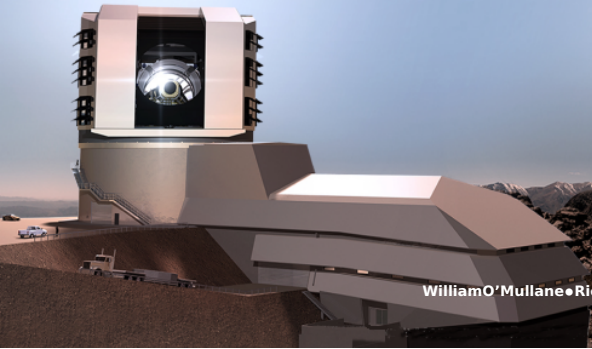
LSST Data Management Overview

DM Project Manager

LSSTRio
25th September 2018



Outline

Introduction

LSST status

Data Management Overview

Data Management Status and Achievements

Conclusion

Milky way analogue

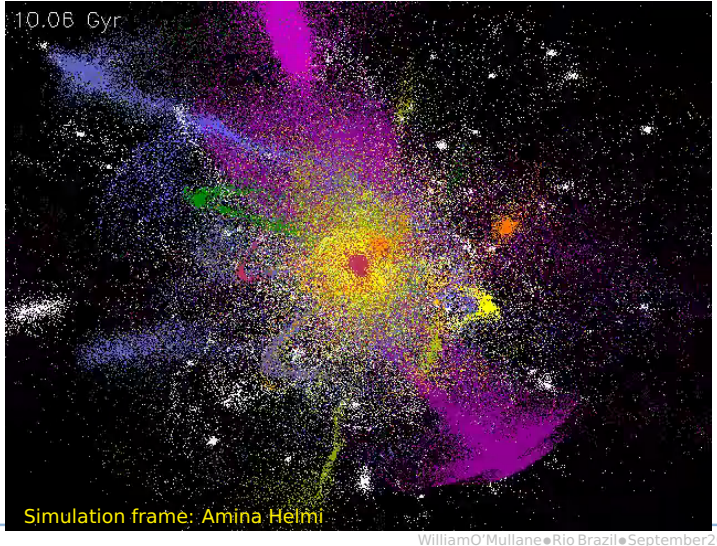
Can we unravel the formation, composition, and evolution of the Galaxy?

Jos de Bruijne

‘Our sun’ →

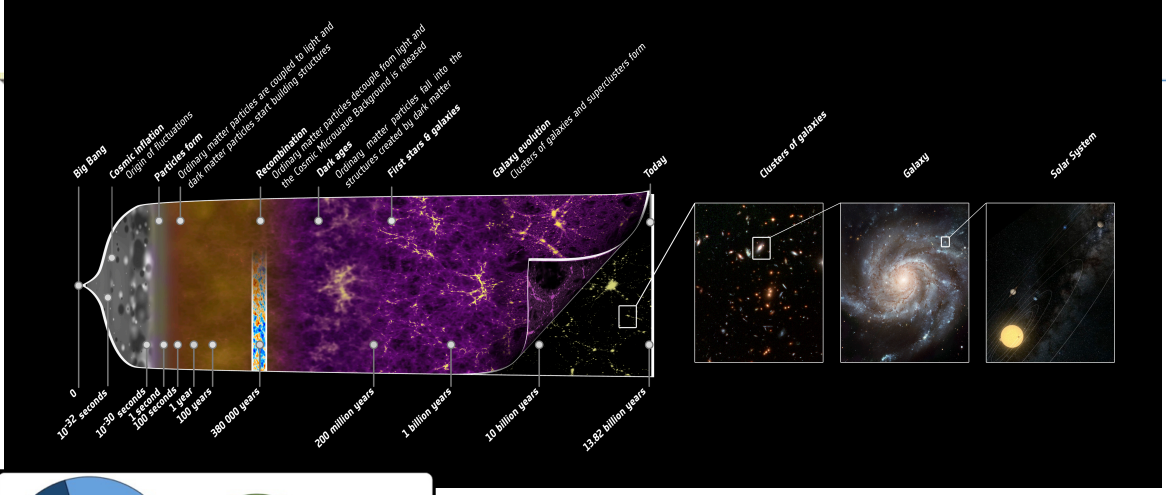
Our view is severely obstructed by the dust in the disk and relatively little is known about the origin, history, and structure of our own Galaxy

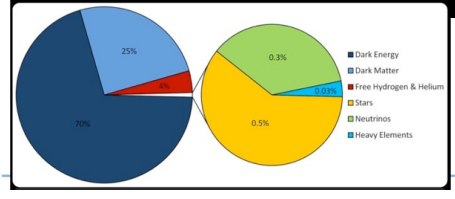
Origin of the Milky Way



Simulation frame: Amina Helmi


Image credit: R. Jay GaBany





The modern cosmological models can explain all observations, but need to *postulate* dark matter and dark energy (though gravity model could be wrong, too)

Killer asteroids: the impact probability is not 0!

LSST is the only survey capable of delivering completeness specified in the 2005 USA Congressional NEO

The Barringer Crater, Arizona: a 40m object 50,000 yr. ago

Outline

Introduction

LSST status

Data Management Overview

Data Management Status and Achievements

Conclusion

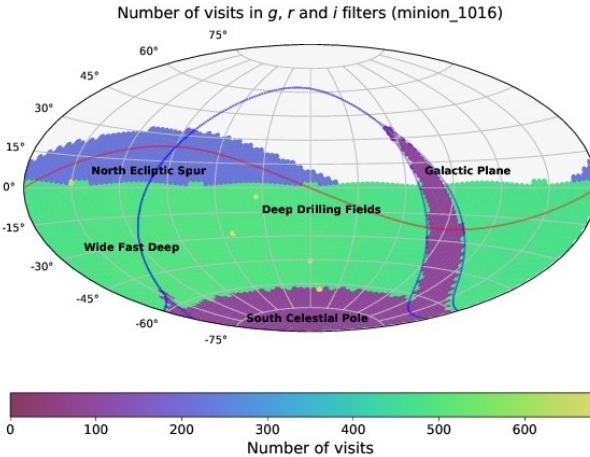
LSST:uniform sky survey

An optical/near-IR survey of half the sky in ugrizy bands to r 27.5 (36 nJy) based on 825 visits over a 10-year period: *deep wide fast*.

- 90% of time spent on uniform survey: every 3-4 nights, the whole observable sky scanned twice per night
- 100 PB of data: about a billion 16 Mpix images, enabling measurements for 40 billion objects!

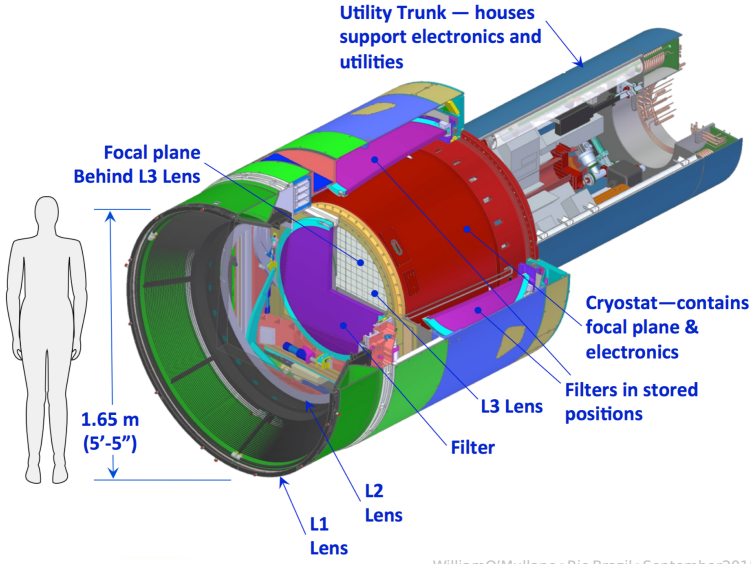
see also `http://www.lsst.org` and Ivezic et al. (2008)-arXiv:0805.2366

Call for white papers - `https://www.lsst.org/call-whitepaper-2018`



10-year simulation of LSST survey: number of visits in u,g,r band

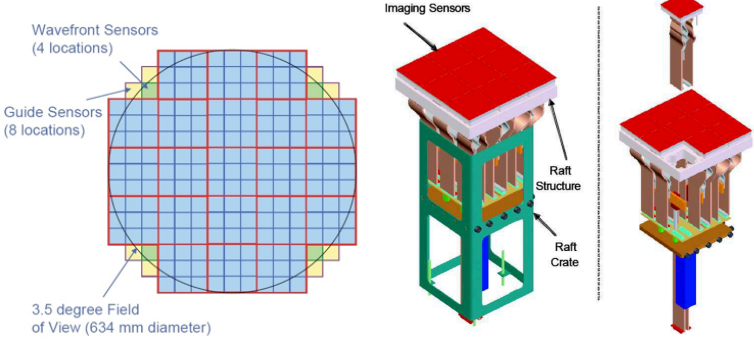
(Aitoff projection of eq. coordinates)

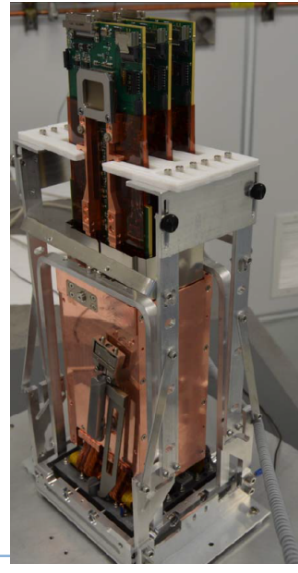


The largest astronomical camera:

- 2800 kg
- 3.2 Gpix

Science rafts





Modular design: 3200 Megapix = 189 x16 Megapix CCD

9 CCDs share electronics: raft (=camera 144 Megapix)

First of 21 rafts $\longrightarrow$

About to accept final raft.

Site shaping up (July 2018)

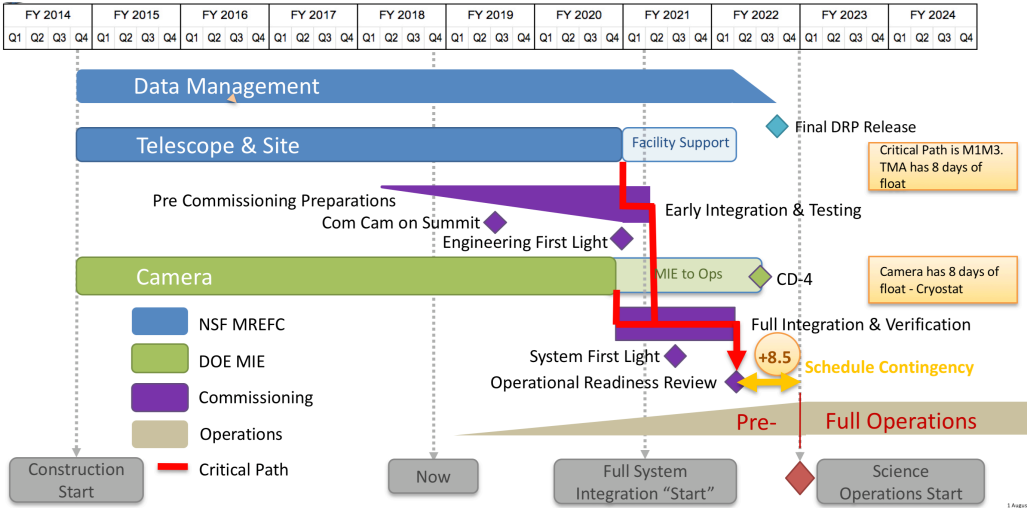
Marine tracking and logistics (1st Sept 2018)

- 9 Pallets of M1M3 lift fixture hardware is MIA! Found in Phoenix warehouse Chapter 7 Trustee Custody

- Valparaiso Express 5 containers held up in Peru, transshipment required to MN Callao Express (already had 10 Containers)
- BBC Arizona Late to Antwerp for Coating vessel & 7 crates
- K&N warehouse Ready in Chile for 15 containers
- Los Angeles Warehouse ready for M2
- 6 EIE Containers in Italy scheduled for Packing in Sept.

Coating Chamber now en Route

LSST Project Schedule



Outline

Introduction

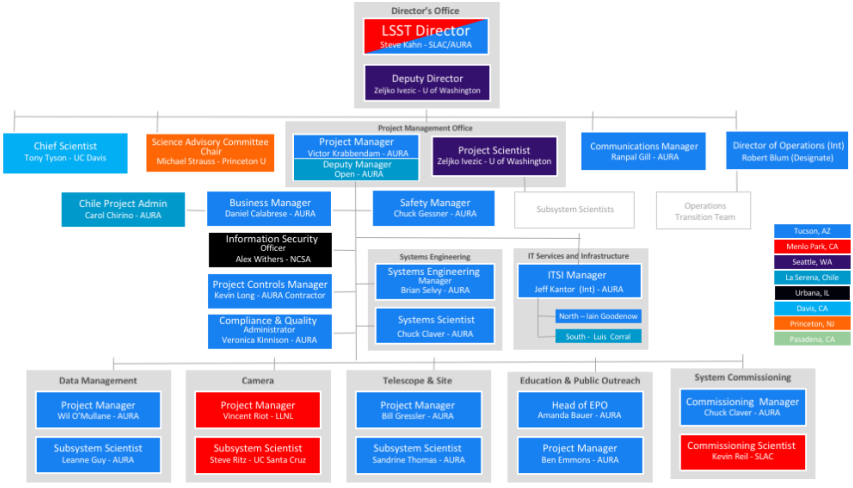
LSST status

Data Management Overview

Data Management Status and Achievements

Conclusion

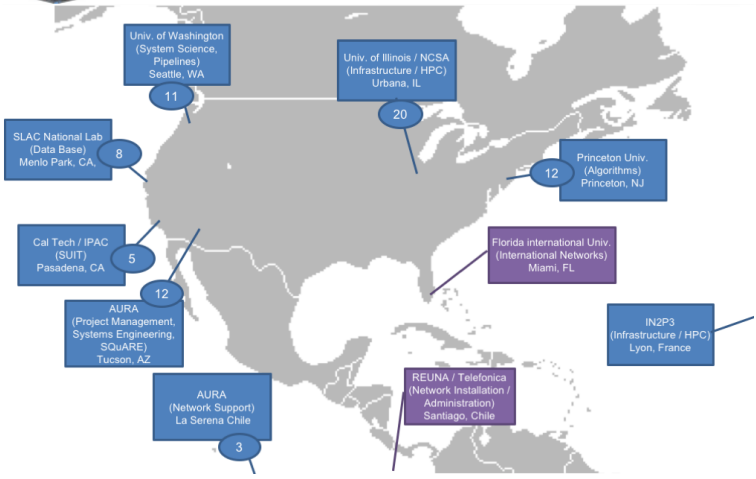
LSST Org Chart: Where DM Fits



LSST Project is large and dispersed.

⇑ Data Management is just one of five subsystems.

Mission Statement



DM's mission:
Stand up operable, maintainable, quality services to deliver high-quality LSST data products for science, all on time and within reasonable cost.

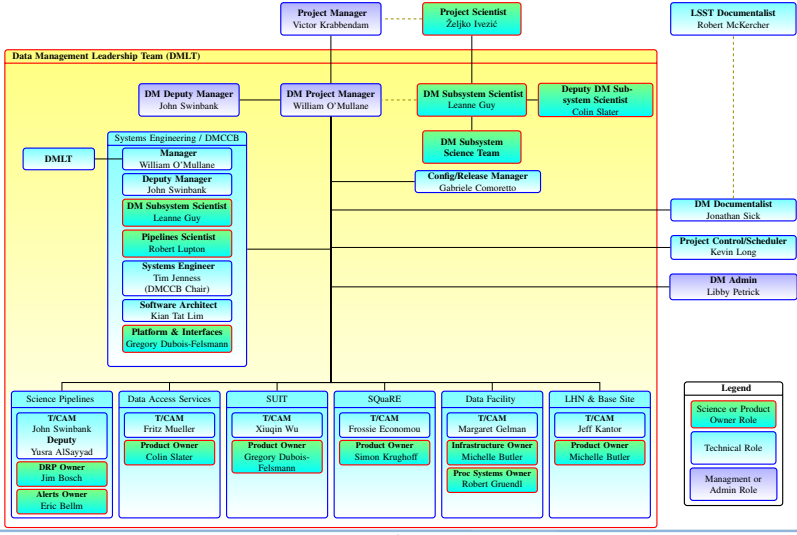
Development is distributed across the Americas.
Plus we have partners like IN2P3.

See Management Plan LDM-294.

Data Management Team

DM with System Engineering Joint All Hands Meeting IPAC 2018

Organization



Leanne Guy replaces Mario Jurić (who stays with the project) as Subsystem Scientist.

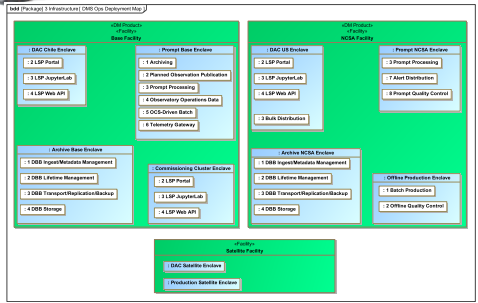
Michelle Butler replaces Don Petravick as Infrastructure Product Owner.

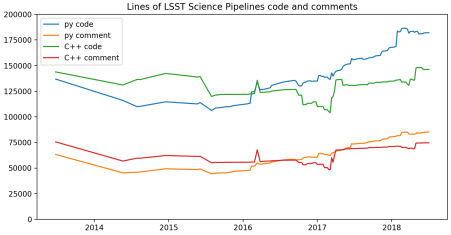
New Release Manager Gabriele Comoretto.

Deputies John Swinbank (PM), Colin Slater (PS), Yusra AlSayyad (Pipelines) and Vaikunth Thukral (DAX).

Toughest thing in any project is communication.

Build and Deploy with a view to Ops





DM must build everything to get LSST data products—as described in LSE-163—to end users.

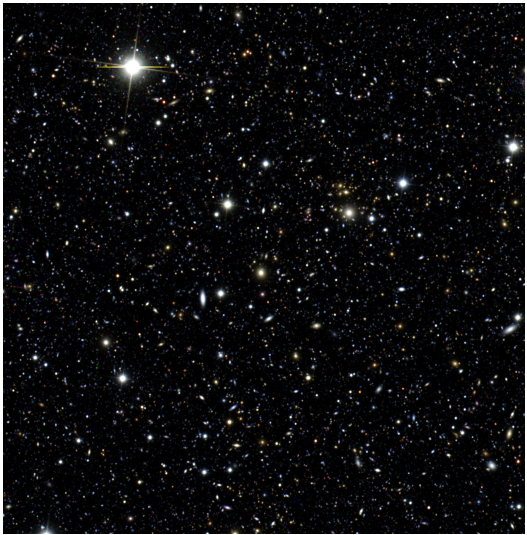
- Large data sets (20 TB/night)
- Complex analysis with small systematics
- Science alerts issued within one minute

$\sim {}^{3}/_{4}$ million lines of code/comments (C++/Python/Java/JavaScript/Kotlin)
See SPIE paper by Jenness et al. (2018)

Architecture and Components LDM-148.
Concept of Operations LDM-230.

Upper diagram courtesy K-T Lim, LDM-148.
Lower diagram by Tim Jenness; covers only the Science Pipelines codebase.

All Comes Down to Images



Simulated image based on three filters
From just one of 189 CCDs

Processed through the
LSST Science Pipelines
`https://pipelines.lsst.io/`
The Pipelines are already in use with other facilities, e.g. Hyper Suprime-Cam.

Still working on performance, algorithmic enhancements, orchestration, etc.

Design LDM-151; Test Specs LDM-533; LDM-534; Test Reports DMTR-52; DMTR-53

Fast start with pipelines

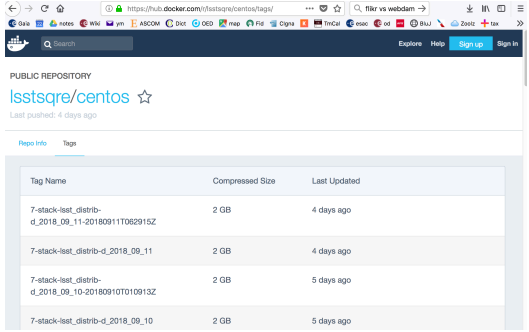
Find release you want at `https://hub.docker.com/r/lsstsqre/centos/tags/`

`>docker run -ti lsstsqre/centos:7-stack-lsst_distrib-v16_0`

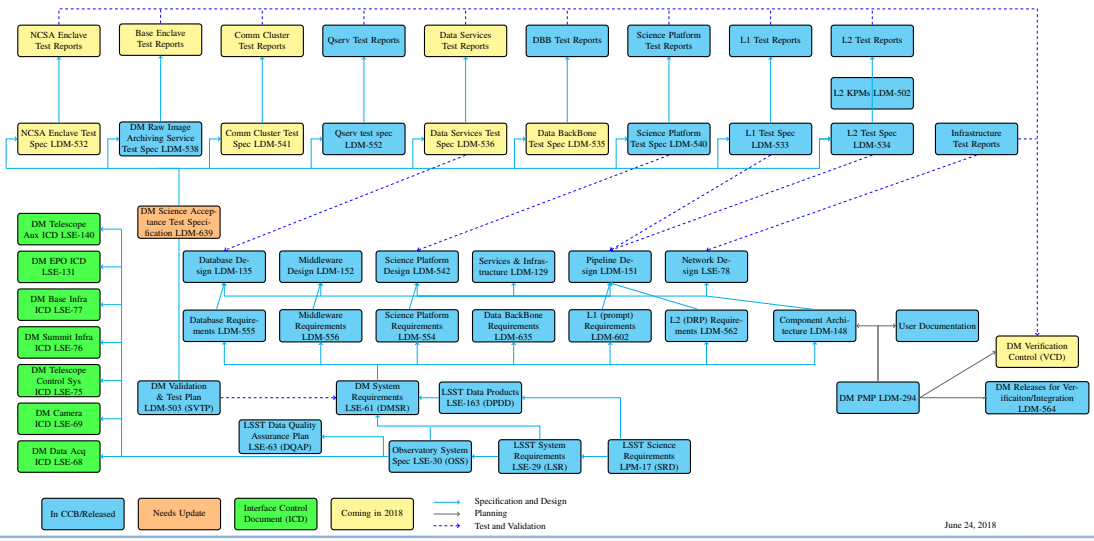
You may see .. `Unable to find image 'lsstsqre/centos:7-stack-lsst_distrib-v16_0' locally`

Setup up the environment ..

`>source /opt/lsst/software/stack/loadLSST.bash`

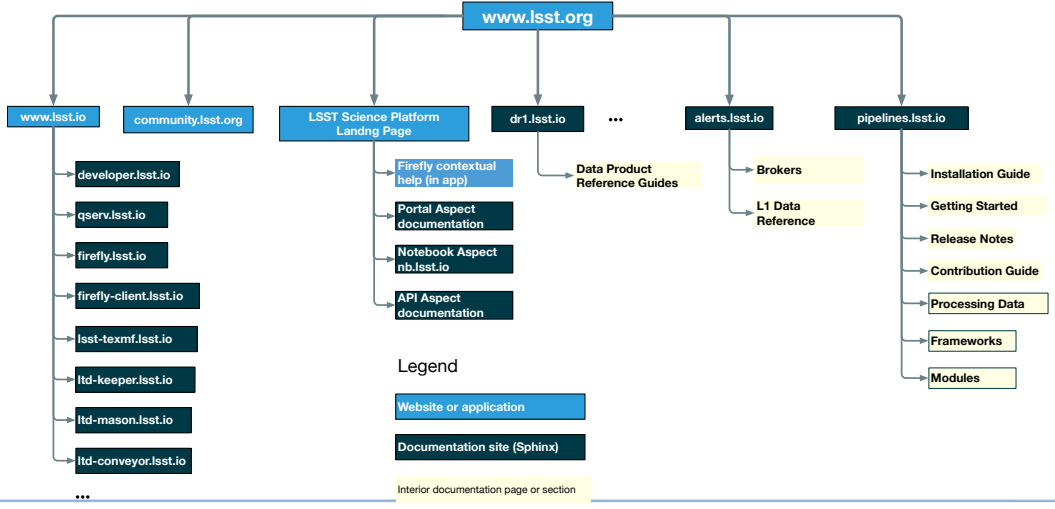


DM Document Tree

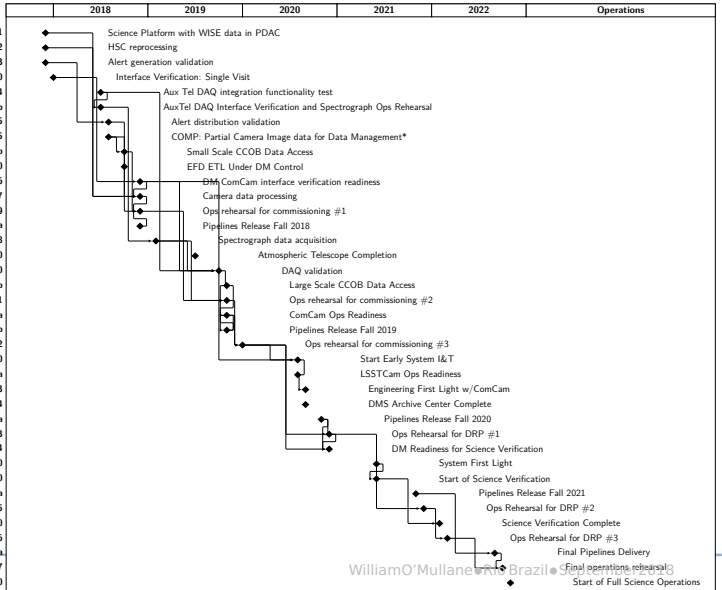


End-User Documentation on the Web

See `lsst.io`, `pipelines.lsst.io`, `nb.lsst.io`



Test Driven Schedule LDM-503



DM adopted a test driven schedule to address verification and progress.

In line with the approach taken by project-level Systems Engineering — adapting our test specs to Jira test manager system.

LDM-503 (test plan) and LDM-564 (releases) now automatically generated from P6

Releases LDM-564

High level features of all releases are listed in LDM-564. These coincide with the test milestones.

Example **LDM-503-7**: Camera Data Processing

- Camera package supporting the LSST Camera. (DM-DRP-5)
- LSSTCam data display and visualization. (DM-SUIT-6)
- Mapping between SUIT systems & NCSA auth system. (DM-SUIT-7)
- SUIT portal integrated with workspace. (DM-SUIT-8)
- Basic instrument signature removal (ISR) capability. (DM-AP-7)
- Camera package supporting the Commissioning Camera. (DM-DRP-38)
- Calibration product generation in support of basic ISR. (DM-DRP-4)

The test specs provide further refinement on the delivered product. The bracketed tag is a level 3 milestone in P6.

Verification and Validation

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 with Commissioning Team: some *rehearsals* will be joint
- **There is no difference between DM Science Validation and Commissioning Science Validation — DM will work closely with Commissioning Team**.

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

DM Verification Details

- Test Specification
  - All tests relating to a DM product (or group thereof) - do we have not done this precisely and now all will be in Jira
- Test planning
  - DM Test Plan (LDM-503) defines a set of test milestones to be fulfilled
  - For each Milestone a Test Plan has to be defined in Jira, with a corresponding Test Run(s)
  - A draft Test Report can be generated and uploaded in DocuShare as a reference for the planned test activity.
- Test Execution - Using the Test Player
  - Comments, Jira issues, and additional information can be added in Jira
  - A final assessment can be included in Jira for the Test Run
- Test Reporting
  - Once the test activity is run the Test Report (DMTR) will be generated from Jira e.g. DMTR-91, and uploaded in DocuShare.

High Level Goals

- 2018: Prototypes for various processes and databases - “Minimum Viable System”
  - Aug: Mountain base network up
  - Oct: “Generation 3” pipeline execution middleware
  - Nov: Ready for 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

LDM-564 describes functionality provided by each milestone.
Test specs to confirm milestone completion under development.

LDM-532; LDM-533; LDM-534; LDM-535; LDM-536
LDM-537; LDM-538; LDM-539; LDM-540; LDM-541

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-503-09 – 2018-11-30: Demonstrate and rehearse commissioning capabilities
- LDM-503-11a – 2019-10-21: Verify that we are ready for commissioning ComCam

Auxiliary Telescope (AuxTel)

Early commissioning is coming soon!!!

- Auxiliary Telescope is in Chile already — the spectrograph will arrive after summer.
- We plan to be taking and processing data early 2019!
  - Official operations late 2019 — milestone LDM-503-08 in preparation for that.
- We will have a pipeline for processing AuxTel data on the commissioning cluster.
- Initially *selected* data will be transferred to NCSA and made available in the Science Platform.
- By summer 2019 we will have an automated processing system for AuxTel data.

Outline

Introduction

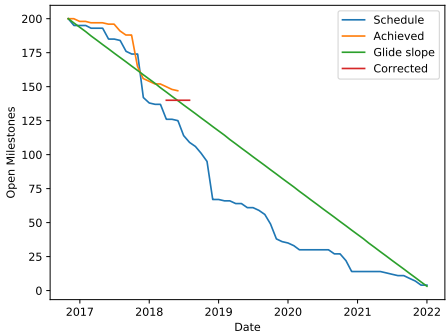
LSST status

Data Management Overview

Data Management Status and Achievements

Conclusion

Milestone Burndown (1.1, 1.2)



- Level 2 milestones achieved (on plan)
  - LDM-503-02 HSC Reprocessing—DMTR-51
  - LDM-503-03 Alert Generation—DMTR-53
  - LDM-503-05 Alert Distribution—DMTR-91
  - LDM-503-04,04b Raw Image Acquisition—DMTR-61
  - DLP-44 Test traffic AURA DWDM LSST fiber
  - DLP-65 AURA DWDM Ready
- Level 2 milestones delayed (but achieved)
  - (+2 months) LSST-1200 Interface Verification: Single Visit `https://confluence.lsstcorp.org/display/SYSENG/1-Visit+Night+with+Images+AuxTel`
  - (+5 months) LDM-503-1 Prototype DAC—DMTR-51

Data Release Production

LDM-503-2: HSC (Hyper Suprime Cam) reprocessing milestone

- First (equal) post-replan, NSF-visible milestone hit by the DM project.
- Joint effort to reprocess (Data Facility team) and analyze (DRP team) HSC data under operations-like conditions
- Milestone successful: DMTR-51!

LDM-503-2: HSC (Hyper Suprime Cam) reprocessing milestone

- First (equal) post-replan, NSF-visible milestone hit by the DM project.
- Joint effort to reprocess (Data Facility team) and analyze (DRP team) HSC data under operations-like conditions
- Milestone successful: DMTR-51!

“Warp Compare” coadds

- New algorithm to robustly reject artifacts when coadding images.
- Now default for HSC processing; stack-wide default to be RFCed soon.

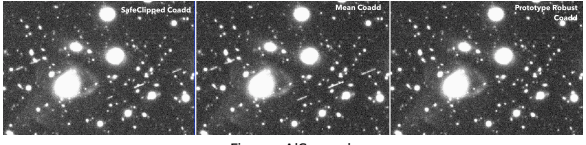


Figure: AlSayyad.

- **LDM-503-3: Alert generation milestone**
  - First(equal) post-replan milestone hit by the DM project DMTR-53!
  - Demonstrating a *end-to-end* alert production pipeline.
- Prototype alert distribution system using Kafka & AVRO; benchmark results on DMTN-028.
- New MOPS (Moving Object) linking algorithm under development and approach to Minor Planet Center DMTN-087.
- Jointcal replaces meas_mosaic
  - Simultaneous astro- and photometric fitting to source lists derived from multiple images.
  - The all new, much improved, more generic replacement for the HSC-specific meas_mosaic.

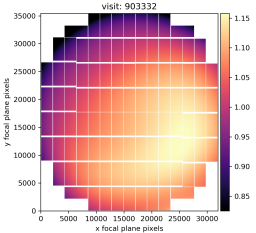


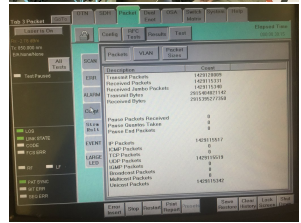
Figure shows the variation in photometric calibration not captured by single frame processing, normalized to 1. This demonstrates fine structure in photometry which Jointcal picks up but per-CCD processing doesn't catch.

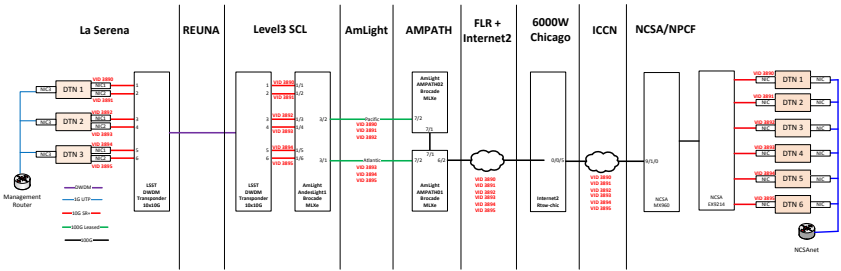
Figure: Parejko.

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.

(Document-28547)



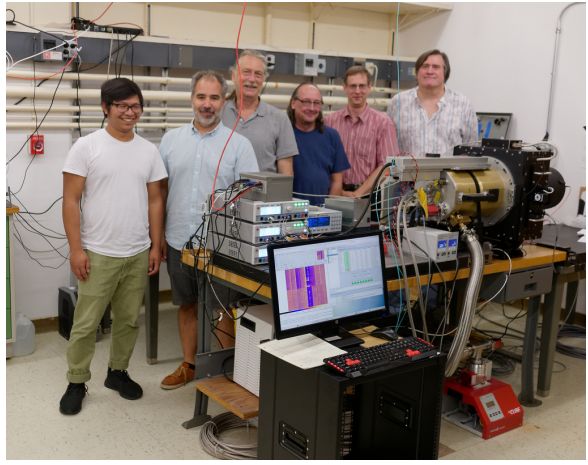


Data Access Services

- Catalog Database (Qserv) to 100 TB range
  - Three 30-node clusters operating:
    - NCSA (PDAC): science dataset (Stripe 82 + AllWISE + NEOWISE)
    - CC-IN2P3 (2 x dev): synthetic dataset
  - 30% DR1 KPM measurements DMTR-17
  - Deployment under Kubernetes
- Gen 3 Data Butler and Supertask in progress

IN2P3 Qserv Cluster - Fritz Muller

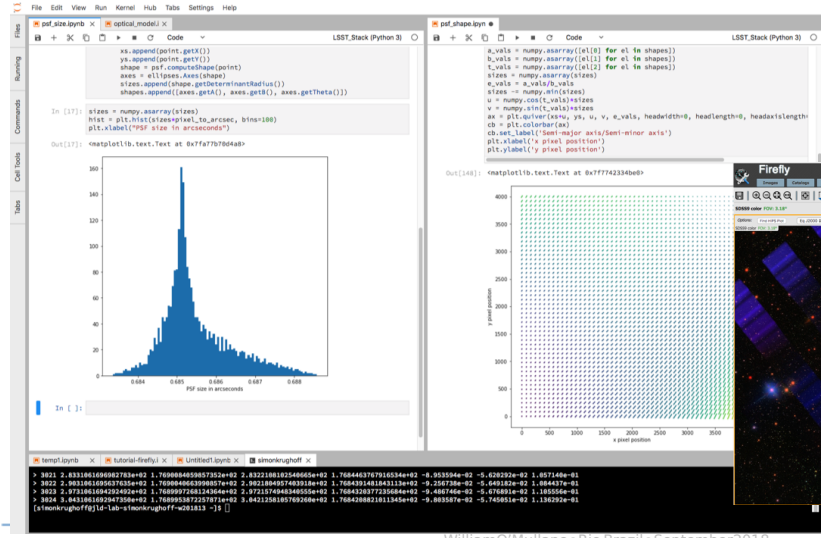
- Observatory Operations Support (Level 1) Services
  - Working within the LSST Systems Engineering Early Pathfinder group, developing and testing integration of T&S, Camera, and DM service software via a series of early integration activities.
  - Initial header service developed and configured for Camera subsystem and AuxTel use cases, ability to acquire pixel data and write FITS files, all commandable by OCS. Demonstrated on Level 1 Test Stand (DMTR-61) and with Spectrograph in Tucson



EIA-341 reading image from the spectrograph in Tucson July 2018

Science Platform Vision to Reality (1.1)

Vision: LSE-319 — Design: LDM-542 — Test: DMTR-51 at NCSA



Portal/Browser
Notebooks
Web API
(Batch)

Images Krughoff(←) and Wu (↑)

Outline

Introduction

LSST status

Data Management Overview

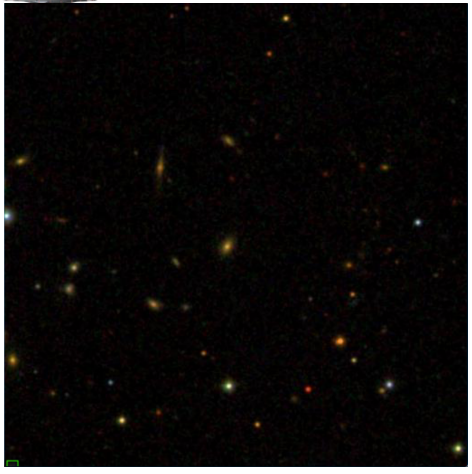
Data Management Status and Achievements

Conclusion

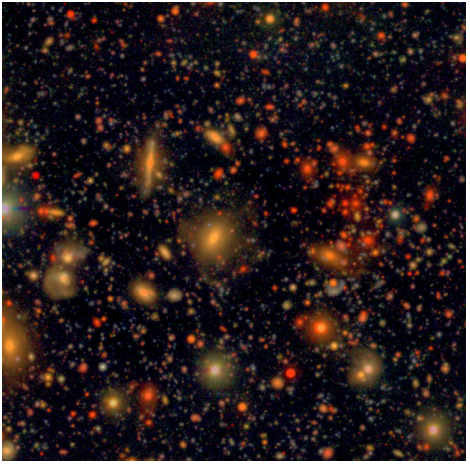
Conclusion

- LSST is getting very real !
- A lot of work on verification for DM
- AuxTel coming very rapidly.
- Looking forward to first data next year.

The END



$\sim 3.5'$ SDSS image



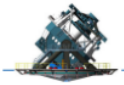
HSC image (COSMOS)

g,r(1.5 hrs) ,i(3 hrs) PSF matched co-add ($\approx$ 27.5)

Images:Lupton and HSC colaboration see also Lupton et al. (2004)

Questions?

http://www.lsst.org http://community.lsst.org



Outline

Reference material

Acronyms I

Acronym	Description
AI	Action Item
AP	Alerts Production
API	Application Programming Interface
AURA	Association of Universities for Research in Astronomy
AVRO	Apache data serialization system
BBC	German shipping company
C	Specific programming language (also called ANSI-C)
CCD	Charge-Coupled Device
DAC	Data Access Center
DAX	Data Access Services
DLP	DM Long Term Plan
DM	Data Management
DMTN	DM Technical Note
DMTR	Data Management Test Report
DRP	Data Release Production
DTN	Data Transfer Node
DWDM	Dense Wave Division Multiplex
EFD	Engineering Facilities Database
EIA	Early Integration Activity
EIE	European Industrial Engineering - Italian engineering company (Dome)
FITS	Flexible Image Transport System

Acronyms II

HSC	Hyper Suprime-Cam
IPAC	Infrared Processing and Analysis Center
IR	Infra Red
ISR	Instrument Signal Removal
K	Kelvin; SI unit of temperature
KPM	Key Performance Metric
L1	Level 1 (ambiguous could mean milestone or processing)
L2	Level 2 (ambiguous could mean milestone or processing)
LDM	LSST Data Management (handle for controlled documents)
LSE	LSST Systems Engineering (Document Handle)
LSST	Large Synoptic Survey Telescope
M2	Second mirror
MIA	Missing In Action
MN	Meeting Minutes
MOPS	Moving Object Pipeline System
N	Newton; SI unit of force
NASA	National Aeronautics and Space Administration (USA)
NCSA	National Center for Supercomputing Applications
NEO	Near-Earth Object
NSF	National Science Foundation
OCS	Observatory Control System
PB	PetaByte
PDAC	Prototype Data Access Center
PM	Project Manager

Acronyms III

PS	Project Scientist
PSF	Point Spread Function
QA	Quality Assurance
Qserv	Query Service, Proprietary LSST Database system
S	Strip (CCD chip along-scan coordinate identifier in focal plane)
SDSS	Sloan Digital Sky Survey
SPIE	the international society for optics and photonics
SUIT	Science User Interface and Tools
T&S	Telescope and Site
TB	TeraByte
US	United States
USA	United States of America
arcmin	arcminute, minute of arc (unit of angle)
kg	kilogram; SI unit of mass
s	second; SI unit of time

References I

[DMTR-91], Bellm, E., 2018, *LDM-503-05 (Alert Distribution Validation) Test Report*, DMTR-91, URL https://ls.st/DMTR-91

[DMTR-53], Bellm, E., Swinbank, J., 2018, *LDM-503-03 (Alert Generation) Test Report*, DMTR-53, URL https://ls.st/DMTR-53

[LDM-533], Bellm, E.C., 2017, *Level 1 System Software Test Specification*, LDM-533, URL https://ls.st/LDM-533

[DMTR-51], Bosch, J., Chiang, H.F., Gower, M., et al., 2017, *LDM-503-02 (HSC Reprocessing) Test Report*, DMTR-51, URL https://ls.st/DMTR-51

[DMTR-61], Butler, M., Parsons, J., 2018, *LDM-503-04 and LDM-503-04b (Raw Image Archiving Service) Test Report*, DMTR-61, URL https://ls.st/DMTR-61

[LDM-538], Butler, M., Parsons, J., Gower, M., 2018, *Raw Image Archiving Service Test Specification*, LDM-538, URL https://ls.st/LDM-538

[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-540], Dubois-Felsmann, G., 2018, *LSST Science Platform Test Specification*, LDM-540, URL https://ls.st/LDM-540

[LDM-542], Dubois-Felsmann, G., Lim, K.T., Wu, X., et al., 2017, *LSST Science Platform Design*, LDM-542, URL https://ls.st/LDM-542

[DMTR-52], Dubois-Felsmann, G.P., 2018, *LDM-503-01 (WISE Data Loaded in PDAC) Test Report*, DMTR-52, URL https://ls.st/DMTR-52

Ivezic, Z., et al., 2008, ArXiv e-prints (arXiv:0805.2366), ADS Link

Jenness, T., Economou, F., Findeisen, K., et al., 2018, In: Software and Cyberinfrastructure for Astronomy V, vol. 10707 of Proc. SPIE, 1070709, doi:10.1117/12.2312157, ADS Link

[DMTN-087], Juric, M., Jones, L., 2018, *Proposed Modifications to Solar System Processing and Data Products*, DMTN-087, URL https://dmtn-087.lsst.io, LSST Data Management Technical Note

[LSE-319], Jurić, M., Ciardi, D., Dubois-Felsmann, G., 2017, *LSST Science Platform Vision Document*, LSE-319, URL https://ls.st/LSE-319

[LSE-163], Jurić, M., et al., 2017, *LSST Data Products Definition Document*, LSE-163, URL https://ls.st/LSE-163

[Document-28547], Kantor, J., 2018, *LSST Network Bandwidth Tests between Chile and the United States* , Document-28547, URL https://ls.st/Document-28547

[LDM-148], Lim, K.T., Bosch, J., Dubois-Felsmann, G., et al., 2018, *Data Management System Design*, LDM-148, URL https://ls.st/LDM-148

References II

Lupton, R., Blanton, M.R., Fekete, G., et al., 2004, PASP, 116, 133 (arXiv:astro-ph/0312483), doi:10.1086/382245, ADS Link

[LDM-564], O'Mullane, W., Economou, F., Jenness, T., Loftus, A., 2018, *Data Management Software Releases for Verification/Integration*, LDM-564, URL https://ls.st/LDM-564

[LDM-294], O'Mullane, W., Swinbank, J., Jurić, M., DMLT, 2018, *Data Management Organization and Management*, LDM-294, URL https://ls.st/LDM-294

[LDM-503], O'Mullane, W., Swinbank, J., Jurić, M., Economou, F., 2018, *Data Management Test Plan*, LDM-503, URL https://ls.st/LDM-503

[DMTN-028], Patterson, M.T., 2018, *Benchmarking a distribution system for LSST alerts*, DMTN-028, URL https://dmtn-028.lsst.io, LSST Data Management Technical Note

[LDM-230], Petravick, D., Butler, M., Gelman, M., 2018, *Concept of Operations for the LSST Data Facility Services*, LDM-230, URL https://ls.st/LDM-230

[LDM-534], Swinbank, J.D., 2017, *Level 2 System Software Test Specification*, LDM-534, URL https://ls.st/LDM-534

[LDM-151], Swinbank, J.D., et al., 2017, *Data Management Science Pipelines Design*, LDM-151, URL https://ls.st/LDM-151

[DMTR-17], Thukral, V., 2018, *Qserv Fall 17 Large Scale Tests/KPMs*, DMTR-17, URL https://ls.st/DMTR-17

[LDM-532], Unknown, 2017, *NCSA Enclave Test Specification*, LDM-532, URL https://ls.st/LDM-532

[LDM-535], Unknown, 2017, *Data Backbone Test Specification*, LDM-535, URL https://ls.st/LDM-535

[LDM-536], Unknown, 2017, *Data Backbone Data Services Test Specification*, LDM-536, URL https://ls.st/LDM-536

[LDM-537], Unknown, 2017, *Data Backbone Infrastructure Test Specification*, LDM-537, URL https://ls.st/LDM-537

[LDM-539], Unknown, 2017, *Data Access Center Enclave Test Specification*, LDM-539, URL https://ls.st/LDM-539

[LDM-541], Unknown, 2017, *Commissioning Cluster Enclave Test Specification*, LDM-541, URL https://ls.st/LDM-541