

LSST Data Products and Data Release Plans

Leanne Guy

Data Management Scientist
AURA/LSST, Tucson, AZ, USA

for the LSST Data Management Team.



Large Synoptic Survey Telescope Corporation
Transients and Variable Stars workshop
Napoli, Italy, 9-11 April 2018



Introduction



- Incoming LSST Data Management Scientist,
 - Taking over from Mario Jurić
 - Not officially started yet
 - so still learning!
- Mid-May 2018 start date, LSST HQ Site, Tucson, AZ, USA.
- Some of you may know me from Gaia, CERN.



The LSST Site is Shaping Up!



Artist's Impression



Photo Dec 2017



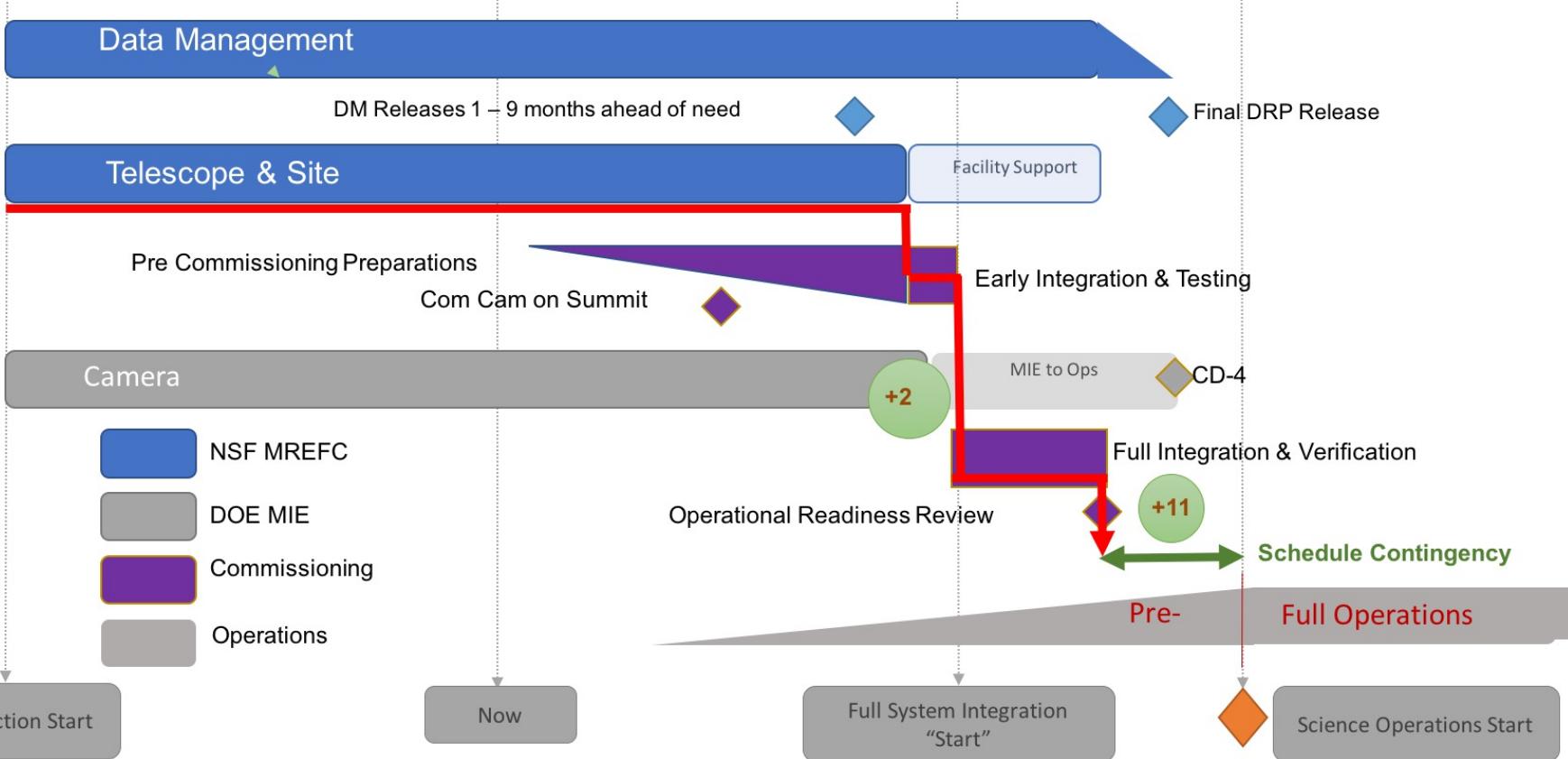
- Dome completion planned mid 2018
- Network partially in place, internal cabling being done
- Follow progress at : [See what's happening on Cerro Pachón](#)



LSST Project Schedule



FY 2014				FY 2015				FY 2016				FY 2017				FY 2018				FY 2019				FY 2020				FY 2021				FY 2022				FY 2023				FY 2024			
Q1	Q2	Q3	Q4																																								



LSST Project Schedule



LSST Data Management



*Data Release Data Products
via annual Data Releases*



20TB raw data/night



*Prompt Data Products
via nightly Alert Streams*



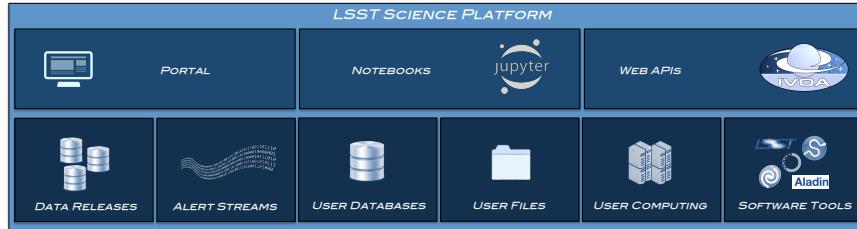
**Average ~ 10 million/night
Real-time latency 60sec**

After 10 years

- 11 Data Releases
- 15PB Database Catalog



INTERNET



**Data access via Data Access Centres & Services:
the *LSST Science Platform***



LSST Data Products



The LSST Data Products are organised into three main categories:

A stream of ~10 million time-domain events per night, detected and transmitted to event distribution networks with 60 seconds of observation.

A catalog of orbits for ~6 million bodies in the Solar System

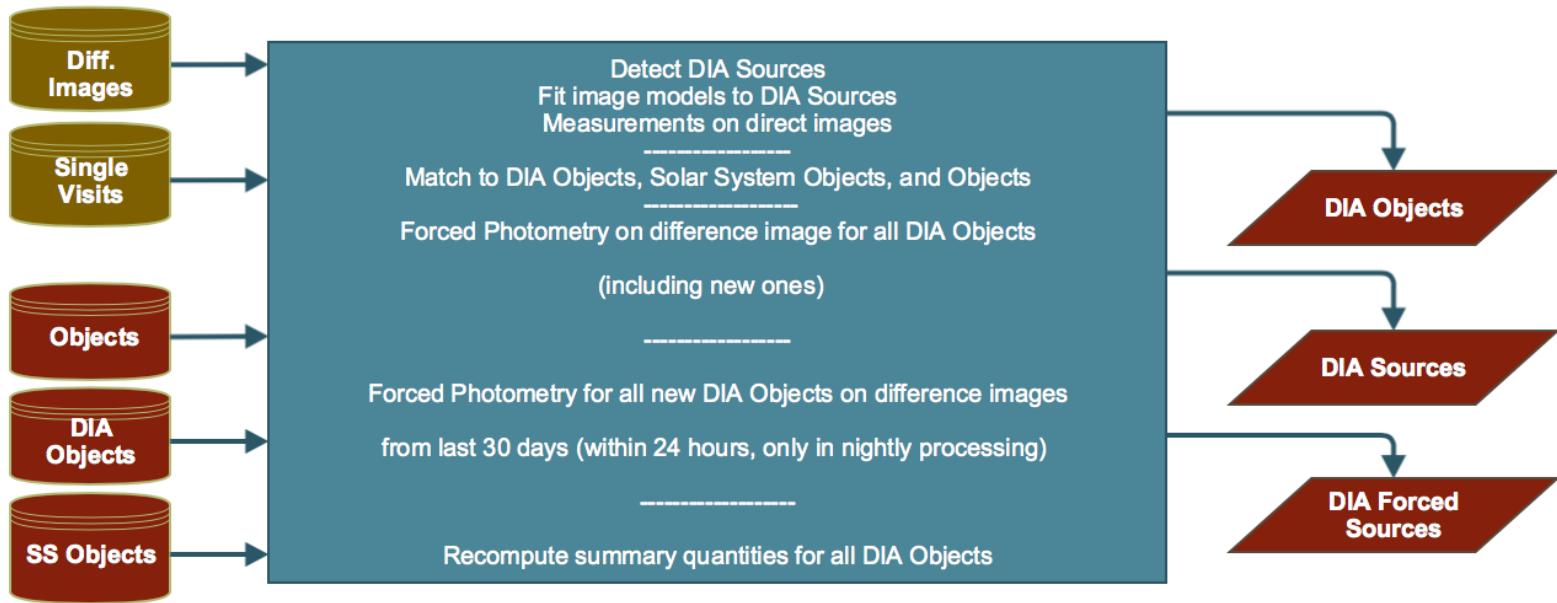
Previously "Level 1" data products

Prompt

[LSST Data Product Categories \(LPM-231\)](#) and <https://www.lsst.org/about/dm/data-products>



Difference Image Analysis (DIA)



Within 60s of shutter close (2nd exposure):

- Alert generated & issued to the stream, one per new DIA Source
- **Prompt** catalogs in the US DAC are updated

Within ~24h:

- Forced photometry in past 30 days of images for all new DIA Objects
- Raw and Processed images available in the US DAC

Prompt processing also includes nightly identification of Solar System Objects.



Prompt Data Products



The ***Prompt*** database is a living database, updated in real-time that contains the objects and sources detected on difference images

DIASources

- signal-to-noise ratio of detection
- coordinates, and association with DIAObject or SSOObject
- time of mid-exposure at location on CCD
- flux in the difference and visit images (PSF and aperture)
- shape parameters (trails, dipoles, FWHM, extendedness)
- parent/child de-blending flags

DIAObjects

- Time-averaged coordinates
- Fluxes by filter, time-averaged
- Periodic and non-periodic variability features
- Nearby objects in the ***Data Release*** catalog'

Non-exhaustive



Alert Packet Contents



Each alert (a VO event packet) will contain at least the following:

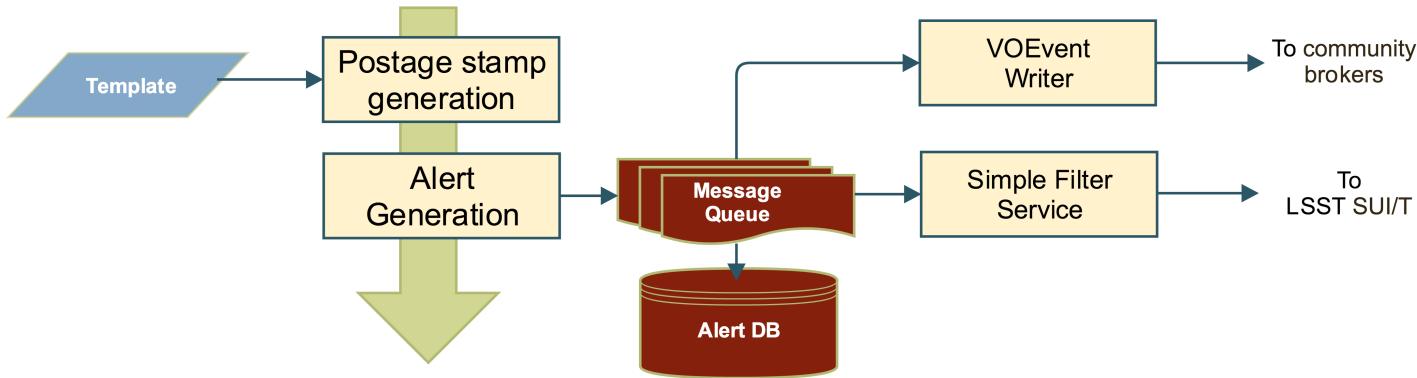
- *alertID*: An ID uniquely identifying this alert. It can also be used to execute a query against the Level 1 database as it existed when this alert was issued.
- ***Prompt database ID***
- Science Data:
 - The DIASource record that triggered the alert
 - The entire DIAObject (or SSObject) record
 - Previous 12 months of DIASource records
 - Matching Object IDs from the latest Data Release, if they exist, and 12 months of their DIASource records
- Cut-out of the difference image centered on the DIASource (10 bytes/pixel, FITS MEF)
- Cut-out of the template image centered on the DIASource (10 bytes/pixel, FITS MEF)



Alert Distribution & Brokers



Alert Distribution



Community Brokers

- Primary end-points of LSST's event streams
- Provided by third parties
- Alerts transmitted in VOEvent format & standard IVOA protocols
- Filter, classify events into subsets according to science goals,
- Trigger follow-up
- Finite number of brokers will be selected by a proposal process to receive the full stream

LSST 'mini-broker'

- User-defined filters that act only on alert packet contents
 - e.g no cross-matching capabilities
- Access to the filtered stream via LSST's Science Platform at the US DAC
- Cap of ~20 alerts per user per visit; some limits on computing capacity

[LSE-151: Science Pipelines Design Document](#)

[LSE-163: Data Products Definition Document](#)



LSST Data Products



The LSST Data Products are organised into three main categories:

A stream of ~10 million time-domain events per night, detected and transmitted to event distribution networks with 60 seconds of observation.

A catalog of orbits for ~6 million bodies in the Solar System

Previously "Level 1" data products

Prompt

A catalog of ~37 billion objects (20B galaxies, 17B stars), ~7 trillion observations ("sources"), and ~30 trillion measurements ("forced sources"), produced annually, accessible through online databases.

Reduced single-epoch, deep co-added images.

Previously "Level 2" data products

Data
Release

[LSST Data Product Categories \(LPM-231\)](#) and <https://www.lsst.org/about/dm/data-products>



LSST Data Products



The LSST Data Products are organised into three main categories:

A stream of ~10 million time-domain events per night, detected and transmitted to event distribution networks with 60 seconds of observation.

A catalog of orbits for ~6 million bodies in the Solar System

Previously "Level 1" data products

Prompt

A catalog of ~37 billion objects (20B galaxies, 17B stars, 10B quasars), observations (“sources”), and ~30 trillion produced annually, accessible through one API.
Reduced single-epoch, deep co-added images.

CU7 Variability Processing

gala

- Statistical Parameters
- Time series Characterization
- Variability Classification (basic)

Data
Release

[LSST Data Product Categories \(LPM-231\)](#) and <https://www.lsst.org/about/dm/data-products>



Data Release Data Products



- Well calibrated, consistently processed catalogs and images
 - Catalogs of objects, detections, detections in difference images, etc.
 - Extensive computation often required - combine information from many exposures
- Made available via a '*Data Release*'
 - Performed yearly (DR2.. DR11)
 - ... with an additional data release for first 6 months of survey data. (DR1)
 - Complete reprocessing of all data to date for each DR with latest pipelines
 -including fully reprocessed *Prompt data products*
- Catalog Access
 - Relational database and via the LSST Science Platform
 - Remote access APIs, VO protocols (e.g., TAP)
- Projected catalog sizes:
 - 18 billion objects (DR1) → 37 billion (DR11)
 - 750 billion observations (DR1) → 30 trillion (DR11)
 - Few PB (DR1) → ~70PB (DR2)



Data Release Catalog Contents



Images:

- Single visit images with IS removed, background, PSF, zero-point and WCS determined
- Difference images, Deep co-adds across the entire survey footprint

Sources & Forced Sources Catalogs

- locations and times of all detections
- aperture and point source flux; PSF moments
- de-blending results (parent/child identifiers)
- model fits (bulge/disk, exponential, petrosian, kron)
- surface brightness, extendedness parameters
- color (seeing-independent)
- Forced photometry performed on all exposures, on all Objects

Objects

- Characterization of objects detected on multi-epoch data
- Photometric redshift
- Periodic and non-periodic features
- Proper motions and parallaxes

Non-exhaustive



Variability Characterization



- ***Data Release Production*** post-processing ***Classification*** pipeline computes
 - Variability summary statistics
 - Probabilistic and/or discrete classification of each Object as a star or galaxy .. and possibly other types (e.g. QSO, SN).
- ... from light curves of *Object* and *DIAObjects*
- Light-curve variability characterized by a series of ***Periodic*** and ***Aperiodic*** summary ‘features’ derived from the light-curve, computed separately for each band
- Initial list of parameters given in LDM-151.



Object Variability Parameters



Variability parameters in the *Object* and *DIAObject* tables

Table 2: DIAObject Table

Name	Type	Unit	Description
diaObjectId	uint64		Unique identifier.
radec	double[2]	degrees	(α, δ) position of the object at time
.....			radecTai.
lcPeriodic	float[6 × 32]		Periodic features extracted from light-curves using generalized Lomb-Scargle periodogram [Table 4, 16] ⁵¹ .
lcNonPeriodic	float[6 × 20]		Non-periodic features extracted from light-curves [Table 5, 16].

Periodic (lcPeriodic) parameters

- 32 parameters calculated for each band
- Lomb-Scargle periodogram
- 3 fundamental frequencies, 4 harmonics
- Linear coefficient, amplitude, etc

APeriodic (lcNonPeriodic) statistics

- 20 parameters calculated for each band
- Median Absolute Deviation, Standard Deviation, Slope trend, Skewness, Welch-Stetson variability index, Percentiles, etc

Transient and Variable Star SC has a task force for reviewing these parameters.



LSST Database Browser



Explore the full schema in the [Database Schema Browser](#)



Data Release Availability & Retention



Once produced & validated, a DR is frozen and archived

- Can be readily accessed in the future
- Facilitates publishing based on specific LSST data releases

Not feasible to keep all DRs loaded and accessible at all times.

- Most users will work with the most recent release
- Older releases will be archived on mass storage
 - Queries against archived releases will not be possible.
 - Available as bulk downloads

Only the contents of the most recent and penultimate data releases will be kept on fast storage and with catalogs loaded into the database.

All raw data used to generate any public data product (raw exposures, calibration frames, telemetry, configuration metadata, etc.) will be kept and made available for download



LSST Data Products



The LSST Data Products are organised into three main categories:

A stream of ~10 million time-domain events per night, detected and transmitted to event distribution networks with 60 seconds of observation.

A catalog of orbits for ~6 million bodies in the Solar System

Previously "Level 1" data products

Prompt

A catalog of ~37 billion objects (20B galaxies, 17B stars), ~7 trillion observations ("sources"), and ~30 trillion measurements ("forced sources"), produced annually, accessible through online databases.

Reduced single-epoch, deep co-added images.

Previously "Level 2" data products

Data
Release

User-produced added-value data products e.g deep KBO/NEO catalogs, variable star classifications, shear maps, etc ...

Enabled by services and computing resources at the Data Access Centers and via the LSST Science Platform

Previously "Level 3" data products

User
Generated

[LSST Data Product Categories \(LPM-231\)](#) and <https://www.lsst.org/about/dm/data-products>



LSST Data Products



The LSST Data Products are organised into three main categories:

A stream of ~10 million time-domain events per night, detected and transmitted to event distribution networks with 60 seconds of observation.

A catalog of orbits for ~6 million bodies in the Solar System

Previously "Level 1" data products

Prompt

A catalog of ~37 billion objects (20B galaxies, 17B stars), ~7 trillion observations ("sources"), and ~30 trillion measurements ("forced sources"), produced annually, accessible through online databases.

Reduced single-epoch, deep co-added images.

Previously "Level 2" data products

Data
Release

User-produced added-value data products e.g. deep KBO/NEO catalogs, variable star classifications, shear maps, etc ...

Enabled by services and computing resources
LSST Science Platform

CU7 Variability Processing

- Variability Classification
- Specific Object Studies

gala

Previously "Level 3" data products

User
Generated

[LSST Data Product Categories \(LPM-231\)](#) and <https://www.lsst.org/about/dm/data-products>



User Generated Products



User-Generated products are created by the community using LSST software & services.

Enable science cases not fully covered by Prompt and Data Release processing, e.g.

- Custom processing of deep drilling fields
- Custom measurement algorithms

LSST will

- ***not*** write unique algorithms for re-processing main survey data
- make available the LSST Software Stack source code, which the community can extend to generate new Level 3 products
- commit ~10% of its computing resources toward enabling end-user analysis and Level 3 data product creation

User Generated software/data products may be migrated to Level 2 (with owners' permission); this is one of the ways how Level 2 products will evolve.



User Generated Example



Scenario: At 1.5 years into the LSST survey, a transient is found in the core of a galaxy. The current *prompt database* has no detections at its coordinates. We need to constrain AGN activity in order to prioritize it as a candidate tidal disruption event (TDE).

Available Data:

'Living' Level 1 DIAObjects: -association with DR1 Objects

-30 days of *precovery* forced photometry

DR1 Objects: -first 12 months of single-visit aperture photometry of host

-variability parameters (currently undefined but e.g., lcNonPeriodic)

DR1 DIAObjects: -difference image limiting magnitudes

-(if SNR>5) 12 months of difference image photometry

Information Gap:

Variability constraints during LSST survey months 13, 14, 15, 16, and 17.

Would not be available until DR2, but could be useful for prioritizing immediate follow-up.

Identified Science Needs:

- a Level 3 pipeline for forced photometry in galaxy cores
- Level 2 variability flags could include an AGN-specific parameter

Courtesy: Melissa Graham



LSST Special Programs



WFD

Wide-Fast-Deep

DDF

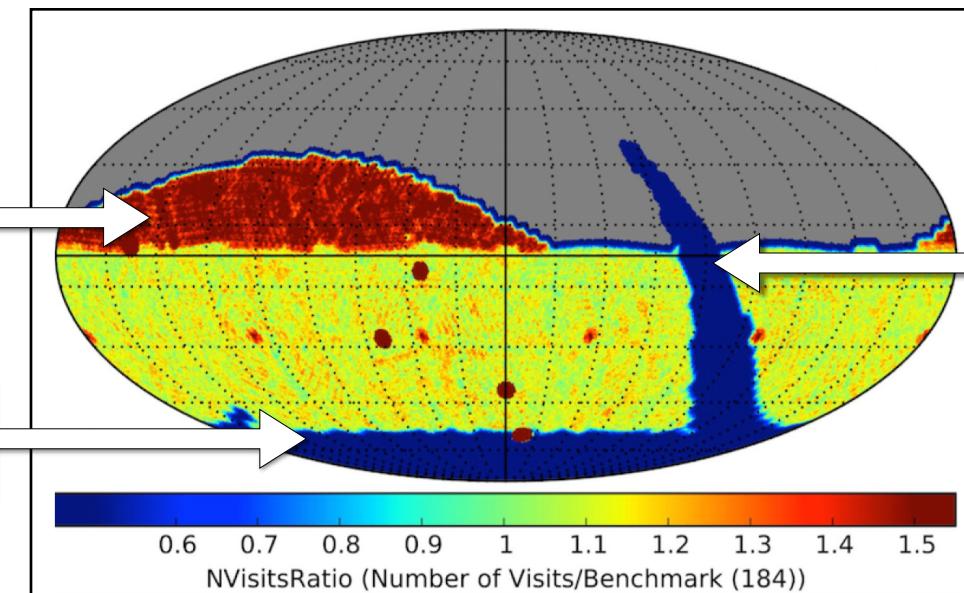
Deep Drilling Field

MS

Mini-Survey

North Ecliptic Spur
(solar system)

South Celestial Pole
(LMC, SMC)



Galactic Plane
(stars and planets)

DDF examples

DDF nightly stacks
to find high-z SN.

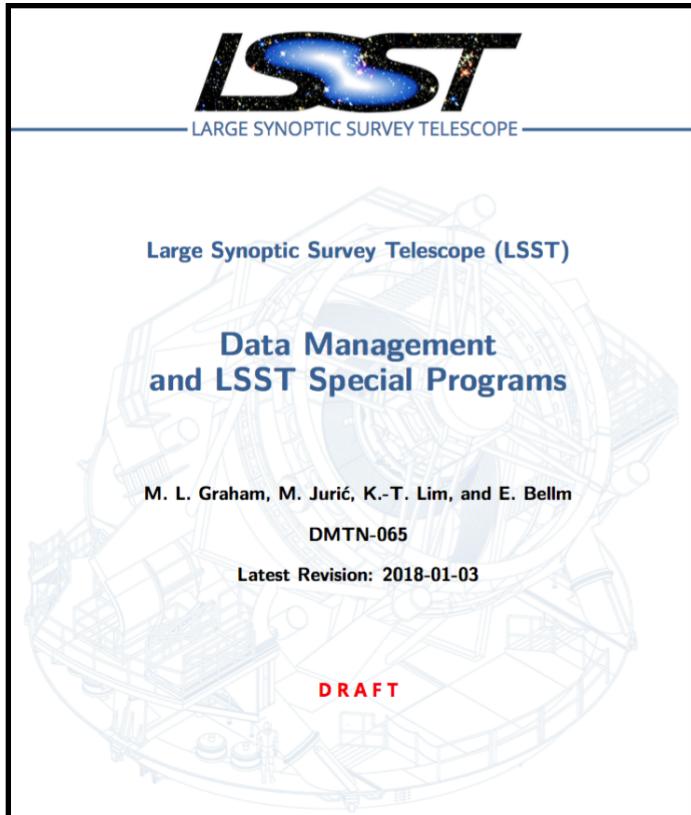
~10% of LSST observing time will be devoted to SPs that extend the areal coverage, depth, and/or sampling cadence to obtain improved coverage of interesting regions.



Data Management & Special Programs



LSST Data Management will



- *not* write unique algorithms for processing special programs data
- Allocate 10% of its computational resources for processing special programs data
- Incorporate special programs data into Alert and Data Release Productions, when scientifically beneficial
- Reconfigure pipelines to generate separate imaging and catalog products for special programs data, whenever possible
- Make the software stack source code available to the community



2018 Call for White Paper Proposals



First four extragalactic deep fields chosen.

First round of white paper proposals for LSST Special Programs.

planning continues through Science Collaborations, Observing Strategy White Paper

October 2018
Deadline for white paper proposals

June 2018
Call for white paper proposals

DMTN-065

Project Science Team defines criteria; Science Advisory Council reviews proposals and makes recommendations.

Latest set of Special Programs incorporated into the Operations Simulator.

planning continues

2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022



With first light expected in 2020, now is the time to undertake the final pre-commissioning optimisation of the LSST Observing Strategy

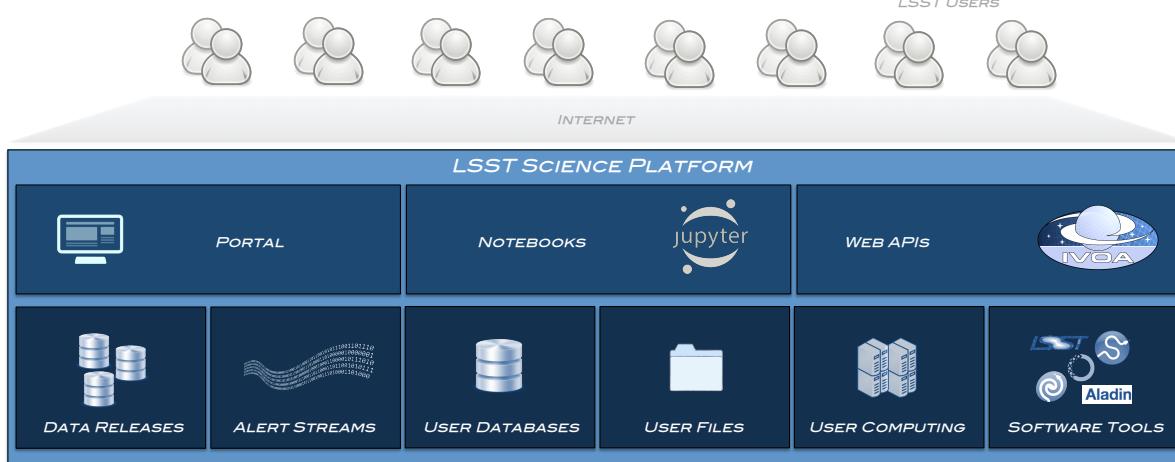
Science driven input for cadence properties such as per-bandpass imaging depth, sky and temporal coverage, etc requested from the science community



LSST Science Platform: Enabling Science



The *LSST Science Platform* is a web-based service available to the scientific community to access, visualize, subset, and perform next-to-the-data analysis of LSST data.



- Set of integrated web applications & services deployed at LSST Data Access Centers (DACs) including software tools and computational resources.
- Underlying DAC services exposed through three primary **user-facing “aspects”**; the Web Portal (novice), the JupyterLab (intermediate), and the Web APIs (expert and remote tools).
- Enables access to **Data Releases & Alert Streams**, and supports ‘next-to-the data’ analysis and **User-Generated** product creation using the computing resources available at the DAC through interfaces that utilize community-accepted standards.

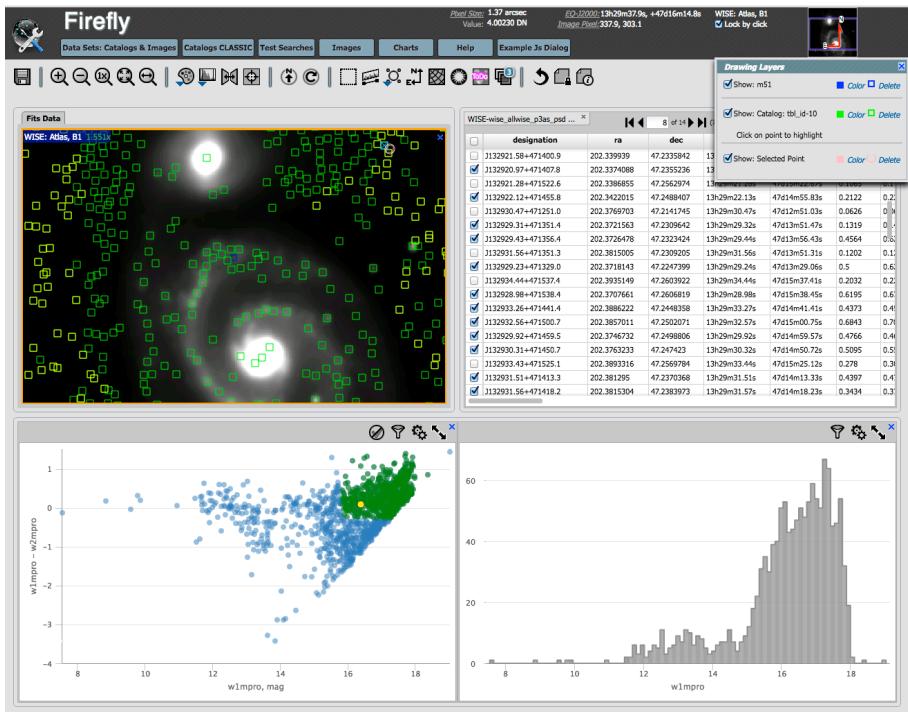
[LSE-319: LSST Science Platform Vision Document](#)



Web Portal Aspect: A window to the LSST Archive



Enable browsing & visualisation of available datasets, Cf. archives such as IRSA, SDSS



The Firefly Web Science User Interface (Wu et al, 2016; ADASS)

- Enables exploratory analysis of the LSST datasets guided by science cases.
- View LSST images
- Request data subsets (forms, SQL)
- Construct simple plots
- Added level of interactivity as compared with existing archives (what? - this is stated in LSE-319)
- Peta-scale capable RDBMS backend



JupyterLab Aspect: Delving Deeper



More sophisticated data selection, analysis & creation of *User-Generated* data products

The screenshot shows the JupyterLab interface. On the left is a sidebar with tabs for Files, Running, Commands, Tabs, and Cell Tools. The main area has two tabs open: 'simon.ipynb' and 'terminal/1'. The 'simon.ipynb' tab shows a code cell with Python code for loading data from LSST's Butler and performing analysis using LSST's afwGeom library. Below the code is a scatter plot of astronomical data points. The 'terminal/1' tab shows a terminal session with LSST command-line tools like eupsBuild and loadSST being used.

```
In [10]: butler = daf.persistence.Butler('singlechip.sample')
exp = butler.get('calexp', visit=410877, ccd=28, filter='r')

In [11]: import lsst.afw.geom as afw_geom
bbox = afw_geom.Box2I(afw_geom.Point2I(1024, 1024), afw_geom.Extent2I(512, 512))
sources = butler.get('src', visit=410877, ccd=28, filter='r')
overlay_masks(exp, bbox=bbox, sources=sources)
```

```
frossie@jld-lab: ~]$ ls /opt/lsst/software/stack/
build eupsBuild.log loadSST.cmd loadSST.zsh stack
eups eupsBuild.log loadSST.cmd loadSST.zsh
[frossie@jld-lab: ~]$ ./opt/lsst/software/stack/loadSST.bash
[frossie@jld-lab: ~]$ which eups
/opt/lsst/software/stack/eups/2.1.3/bin/eups
[frossie@jld-lab: ~]$
```

- Identical to working with Jupyter notebooks
- Computation and analysis at resources provided by LSST DACs
- Enables science discovery by ...
- .. **'bringing the analysis to the data'**, avoiding the need to download store and process large volumes of data.
- User environments with pre-installed libraries: AstroPy, LSST science pipelines, Anaconda, etc
- Users can install own tools

LSST Science Platform Demos: <http://ls.st/bgt>

Integrated environment:

- Data queries will be shareable across the Portal and the JupyterLab aspects
- Submit a complex SQL query in the Notebook, browse & visualize results in the Portal.

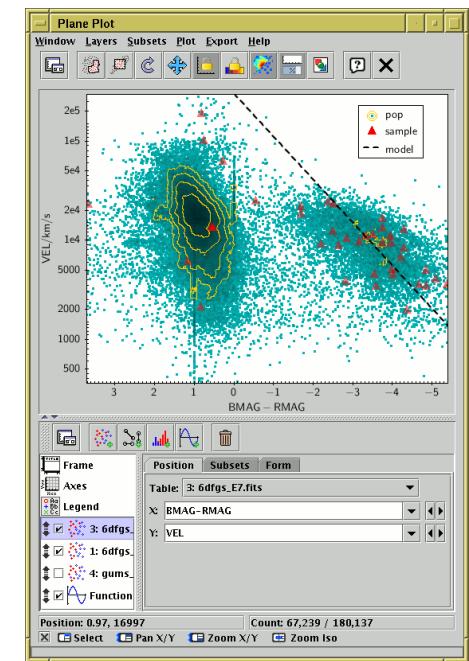
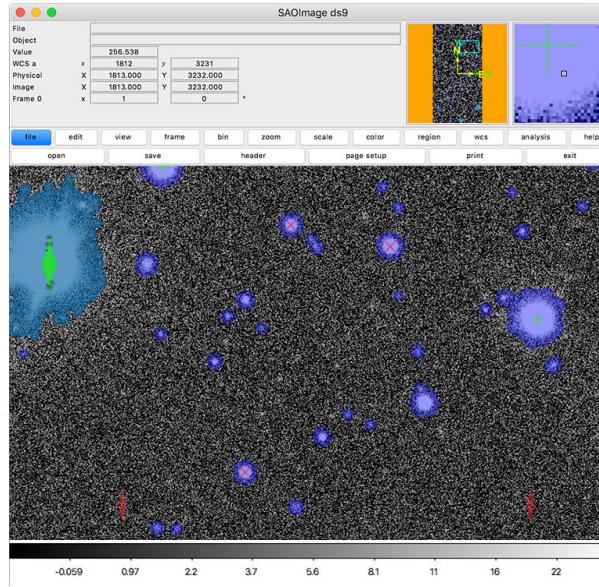


Web API Aspect: Integration with existing tools



The **Web API** aspect allows integration with familiar tools, enabling remote access to LSST DAC services via APIs using community-accepted formats and protocols.

- LSST Data exposed via VO interfaces enables the use of familiar tools such as TOPCAT, DS9, etc.
- VO Simple Cone Search, TAP (for catalogs) and SIAP (for images) will be supported.



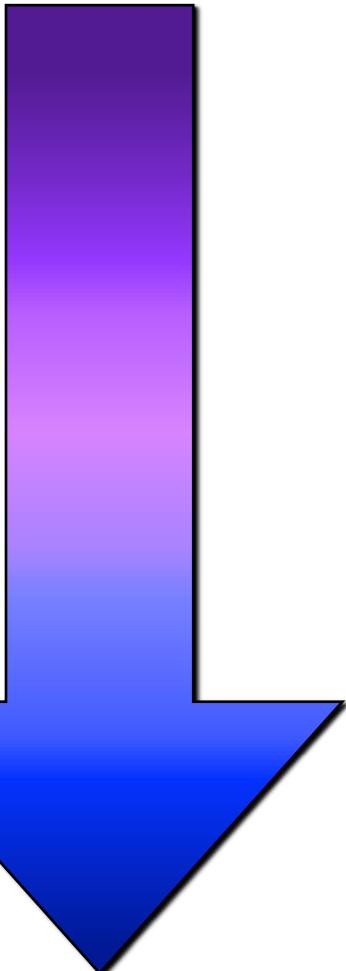
Build a query using the **Portal** query builder and access results remotely via JupyterLab or TOPCAT



Documentation



Requirements Documentation Flow-Down



“From Science Drivers to Reference Design”
Ivezic et al. (2008), arXiv:0805.2366

“LSST Science Requirements Document (SRD)”
[ls.st/lpm-17](#)

“LSST DM Subsystems Requirements (DMSR)”
[ls.st/lse-61](#)

“LSST DM Science Pipelines Design (DMSP)”
[ls.st/ldm-151](#)

“LSST Data Products Definitions Document (DPDD)”
[ls.st/lse-163](#)

“LSST Data Product Categories (DPC)”
[ls.st/lpm-291](#)

“LSST Science Platform Vision Document (LSP)”
[ls.st/lse-319](#)



Getting Involved : LSST Community Forum



Secure | <https://community.lsst.org>

LSST Community

Any opinions, statements (including statements about LSST and what it will deliver), or recommendations expressed on this forum are those of the author and do not necessarily reflect the views of the LSST Project.

Please take a moment to review our [community guidelines](#).

[all categories](#) ▾ [all tags](#) ▾ [Categories](#) Latest Bookmarks Unread My Posts [+ New Topic](#)

Category	Topics
Science Public discussions about LSST science. Data Q&A LSST2017 LSST2016 Milky Way (Open)	87
Support Community support venue for using the LSST software, services and data.	172
Data Management Discussions with LSST Data Management developers about LSST Stack development. DM Notifications DM Meetings DM IN2P3 DM RFD	674
Simulations Discussions related to LSST Simulations. Sims Announcements	48
Camera For discussions related to the LSST Camera.	1
Cross-System The Cross-Systems category hosts sub-categories for specific collaborations between LSST's subsystems. Camera-DM Visualization Calibration Operations	111
EPO	21

Latest

Welcome to community.lsst.org Meta	2	Aug '15
Stack release 15.0 - Status and discussion DM Notifications	0	20h
Deep drilling whitepapers Simulations opsim, special-surveys	9	2d
Notes from March 2018 SMWLV telecon Milky Way (Open)	0	7d
Reducing Suprime-cam Data Support	11	8d
Pipelines package docs hack day instructions Data Management doc-dev	6	8d
lsst::utils::Cache DM Notifications cache, dm-dev	2	9d
Webcast links for Precision Astrometry Conference	~	

<https://community.lsst.org>

community.lsst.org



References & Contacts



- LSST Data Products Definition Document [LSE-163](#)
- LSST Science Pipelines Design Document [LDM-151](#)
- LSST Science Platform Documents: [LSE-319](#)

- Ask questions at LSST Community
 - <https://community.lsst.org/c/sci/data>
 - Data pipelines or products related questions should go to the “Data Q&A” channel
 - moderated by the DM-SST (Subsystem Science Team)
 - approved answers will be generated and posted in a timely fashion

- Points of contact
 - Transient & Variable Stars Science Collaboration points of contact:
 - Melissa Graham & Eric Bellm
 - DM Subsystem Scientist: Leanne Guy

- Many thanks to Melissa Graham, for much of this material!