Data Management Monthly Report
May 2017

High-level Summary/Status

The DMLT got together in Seattle and had a productive face-to-face meeting. Work continues toward the July review. We now have the first versions of a product tree, architecture diagrams, and component descriptions. A tracing of components to newly-updated requirements has been put in place, uncovering some requirement updates.

The SuperTask working group is wrapping up formal requirement documentation. An internal review of the draft requirements preceded a "hack week" to produce in working code. There are still some requirements on other components to be resolved.

Significant improvement to the scientific algorithms include:

- The WCS implementation for Alerts was significantly improved as well as bringing the Zackay, Ofek & Gal Yam (2016) image differencing algorithm to the LSST DM Stack.
- Work on PSF homogenization is now complete, making PSF-matched coadds a first-class data product in the LSST stack.
- The stack now also includes a pipeline capable of fitting out-of-focus images, an important step towards high-quality PSF modeling.
- The prototype Non-negative Matrix Factorization (NMF) deblender continues to show promising results.

Highlights from the PDAC, data processing and visualization:

- All initial WISE data is now available in the PDAC.
- A discrepancy in the overlap region between NCSA and IN2P3-processed regions of the SDSS Stripe 82 data available in the PDAC has been resolved.
- We successfully processed the full HSC Strategic Survey Program (SSP) Public Data Release 1 (PDR1) using the LSST stack on hardware at NCSA.
- A time series visualization tool was was released along with the NEOWISE 2017 data release: LSST benefits from many bug fixes and improvements.

Networks and infrastructure news:

- The Summit Base ITC Design Document (SBIDD) and Summit Computer Room Requirements Documents have been submitted for baselining.
- We accepted/tested La Serena - AURA Gatehouse fiber and summits - Gatehouse fiber.
- Tarballs for eups-based binary distribution are now in the automated weekly distributions.

Risk Management

The DM Risk Register was reviewed in detail ahead of the Risk Management meeting next month which will focus on DM.

Detailed Project Progress and Status

LSST Program Office

DM Project Management and Control

Current accomplishments:

O'Mullane visited SLAC to discuss Qserv and potential alternative technologies. Very collaborative discussions culminated in an agreement that we should investigate and reconfirm approach.

O'Mullane & Swinbank were introduced to PST at UW. O'Mullane presented the current status of work within DM and raised the question of user expectation for query systems, relating to the Qserv discussion.

Lead successful a DMLT F2F meeting at UW. Introduced John Swinbank as deputy DM project manager.

Swinbank visited Tucson to discuss various DM topics and strategy and to better understand EVMS in the context of DM. An afternoon was spent examining milestones in Primavera with Kevin Long. Both O'Mullane and Swinbank now have access to Primavera and a basic understanding of how it works. A series of test milestones were discussed.

Planned activities:

- Reduce risk exposure in the risk review, get risk report for use in the NSF review in July.
- Formalise test milestones as part of the verification strategy for DM and align other milestones with those.
- Keep moving on the review data pack.

DM Science

Current accomplishments:

Summary of key accomplishments: The DM Science Team has completed internal studies on photometric redshifts (Graham; http://ls.st/DM-6367 and http://ls.st/PUB-44), initial examination of use cases for potential Hadoop-style processing (Slater), and the science validation of the Prototype Data Access Center (Suberlak; Portal and Database components; https://github.com/lsst-dmsst/PDAC_report/blob/master/main.pdf). The Science Platform Vision document has been completed and ready for submission to the Change Control Board (Juric; https://github.com/lsst-dmsst/LSE-319/blob/master/science-platform.pdf)

Graham: The journal manuscript "Photometric Redshifts with the LSST: Evaluating Survey Observing Strategies with a New Nearest-Neighbors Estimator" (Graham et al.) has passed the LSST internal review process with several rounds of minor revisions, and is in the two-week phase of identifying builder co-authors. The internal document evaluating the "LSST DRP Level 2 Catalog Photometric Redshifts" was preliminarily presented in a DM-SST telecon and discussed. Work on the internal document "Data Management and LSST Special Programs" continued, and initiated a more in-depth study of the minimum exposure time that will be suitable for the AP pipeline. An oral presentation on LSST Data Products was created and delivered at the "Supernovae: The LSST Revolution" at Northwestern.

Slater: Presented to the relevant DM science leads the results of a first analysis of using Spark for storing and processing light curve data. Identified which Spark use-cases could be easily met and which would require more detailed investigation in order to meet appropriate performance requirements. Finished introduction and conclusions for star/galaxy separation paper, then circulated it to coauthors.

Suberlak: Completed the source number density comparison between PDAC and UW- hosted S82 dataset, and finished the user report summarizing the work. This report led to direct PDAC interface improvements bundled in epic DM-10432 , specifically JIRA tickets DM-7990 ,

DM-10477, DM-10431, DM-10433, DM-10463 , DM- 10465, and DM-10466.

Planned activities:

- Continuation of work on finalizing requirements for Special Programs (Graham).
- Begin organizing and collecting datasets for tests of crowded field photometry (Suberlak and Slater), to investigate the pipeline performance, pursue the best directions for improvement.
- Develop proof-of-concept implementation of light curve retrieval and analysis in Spark (Slater).
- Revisions and cleanup of Star-Galaxy separation paper (Slater).

DM System Engineering

Current accomplishments:

DM System Architect and Deputy System Architect:

- Attended face-to-face DMLT meeting.
- Completed first version of product tree (in LDM-294), architecture diagrams, and component descriptions (in LDM-148).
- Traced LSE-61 Data Management requirements to components and vice versa.
- Supported visit to SLAC by new DM Project Manager.
- Evaluated options for storing processed visit images or equivalents (RFC-325).
- Pushed LCR-943 through change control. This was minor updates to LSE-63.
- Processed LDM-523 and LDM-151 and released them to DocuShare.
- Submitted LCR-962 (DMSR/DPDD requirements update).
- Updated and corrected the MagicDraw model of LSE-61 and succeeded in creating an LSE-61 document from the model using LaTeX.

DM System Interfaces Scientist:

- Attended face-to-face DMLT meeting.
- Presented SuperTask requirements (based on working group activities) to DMLT.
- Brought SuperTask requirements through a DM-level review.
- Convened a one-week hack session on SuperTask that demonstrated most of the key ideas.

Planned activities:

DM System Architect and Deputy System Architect will:

- Complete documents including LDM-148 for DM review.
- Give talk on "Managing Projects with git and GitHub" at the 2017 ASTERICS/OBELICS Summer School in Annecy.
- Attend the Lyon LSST meeting.
- Work on OSS to DMSR flowdown and LSE-61 model additions for review.

DM System Interfaces Scientist will:

- Produce a formal requirements document for SuperTask.
- Refresh the summary presentation on all system-level DM ICDs.

DM Science Quality and Reliability Engineering (SQuaRE)

Current accomplishments:

- Work on our new verification framework was completed with an informal review. Many thanks to John Parejko (UW) for serving as Product Owner for this work. This framework gives us the ability to create machine-parsable specification libraries in a modular way. It includes metric definitions, subsets of data that are appropriate to those metrics and an extensible class for specifying boundaries (eg thresholds) for those specifications. A summary of this work is in the slides attached to DM-10747 and documented in SQR-019 .
- The JupyterLab prototype work proceeded with the addition of many features such as a py3 stack pod, a py2/py3 selector (as a testbed for the selector, only py3 will be supported per the DM System Engineer), an upgrade to JL 0.20, the addition of some persistent storage to facilitate testing, and Hub logout usability improvements. SQR-018 was updated to reflect the architectural changes.
- Brian Granger (co-founder of JupyterLab) visited Tucson for a hack session and to give us feedback to our architectural approach. It was very useful to us and we appreciate his interest and support.
- Tarball builds to allow for eups-based binary distribution are now automatically produced as part of the weekly release.
- Work on a Kubernetes-based RabbitMQ deployment to allow scalable bulk shipping of logs to logging.lsst.codes has been completed and tested.
- SQuaSH has been switched to a responsive layout design to facilitate mobile use and

improve user experience; this is in the developer sandbox awaiting release.
- The commercial 3rd party jira bot was replaced by the in-house sqrbot support due to complaints about the service and a change in their pricing structure. Thanks to BvK (SLAC) for a useful patch.
- FE traveled to Seattle to participate in the DMLT meeting.
- AF traveled to a DESI review as part of his LineA time.
- JS was on loan to DM Architecture to work on extending the documentation to handle LaTeX documents with browser-native PDF landing pages. An example of this work can be seen at DMTN-044.
- Usual end of cycle activities and planning for FY17.
- Pittsburgh contract: MWV attended the DMLT meeting and made improvements to validate_drp.

Planned activities:

- F17 epic bootstrap.

Recruitment update:

- None planned.

University of Washington

Current accomplishments:
02C.03.00 - Alert Production Management Engineering and Integration

- Rawls went to Python in Astronomy
- Patterson went to IVOA in China
- Patterson, Bellm, and Krughoff attended the brokers meeting in Tucson
- Krughoff and Bellm attended the DMLT meeting in Seattle
- Rawls extended the prototype through the difference imaging stage (DM-9654, DM-9652)

Hiring

- No activity.

02C.03.01 - Single Frame Processing Pipelines

- Findeisen began work on improving the single frame PSF fitting system by taking a census of the current system (DM-10002).
- Owen reached the significant goal of implementing the replacement Wcs class (DM-8440). Along the way he implemented a new version of the image warper that uses the new Wcs and showed that it is comparable or better than the current version (DM-10429).

02C.03.02 - Association Pipelines

- No advancements in this area.

02C.03.03 - Alert Generation Pipelines

- Patterson implemented a proof-of-concept filtering system (DM-10026).

02C.03.04 - Image Differencing Pipeline

- Sullivan butlerized the DCR corrected template creation (DM-10522, DM-10438, DM-9851).
- Reiss implemented the spatially varying ZOGY algorithm in the stack (DM-8145, DM-10508).

02C.03.05 - Application Framework for Exposures

- Findeisen improved documentation in the Developer Guide (DM-9024).
- Parejko took care of various pieces of technical debt (DM-10288, DM-10401).
- Owen resolved a wide array of technical debt (DM-10496, DM-10497, DM-10485, DM-10410, DM-10394, DM-10116, DM-9939, DM-9765, DM-10274, DM-10281, DM-8606).

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

- Morrison fixed bugs in the new matcher found during the validation process (DM-10453).
- Parejko did significant cleanup of the jointcal system (DM-9495, DM-9297, DM-6626) as well as making the astrometry model configurable (DM-9506).

Planned activities:

02C.03.00 - Alert Production Management Engineering and Integration

- Sullivan will travel to Lyon for the LSST-IN2P3 meeting.
- Patterson will travel to the Women in Astronomy meeting.
- Rawls, Reiss, Findeisen, Bellm and Krughoff will work on defining the AP minimum viable system requirements and design.

02C.03.01 - Single Frame Processing Pipelines

- Morrison will finish validation of the new matcher.

02C.03.02 - Association Pipelines

- Morrison will continue to work on the minimum matching system.

02C.03.03 - Alert Generation Pipelines

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

02C.03.04 - Image Differencing Pipeline

- Reiss will tidy up the diffim work to transition to working on the minimum viable system.
- Sullivan will work on the write-up of the chosen DCR algorithm

02C.03.05 - Application Framework for Exposures

- Various technical debt.

02C.03.06 - Moving Object Pipelines

- No planned work.

02C.03.07 - Photometric Calibration Pipeline

- Parejko will work on improving the spatially variable photometric model in jointcal.

02C.03.08 - Astrometric Calibration Pipeline

- No planned work.

Recruitment update:

- No activity.

Princeton University and University of California, Davis

Current accomplishments:

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

- Travel & meetings:
    - Rykoff visited Princeton for the week of 1 May for discussions on integration of the Forward Global Calibration Method (FCGM; Burke et al, 2017) with the LSST software stack (discussed in more detail below).
    - Bosch & Swinbank travelled to Seattle during the week of 8 May for a DM Leadership Team face-to-face meeting.
    - Swinbank travelled to Shanghai, China during the week of 15 May for the International Virtual Observatory Alliance's Northern Spring 2017 Interoperability Meeting. This meeting marked the end of his tenure as chair of the IVOA's Time Domain Interest Group.
    - TDIG discussions at this meeting focused on the future of the VOEvent system for distributing notifications of transient celestial events, including a presentation on the proposed LSST alert system by Patterson (University of Washington), and on the representation of time series data within the Virtual Observatory.
    - Swinbank travelled to Tucson during the week of 22 May for project management discussions with O'Mullane (DM Project Manager) and Long (LSST Project Controls Specialist). These discussions focused on our approach to tracking and reporting earned value and on preparations for the review of the Data Management subsystem scheduled for July.
- Management & planning:
    - This being the last month of the Spring 2017 (S17) development cycle, efforts focused on planning for the next (F17) cycle.
    - Bosch and Swinbank met with Lupton (Pipelines Scientist) and Bellm &

Krughoff (University of Washington group leaders) to discuss priorities for the F17 cycle. This followed by management effort to schedule and resource load the work.
    - The resulting plan was discussed at the May 2017 DMLT face-to-face meeting, and has been loaded LSST Project Management & Controls System.
- Data Management replan [DM-8494]:
    - Minor changes and updates to LDM-151, the Science Pipelines design document, were added to the in-progress draft. [DM-8497, DM-8498, DM-9356, DM-9361, DM-10511]
    - The version of LDM-151 resulting from the replan was accepted by the DM Leadership Team, and has now been baselined.
    - Following discussions at the May 2017 DMLT face-to-face meeting, Swinbank has agreed to accept responsibility for maintaining LDM-151, the design document for the LSST Science Pipelines.
- Quality assurance [DM-8299]:
    - The DRP team has continued to cooperate with colleagues at NCSA to prototype large-scale processing of HSC PDR1 on LSST owned hardware.
    - Work has begun on designing an upgraded suite of visualization tools which will help us monitor the quality of LSST data through the next phase of development. [DM-10044, DM-10045]

02C.04.01 - Application Framework for Catalogs

- Middleware evolution [DM-1109]:
    - The DRP team produced a pair of documents describing the state of the interface between the SuperTask design, as currently envisaged by the SuperTask Working Group, and the Science Pipelines. [DM-10329]
        - The Science Pipelines Perspective, and
        - A description of the definition of SuperTask quanta.
    - In addition, team members participated in the SuperTask Working Group "hackathon", building a prototype SuperTask design. This work was still undergoing review in early June, and will be fully reported next month. [DM-10600]
    - Finally in this work package, an interface between the LSST codebase and the "MEDS" (Multi-Epoch Data Structure) file format, as used in multi-epoch fitting by the Dark Energy Survey. This will enables the use of DES-developed shear measurement codes with LSST, providing a convenient baseline and point of comparison for LSST-developed codes. [DM-9603]
- Emergent work and reduction of technical debt [DM-8136]:

- A number of minor usability enhancements were made and bugs and typos were fixed across the codebase. [DM-2009, DM-9866, DM-10289, DM-10489, DM-10510, DM-10451]
- In conjunction with the Alert Production team, a significant overhaul of the way memory is managed in the Jointcal package — used for simultaneously fitting for astrometry and photometry across multiple visits — was undertaken. This has resulted in a more robust and reliable codebase, which will be easier to work with in future. [DM-4043]

02C.04.02 - Calibration Products Pipeline

- Auxiliary telescope development [DM-8151]:
  - Work is underway to convert the prototype pipeline for processing spectroscopic data from the 0.9m telescope at CTIO to run within the framework of the LSST stack. This work will continue into next month. [DM-10253, DM-10565]
- Photometric Calibration [DM-8276]:
  - Rykoff visited Princeton for the week of 1 May to discuss converting the Forward Global Calibration Method codebase to run within the framework provided by the LSST stack. This will provide a powerful method of providing a "global" photometric calibration typing together photometry over large areas of sky. Although this approach may not be necessary in operations (our hope is tying LSST's photometry to Gaia will render it unnecessary), it will remain an essential tool for development and testing with precursor surveys.
  - These productive discussions provided a clear path forward, and work is now underway to complete this conversion. This work is expected to continue into the F17 cycle.

02C.04.03 - PSF Estimation

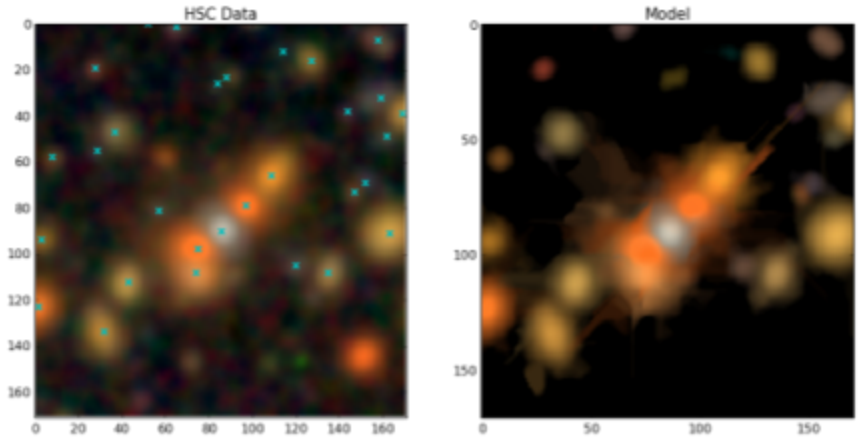
- Wavefront measurement and PSF reconstruction [DM-1909]:
  - Completed development of the "donut fitting" pipeline, which enables fitting Zernike polynomials to out-of-focus images within the framework provided by the DM Stack. [DM-9567]
  - A suite of visualization tools for interacting with the results of the above were undergoing review at the end of the period being reported upon. They are expected to be accepted in early June. [DM-9987]
  - That completes the work scheduled in this area for the S17 cycle. However, this effort will continue in F17.

02C.04.04 - Image Coaddition Pipeline

- PSF homogenization [DM-8289]:
    - PSF-homogenized coadds are now available as a standard data product in LSST process. [DM-8491]
    - Documentation has been updated to describe the process of producing PSF-homogenized coadds. [DM-10004].
    - That completes this work package.

02C.04.05 - Object Detection and Deblending

- Deblender development [DM-8140]:
    - The prototype Non-negative Matrix Factorization (NMF) deblender was updated to use the new Footprint system, which was merged to the stack last month.
    - The update scheme used by the NMF deblender has been changed so that it always enforces a strict monotonicity constraint (the monotonicity constraint requires that sources never increase in brightness away from their center; implementing it strictly means that this is enforced throughout the fitting process). [DM-10612]
    - A new test has been added to demonstrate that the NMF deblender correctly quantifies the overlap between a pair of galaxies. It performs substantially better in this respect than the old codebase. [DM-9784]



A comparison of real HSC data from the COSMOS field (left) with the deblend model (right) which was derived from the peaks marked with blue crosses, demonstrating the NMF deblender's ability to handle overlapping galaxies.

02C.04.06 - Object Characterization Pipeline

- Enhancements and upgrades to CModel flux measurement [DM-1111]:
    - Completed work started last month to enable simulated source injection during image processing. This is an essential technique in debugging galaxy model fitting code. [DM-5310, DM-9998]
- Emergent work and reduction of technical debt [DM-8306]:
    - A number of minor usability enhancements and documentation upgrades were made across the codebase. [DM-9050, DM-10381, DM-10386]

Planned activities:

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

- Coordinate the start of the new (F17) cycle.
- Prepare for the NSF/DOE review of LSST Data Management in July.
- Members of the team will attend the LSST@Lyon workshop during the week of 12 May.
- Paul Price of the Hyper Suprime-Cam and Subaru Prime Focus Spectrograph projects will visit Princeton for discussion starting 28 June.

02C.04.01 - Application Framework for Catalogs

- Design & implement a replacement for the statistical framework currently available in the LSST stack.
- Continue with SuperTask Working Group follow-up work.

02C.04.02 - Calibration Products Pipeline

- Continue conversion of the prototype Auxiliary Telescope Pipeline to LSST stack conventions.
- Continue conversion of FGCM to LSST stack conventions.
- Restart work on interfacing the LSST stack with data from the camera test stands.

02C.04.03 - PSF Estimation

- Include an optical distortion model in the “donut” fitting code.

02C.04.04 - Image Coaddition Pipeline

- Work will continue on warped image comparison, the method by which defects appearing in only a subset of images can be identified and masked during coaddition. A number of approaches to this will be prototyped over the upcoming months.

02C.04.05 - Object Detection and Deblending

- Work will continue on the NMF deblender.

02C.04.06 - Object Characterization Pipeline

- Begin experimenting with shear measurement algorithms on coadds. If this work is successful, it may be possible to meet some of our science goals by working directly on the coadds, rather than resorting to multi-epoch fitting: this could result in a significant resource saving.

Recruitment update:

- No changes during this month.

IPAC / California Institute of Technology

02C.05.00

- Continued to work with IPAC IRSA group on collaboration in Firefly development, plan and schedule coordination. IRSA released the first version of its Time Series Tool along with the NEOWISE 2017 data release. The Time Series tool can upload WISE time series data table to allow users to interactively plot NEOWISE time series data, view the associated images, and search for any underlying periodicity.
- Continued to work on SUIT requirements in the text of Science Platform.
- David Shupe gave a talk at Python in Astronomy 2017, May 8-12, Leiden, NL. Please see slides here: https://confluence.lsstcorp.org/display/DM/SUIT+presentations.
- Xiuqin, Gregory, and David C. attended DMLT face to face meeting at UW, 5/9-5/12, 2017.
- Finished F17 plan with resource loaded.

02C.05.01 Basic Archive Access Tools

- Upgrade React from 15.3 to 15.5 and other needed packages DM-10488.
- PDAC:

    - Worked on adding WISE data search and display function in PDAC portal
    - worked on setting up user login page in PDAC portal
    - Started work on docker deployment of Firefly based applications

02C.05.02 Data Analysis and Visualization Tools

- Various bug fixes and improvements in Firefly reported by testers and Chris Suberlak:
    - Error message improvement.
    - Fix a bug to enable Firefly working in IE11.
    - Coordinate grid overlay on image improvements DM-6501.
- Python API and Jupyter widgets:
    - Add more examples to the firefly-client repository.
    - Add zoom and pan capabilities.
- Plotting capabilities:
    - Plotting architecture change to enable plotting multiple data traces in a same chart space DM-9945.
    - Adapted the line chart to use the use architecture mentioned above DM-10462.
    - New capability to save chart as PNG image DM-10275.
- Support Camera visualization by attending weekly meeting and answer questions.
- Image handling server side code refactoring to reduce memory usage DM-10370

02C.05.03 Alert/Notification Toolkit

- No new work done.

02C.05.05 User workspace

- No new work done.

<u>Planned activities: (for June., 2017)</u>

02C.05.00

- Continue to work with IPAC IRSA group on collaboration in Firefly development, plan and schedule coordination.
- Refine the F17 plan.
- Prepare for July NSF/DOE review.

02C.05.01 Basic Archive Access Tools

- Continue working on science platform vision as part of SUIT design.
- PDAC development:
    - Deploy PDAC with WISE data and login page.
    - Forced source search improvement, enable access by object ID in Stripe82 data.

- Build SUIT portal docker for deployment.

02C.05.02 Data Analysis and Visualization Tools

- New functions:
    - Change Firefly logarithm definition to be in sync with other languages DM-10349.
    - Add VOtable file upload function.
- Support Camera visualization work by attend the weekly meeting.

02C.05.03 Alert/Notification Toolkit

- No work planned.

02C.05.05 User workspace

- Some work as part of the Science Platform vision document.

Recruitment update:

- No activities.

SLAC / Stanford University

Current accomplishments:

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

- Gaponenko repaired “hole” between NCSA and IN2P3 halves of the Stripe82 dataset in the PDAC [DM-8241]
- Gaponenko completed load of initial WISE datasets (ALLWISE Source and MEP) in the the PDAC [DM-9372, DM-9373]
- Gaponenko documented the Stripe82 dataset loading process [DM-8406]
- Lo corrected a bug in PDAC where cutout images were returned without an attached WCS [DM-10364]
- Lo corrected a bug in PDAC where cutout images were off-center when the size was specified in angular degrees [DM-10441]
- Van Klaveren completed reload of the recovered the Stripe82 metadata into metaserv on the PDAC [DM-10374]
- Van Klaveren tweaked the SQLAlchemy pool_recycle parameter in the PDAC to reduce incidence of “mysql server has gone away” errors [DM0-10455]
- Mueller worked with team to complete F17 cycle planning
- Mueller attended DMLT Face to Face meeting at UW
- Van Klaveren and Lo to attended IVOA Interop meeting in Shanghai

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

- Lo read/researched relevant IVOA standards for imgserv V1 APIs

02C.06.02.01 Data Access Client Framework

- Pease added Swift backend support ("formatter") to the Butler. [DM-10528, DM-10529]
- Pease corrected a Butler bug where filters were somtimes set to None when they should have been left unset. [DM-10530]
- Pease removed caching code from the Butler because it was causing confusion/issues with unexpectedly shared objects. [DM-10558]
- Pease corrected an issue with the Butler where output repos could no longer be read if the repo path was from a symlink. [DM-10268]
- Salnikov and Pease attended Supertask working group meetings and participated in the SuperTask working group hack week.

02C.06.02.02 Web Services

- Van Klaveren improved error handling in dbserv so non-standard SQL errors thrown from MariaDB are now reported with sensible error codes and messages instead of the default SQLAlchemy "-1 totally whack." [DM-10576]
- Van Klaveren did groundwork coding for V1 metaserv, based on IVOA RegTAP. [DM-10375]

02C.06.02.03 Query Services

- Salnikov designed the Qserv API to be used for disconnected queries. [DM-3480]
- Salnikov implemented SHOW PROCESSLIST and SELECT … FROM INFORMATION_SCHEMA.PROCESSLIST which can display a list of currently running queries. [DM-2840]
- Gates corrected a long-standing temporary-table bug in Qserv that was preventing cross-database joins from working correctly. [DM-10283]
- Gates added seatbelt code to qserv czar to cancel queries if the returned result sets are becoming too large. [DM-10602]
- Gates corrected a query-id race condition in the qserv results mergine code. [DM-10734]
- James continued qserv containerization/deployment prototyping using Kubernetes on the IN2P3 clusters. [DM-10212, DM-10392, DM-10426, DM-10514]
- Mueller updated SciSQL tar-and-patch eups package from upstream, to address Python3 build issues. [DM-10569]

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.
- Gaponenko to complete load of WISE n-band catalogs into PDAC.
- Mueller to attend LSST Lyon workshop.
- Mueller to participate in document prep efforts for July DM review.

02C.06.01.01 Catalogs, Alerts and Metadata

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

02C.06.01.02 Image and File Archive

- Lo to do refactoring work within imgserv, and provide initial implementation of IVOA SOAD-based v1 API (current level of functionality only; no new cutout functionality).

02C.06.02.01 Data Access Client Framework

- Van Klaveren to port Pandas read_sql and to_sql to afw.table, in support of Butler DB backend.
- Pease to address issue in Butler where parent repository properties are dropped when loaded via child repositories.
- Pease to update outdated Butler documentation relating to composite datasets.

2C.06.02.02 Web Services

- Van Klaveren to complete port of DAX web services to Python3.
- Van Kalveren to complete initial implementation of metaserv v1 API.

02C.06.02.03 Query Services

- Gates and Hanushevsky to continue investigation into abnormally high memory usage on

the czar when processing large result sets.
- Gates to work on tying up some loose ends and instabilities with query cancellation.
- Gaponenko to continue work on data distribution design and prototyping.
- Hanushevsky to convert czar code to use XRootD SSI v2 API.
- Jammes to continue containerization/deployment research using Kubernetes.

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.

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, 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 started construction of a prototype IT configuration management database (CMDB) for the LSST Data Facility. Initial work was focused on mapping service dependencies to lower-tiered service layers as tuples and identifying service components (software, configurations, processes and scripts, documentation, verification tests, etc.). Progress was made on developing and refining the CMDB schema defining the relationships (e.g., “is a component of,” “is dependent on,” “is verified using,”). The CMDB provides a basis for phasing of service integration, sizing of physical system components, supporting documentation of DM architecture and enumeration of the product tree, and enumerating the documentation and verification tests for incorporation into the overall DM and commissioning integration, test, verification, and validation plans.

We made updates to and refactored LDM-144, the technology forecasting and costing model,

and updated elements (e.g., local networking) with current baseline technologies. This work is ongoing.

Two managers travelled to the DMLT face-to-face meeting in Seattle, where we presented on current status and near-term plans and discussed preparations for the DM review.

We prepared for the next cycle plan, including:

- Near-term storage use cases and proposed improvements to filesystem organization.
- Improved high-level specification of containerized application management operational model and near-term technical work.
- Current status of metadata and provenance models and mechanisms.
- Specification of header writing device, including clarification on data sources, contact, and mechanisms for retrieving metadata information.
- Clarification of process for shipping equipment to Chile.

Participation in the Chilean ITC Tiger Team and the DM Subsystem Science team continued.

DM-9656 Replanning, Commissioning, Operations Planning (S17b)

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 refined the mapping of Service Catalog dependencies. These mappings are required input into a service management platform to track dependencies among system services and components. We also continued work on forms and processes for Incident and Change Requests. For service monitoring, we configured and tested key performance thresholds in the prototype monitoring dashboard.

DM-8515 Service Management Definition & Refinement (S17b)
DM-8516 Enterprise Monitoring Integration – Phase 1
DM-9658 Enterprise Monitoring Design - Phase 2

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 deployment of the AA and network-based security services endpoints in Chile that will occur in the fall, coordinating with staff in Tucson and Chile, and continuing updating quotes with vendors.

We used the DM replan WBS as the basis for beginning a revision of the sizing/costing model of physical system components, including computing, storage, and networking components.

DM-8512 Planning Deployment in Chile
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 May, we continued to enhance this service façade with the ability to orchestrate and run more complex pipelines in the planned workload management framework using Pegasus and HTCondor. We successfully processed the full HSC Strategic Survey Program (SSP) Public Data Release 1 (PDR1) data using the legacy ctrl_pool processing framework. We continued our involvement in the SuperTask/Butler Working Group, providing feedback for requirements and design documentation.

NCSA personnel attended the annual HTCondor Week in Madison, Wisconsin, presenting a talk on use of HTCondor in the LSST Project.

DM-8333 Run selected existing science codes with initial Batch Production Service façade
DM-9662 Workflow creation improvements & expansion
DM-9664 SuperTask Working Group Activities

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 integrating the archiving and prompt processing Level 1 services, including work on our interface to the OCS system to continue support of the series of Early Pathfinder Software Integration Activities. We continued 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.

DM-8317 Enhancements for OCS Bridge components
DM-9659 Header construction and distribution architecture
DM-9661 Reintegration of archiving and prompt processing

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 for May involved adding timestamps to existing library code in order to better support collection of metrics for HSC reprocessing.

DM-9663 Emergent Middleware Work (S17b)

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 documented work on a prototype data backbone façade that used the existing DESDM framework. 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 are evaluating the Rucio data distribution package, used by the ATLAS experiment, distributing data within the Data Backbone to service access points (between the Base Facility, NCSA, CC-IN2P3, and other sites, including camera test stands at SLAC and Tucson).

DM-9665 Prototype Data Backbone Façade

02C.07.03 Environment and Tools

QA Environment

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

We continued work to stand up a Jupyter hub instance to support upcoming QA activities, providing access to common datasets one the GPFS filesystem, and documenting configuration and use of the service.

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.

We concluded migration of users from NFS to GPFS file systems, including decommissioning related computing resources. Implementing of disaster recovery procedures for datasets continued. We completed procurement of the systems to be used in raft-scale testing of the L1 Services, and neared completion of acquisition activities for the Authentication and Authorization deployment in Chile.

DM-8502 Disaster Recovery for Science Datasets
DM-8503 Migrate Users from NFS
DM-8505 Decommissioning Hardware
DM-9685 Procurement Activities for L1 Test Systems

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.

We continued refining prototype system to monitor the services provided on the development cluster, testing group- and service-based authorization mapping for use with individualized dashboards, and implementing a data retention policy for log files and related data. Emergent work included setting up a cluster to investigate the use of elastic services.

DM-9666 Service Management & Emergent Work (S17b)
DM-9667 Cluster Service Monitoring Development

Planned activities:

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

Planning and Management

We will continue preparations for the DM review, including integration of L3 milestone dependencies, documentation, and presentations. Documentation includes porting concept of operations service documentation to the new DM document framework.

We will continue development of the CMDB and mapping of service dependencies. This provides a basis for phasing of service integration, sizing of physical system components, supporting documentation of DM architecture and enumeration of the product tree, and enumerating the documentation and verification tests for incorporation into the overall DM and commissioning integration, test, verification, and validation plans.

We will continue work on a capacity model for the pre-operations phase of the project based on use cases of archiving raw data from test stands and from instruments as they are commissioned, as well as processing use cases for continued development support, service integration, and production. Requirements from DM, the commissioning team, pre-operations plan, and EPO will be gathered and synthesized.

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

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

Service Management and Monitoring

We will finish planned work integrating Nagios data into the service-level monitoring dashboard prototype. We will continue work improving service management processes, including developing single entry points for incident response and request response. We will continue maintaining and evolving the service management framework, including evaluation of a formal enterprise service management tool. We will attend the ServiceNow conference for evaluating an industry-standard a service management framework and schema.

DM-8516 Enterprise Monitoring Design – Phase 2

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 submit plans for the AA hardware installation in Chile to the project for final approval. We will resume 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 procurements. We will also begin acquisition planning for fiscal year 2018, and begin the semi-annual refresh of the FY17 sizing model.

DM-8512 Planning Deployment in Chile
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

02C.07.01 Processing Control

Batch Production Services

We will continue development of the workflow-based Batch Production System, designing and beginning implementation of data staging and the ability to orchestrate and run more complex pipelines in the planned workload management framework using Pegasus and HTCondor . We will continue to enhance the DESDM-based service façade, scaling up to larger datasets and setting up a test database instance for querying metadata. We will finish the report on reprocessing HSC Strategic Survey Program (SSP) Public Data Release 1 (PDR1) data, and make the data available to developers. We will continue reprocessing HSC data based on DM needs, to both verify the software stack and to test the development of science algorithms.

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

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

Common Middleware and Other Tasks

Work for June 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 work gathering requirements for file services in the 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. We are evaluating the Rucio data distribution package, used by the ATLAS experiment, distributing data within the Data Backbone to service access points (between the Base Facility, NCSA, CC-IN2P3, and other sites, including camera test stands at SLAC and Tucson). We will also evaluate candidates for consolidated database services, and begin modeling data flow to properly determine capacity and throughput requirements for database services.

DM-8331 Gather requirements for and specify design of file portion of Data Backbone services
DM-10707 File distribution and management framework
DM-10708 Consolidated database technology evaluation
DM-10709 Operational concepts of databases supporting production data flows

02C.07.03 Environment and Tools

QA Environment

We plan to finish work providing a Jupyter hub instance and access to datasets to support upcoming QA activities will continue with access to common datasets one the GPFS filesystem, and documenting configuration and use of the service.

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

02C.07.04 Site Infrastructure

Hardware Acquisition, Deployment and Provisioning
Work on implementing disaster recovery procedures for datasets will be completed. We will procure and provision the systems to be used in raft-scale testing of the L1 Services, and begin acquisition and provisioning activities for the AA deployment in Chile. To prepare for planned FY17 acquisitions and upgrades, we will begin the process of refreshing and reorganizing the computing lab at NCSA-3003.

DM-8502 Disaster Recovery for Science Datasets
DM-10712 Chile Base AA Acquisition & Installation
DM-10713 Level 1 Test System Provisioning Phase 1
DM-10715 NCSA 3003 Lab Refresh

Service Management for LSST development resources, Emergent and Miscellaneous Work
Work for May is anticipated to include addressing user- and developer-driven requests, and related system needs. We will continue to investigate and prototype additional means to monitor the services provided on the development clusters, including adding monitoring to the PDAC based on developer needs.

DM-10716 System Monitoring Refinement (F17a)

Recruitment update:
Kay Avila began work as a Security Engineer in NCSA's Incident Response and Security group. Jake Rundall was hired as System Engineer, and will assist with administering LSST development clusters.

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: Submitted LCR-963 to start review for baselining LSE-309 Summit Base ITC Design Document (SBIDD). As previously reported, the initial baseline, while complete for the Summit and Summit - Base, will still be incomplete for the Base. This is unavoidable due to the more advanced schedule for Summit Facility and Network and Summit - Base Networks, which are already being procured/constructed. To complete the Base section, we will need to have a complete Control Room design, expected after an 8-month INRIA design contract. When the Tiger Team disbands in August, 2017 SBIDD responsibility will migrate to "regular" System Engineering team. In parallel LCR-964 to baseline LSE-299 Summit Computer Room Requirements Document.
- Summit and Base Networking and Computing: The selection committee ranked the bids for the Summit LAN and selected Dimension Data/Cisco. The selection report was sent to procurement indicating the results. Delivery is in 45 days so we have a month to work out Final Bill of Materials/contract. The new fibers were installed by Telefonica into the NOAO computer room. Detailed planning of racks and network equipment installation continued this month. Completed support for EIE dome team and restoration of phones and internet in their containers on Pachon after a severe storm.

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: Fiber measurements made in same test as La Serena - AURA Gatehouse below.
- DWDM Equipment: Coriant is ready to ship the AURA/LSST DWDM equipment from Berlin. The REUNA-CORIANT/Raylex contract was signed. The PO of the equipment for La Serena – Santiago was issued. A kick-off meeting was done with CORIANT/Raylex to review the equipment architecture, training activity and the deployment plan.
- Santiago-La Serena: The site survey has been completed, in preparation for deploying the DWDM equipment. Telefonica is making final corrections to some housing elements.
- La Serena – AURA Gatehouse: The deployment of the local loops is finished. Report #2 “Report of delivery of fiber from Telefonica” to Contract 1A Annex III B has been delivered. The report includes the fiber measures in term of attenuation and length. All measurements are very good and meet requirements.

02C.08.03.02 International Chile - US WAN

- 100 Gbps Managed Ring: No planned activity in this period. The ring is performing as expected.
- Management and Coordination Contract:
  Conducted the regular monthly meeting of the LSST Network Engineering Team (NET), in which the NET revised the Network Utilization sheet of LDM-142, discussed updates to LSE-78, the Network Verification Plan and Matrix, and the Network End to End Test Plan.
  Members of FIU attended several network conferences and symposia relevant to LSST networking:
  Terena Networking Conference (TNC) in Linz, Austria; Global Lambda Integrated Facility (GLIF) and Global Network Architecture (GNA) meetings; 35th Brazilian Symposium on Computer Networks (SBRC); 2017 Brazilian NREN (RNP) Workshop; 8th Workshop on Experimental Research of Future Internet (WPEIF); 27th LACNIC and 34th Brazilian Network Operators Group (GTER) meeting.
- Spectrum Contract: AmLight engineers finished a document containing questions to be addressed by Angola Cables, in order to move forward with the DWDM system design and specification to support the spectrum utilization. The questionnaire consists in three parts: 1) Minimum information required per optical route; 2) Additional information required; and 3) Key Parameters about the Alien System that may improve the accuracy of the design. This approach was chosen to facilitate the understanding.
- US National WAN: The PI at FIU, Julio Ibarra, met with the ESNet Director Inder Monga at the TNC17 conference in Austria this month. Inder is reworking the costs again lower. We remain confident in the approach LSST is taking where any expenses are addressed through DOE accounting. The overall timeline may slip a few months from 12/17, but we remain confident that this work now will be best for LSST in the long term- including after the initial operating period.

Planned activities:

02C.08.01 Base Center

- Summit Base ITC Tiger Team: LSE-309Summit - Base ITC Design Document and LSE-299 Summit Computer Room Requirements Document to CCB.
- Summit and Base Networking and Computing: Meet with Cisco to finalize the Bill of Materials. Continue planning the installation of the summit network and arrangements to bring the Luis Corral up to speed with Cisco ACI technology. Visit the summit to plan ducts and cable trays.

02C.08.03 Long-Haul Networks

02C.08.03.01 Chile National WAN

- Summits - AURA Gatehouse Network: Complete measurements of the fibers between the Summits and Gatehouse.
- DWDM Equipment: We expect to receive AURA/LSST DWDM from Coriant by end of June. REUNA and Coriant/RAYLEX finish deployment plan.
- Santiago-La Serena: Continue the supervision of Telefonica's corrections to housing nodes to be ready for the installation of the equipment.
- La Serena - AURA Gatehouse: Receive the measurements of the fiber and accept the fiber.

02C.08.03.02 International Chile - US WAN

- 100 Gbps Managed Ring: Next planned activities are scheduled for June 2017, when the Pacific 100G circuit will be opened in Chile.
- Management and Coordination Contract: Complete updates by LSST Network Engineering Team (NET) to resolve issues/comments on Network Verification Plan and Matrix. Complete a revision to the Network End to End Test Plan. Submit DM change request to adjust bandwidth allocations in LDM-142 based on NET comments.
- Spectrum Contract: FIU and Angola Cable continue to work out operations and maintenance payment timing.
- US National WAN: Negotiations to refine and improve the ESNet service and costs will continue during June.

Recruitment update:

- Of the two Network Engineers who were interviewed in La Serena, one accepted. Luis Corral will start work June 19th. We will onboard the new hire and arrange for courses. We will revert to hiring an IT/Network Technician and continue with the hiring process for the System Engineer/Admin (F2F interviews with Technicians. phone interviews with System Engineer/Admins).