



LSST Data Management Status

William O'Mullane, AURA/LSST
DM Project Manager

DM All Hands
6th March 2018





Outline



Data Management Overview

Risks and Opportunities

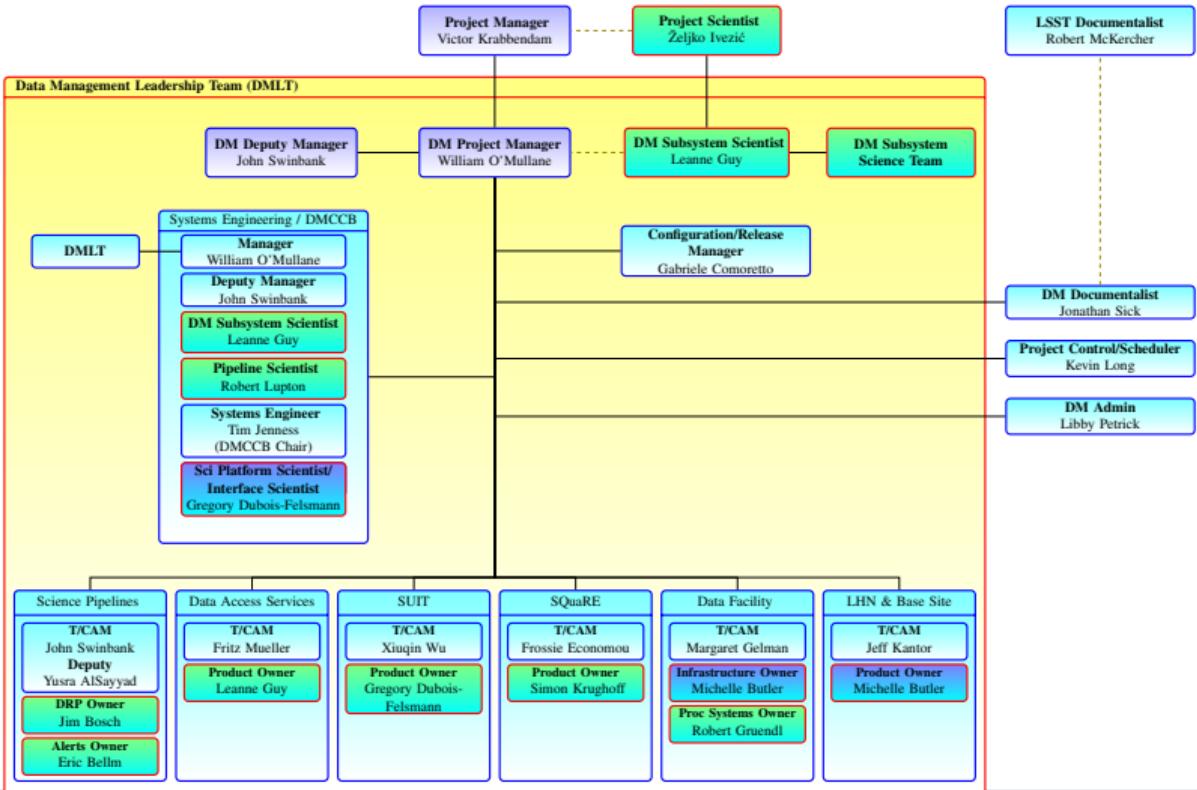
Status

Conclusion

Conclusion



Organization



Welcome Leanne Guy and thanks to Mario (who stays with us)!!

Welcome Michelle Butler and thanks to Don

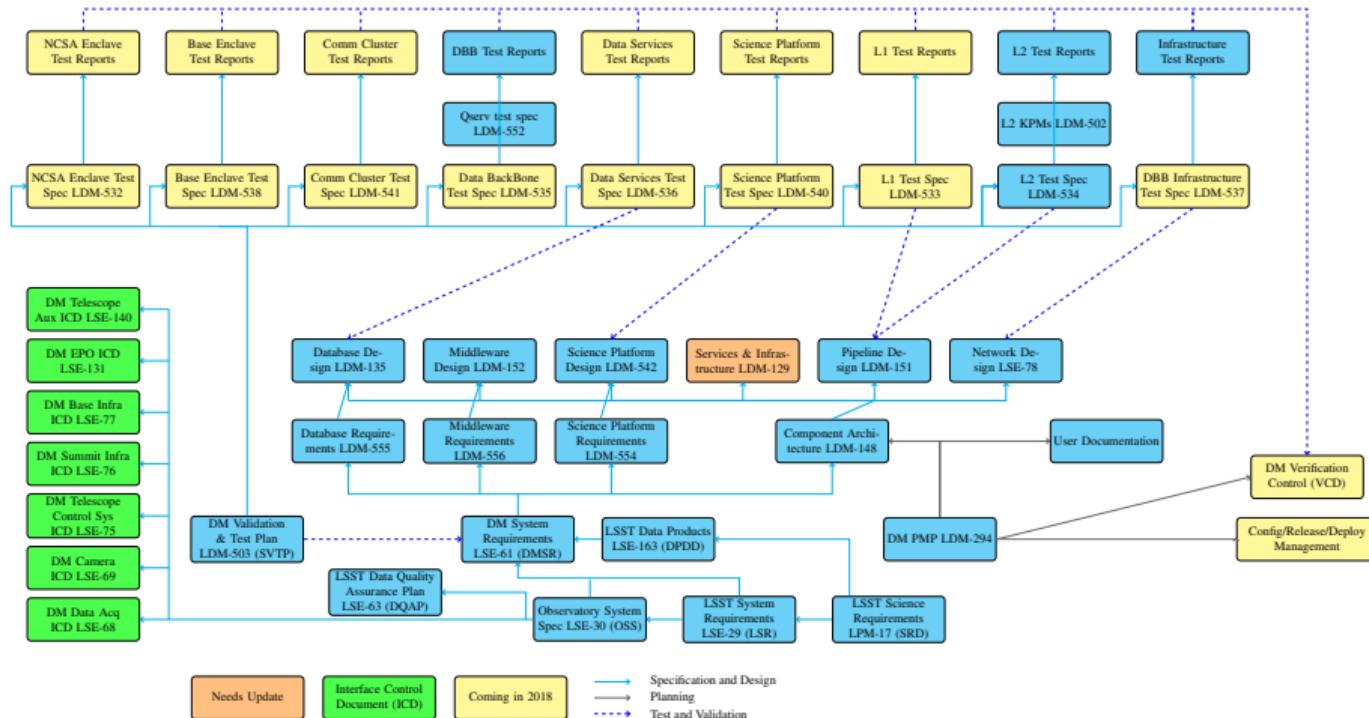
Welcome Gabriele Comoretto.

Deputies John Swinbank (PM) and Yusra AlSayyad (Pipelines)

Toughest thing in any project is communication.

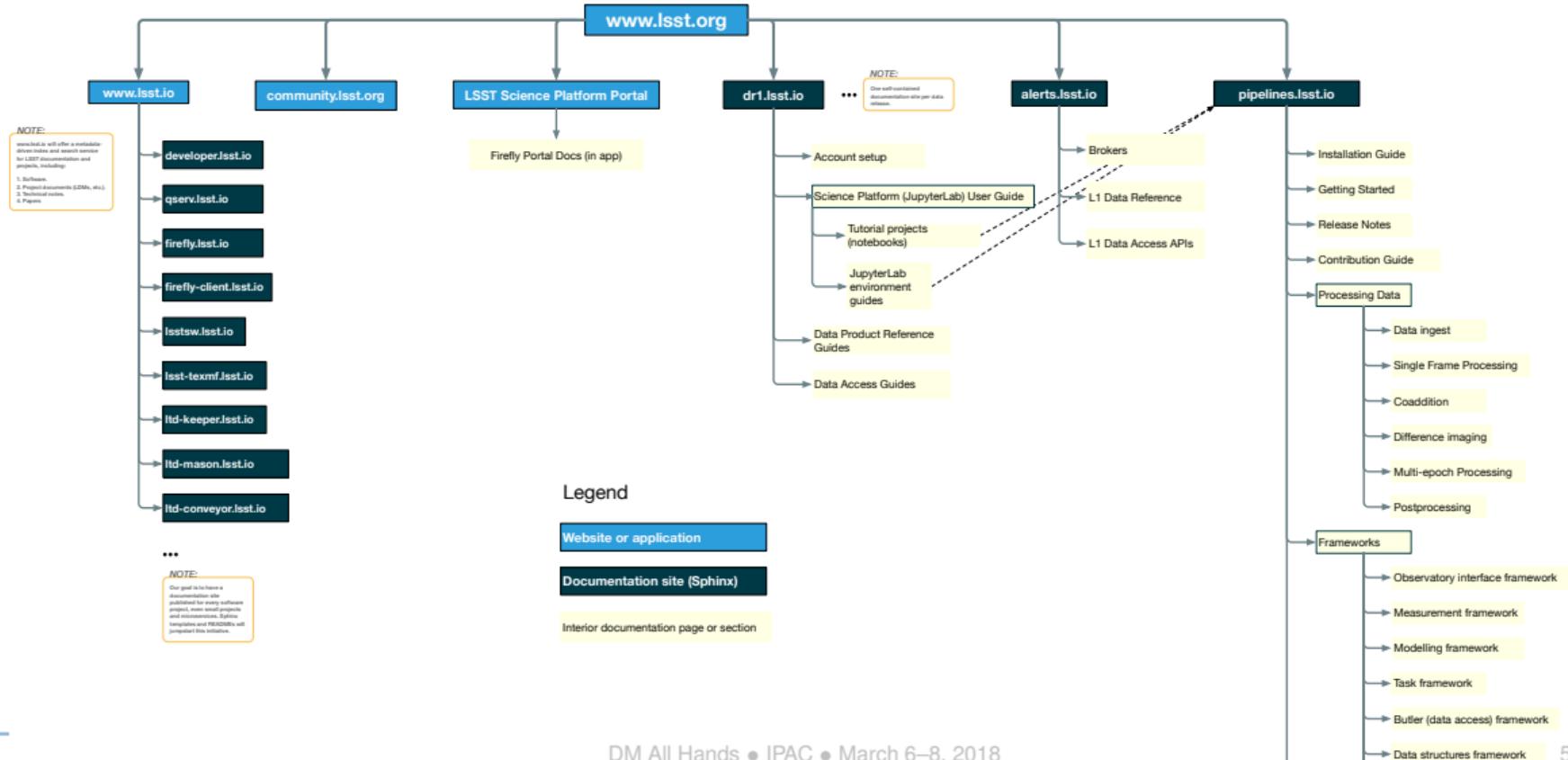


DM Document Tree



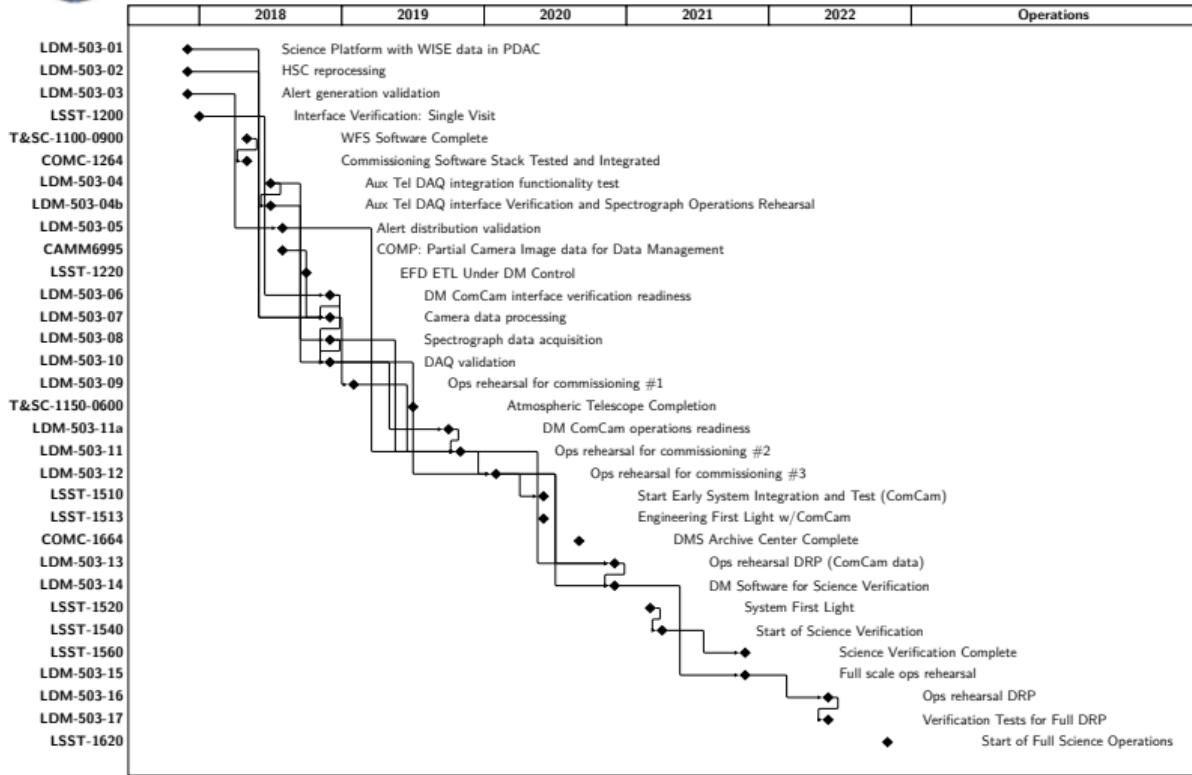


End-User Documentation on the Web





Verification is a Priority



Across all of LSST verification is a big topic right now.

DM is adopting a test driven schedule to better address this.



Verification: Have we built everything we are supposed to build?

- In line with the Project's System Engineering approach
- Demonstrate that we cover all requirements on DM
- LDM-503 shows the DM verification matrix

Validation: Have we built the right thing and does it work as expected?

- Must tackle *both* Scientific and Operational Validation
- Talking with the Commissioning Team: some *rehearsals* will be joint

LDM-503 addresses DM's plans for verification & validation.



High level status



- The DM Basline is now the reviewed plan (LCRs processed)
- S18 detail plans submitted
- Milestones now being reported monthly
- Level 2 milestones achieve in December
 - LDM-503-2 Test report: HSC Reprocessing -DMTR-51
 - LDM-503-3 Test report: Alert Generation Validation -DMTR-53
- Level 2 milestones delayed in December
 - LDM-503-1 Test report: Science Platform with WISE data in PDAC
 - Instability PDAC hardware lack of personnel,
 - Testing command last week



High Level Goals



- 2017: Prototyping data access and first access to hardware
 - Jun: Prototype Data Access Center with SDSS + WISE Data
 - Aug: Working with camera test stand data
 - Dec: Prototype notebooks, private databases for Science Platform
- 2018: Prototypes for various processes and databases - “Minimum Viable System”
 - Jun: Calibration Products accessible through Butler
 - Aug: Mountain base network up
 - Oct: Spectrograph data acquisition
 - Dec: Prototype QA/Commissioning Environment
- Dec 2019: ComCam L1, L2 Production
- Dec 2019: Base Center Integration Complete
- Jun 2020: Camera L1, L2 Production
- Jul 2021: US Data Access Center Integrated

Test plans to confirm milestone completion are under development.



Commissioning Start Requirements



November 2019: DM for Commissioning (minimum required for start of commissioning with ComCam): **(See LSE-79 §3.3 and table 8)**

- Pipeline: single-frame measurement including ISR, ghost masking, cosmic ray detection, PSF estimation, astrometric and photometric calibration, background estimation, single-frame deblending, master calibration image generation, atmospheric characterization
- Services: archiving, EFD transformation, Data Backbone for files (Base/NCSA), telemetry gateway, OCS-controlled batch, offline processing
- LSST Science Platform on Commissioning Cluster: Notebook Aspect, image access, user file storage, batch computing

Milestones:

- LDM-503-9 – 2018-11-30: on the right track with beta software.
- LDM-503-11 – 2019-10-31: verification of ComCam commissioning requirements.



Outline



Data Management Overview

Risks and Opportunities

Status

Conclusion

Conclusion



Realized Risks



As part of the replan three risks were realized:

- DM-062 Programming team productivity below estimate due to geographical distribution/competing priorities.
- DM-085 SUI workload underestimated.
- DM-087 SUI requirements change.

These risks are being addressed in the new DM plan and the associated request to draw on contingency.

New management and organization allowed us to reduce exposure on some other risks. This plus assessment of exposure on other risks reduced our risk exposure at the level of ~\$12M (to ~\$12M).



Top risks



DM-018 Computing power required for Data Release Production exceeds estimates

- Lots of verification testing (e.g. HSC processing)

DM-023 Unanticipated characteristics of real data result in poor MultiFit performance (computational)

- Lots of verification testing (e.g. HSC processing)

DM-042 Loss of key personnel

- Management structure reducing ‘single points of failure’
- Focus on written design documentation & verification plans

DM-021 Object counts exceed expectations, leading to insufficient compute

- Improved modeling: work in progress (Jurić)

DM-032 LSST DM hardware architecture becomes antiquated

- Core algorithmic code is flexible and hardware-agnostic
- Actively tracking and anticipating the state of the art



Outline



Data Management Overview

Risks and Opportunities

Status

Conclusion

Conclusion



Pipelines





Architecture





Data Access Services





- Observatory Operations Support (Level 1) Services
 - Working closely within the LSST Systems Engineering Early Pathfinder group, developing and incrementally testing integration of T&S, Camera, and DM service software via a series of early integration activities.
 - Initial header service developed and configured for Camera subsystem and AuxTel use cases, ability to acquire pixel data and write FITS files, all commandable by OCS. Demonstrated on Level 1 Complete Test Stand.
- Offline Campaign Processing and Batch Production Services
 - Production processing of HSC data for 503-2 milestone, working with DRP group.
 - Upgrades to existing production framework based on DES data management system, including migration to Python 3.
 - Continued regular dataset reprocessing to provide datasets for developers based on biweekly stack updates.



- Data Backbone Services
 - Single-node consolidated database operational, supporting integration and testing of the batch production service.
 - Provided operations use cases and developed requirements for Gen3 middleware working groups.
- Container Application Management Services
 - Investigated deployment documentation and service level configurations requirements for FY18 Kubernetes deployment.
 - Worked with SLAC, SQuaRE, and IPAC for requirement investigation and deployment configurations for initial use case support.
- Developer Support Services
 - Upgrades to PDAC, including new dedicated Qserv head node provisioned in PDAC and interim Kubernetes installed on single PDAC node for special testing



- LSST-DB (MySQL developer database) replacement provisioned.
- Expand Level 1 Complete Test Stand to support early pathfinder testing and other cross-subsystem use cases.
- ITC and Facility
 - 11 of 18 FY17 and FY18 capabilities provisioned including: initial production Kubernetes cluster, infrastructure for Chilean Base network security and identity management endpoint (shipped, installed, tested, and shipped), 3PB GPFS expansion for initial production file systems, consolidated database (Oracle cluster), Level 1 Complete Test Stand expansion, system for Tucson ATS Test Stand, data transfer nodes, disaster recovery capability for site file systems, central core network upgrade supporting large capability expansion.
 - Matured system configuration management and operations with xCAT and Puppet.



LSST Data Facility (NCSA) IV





- AuxTel Operations (Fall) (LSST- and DM-level milestones)
 - AuxTel Test Stand operations (April).
 - Provisioning of Chilean ITC (Summit facility) for AuxTel operations.
 - Spectrograph Archiving Service building raw images, Header Service.
 - Data transfer to NCSA, ingestion and archiving into initial version of production Data Backbone (single site).
 - CPP production processing.
 - EFD ETL Service operational, including database and Large File Annex.
- Chilean Base AA and security monitoring systems installed, services deployed, and monitored (April 2018) in NOAO facility.
- Application-level integration with service monitoring framework, initial dashboards for service operations and capacity management
- Supporting ongoing development
 - Initial Kubernetes Service in VERY friendly user mode March 2018. Integration for developers use cases (e.g., hosting Jupyter notebooks, Alert Distribution system testing, etc.) throughout CY18.



Base & Network:DWDM



Installed Dense Wave Division Multiplex (DWDM) equipment on Cerro Pachon and in La Serena



Base & Network:Summit Computer Room



Coordinated implementation of LSE-299 Summit Computer Room Requirements Document with Summit Construction Team



Base & Network:Fiber First Light



Successful transfer of digital data over LSST/AURA fiber optic networks from the Summit Site on Cerro Pachon to NCSA. A set of 6 x 10 Gbps Network Interface cards on Data Transfer Nodes (DTN) configured with iPerf3 generated a sustained data rate of approximately 44 gigabits per second, over a period of 24 hours, exceeding the target of 40 gigabits per second.

- Initial Summit Network Ready (DM-NET-3)
- Summit Computer Room installations (Aux Tel, EFD, ECS, etc.)
- Base LAN Design Complete
- LSE-239 Base Data Center Requirements updated/rebaselined
- 100 Gbps test between Chile and NCSA
- Activation of Spectrum Link (Sao Paolo → Boca Raton)



SQuaRE





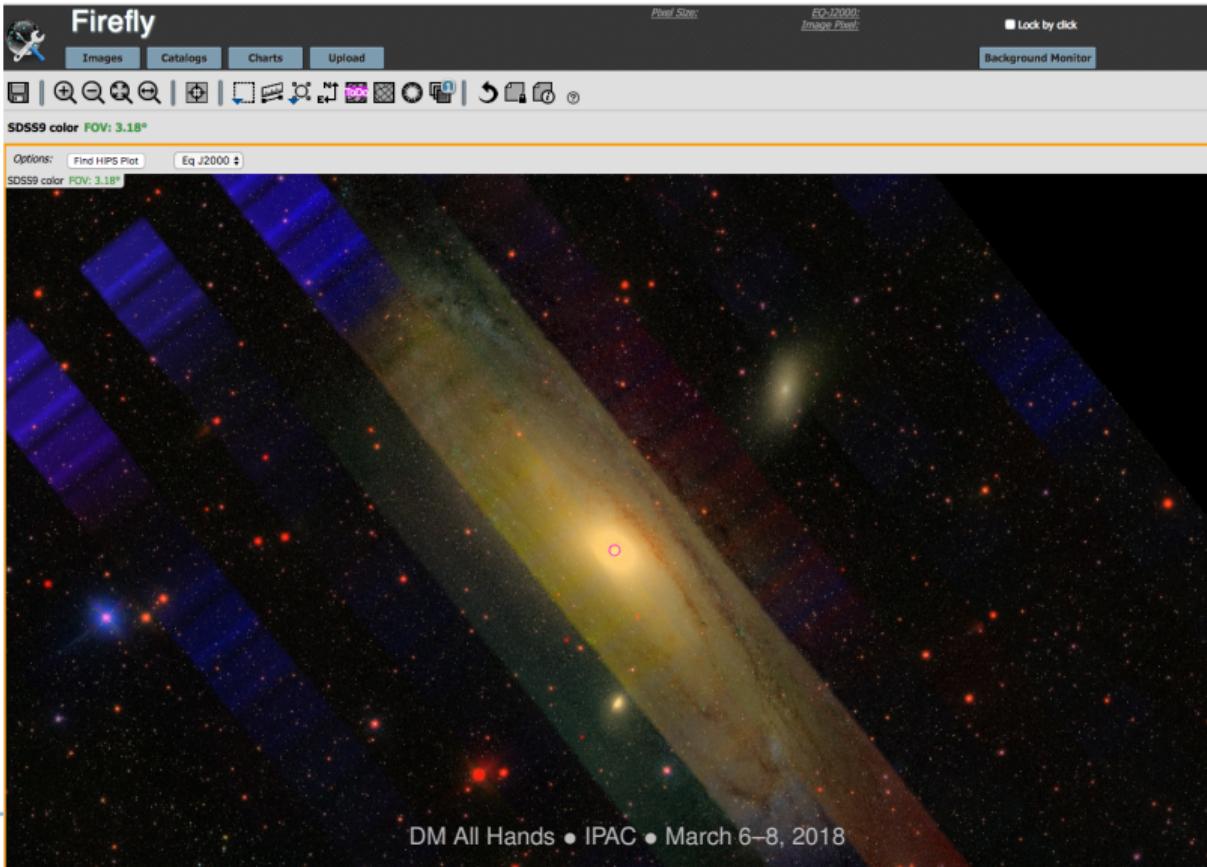
SUIT: Major Achievements



- New Features in PDAC
 - Search and display WISE single epoch sources
 - Incorporated the DAX imgServ v1 in SUIT
- Visualizaiton
 - Search and display HiPS images
 - Multiple scattered plots within one plotting area
- Build and deployment
 - Experimented with docker and Kubernetes
 - Added Jenkins task to build Firefly docker image
 - Added Jenkins tasks to deploy/destroy Firefly application as Kubernetes pod
- Implemented embedded database HyperSQL for table data support in Firefly
- Updated 3rd party packages: React, Webpack ...
- Various improvements and bug fixes



SDSS HiPS images in Firefly





SUIT: Plan for Next 6 Months



- New Features in PDAC
 - Search and display HSC data reprocessed by LSST pipeline
 - Use the new version of DAX APIs: metaServ, imgServ, dbServ
 - Integrate the authentication system
 - Access workspace when available
- Improve overall user experience in SUIT
- Make Firefly npm installable to support using Firefly API in JupyterLab



- Data Access Policy LSE-349 draft prepared; now under review by project management
- Alerts Broker Policy LDM-612 draft is underway; UW broker team focused on ZTF Alerts
- DM & Special Programs DMTN-065 spawned RFC-412 and Jira tickets under discussion (to close in next month)
- Initial testing of pipeline performance in crowded fields is complete; written report available soon.



Outline



Data Management Overview

Risks and Opportunities

Status

Conclusion

Conclusion



Outline



Data Management Overview

Risks and Opportunities

Status

Conclusion

Conclusion



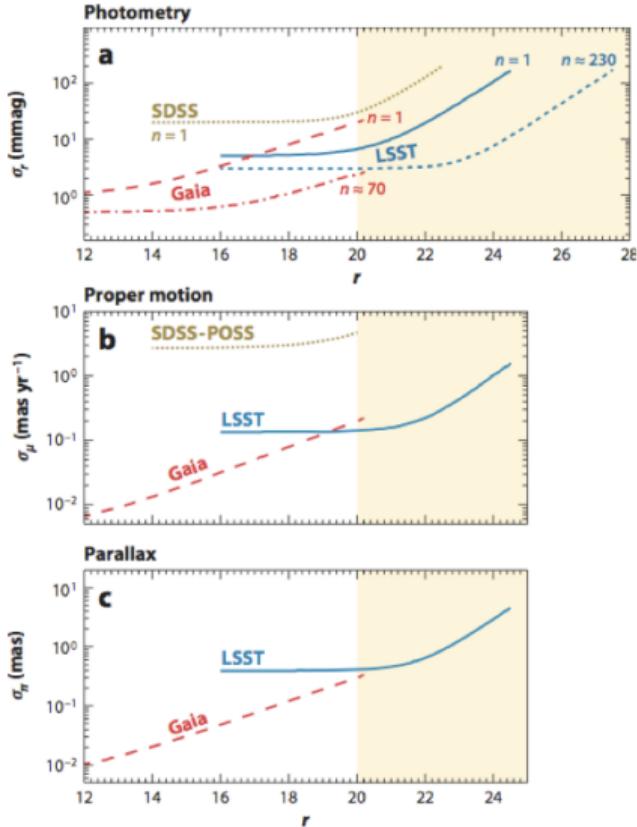
Conclusion



- New DM Project Management is in place since April 2017.
- Organization, prioritization and specification have been the first priority.
- To date encountered a real willingness to cooperate on DM and *get the job done*.
- Will make pragmatic and timely decisions to see DM works.
- We had a successful NSF/DOE Review in July. Got some recommendations on how to make an *even better* DM.
- LSST has the potential for a *huge* impact in astronomy; DM is fundamental to this.
- Looking forward to the first LSST images!



The END



The Milky Way disk *belongs* to Gaia, and the halo to LSST (plus very faint and/or very red sources, such as white dwarfs and LT(Y) dwarfs).

Željko Ivezić

Questions ??



Outline



Reference material



Acronyms I



Acronym	Description
IPAC	Infrared Processing and Analysis Center



References I



[LSE-79], Claver, C., The LSST Commissioning Planning Team, 2017, *System AI&T and Commissioning Plan*, LSE-79, URL <https://ls.st/LSE-79>

[LDM-503], O'Mullane, W., Jurić, M., Economou, F., 2017, *Data Management Test Plan*, LDM-503, URL <https://ls.st/LDM-503>

[LSE-239], Petravick, D., Hoblitt, J., Lim, K.T., et al., 2016, *Base Facility Data Center Design Requirements*, LSE-239, URL <https://ls.st/LSE-239>