continue maintaining and

evolving the service management framework, refinement of Change Control and Incident Response frameworks.

- DM-10693 Service Management Service Operation Workflow Development
- DM-10694 Service Management Framework Implementation
- DM-10695 Service Management - Emergent (F17a)

Hardware acquisition and provisioning planning

We will continue planning for procurement of the full-scale L1 test system, as well as update plans for expected remaining FY17 hardware expansion as described in the FY17 Annual Acquisition Strategy Document. We will also begin acquisition planning for fiscal year 2018, and begin the semi-annual refresh of the FY17 sizing model.

- DM-8506 Costing Model Redesign
- DM-10696 FY17 Procurement Activity #2 Planning
- DM-10697 FY17 Procurement Activity #3 Planning
- DM-10699 FY17 Sizing Model Refresh 2
- DM-10698 FY18 Acquisition Planning

02C.07.01 – Processing Control

Batch Production Services

We will continue development of the workflow-based Batch Production System, with incremental implementation of data staging and the ability to orchestrate and run more complex pipelines in the planned workload management framework using the Pegasus workflow engine, with consideration of the operational use cases and Pegasus functionality. We will continue to enhance the DESDM-based service façade based on the need for a stable production-like system in the near term before commissioning ramps up. We will continue the planned reprocessing of HSC data based on DM needs, to both verify the software stack and to test the development of science algorithms. We will also work on collecting and summarizing application-level performance metrics, which support development of service-level monitoring, capacity planning, and operational improvements.

- DM-10700 Batch Processing System Phase 1b
- DM-10701 Batch Processing System Phase 1a
- DM-10702 HSC reprocessing campaigns

Level 1 Services

We will continue refining the archiving and prompt processing Level 1 services, using the established Early Pathfinder Software Integration Activity schedule to drive development priorities. We will also continue working with Camera and Telescope & Site developers to identify sources and contents of the EFD, OCS metadata available to build headers, and contents of headers required by CCS developers and downstream DM processing. We will

continue converting to python 3, as well as working with the installed DAQ simulator, and integrating and testing our framework with the provided API.

- DM-10703 L1 Services & Integration Activity Support
- DM-10704 Header Service Design & Implementation

Common Middleware and Other Tasks

Work for August 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-10705 – Emergent Middleware Work (F17a)

02C.07.02 – Infrastructure Services

Data Backbone

We will continue investigating methods for data movement over a WAN supporting the spectrograph test stand in January 2018, as well as long-term solutions for data distribution in the Data Backbone. Current prototyping involves using the Rucio data distribution package to distribute data within the Data Backbone to service access points (between the Base Facility, NCSA, and CC-IN2P3). We will also continue work on the architectural framework for providing consolidated database services, developing the architecture of the first planned Oracle platform.

We will begin participation in the Butler working group, representing Data Facility operations, central archive management and data distribution, and production requirements.

- DM-10707 File distribution and management framework
- DM-10708 Consolidated database technology evaluation

Other Infrastructure

Work continues on investigating and prototyping the containerized application management framework, including the interaction between Kubernetes, Docker, and HTCondor, as well as the resource management capabilities of elastic provisioning. We will be meeting with the HTCondor team to discuss planned implementation and concerns. We will also begin discussions with SQuaRE towards integrating Jupyter notebook containers.

- DM-10711 Containerized Application Management Services

02C.07.03 – Environment and Tools

No work is planned in this WBS for the F17 cycle.

02C.07.04 – Site Infrastructure

Hardware Acquisition, Deployment and Provisioning

In August, the Chilean Base instance of network-based security and authentication & authorization systems will be provisioned locally, configured, and verified, in preparation for shipment and installation in Chile. NCSA 3003 lab refresh activities will be completed. Acquisition of hardware for upgrades for database services and qserv development resources will begin.

- DM-10712 Chile Base AA Acquisition & Installation
- DM-10715 NCSA 3003 Lab Refresh
- DM-10719 Elastic Services Testbed Provisioning

Service Management for LSSST development resources, Emergent and Miscellaneous Work

We plan to resolve issues with Disaster Recovery software, completing this phase of DR implementation. Work will continue on system monitoring improvements, with the goal of providing high-level service and system monitoring displays to support FY18 project integration activities, and supporting development of service-level monitoring and management framework and process development.

- DM-8502 Disaster Recovery for Science Datasets
- DM-10716 System Monitoring Refinement (F17a)

Recruitment update

A requisition for a full-time database administrator is in progress.

1.02C.08: International Communications & Base Site

Current accomplishments

02C.08.01 – Base Center

- Summit Base ITC Tiger Team: LCR-963 was approved baselining LSE-309 Summit Base ITC Design Document (SBIDD). A few additional updates regarding the Clean Room and White Room will be addressed in a new LCR. The CCB reviewed LCR-964 requesting to baseline LSE-299 Summit Computer Room Requirements Document, including cost analysis of impact of adding third Computer Room Air Conditioner (CRAC). The document was imported into Magic Draw and formatted according to System Engineering standards. A new version will be issued and re-reviewed.
- Summit and Base Networking and Computing: Coordinated with NCSA regarding the equipment they are planning to install in Oct/Nov. We need to purchase a UPS unit for all the equipment including the NCSA equipment and the DWDM. This is a

small unit to be installed in the existing Data Center, to address the need until the Base Center is available for occupancy. The Summit LAN procurement was awarded to Dimension Data and the initial order was placed. A requisition for an additional time server to support commissioning activities was also placed. Submitted an LCR to update PMCS plans for the summit network installation reflecting the slip in the Summit Facility occupancy dates due to severe winter weather. There were problems encountered with the installation of the new firewalls in La Serena and this work continued.

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: No activity. This segment is accepted.
- DWDM Equipment: The AURA/LSST DWDM equipment arrived in Chile and is being stored in the installer’s (Raylex) warehouse until access to the site and AURA staff return to Chile; this will occur in September. A training workshop named “hiT 7300 System Architecture, Provisioning and Troubleshooting” was done from the 24th to 28th of July. There is a delay in the reception of REUNA DWDM equipment, the new dates provided by CORIANT are:
 - Reception of Intermediate nodes (amplifier nodes): middle of August
 - Reception of terminal nodes (chassis of La Serena – Santiago backbone and Santiago metropolitan ring): end of August (to be confirmed).
 - Reception of line cards (Flexigrad transponders): end of September
- Santiago–La Serena: This segment was accepted. Telefonica is still working on the correction of some elements missing in the housing sites (final electrical connections mainly).
- La Serena–AURA Gatehouse: This segment is accepted.

02C.08.03.02 – International Chile–US WAN

- *100 Gbps Managed Ring*: No planned engineering activity in this period. The ring is performing as expected. Currently, AmLight has two 100G links in a ring topology provided by TI Sparkle connecting Miami to Sao Paulo. These 100G links were configured through two diverse paths: one 100G link uses a submarine system in the Atlantic Ocean and the second 100G link uses a combination of terrestrial fibers plus submarine systems. The terrestrial segment goes from Sao Paulo/Brazil to Valparaíso/Chile where the TI Sparkle landing station is. Before getting to Valparaíso, TI Sparkle opens the fiber at its facility inside the Level3 data center, in Huechuraba/Great Santiago/Chile. From Valparaíso to Miami the 100G link goes through a submarine system in the Pacific Ocean to Panama, where it then enters

the Caribbean Sea. This fiber layout was chosen to be resilient and to allow natural expansion of the AmLight network. AmLight's point of presence in Chile is also inside Huechuraba Level3 data center. To use the 100G link in Chile, AmLight needs to connect to TI Sparkle via the data center's Meet-Me-Room, where all cross-connects are established. TI Sparkle's network is ready to connect to AmLight, and a request was submitted by TI Sparkle to Level3 to install two fibers pairs between TI Sparkle and the Meet-Me-Room. The Contract Delivery #3 invoice is long overdue and the PMCS continues to show positive EV cost variance. This is because the provider of this leased circuit has not invoiced FIU, and therefore FIU could not invoice LSST. There has been a lot of back and forth as the provider had to update their company information in the FIU vendors system and in sam.gov. The invoice will be processed at FIU in early August.

- *Management and Coordination Contract:* Conducted the regular monthly meeting of the LSST Network Engineering Team (NET). Completed updates to resolve issues/comments on Network Verification Plan and Matrix. The Networks Verification Matrix content has been captured in Magic Draw for all networks except the Summit Network. A First Fiber Optic Light demonstration is being planned for the end of CY17. The goal of the demonstration is to show continuity of the fiber path at a higher bandwidth than formerly available and to have data transferred over the links. This will be one time event and not a formal start of operation. Acceleration of planned purchases of equipment to support the demonstration will take place only in case the equipment will be for the future links. If the demonstration is not ready for Cerro Pachon to NCSA by ADASS in October, REUNA may coordinate a local demonstration just in Chile. Dr. Julio Ibarra attended a meeting in Miami with the FLR board members. FLR will be utilized as part of the LSST networking in Florida. The next meeting of the South American Astronomy Coordination Committee (SAACC) will be held in Santiago, Chile in conjunction with the ADASS conference, on October 18, 2017.
- *Spectrum Contract:* Angola Cables provided the answers to the technical engineering document detailed in the May 2017 monthly report, which provided the AmLight engineers with the information they needed to begin working to adjust the DWDM system specification accordingly.
- *US National WAN:* ESNNet collaboration negotiations with DOE continued. The draft request to ESNNet was revised by Donald Petravick and Jeff Kantor based on discussions with DOE. ESNNet contribution will be for operations not construction, phased with some level of support by 2020. Discussions with FNAL were initiated to define a collaborative role with NCSA in network operations and management during the LSST operations phase. FIU is working with FLR and other fiber providers to determine the most cost effective and greatest throughput for transit from AMPATH north.

Planned activities

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

- Luis Corral and Ron Lambert will attend the LSST 2017 meeting in Tucson.
- Luis Corral will attend a Cisco ACI and ISE course in the US.
- Guido Maulen will attend a certification course in Santiago on Data Center cabling

02C.08.01 – Base Center

- Summit to Base ITC Tiger Team: LSE-299 Summit Computer Room Requirements Document to be approved by CCB. New LCR to be prepared to update LSE-309 Summit to Base ITC Design Document with Clean Room and White Room information.
- Summit and Base Networking and Computing: Ron Lambert will visit Tucson to make detailed summit network cable raceway drawings with Bill Schoening, and attend a meeting with Cisco. Complete installation of new firewalls in La Serena.

02C.08.03 – Long-Haul Networks

02C.08.03.01 – Chile National WAN

- Summits - AURA Gatehouse Network: No activity. This segment is accepted.
- DWDM Equipment: We expect to receive REUNA DWDM equipment from Coriant. The equipment will be delivered to the installer (RAYLEX) and held there until installation. Coordinate training courses, “TNMS Operation and Administration for hiT7300” and “Tailored TransNet training for hiT 7300”. The following is a very preliminary deployment plan, is pending the actual delivery date of the equipment:
 - AURA summit-LS - September
 - REUNA-LS - September
 - REUNA-Stgo to Level3-Stgo - August
 - Physical installation of Intermediate nodes (6 sites) - September - October
 - Test and Start up of long haul segment - November
 - Santiago ring (metropolitan network) - November - December
 - Test and start up of Santiago ring - November - December
- Santiago-La Serena: Telefonica will complete the remaining corrections 48 hrs before the installation of the DWDM equipment. REUNA to receive the measurements of the local loops and provide to AURA.
- La Serena - AURA Gatehouse: No activity. This segment is accepted.

02C.08.03.02 – International Chile - US WAN

- 100 Gbps Managed Ring: Once the fiber pairs in Huechuraba become available, CLARA (AmLight's host in Chile) will order two cross-connects from Level3. CLARA

will use its fiber pairs to connect TI Sparkle to AmLight's network devices. In the end, the 100G link from Sao Paulo to Miami via Pacific will become two links: one 100G link from Sao Paulo to Santiago and one 100G link from Santiago to Miami.

- Management and Coordination Contract: Ron Lambert will visit FIU/Ampath to coordinate with Jeronimo Bezerra over QoS and testing. Verification Matrix content for Summit Network to be captured in Magic Draw. Submit DM change request to adjust bandwidth allocations in LDM-142 based on NET comments.
- Two abstracts were accepted for the ADASS Conference, which will take place in Santiago, Chile in October 22-26, 2017: International Networking in support of Extremely Large Astronomical Data-centric Operations; Challenges of Standardizing and Supporting ITC Services in a Widely Distributed Project: ITC Design of the LSST Summit - Base Complex.
- Spectrum Contract: By the end of August AmLight engineers will present the final configuration document to the LSST NET engineering for a final discussion. The Operations and Maintenance Agreement is expected to be complete as well.
- US National WAN: Discussions with ESNet for LSST within DOE to define the ESNet service and costs will continue during August. FIU will continue working with FLR and other fiber providers to determine the most cost effective and greatest throughput for transit from AMPATH north.

Recruitment update

- There has been no progress on the hiring of the IT System Admin/Eng due to an absence of the HR person in Chile. Since we have already conducted multiple phone interviews, we are exploring the possibility of making an offer to the top candidate without further interviews. Guido Maulen, the new IT/Network Tech accepted and is due to start work on Aug 1.

1.02C.10: Science Quality and Reliability Engineering

Current accomplishments

- The macOS CI builds are back online following the relocation of the hardware in a rack in the NOAO computer room and under remote switch management. The automated jobs keeping disk usage low seem to be working well. We have not seen any stability issues since the move.
- Work on the pipelines documentation is progressing well, with a getting started tutorial based on DMTN-023 that provides introductory level documentation of the stack and other improvements that are slowly making their way through careful review.
- We brought our JupyterLab deployment up to date with the latest alpha and resolved some issues. The prototype notebook service was extended to offer a choice of the last four weekly stack containers. Improvements were made to the

Save/Exit menu and pandoc was added to our pods to allow some of the notebook export functions to work.

- Our migration of the SQuaSH backend deployment to kubernetes is now well underway with two microservices now able to stand up under k8.
- FE attended and presented at the NSF/DOE review panel hosted by NCSA July 25-2017.

Planned activities

- Stand up JupyterLab cluster to support the stack tutorial at LSST-2017
- AT to attend JupyterCon

Pittsburgh contract

- We have renewed our contract with the University of Pittsburgh for a fractional FTE contribution for Michael Wood-Vassey during FY 2018.
- Work performed under this contract included improvements to validate_drp (DM-11215, DM-9961, DM-11388, DM-11300).

Recruitment update

None planned