LSST Data Management Status

William O'Mullane, AURA/LSST
DM Project Manager

DM All Hands
6th March 2018

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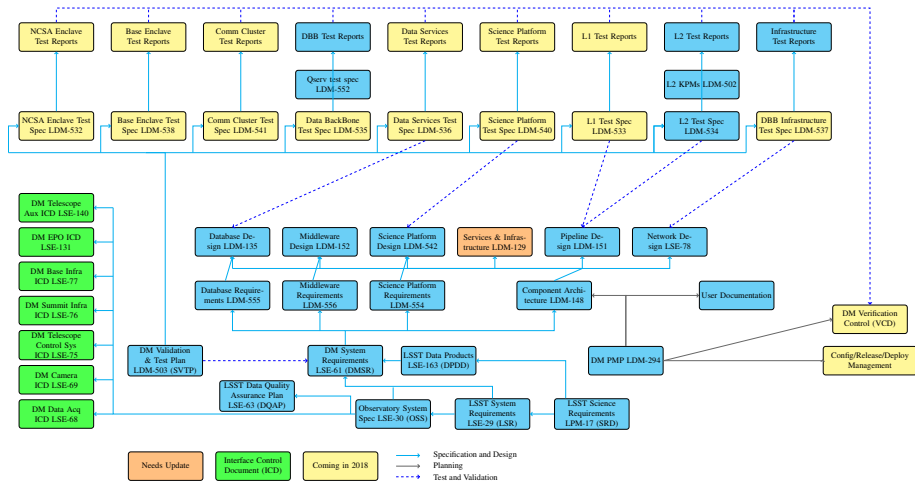
Welcome Michelle Butler and thanks to Don

Welcome Gabriele Comoretto.

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

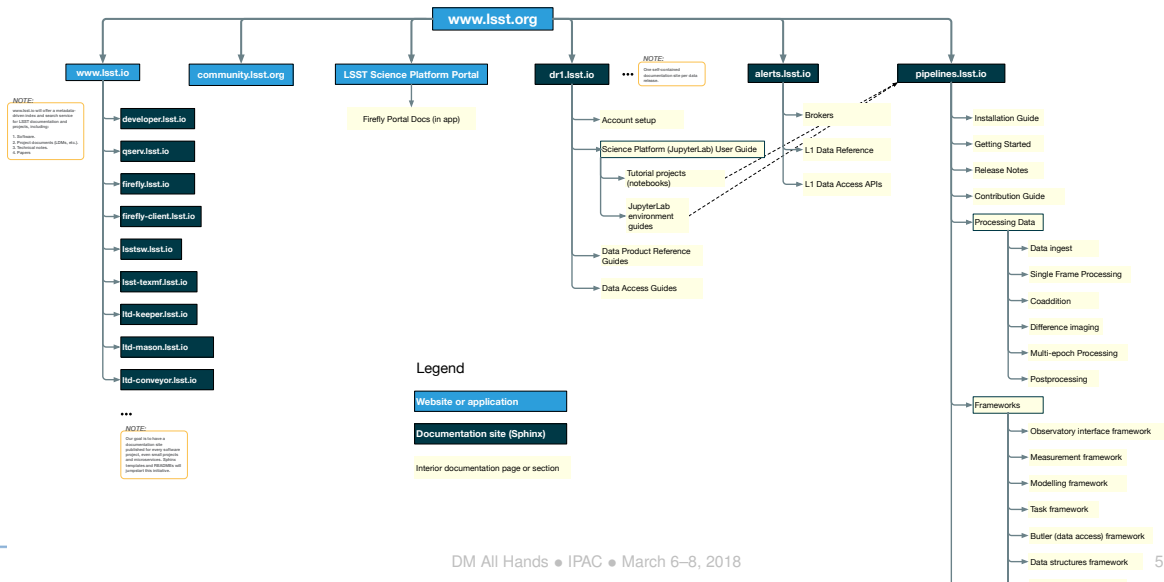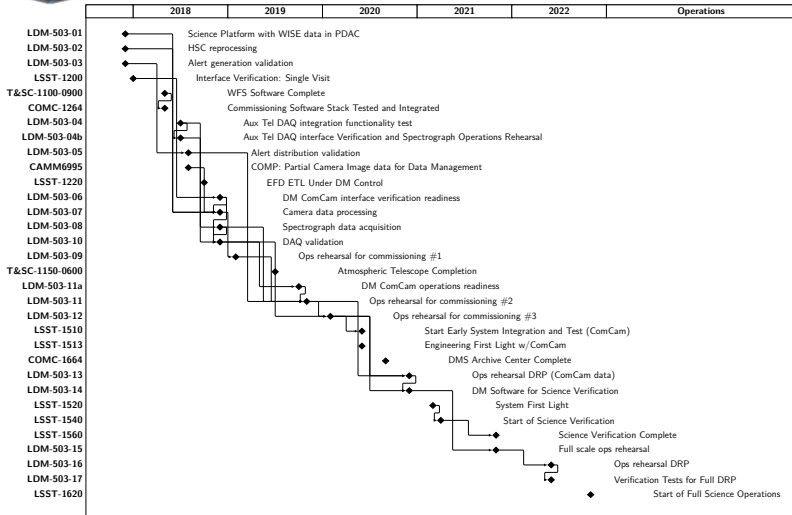Toughest thing in any project is communication.

DM Document Tree

DM All Hands ● IPAC ● March 6–8, 2018   4

End-User Documentation on the Web

www.lsst.org

- **www.lsst.io**

  NOTE:
  www.lsst.io will offer a metadata-driven index and search service for LSST documentation and projects, including:
  1. Software
  2. Project documents (LDMs, etc.)
  3. Technical notes
  4. Papers

  - developer.lsst.io
  - qserv.lsst.io
  - firefly.lsst.io
  - firefly-client.lsst.io
  - lsstsw.lsst.io
  - lsst-texmf.lsst.io
  - ltd-keeper.lsst.io
  - ltd-mason.lsst.io
  - ltd-conveyor.lsst.io

  ...

  NOTE:
  Our goal is to have a documentation site published for every software project, even small projects and microservices. Sphinx templates and READMEs will jumpstart this initiative.

- **community.lsst.org**

- **LSST Science Platform Portal**
  - Firefly Portal Docs (in app)

- **dr1.lsst.io** ...

  NOTE:
  One self-contained documentation site per data release

  - Account setup
  - Science Platform (JupyterLab) User Guide
    - Tutorial projects (notebooks)
    - JupyterLab environment guides
  - Data Product Reference Guides
  - Data Access Guides

- **alerts.lsst.io**
  - Brokers
  - L1 Data Reference
  - L1 Data Access APIs

- **pipelines.lsst.io**
  - Installation Guide
  - Getting Started
  - Release Notes
  - Contribution Guide
  - Processing Data
    - Data ingest
    - Single Frame Processing
    - Coaddition
    - Difference imaging
    - Multi-epoch Processing
    - Postprocessing
  - Frameworks
    - Observatory interface framework
    - Measurement framework
    - Modelling framework
    - Task framework
    - Butler (data access) framework
    - Data structures framework

Legend

Website or application

Documentation site (Sphinx)

Interior documentation page or section

		2018	2019	2020	2021	2022	Operations

- **LDM-503-01** — Science Platform with WISE data in PDAC
- **LDM-503-02** — HSC reprocessing
- **LDM-503-03** — Alert generation validation
- **LSST-1200** — Interface Verification: Single Visit
- **T&SC-1100-0900** — WFS Software Complete
- **COMC-1264** — Commissioning Software Stack Tested and Integrated
- **LDM-503-04** — Aux Tel DAQ integration functionality test
- **LDM-503-04b** — Aux Tel DAQ interface Verification and Spectrograph Operations Rehearsal
- **LDM-503-05** — Alert distribution validation
- **CAMM6995** — COMP: Partial Camera Image data for Data Management
- **LSST-1220** — EFD ETL Under DM Control
- **LDM-503-06** — DM ComCam interface verification readiness
- **LDM-503-07** — Camera data processing
- **LDM-503-08** — Spectrograph data acquisition
- **LDM-503-10** — DAQ validation
- **LDM-503-09** — Ops rehearsal for commissioning #1
- **T&SC-1150-0600** — Atmospheric Telescope Completion
- **LDM-503-11a** — DM ComCam operations readiness
- **LDM-503-11** — Ops rehearsal for commissioning #2
- **LDM-503-12** — Ops rehearsal for commissioning #3
- **LSST-1510** — Start Early System Integration and Test (ComCam)
- **LSST-1513** — Engineering First Light w/ComCam
- **COMC-1664** — DMS Archive Center Complete
- **LDM-503-13** — Ops rehearsal DRP (ComCam data)
- **LDM-503-14** — DM Software for Science Verification
- **LSST-1520** — System First Light
- **LSST-1540** — Start of Science Verification
- **LSST-1560** — Science Verification Complete
- **LDM-503-15** — Full scale ops rehearsal
- **LDM-503-16** — Ops rehearsal DRP
- **LDM-503-17** — Verification Tests for Full DRP
- **LSST-1620** — Start of Full Science Operations

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
- – **?** 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

? addresses DM's plans for verification & validation.

– The DM Basline is now the reviewed plan (LCRs processed)

– S18 detail plans submitted

– Milestones inow being reported monthly

– Level 2 milestones achieve in December

- LDM-503-2ăăăTest report: HSC Reprocessing -DMTR-51
- LDM-503-3ăăăTest report: Alert Generation Validation -DMTR-53

– Level 2 milestones delayed in December

- LDM-503-1ăăăTest report: Science Platform with WISE data in PDAC

  - Instability PDAC hardware lack of personnel,
  - Testing commend last week

– 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.

November 2019: DM for Commissioning (minimum required for start of commissioning with ComCam): **(See ? §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.

Data Management Overview

Risks and Opportunities

Status

Conclusion

Conclusion

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).

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

–

–

–

–

– The Network Engineering Team completed the first successful transfer of digital data over LSST/AURA fiber optic networks from the Summit Site on Cerro Pachon to the Base Site in La Serena and on to the Archive Site at NCSA in Champaign. 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.

–

–

–

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.
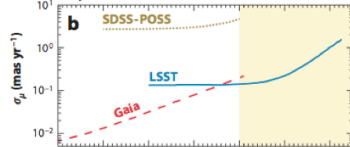
- Looking forward to the first LSST images!

Photometry / a: σ_r (mmag) vs r — SDSS, $n=1$; LSST, $n=1$, $n\approx230$; Gaia, $n=70$

Proper motion / b: σ_μ (mas yr^{-1}) vs r — SDSS-POSS, LSST, Gaia

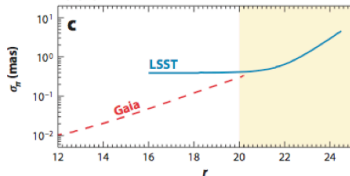Parallax / c: σ_π (mas) vs r — LSST, Gaia
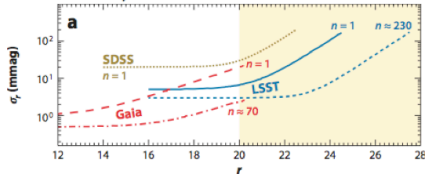
The Milky Way disk *belongs* to Gaia, and the halo to LSST (plus very faint and/or very red sources, such as white dwarfs and LT(Y) dwarfs). Željko Ivezić

Questions ??

Reference material

Acronym	Description
IPAC	Infrared Processing and Analysis Center