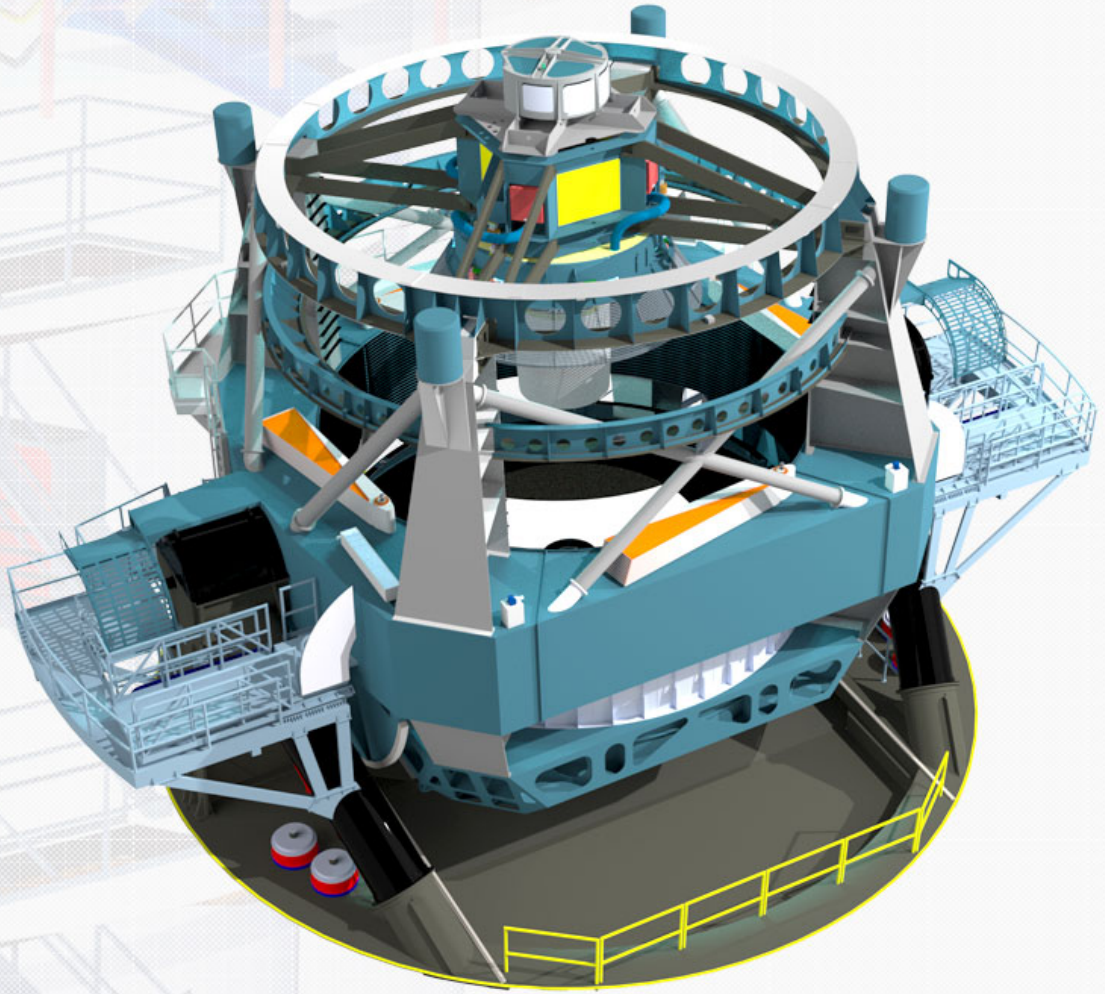
Data Management Project Status

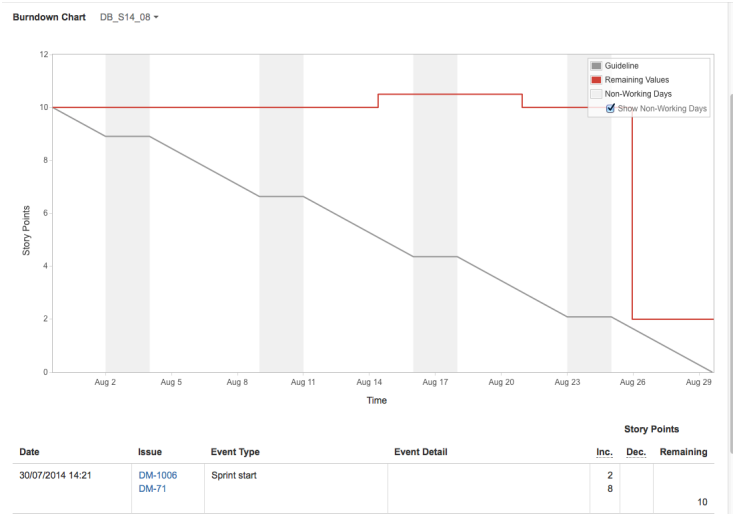
Jeffrey Kantor
Data Management Project Manager

November 13, 2014



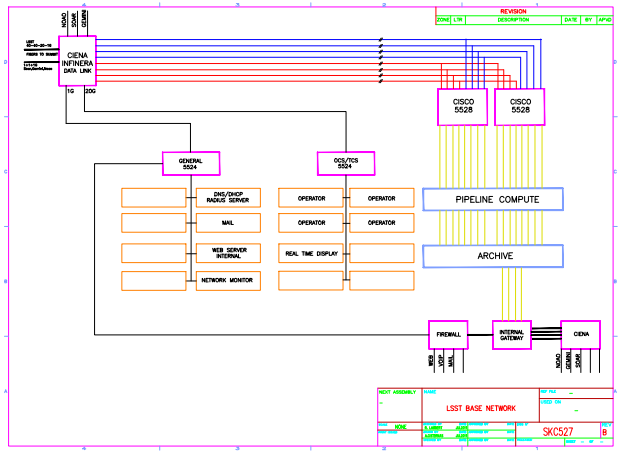
Past 6 months Accomplishments – DM System

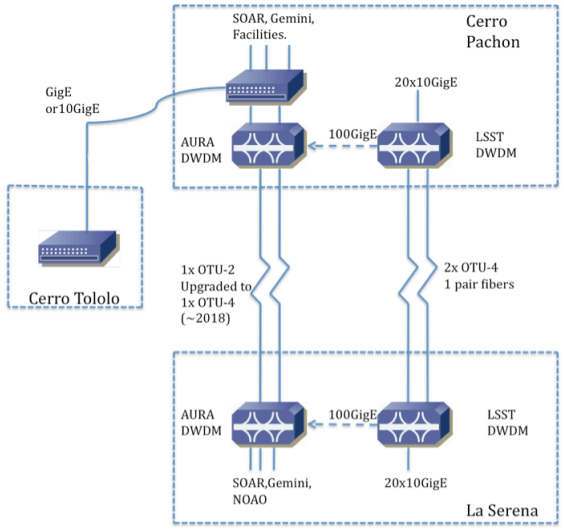
- Summer 2014 Complete
  - Primarily a refactoring and clean-up of the R&D code base to prepare for construction
  - No significant new functionality added, no new data sets processed
- Winter 2015 Planned
  - 75 detailed activities, 2 – 4 week granularity, resources loaded and leveled
  - Implemented JIRA Agile for bottom-up planning, imported into PMCS
  - Highlights include:
    - qserv global catalog for production tracking
    - image cutout service
    - automation of continuous integration
    - pipeline fault tolerance
    - multifit plug-in framework
    - SUI prototypes
    - CameraGeom software for Camera team
    - improved astrometry solver
  - Detailed descriptions in JIRA Agile

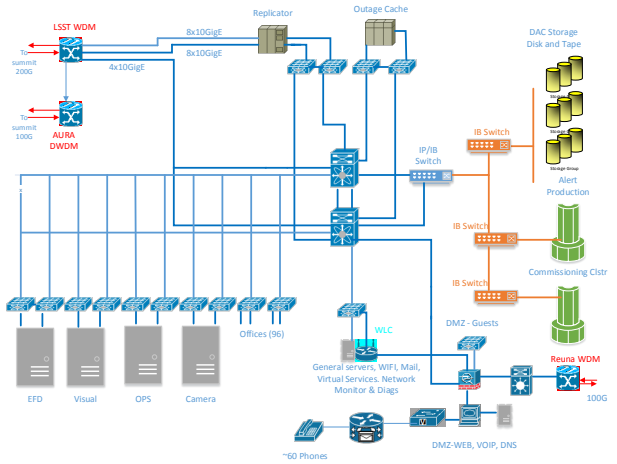


Past 6 months Accomplishments - Interfaces

- DM-OCS
  - New LSE-72 written from scratch to match design discussions in previous cycle
- DM-Camera
  - LSE-69 modified to align with OCS-DM-Camera interaction design from LSE-70/71/72
  - LSE-130 to be presented to CCB in November for Phase 1 baseline
- DM-Telescope
  - LSE-140 created to cover DM interactions with auxiliary instrumentation
  - LSE-77 / Base Facility design process rebooted, NCSA took over responsibility
- LSST Long Haul Networks
  - LSE-78 updated with latest design/plan for La Serena – Santiago fiber, summit and base networks







Past 6 months Accomplishments - Management

- Sub awards
  - Contracts prepared and in final discussions with all 7 institutions
  - Letters of authorization to proceed sent
- Recruiting/Staffing
  - 12 of 18 positions filled
  - Established charter/plan for Science Quality and Reliability Engineering (SQuaRE) Group

- Miscellaneous Support Activities
  - Developed plan for expanding and updating LSE-99 LSST Cyber-Security Plan
  - Supported NSF Cost Sufficiency Analysis and Contingency Analysis with ~ 700 pages of additional supporting documentation
  - Established process for large acquisitions of infrastructure via NCSA
  - Contributed to draft MOA and letters for IN2P3 DM operations role

Recruiting and Staffing – FY14/FY15

Caltech IPAC: 3 of 3 positions filled

NCSA: 2 of 2 positions filled

NOAO: 1 of 1 positions filled

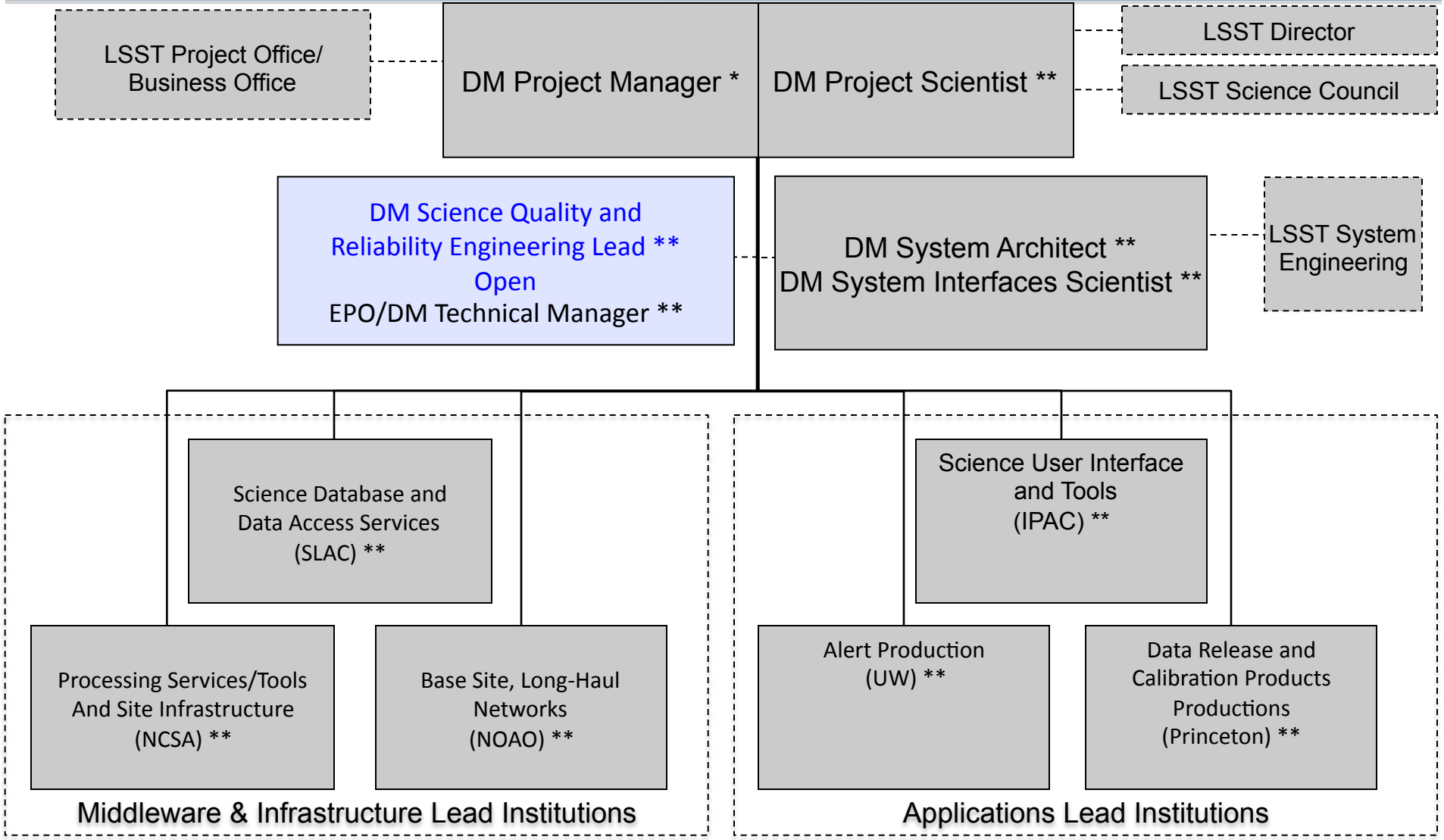
Princeton: 2 of 3 positions filled

SLAC National Accelerator Laboratory (DM): 1 of 3 positions filled, interim staffing via contract resources)

University of Washington (DM): 1 of 3 positions filled

LSST Project Office – Data Management: 2 of 3 positions filled, 1 position opening next month, DM Project Scientist relocated to UW

Construction/Commissioning Phases Organization



* DM LeadershipTeam

** and DM Science/Architecture Team

SQuaRE - Overview

- SQuaRE stands for **S**cience **Qu**ality **a**nd **Re**liability **E**ngineering)
- A Tucson-based (LSST/AURA-led) DM group
- SQuaRE's organizational purpose:

Provides scientific and technical feedback to the LSST DM Manager that demonstrates LSST/AURA DM is fulfilling its commitments as expected by the NSF with regards to science quality and software/IT performance and reliability

- Reports to DM Project Manager & Scientist
- Advises DM System Architect
- 2-in-a-box lead + 4 staff (current: 1 + 1) [5.3 total FTEs]
- Recruitment for other positions in various stages is targeting "generalists" (astronomical data reduction/software engineering/IT)

Top 5 DM Risks

- Programming team productivity below estimate due to geographical distribution/competing priorities (DM-062)
  - Exposure Cost: $ 1807 K Probability: 10 – 25%
- Computing power required for Data Release Production exceeds estimates by large factor (DM-018)
  - Exposure Cost: $ 1348 K Probability: 10 – 25%
- Unanticipated characteristics of real data result in poor MultiFit performance (computational) (DM-022)
  - Exposure Cost: $ 962 K Probability: 25 – 50%
- Loss of key personnel (DM-042)
  - Exposure Cost: $ 888 K Probability: 25 – 50%
- Object counts exceed expectations, leading to insufficient compute (DM-021)
  - Exposure Cost: $ 823 K Probability: 10 – 25%

Next 6 months - Technical

- Winter 2015 complete
- Summer 2015 detailed plan complete
  - Including testing of FY15 Key Progress Metrics from LDM-240
- Exercise process for initial acquisitions of infrastructure via NCSA
  - Initial acquisition of development cluster
- Conduct reviews of Infrastructure design, Sizing Model
- Interface Control Documents
  - Phase 2 baseline, phase 3 draft LSE-68, - 69, -72, -77, -78
  - Phase 1 baseline, phase 2 draft LSE-130, -140

Next 6 months - Management

- Sub awards
  - Contracts completed with all 7 institutions
- Recruiting/Staffing
  - 7 of 7 currently open positions filled
  - 2 new positions opened and filled
- Miscellaneous Support Activities
  - Deliver updated LSE-99 Cyber-Security Plan (January 2015)
  - Continue to support reporting and analyses for NSF, with Earned Value
  - Complete MOA for IN2P3 DM operations role, conduct due diligence meetings with NCSA and IN2P3, develop implementation plan
  - Complete MOA for Brazil networks operations contributions

Potential International Partners

- IN2P3
  - Memorandum of Understanding in work since PDR
  - Computing Center in Lyon will do half of annual Data Release Processing, or other activities that offset operations costs
  - Will acquire data rights beyond those already covered by Camera MOA
- Brazil
  - Currently working on 2-phase agreement
    - Phase 1 – Provide 100 Gbps network from Santiago, Chile to Miami, FL in return for data rights for approximately 20 PIs and associated post-docs, students
    - Phase 2 – Provide ~$500k/yr operations funding in return for right to host a Data Access Center for Brazil