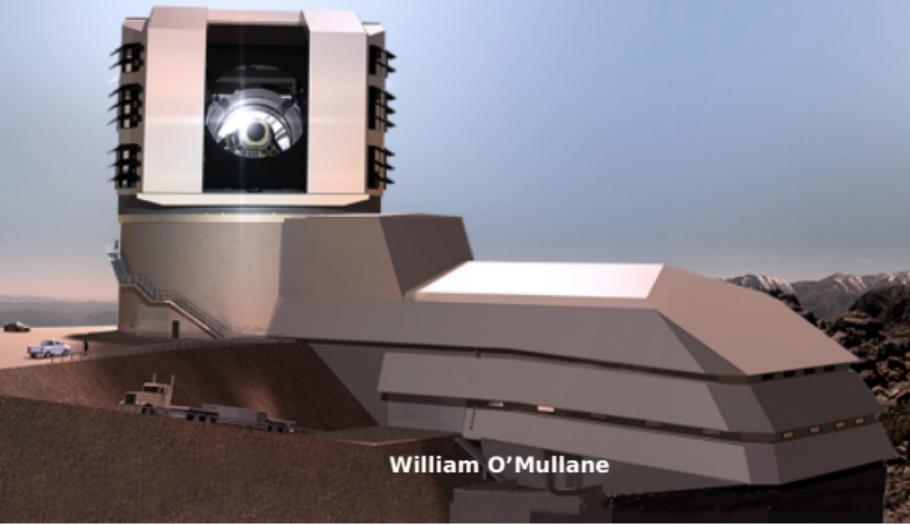




# LSST Data Management

William O'Mullane input from Željko Ivezic  
DM Project Manager

LSP Observatory Review 2017  
Paranal 13<sup>th</sup> December 2017



William O'Mullane





# The History of LSST



## A New Kind of Telescope Optimized for Surveys

~2000

Modified 3-mirror Paul-Baker  
Design Seeing limited over  
3.5 deg field of view *Dark  
Matter Telescope*

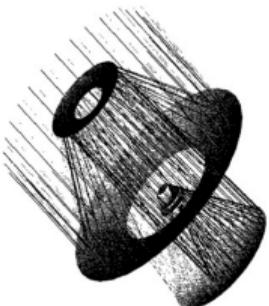
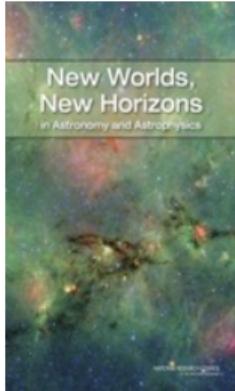


Figure 1. Optical layout with rays at  $\pm 1.5^\circ$  field angle.

~2010

LSST selected as the highest  
priority ground-based  
instrument for the coming  
decade



~2014

Formal construction start!  
Joint DOE + NSF project



**Beth Willman**



# Outline

---



Large Synoptic Survey Telescope status

LSST Operations

Data Management

Data Access/ Science Platform

Conclusion



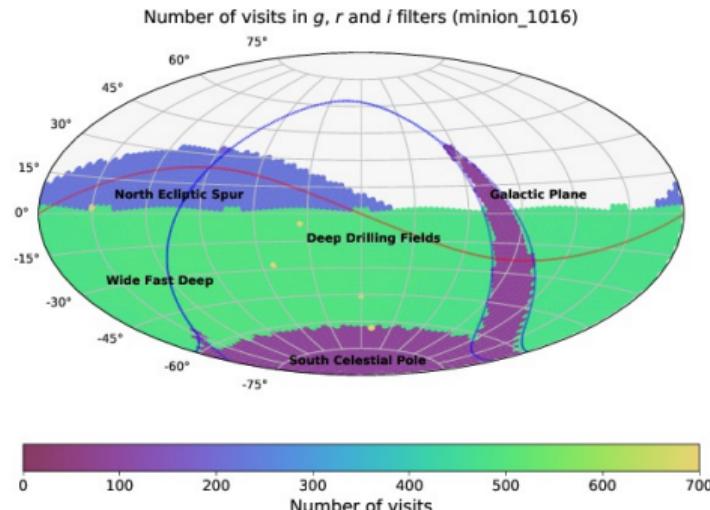
# LSST:uniform sky survey



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!**

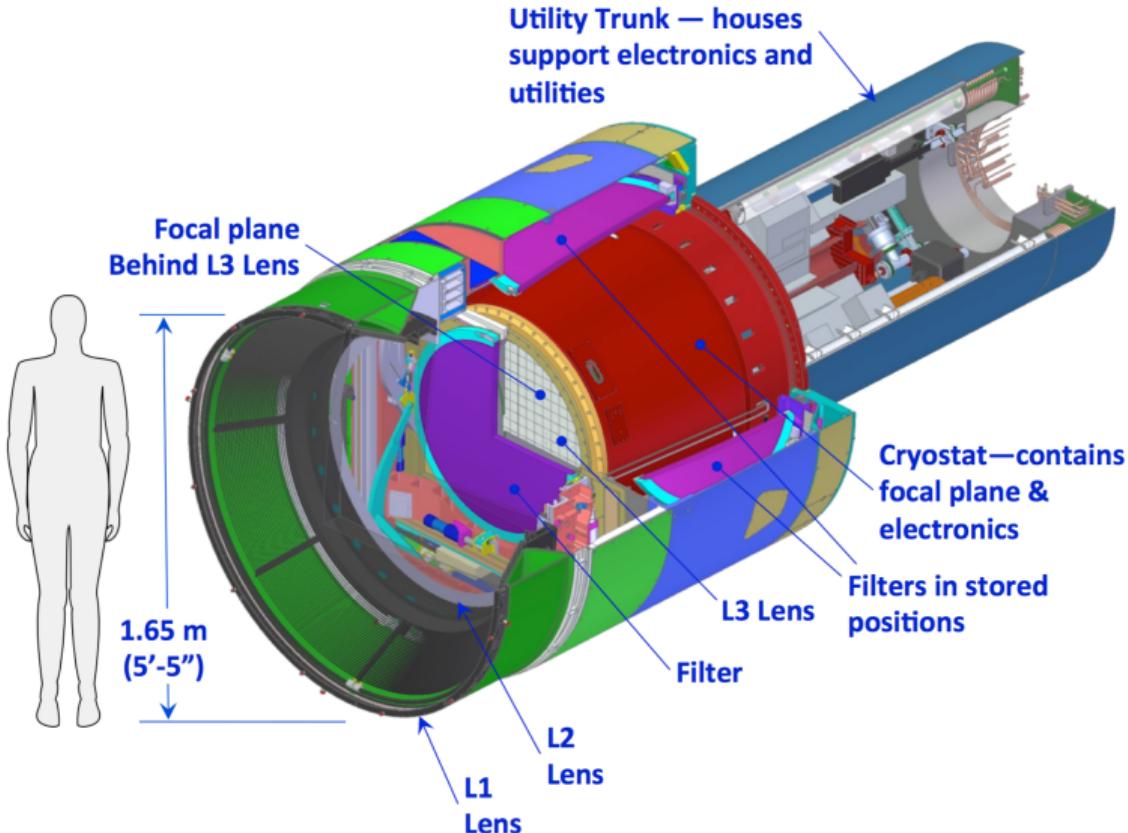
see also <http://www.lsst.org> and arXiv:0805.2366



**10-year simulation of LSST survey:  
number of visits in u,g,r band (Aitoff  
projection of eq. coordinates)**



# LSST Camera

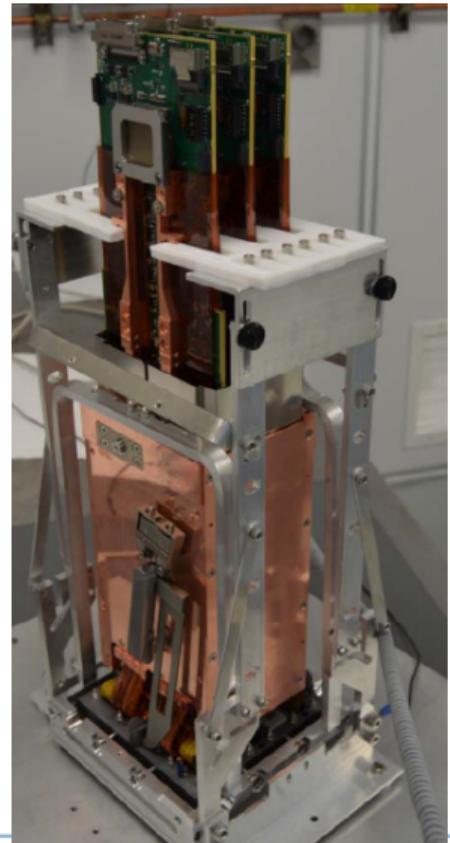
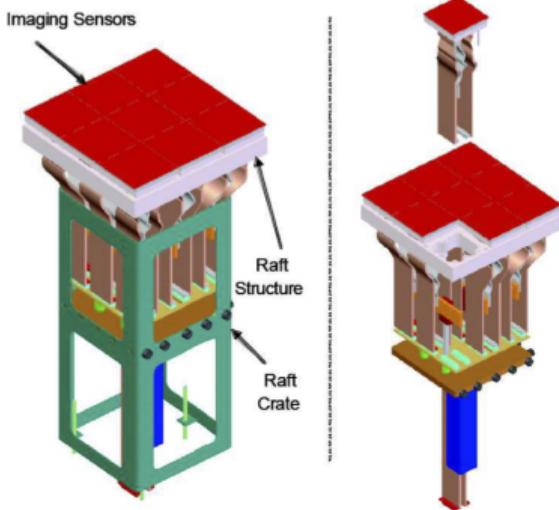
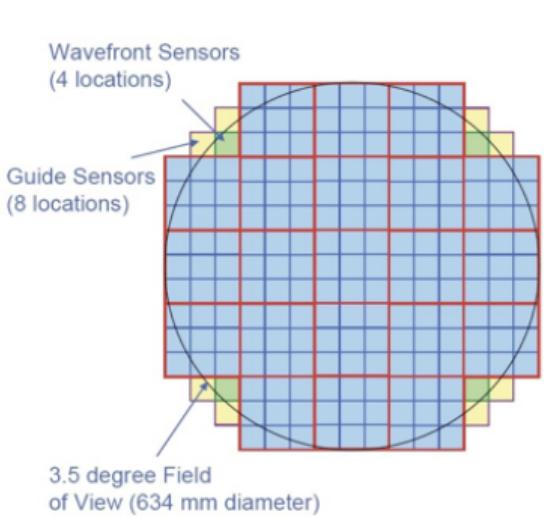


**The largest astronomical camera:**

- 2800 kg
- 3.2 Gpix



# Science rafts



Modular design: 3200 Megapix = 189 x16 Megapix CCD  
9 CCDs share electronics: raft (=camera 144 Megapix)

**Eight of 21 rafts completed, one accepted →**



# Hardware arriving



- Coating plant in Germany
- Stainless steel vessels fabricated
- Successful leak tests
- M1M3 Cell integration
- Cell weldment/vacuum testing successful
- Surrogate mirror integrated, 355 supports
- Actuator system components being installed



# Site shaping up



**Artist impression**

- Prime contract to finish Jan 2018.
- Azimuth rail sections aligned and grouted - Dome completion planned mid 2018
- Network partially in place, internal cabling being done



**Photo December 2017**



# Break Bulk and $\approx$ 70 containers



\$96M of equipment to site in 2018

Identification	Items	Dimensions Metres (each)
M1M3 Mirror	1	9.15 x 9.15 x 3.05
M1M3 Primary Cell	1	9.20 x 9.20 x 3.05
Vaculift (containerized)	1	9.20 x 9.20 x 1.00
M1M3 Surrogate Mirror (containerized)	1	9.20 x 9.20 x 1.53
M1M3 Cell Cart (containerized)	1	9.20 x 9.20 x 1.53
<b>Plus estimated 10 FEU</b>		
Coating Plant - To be Disassembled at Coquimbo	1	9.00 x 9.50 x 5.00
<b>Plus estimated 30 FEU</b>		
TMA - Azimuth Ring Sector	4	11.20 x 2.70 x 0.54
TMA - Main Support A & B	2	13.93 x 6.61 x 4.58
TMA - Circular Beam CRA & CRB	2	8.66 x 2.64 x 1.03
TMA - Elevation Brace A & B	2	10.48 x 3.90 x 2.53
TMA - Trunion A & B	2	2.90 x 2.88 x 1.71
TMA - Cradle A & B	2	7.95 x 3.09 x 0.98
TMA - Top End Pier 1 thru 4	4	5.58 x 1.70 x 4.02
TMA - Spider Spindle + TBR + Integrator 1	1	10.20 x 8.70 x 2.80
TMA - Integrator 2 & 3	2	2.61 x 2.61 x 2.51
TMA - M2 Surrogate Mass	1	3.50 x 3.50 x 1.50
TMA - Camera Lifting Fixture	1	7.80 x 3.40 x 2.60
TMA - Cover Trunions	1	3.20 x 3.20 x 1.00
<b>Plus estimated 30 FEU</b>		

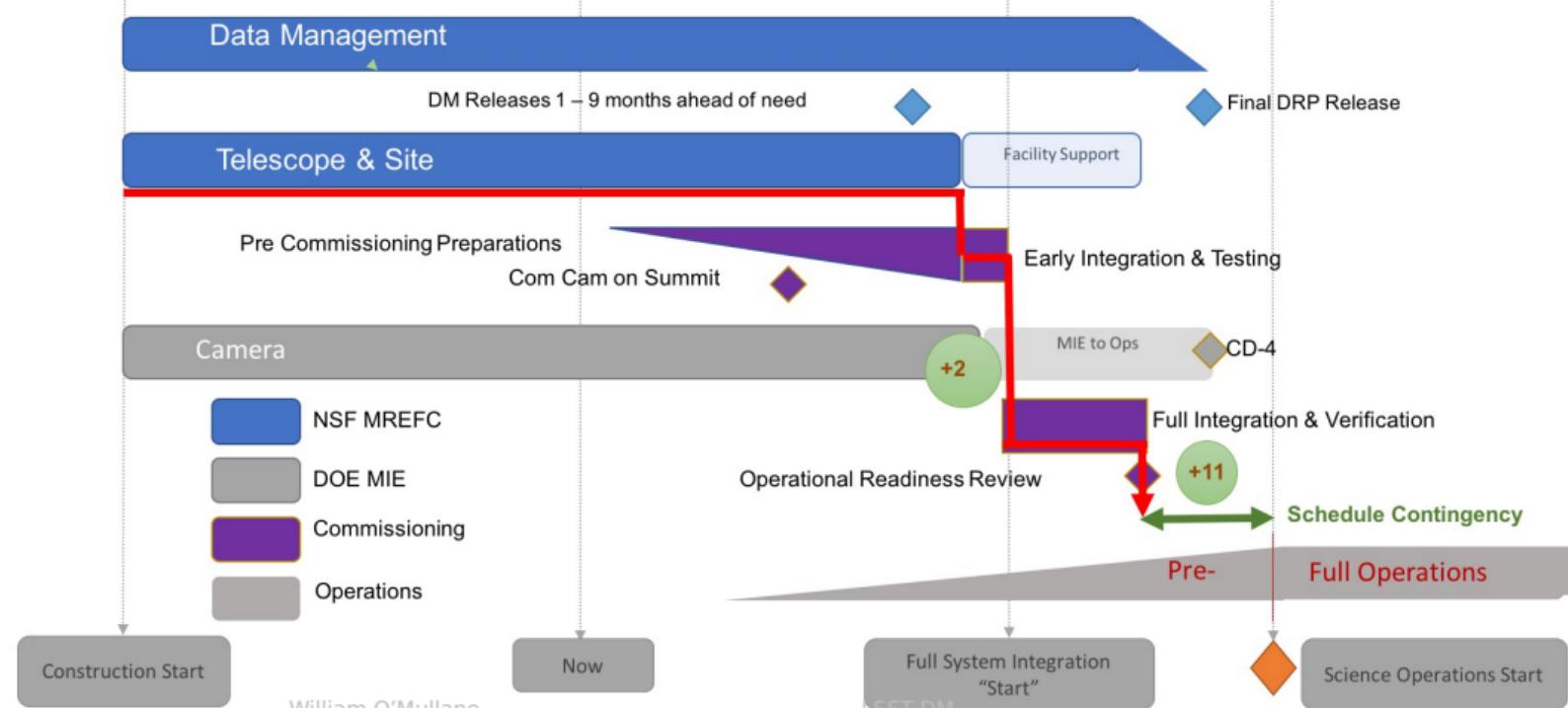




# High level plan



FY 2014				FY 2015				FY 2016				FY 2017				FY 2018				FY 2019				FY 2020				FY 2021				FY 2022				FY 2023			
Q1	Q2	Q3	Q4																																				





# Potentially lots of data for DM



Data Production Milestone	Completion Date
First calibration data from Auxiliary Telescope	02 Aug 2018
First on-sky and calibration images with ComCam	29 Jan 2020
Sustained scheduler driven observing with ComCam	11 May 2020
Images from Camera re-verification at Summit Facility	16 Jun 2020
First on-sky and calibration data from Camera+Telescope	18 Nov 2020
Sustained scheduler driven observing with Camera+Telescope	08 Feb 2021
Start Science Verification mini-Surveys	30 Mar 2021



## DM Challenges



LSST will provide A large (100 PB) database and sophisticated analysis tools: **for each of 40 billion objects there will be about 1000 measurements (each with a few dozen measured parameters)**

1. Large data volume
2. Large number of objects
3. Highly multi-dimensional space
4. Unknown statistical distributions
5. Time-series data
6. Truncated, censored and missing data
7. Unreliable quantities (e.g. unknown systematics and random errors)



# Outline

---



Large Synoptic Survey Telescope status

LSST Operations

Data Management

Data Access/ Science Platform

Conclusion



# LSST operations proposal



- Proposal was submitted in summer to NSF/DOE to fund LSST operations
- A joint agency review was completed with positive feedback Dec 7<sup>th</sup> 2017.
- LSST Operations key points:
  - Distributed over SLAC, Tucson, La Serena, NCSA Illinois
  - Large data facility element
  - To fit in the new National Center for Optical and infrared Astronomy (NCOA) framework



# LSST Operations - Distributed



100 up to  
200Gbps  
international  
links

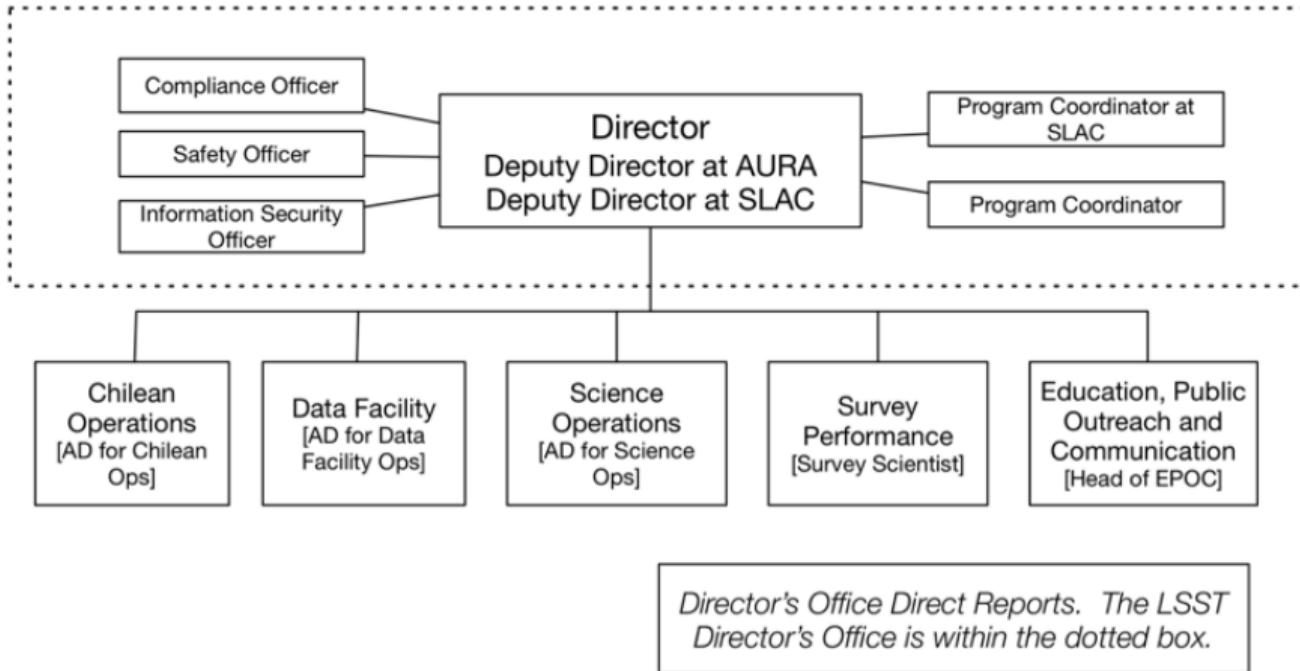
100(40)Gbps  
summit base See  
(LSE-78)

**Jeff Kantor**





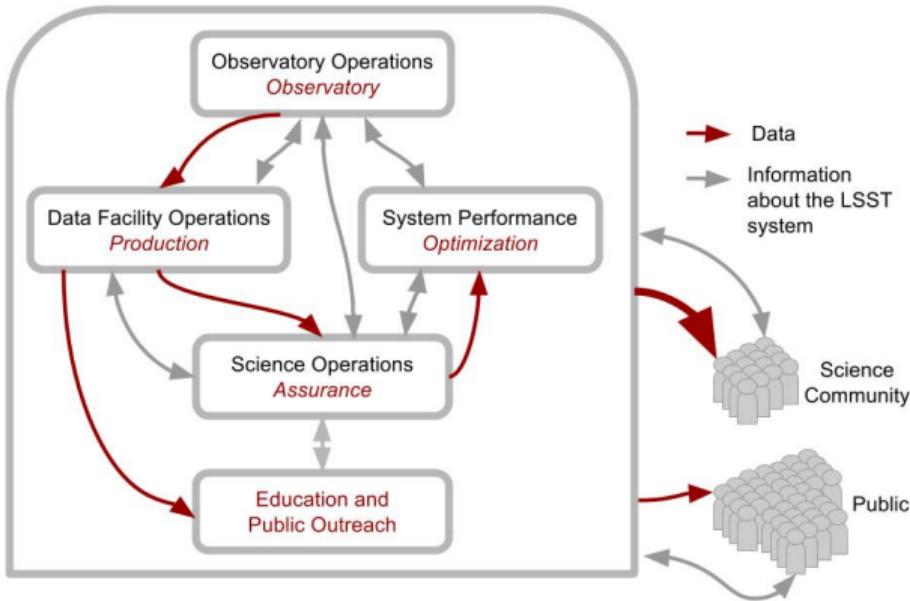
# LSST Operations - Organisation



**Beth Willman**



# LSST Operations - Communications

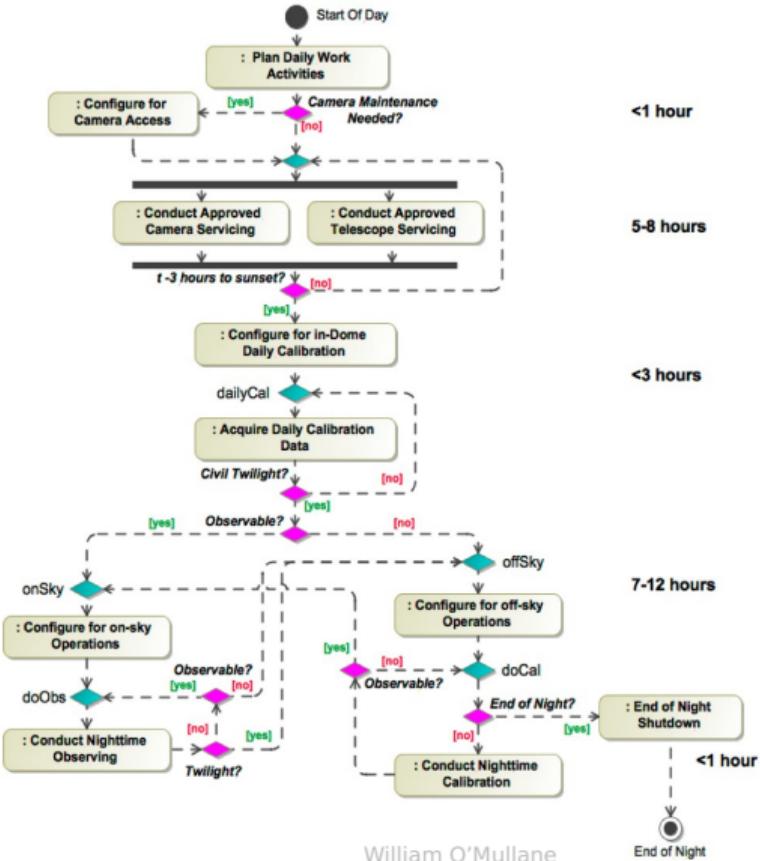


- Formal and informal channels in place
- Regular weekly daily meetings as well as Jira for tracking
- Slack, community.lsst.org, etc as well

Phil Marshall



# Observatory Operations Activities (on summit)



High level 24 hour activity (50FTE):

- Regular maintenance
- Evening calibrations
- Nightly observations
- Day crew, night crew shift 1, night crew shift 2
- Software
- ITC supports daily data transmission



On a daily basis Science Operations Staff are looking at data quality from both instrument and software perspectives asking many questions (28FTE):

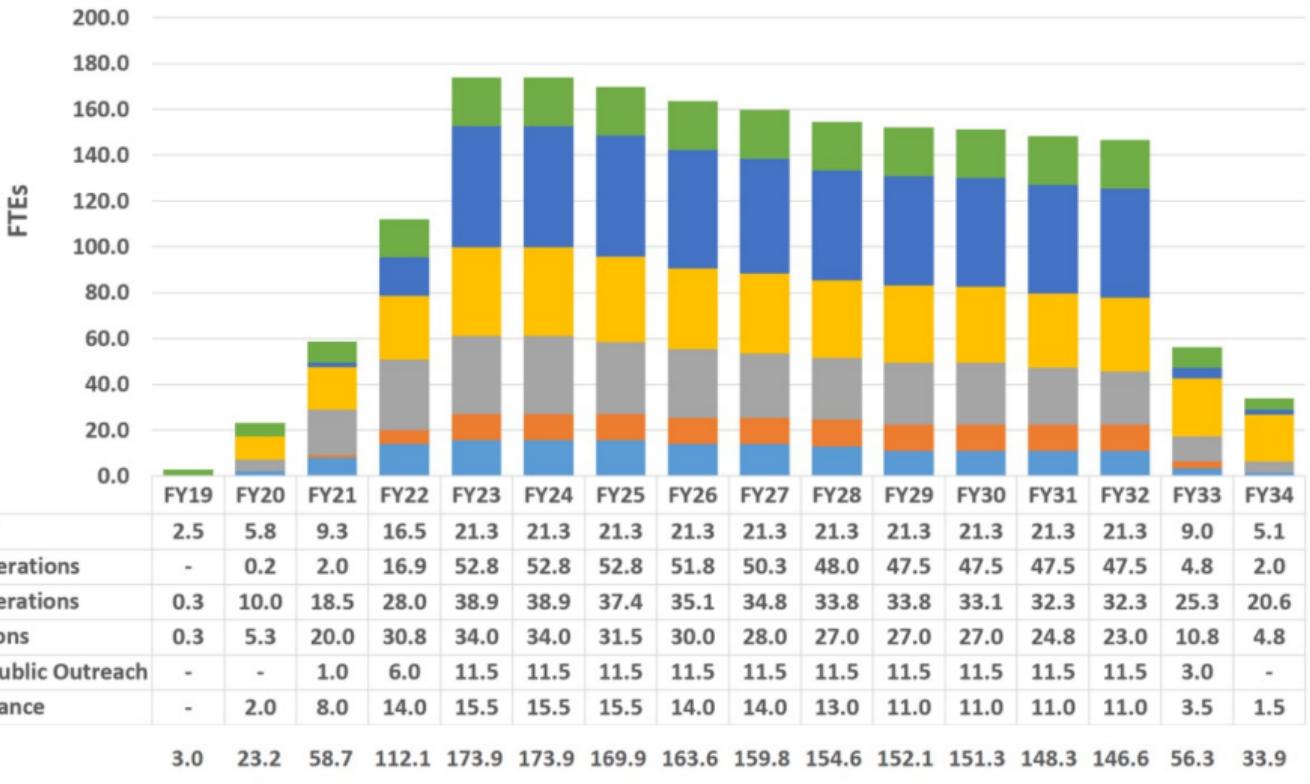
- Are the alerts as good as we can make them ?
- Are there any data products we could deliver/improve?
- Are the changes we should request on the telescope ?
- Was there some event (weather/hardware) affecting data we should be telling the community about ?

Longer term :

- Are there any disturbing trends in Key Performance Metrics?
- How is the Data Release Product quality ?



# LSST Operations - Staffing profile





# Outline

---



Large Synoptic Survey Telescope status

LSST Operations

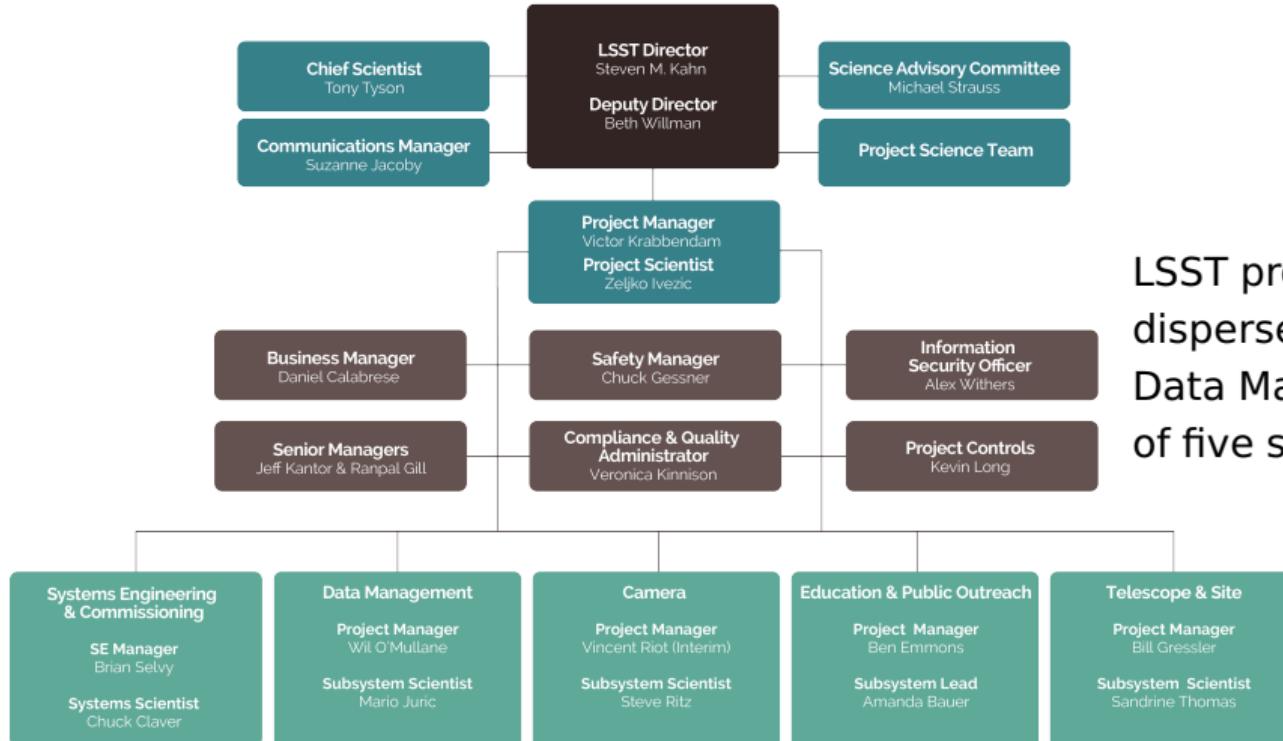
Data Management

Data Access/ Science Platform

Conclusion



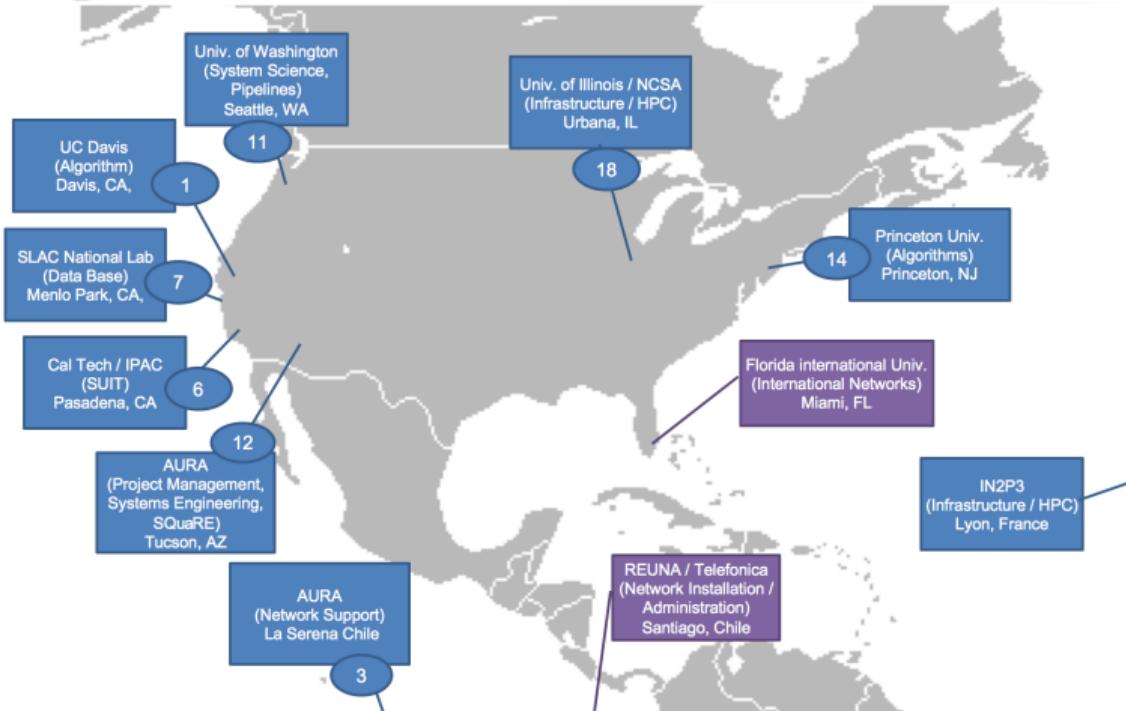
# LSST org chart - where DM fits



LSST project is large and dispersed  
Data Management is just one of five subsystems.



# Data management



DM Mission :

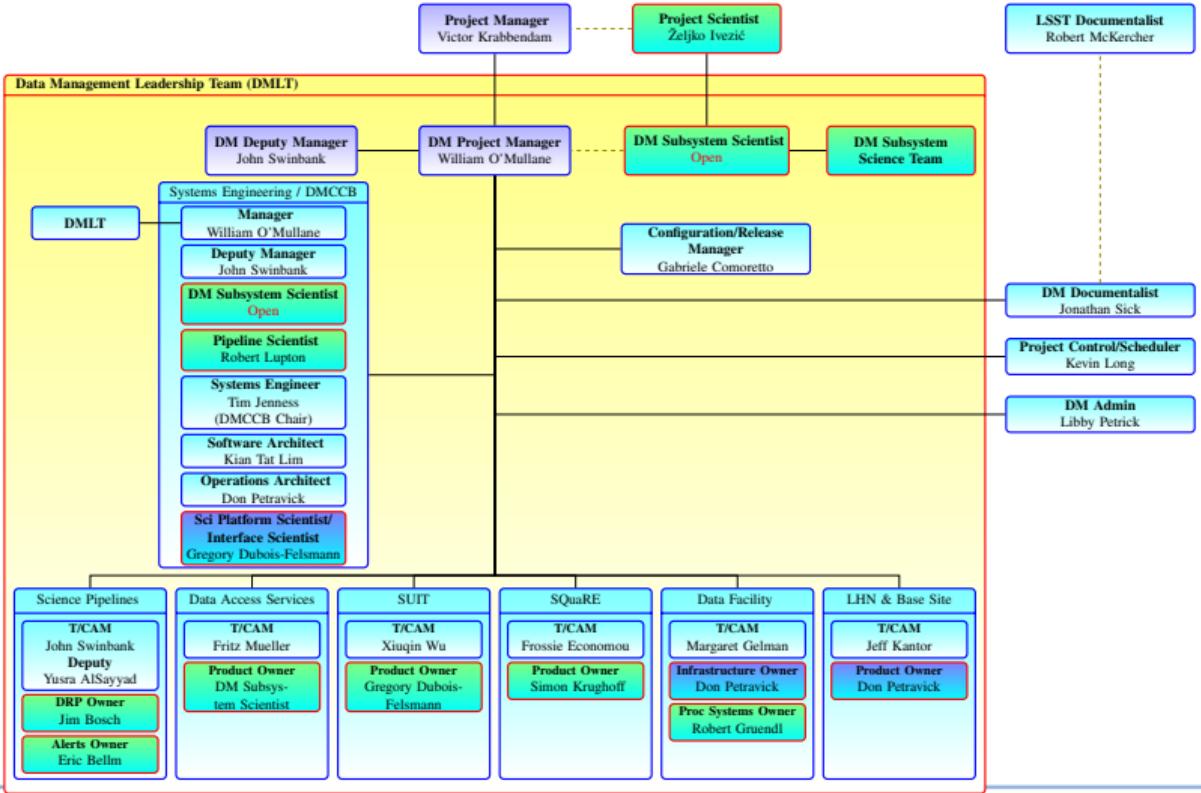
*Stand up operable,  
maintainable, quality services  
to deliver high-quality LSST  
data products for science, all  
on time and within reasonable  
cost.*

LSST DM development is  
distributed across the  
Americas.

Plus we have partners like  
IN2P3



# DM Organization



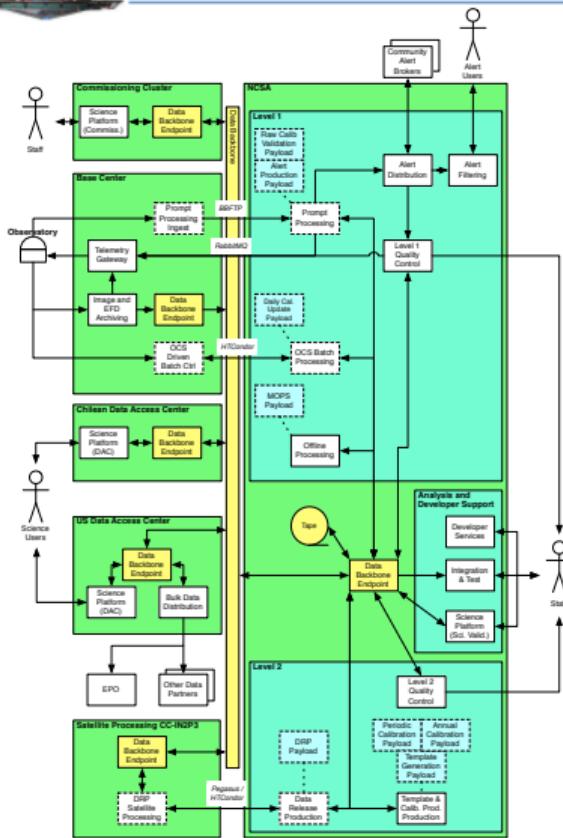
DM leadership meet two times a year and have a weekly telecon.

Technical mangers have a *standup* every Tuesday and Friday.

Toughest thing in any project is communication.



# DM build and deploy



DM must build everything to get LSST products (see <http://ls.st/dpdd>) to the users.

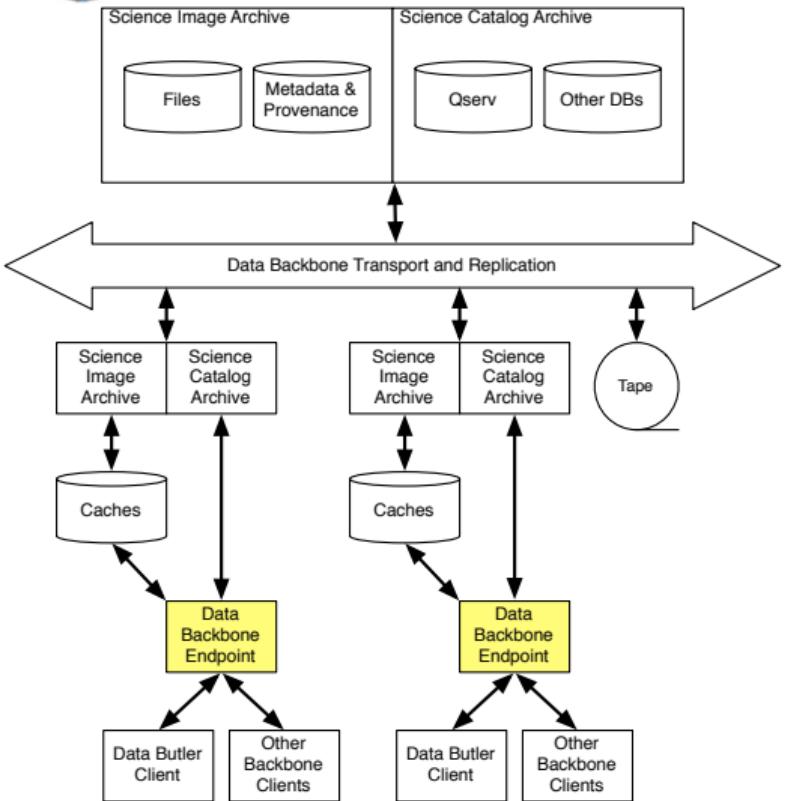
- large data sets (20TB/night)
- complex analysis
- aiming for small systematics
- Science Alerts in under 2 minutes .. (aiming for 1 minute)

About  $\frac{1}{2}$  million lines of code (C++/python)

diagram K.T. Lim



# Data Backbone



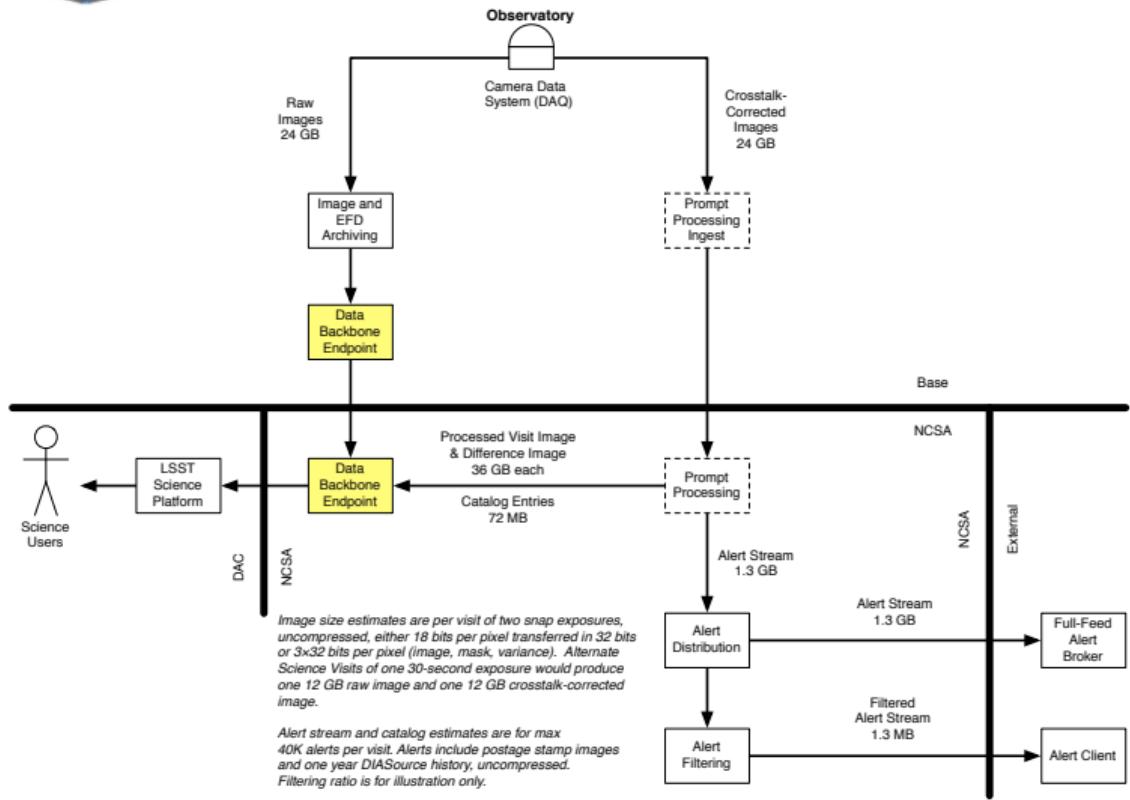
One small box on the previous slide was Data Backbone.  
That hides several things

- Qserv - the LSST end user database
- All the networks : we now have fiber to the mountain and from La Serena to NCSA (two routes)

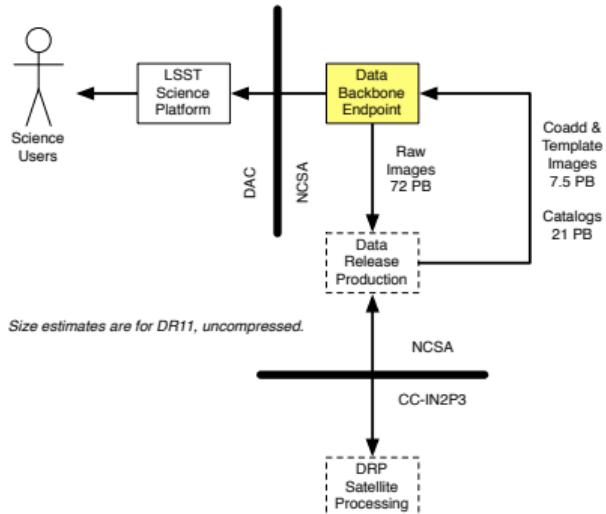
diagram K.T. Lim



# Data flow



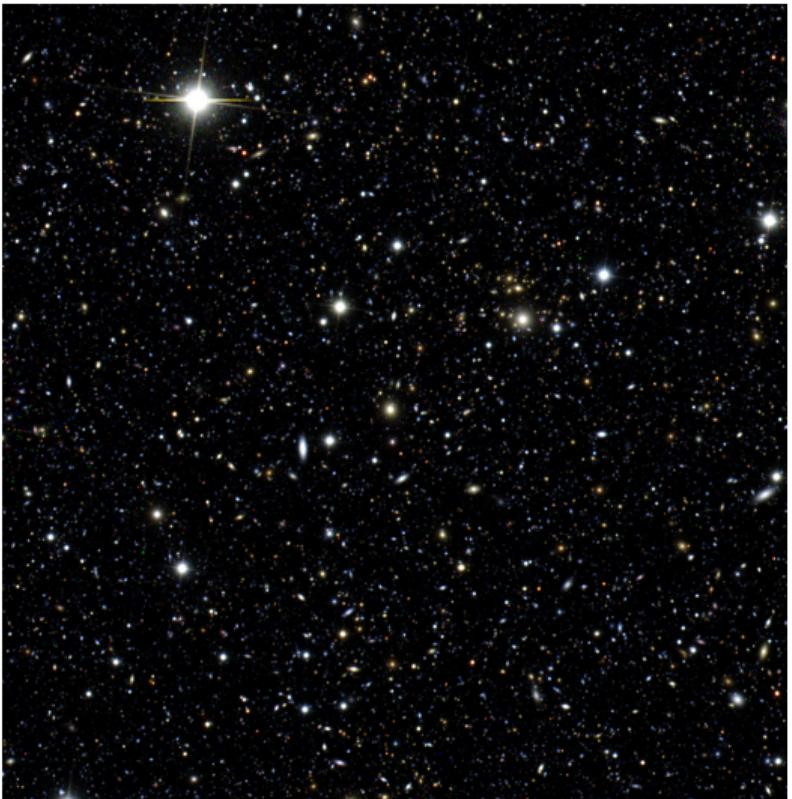
Lots to do every night ..  
Plus annually there is a data release



Images from K.T. Lim



# All comes down to images ...



False color from 3 simulated filter images  
From just one of 109 CCDs

To work with that there is the *LSST Stack*  
<https://pipelines.lsst.io/>  
Current version 14.0 release in November  
2017.  
Python 3 and C++ 14.



# Outline

---



Large Synoptic Survey Telescope status

LSST Operations

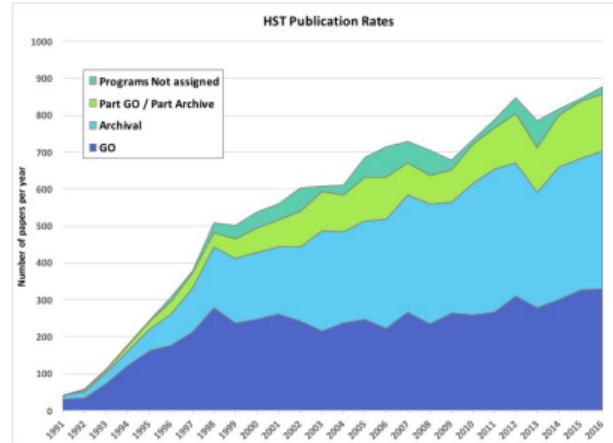
Data Management

Data Access/ Science Platform

Conclusion



# The Era Of Surveys





# First serious go at code to data



Early in the millennium - CasJobs (C#)(OMullane et al., 2005)

View Jobs - Mozilla Firefox

File Edit View Bookmarks Tools Help

## SDSS Query / CasJobs

SDSS

Help Tools Query History MyDB Import Groups Output Profile Admin SkyServer Logout thakar

Refresh this page to get latest info

Status	Target	Name Like
ANY	DR6	

Apply

Name	Query	Target	Submitted	Time(h:m:s)	Rows	Status
Info Quasar Search in Imaging	SELECT TOP 1000 run,	DR6/600	7/15/2007 7:56:02 PM			Ready
Info Quick Query	-- Galaxies meeting multiple s	DR6/1	7/15/2007 7:55:02 PM	0:0:0	10	Finished
Info Sample Query 1	SELECT top 1000 G.ObjID, G.u,	DR6/600	7/15/2007 7:54:13 PM			Started
Info My Query	select objid,ra,dec,modelmag_g	DR6/600	7/15/2007 12:17:42 AM	0:0:4	11308	Finished
Info My Query	SELECT count(*) as 'total'	DR6/600	7/8/2007 11:28:37 AM	0:1:1		Failed
Info My Query	SELECT TOP 1000 P.ObjID into m	DR6/600	7/8/2007 11:27:47 AM	0:7:53	1000	Finished
Info My Query	SELECT TOP 1000 run,	DR6/600	7/8/2007 11:27:22 AM	0:0:1		Failed
Info My Query	SELECT TOP 10000 run,	DR6/600	7/8/2007 11:27:06 AM	0:0:2	10000	Finished
Info My Query	select top 10 * into mydb.MyTa	DR6/1500	6/26/2007 4:30:35 PM	0:0:0	10	Finished
Info My Query	select top 10 * from dr5quasar	DR6/1	6/26/2007 1:08:12 PM	0:0:0	10	Finished
Info My Query	select top 10 * from dr3quasar	DR6/1	6/26/2007 1:08:03 PM	0:0:0		Failed

Contact: \$Name: v3\_1\_0 \$,\$Revision: 1.9 \$, Last modified: Thursday, September 14, 2006 at 9:54:32 PM

DR3

Selected object

ra	18.67664
dec	-0.66097
type	GALAXY
u	14.83
g	13.74
r	13.19
i	12.92
z	12.92

Get Image

Parameters

Get Image

Use query to mark objects

Drawing options

- Grid
- Label
- PhotoObjs
- SpecObjs
- Targets
- Outline
- BoundingBox
- Fields
- Masks
- Plates
- InvertImage

Explore

Register

Add to notes

Show notes

N  
E  
W  
S

evolved

<http://www.sciserver.org/>

William O'Mullane

LSST DM

31



# LSST Science Platform



For DR2:

- Computing: 2,400 cores ( $\approx 18$  TFLOPs)
- File storage:  $\approx 4\text{PB}$  (VOSpace)
- Database storage:  $\approx 3\text{PB}$  (MYDB)

The Science Platform has three user facing aspects: the Portal (novice), the JupyterLab (intermediate), and the Web APIs (expert and remote tools).

We enable access to the Data Releases and Alert Streams, and support next-to-the data analysis and Level 3 product creation using the computing resources available at the Data Access Center (DAC).

Mario Juric

Concept has been around a year or two - name suggested in Nov 16



# PDAC - Portal, Notebook



Add a URL for downloading the image by ID

```
In [16]: df_coadds['img_url'] = df_coadds.deepCoaddId.map(lambda x:  
    'http://lsst-qserv-dax01.ncsa.illinois.edu:5000/image/v0/deepCoadd/id?id=' + str(x))
```

Verify that df\_coadds has at least five entries for the five filters

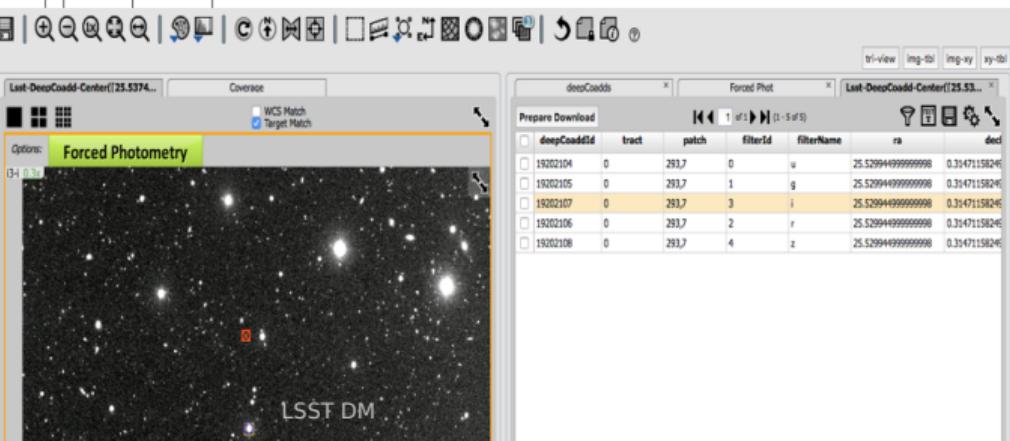
```
In [15]: df_coadds
```

```
Out[15]:
```

	deepCoaddId	tract	patch	filterId	filterName	ra	decl	htmlId20	equinox	raDeSys	...	corner3Ra	corner3Decl	corner4Ra
0	19202104	0	293,7	0	u	25.529945	0.314712	17147177442654	2000.0	ICRS	...	25.416645	0.422899	25.416645
1	19202105	0	293,7	1	g	25.529945	0.314712	17147177442654	2000.0	ICRS	...	25.416645	0.422899	25.416645
3	19202106	0	293,7	2	r	25.529945	0.314712	17147177442654	2000.0	ICRS	...	25.416645	0.422899	25.416645
2	19202107	0	293,7	3	i	25.529945	0.314712	17147177442654	2000.0					
4	19202108	0	293,7	4	z	25.529945	0.314712	17147177442654	2000.0					

5 rows x 34 columns

- Results available in portal and notebook
- Notebook can drive portal
- Forced Photometry
- NeoWise and Stripe82 currently loaded





# Evolution of the platform



- Documentation
  - The direction is now documented in LSE-319
  - More formal requirements in LDM-554
  - Design is in LDM-542
  - First user reports in DMTR-22
- Coming soon
  - Portal focus on exploration/discovery of data products
  - full sky browsing e.g. something like HiPS <http://www.ivoa.net/documents/HiPS/>
  - Authentication integrated with existing credentials
  - Quotas on CPU and Disk .. has to be done..
  - Data Upload to user space
  - Scheme to include collaborator code in Stack (thus in platform)
    - proper way to pull any module or code into notebook kernel from say github
- To be used in commissioning by the commissioning team
  - Boot camps soon



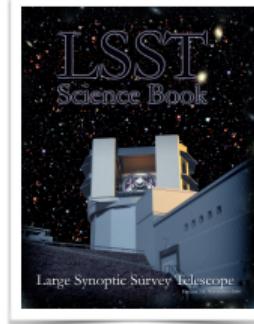
# Preparing for LSST: Science Collaborations



- 9 LSST science collaborations, covering the science topics envisioned for LSST:
  - AGN; Dark Energy; Galaxies; Stars, Milky Way and Local Volume; Solar System; Statistics and Informatics; Strong Lensing; Supernovae; Transients/Variable Stars
- Preparing the LSST for the community, and the community for LSST
- More than 400 unique members!

**All Contributors with data rights can join !**

Initial collaboration work: LSST Science Book  
(LSST Science Collaboration, 2009)





# Outline

---



Large Synoptic Survey Telescope status

LSST Operations

Data Management

Data Access/ Science Platform

Conclusion



## Conclusion



- Current astronomical surveys are changing the way we do astronomy
- LSST will soon follow perhaps ushering a complete change in how we approach end user data interaction
- LSST is on track and data is coming sooner than you might think
- There are plenty of challenges
- Verification and Validation on radar for now
- Looking forward to the first LSST images !

Motto for the future from LSST Project Scientist:

**Ask Not What Data You Need To Do Your Science,  
Ask What Science You Can Do With Your Data.**

Željko Ivezić



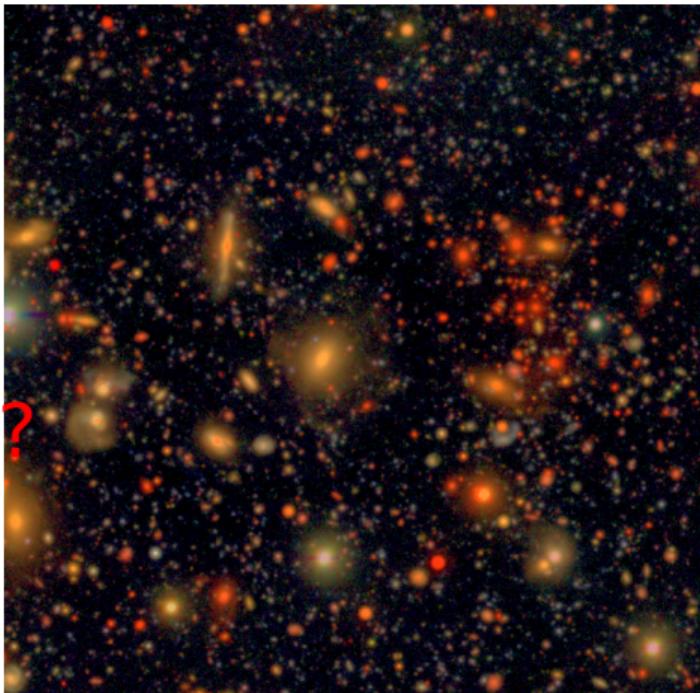
# The END



Questions?

~ 3.5' SDSS image

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



HSC image (COSMOS)  
g,r(1.5 hrs) ,i(3 hrs) PSF matched co-add ( $\approx 27.5$ )  
Images:Lupton and HSC collaboration see also Lupton et al. (2004)



# Outline



## Reference material



# Acronyms I



<b>Acronym</b>	<b>Description</b>
ADQL	Astronomical Data Query Language
API	Application Programming Interface
CCD	Charge-Coupled Device
CPU	Central Processing Unit
DAC	Data Access Center
DM	Data Management
DOE	Department of Energy
FOP	Flight Operation Procedure (Plan)
ITC	Information Technology Center
IVOA	International Virtual-Observatory Alliance
JHU	Johns Hopkins University
LSST	Large Synoptic Survey Telescope
MPP	Massive Parallel Processing
NCOA	National Center for Optical and infrared Astronomy
NCSA	National Center for Supercomputing Applications
NSF	National Science Foundation
PB	PetaByte
PDAC	Prototype Data Access Center
PSF	Point Spread Function
SDSS	Sloan Digital Sky Survey
SIA	Sagnac Interferometer Assembly



## Acronyms II



SLAC	Stanford Linear Accelerator Center
SQL	Structured Query Language
TAP	Table Access Protocol
TFLOP	Tera FLOP
VO	Virtual Observatory



# References



- [LDM-554], Ciardi, D., Dubois-Felsmann, G., 2017, *Science Platform Requirements*, LDM-554, URL <https://ls.st/LDM-554>
- [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>
- [LSE-319], Jurić, M., Ciardi, D., Dubois-Felsmann, G., 2017, *LSST Science Platform Vision Document*, LSE-319, URL <https://ls.st/LSE-319>
- [LSE-78], Lambert, R., Kantor, J., Huffer, M., et al., 2017, *LSST Observatory Network Design*, LSE-78, URL <https://ls.st/LSE-78>
- LSST Science Collaboration, 2009, ArXiv e-prints (arXiv:0912.0201), ADS Link
- Lupton, R., Blanton, M.R., Fekete, G., et al., 2004, Publications of the ASP, 116, 133 (arXiv:astro-ph/0312483), doi:10.1086/382245, ADS Link
- Ness, J.U., Parmar, A.N., Valencic, L.A., et al., 2014, Astronomische Nachrichten, 335, 210 (arXiv:1311.5751), doi:10.1002/asna.201312001, ADS Link
- O'Mullane, W., Li, N., Nieto-Santisteban, M., et al., 2005, *Batch is back: CasJobs, serving multi-TB data on the Web*, Tech. rep., Microsoft,  
Microsoft Technical Report MSR TR 2005 19 (arXiv:cs/0502072), ADS Link
- [DMTR-22], Suberlak, K., Ivezić, Ž., The PDAC Team, 2017, *Prototype Data Access Center: User Report*, DMTR-22, URL <https://ls.st/DMTR-22>