

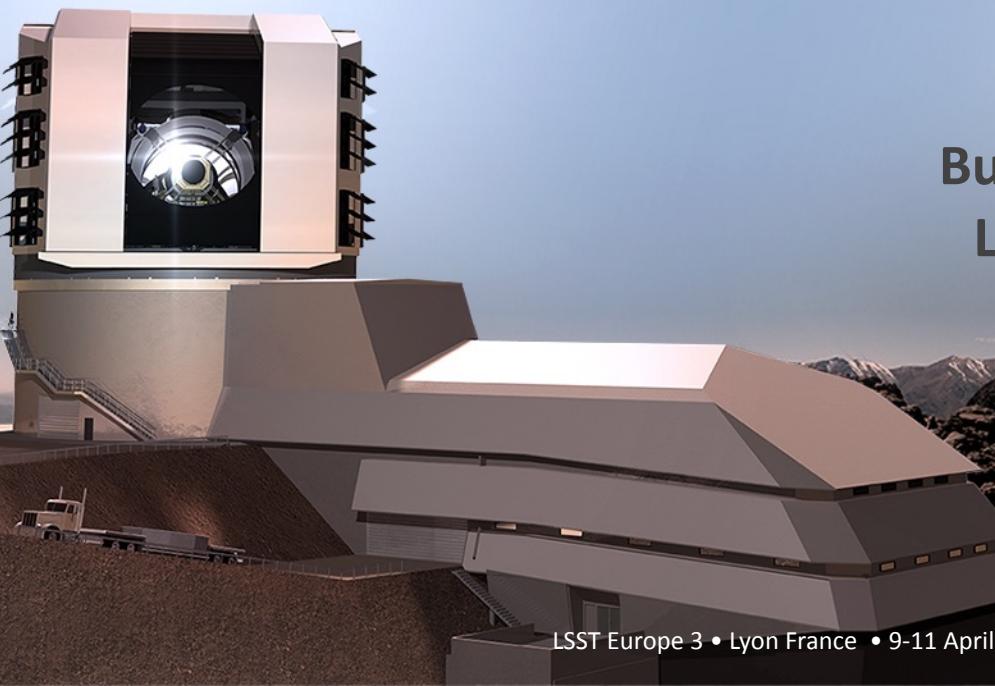
LSST Data Products

Leanne Guy

Data Management Scientist

AURA/LSST,

Tucson, AZ, USA



LSST @ Europe 3
Building Science Collaborations
Lyon, France, 11-15 June 2018



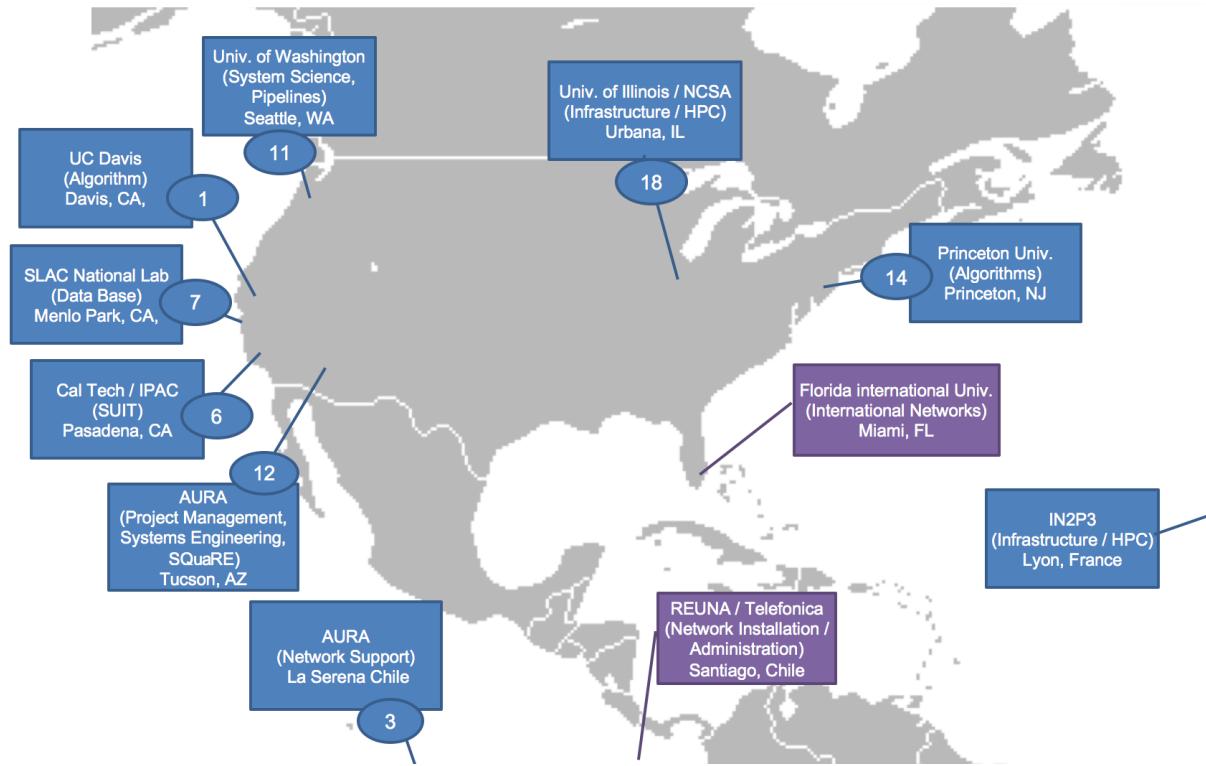
Introduction



- LSST Data Management Scientist
 - Taking over from Mario Jurić
 - Started 14 May 2018
- Based at AURA, LSST HQ Site in Tucson, AZ, USA The AURA logo consists of the word "AURA" in blue capital letters, with a black swoosh graphic above it that ends in a blue circle.
- Previously worked on the ESA Gaia mission for over 10 years, on the ATLAS experiment and LHC Computing @ CERN



LSST Data Management



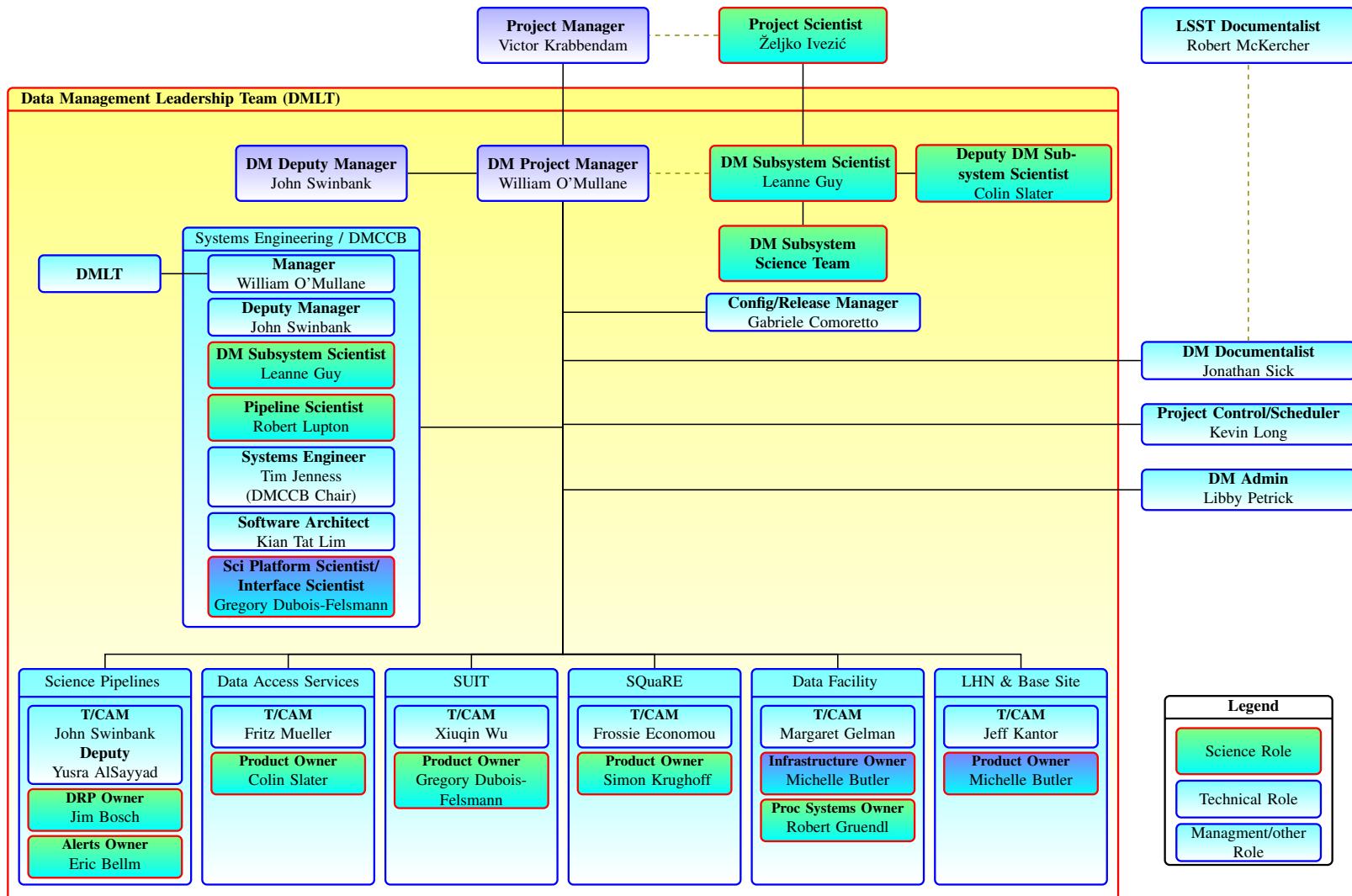
Mission Statement:

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

The ultimate deliverable of LSST is not the telescope or camera but the final science-ready data products.

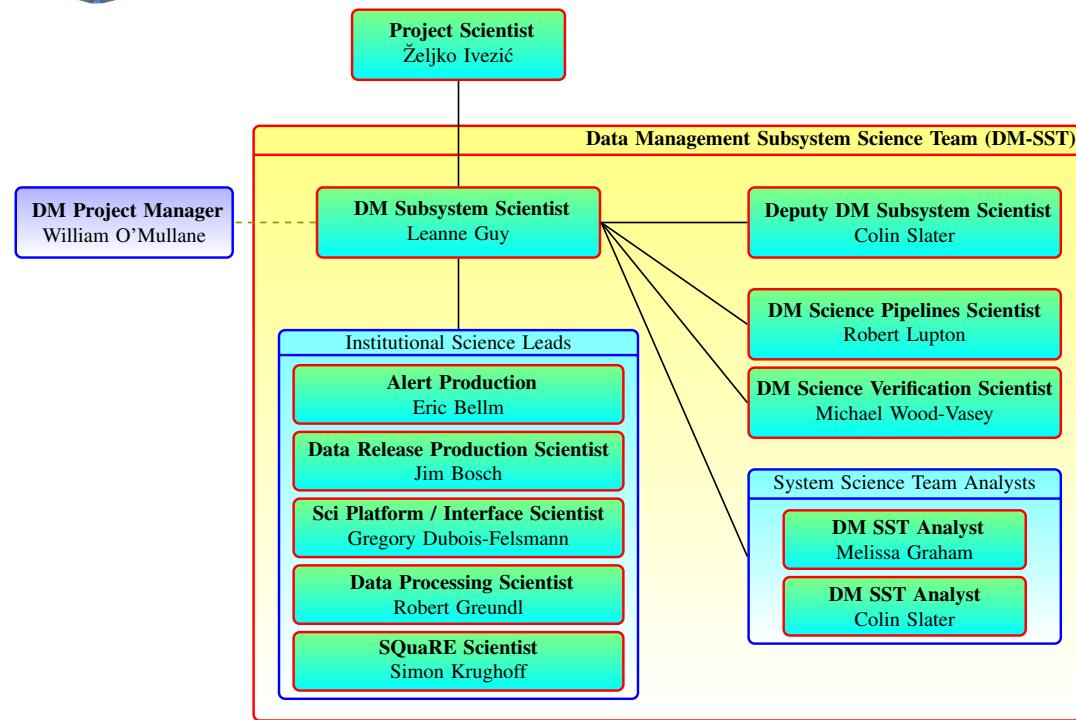


Data Management Organization





DM Subsystem Science



Mandate: to ensure that the DM system's pipelines and data products provide solutions that meet the overall LSST scientific goals

- Work with the science community to understand their needs and how they will be met by the DM system
- Identify scientific opportunities, identify risks, evaluate impact and initiate change
- Lead the Science Validation of the deliverables of the LSST DM System.



LSST Data Management



*Data Release Data Products
via Annual Data Releases*



*11 Data Releases in 10 years
Final database catalog: 15 PB*

*20TB raw data/night
(with calibration exposures)*

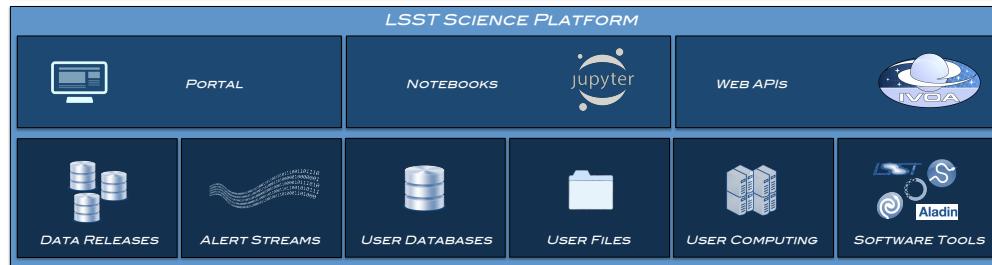


*Prompt Data Products
via nightly Alert Streams*

A binary code stream consisting of a series of 0s and 1s, representing the data in the alert streams.

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

LSST Science Platform



**current & penultimate
Data Releases**

**alerts database &
mini-broker**

Data access via Data Access Centres & Services



LSST Data Product Categories



Prompt

Formerly “Level 1 data products”

Real Time Difference Image Analysis (DIA)

A stream of ~10 million time-domain events per night (Alerts), detected characterized and transmitted to event distribution networks with 60 seconds of shutter close.

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

Data Release

Formerly “Level 2 data products”

Reduced single-epoch & deep co-added images, reprocessed DIA products

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

User Generated

Formerly “Level 3 data products”

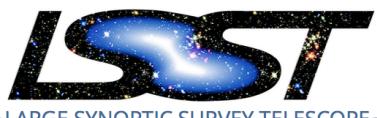
User-produced added-value data products

Deep KBO/NEO, variable star classifications, shear maps, etc ...

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



LSST Data Products Definition Document



Large Synoptic Survey Telescope (LSST)

Data Products Definition Document

M. Jurić, T. Axelrod, A.C. Becker, J. Becla, E. Bellm, J.F. Bosch,
D. Ciardi, A.J. Connolly, G.P. Dubois-Felsmann, F. Economou,
M. Freeman, M. Gelman, R. Gill, M. Graham, Ž. Ivezić, T. Jerness,
J. Kantor, K.S. Krughoff, K-T Lim, R.H. Lupton, F. Mueller,
D. Nidever, W. O'Mullane, M. Patterson, D. Petravick, D. Shaw,
C. Slater, M. Strauss, J. Swinbank, J.A. Tyson, M. Wood-Vasey, and
X. Wu

LSE-163

Latest Revision: 2018-02-09

This LSST document has been approved as a Content-Controlled Document. Its contents are subject to configuration control and may not be changed, altered, or their provisions waived without prior approval. If this document is changed or superseded, the new document will retain the Handle designation shown above. The control is on the most recent digital document with this Handle in the LSST digital archive and not printed versions.

A high-level description data products and processing services to be delivered by LSST

[LSE-151: Science Pipelines Design Document](#) & [LSE-163: Data Products Definition Document](#)

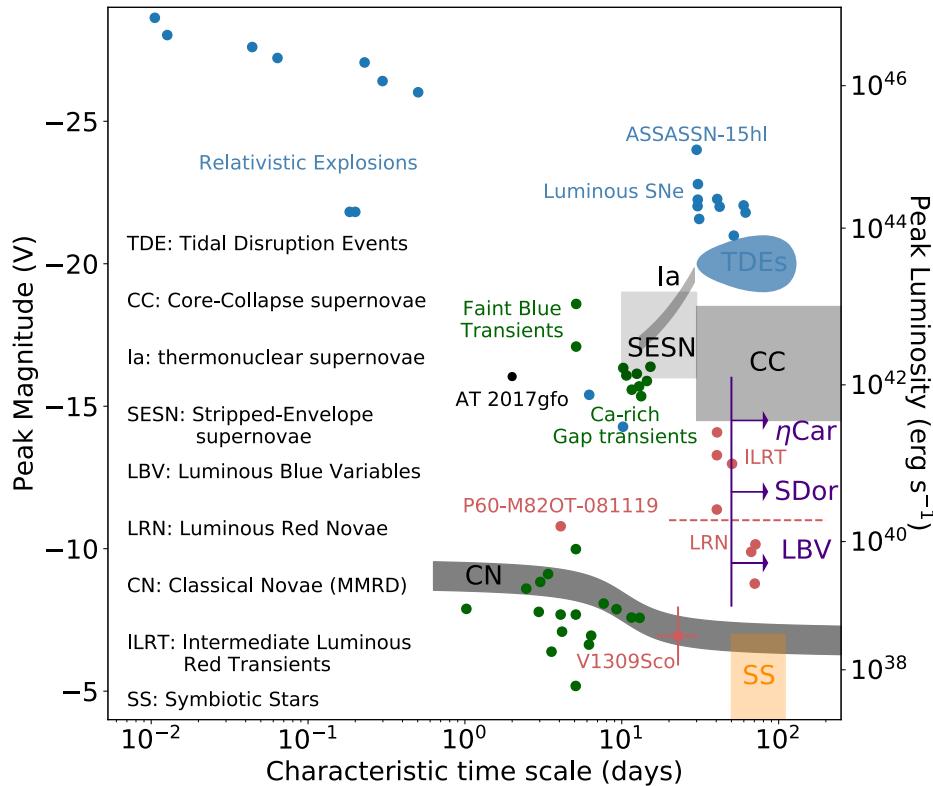
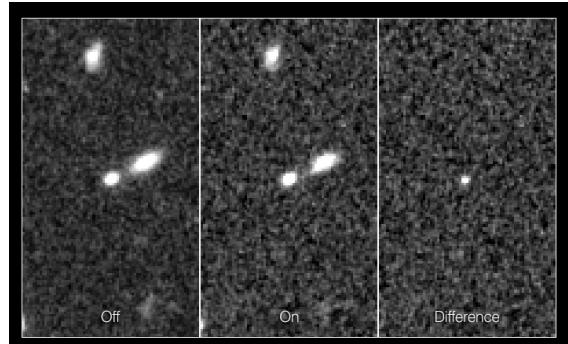


Prompt Data Products



Enabling discovery and rapid follow-up of time domain events

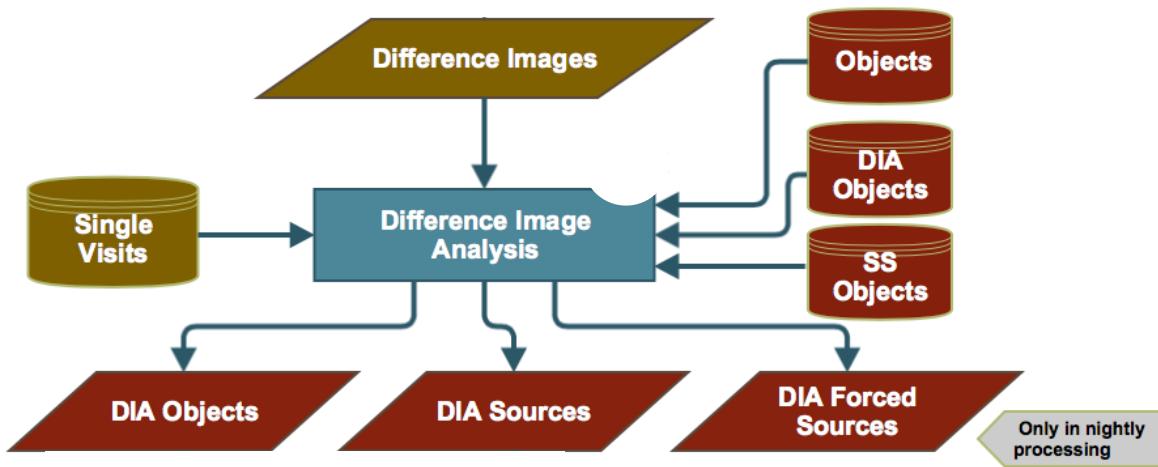
- Transient science
 - Nova, supernova, GRBs
 - Source characterization
 - Instantaneous discovery
- Nearby Solar System Objects
 - NEOs, PHAs, etc



LSST: from Science Drivers to Reference Design and Anticipated Data Products, Ivezic et. al.



Difference Image Analysis (DIA)



Prompt data products
are produced by the
DIA pipeline and
distributed via Alert
Production

Within 60s

- Single-visit images processed
 - Differencing image analysis performed
 - DIA Sources detection and associated
 - Alerts generated & issued to the stream
(one per new DIASource)
 - ***Prompt*** catalogs in the US DAC updated

Within ~24h

- Forced photometry in past 30 days of images for all new DIAObjects
 - Moving Object Pipeline (MOPS) runs on DIASource catalog
 - Raw and Processed images available in the US DAC



Prompt Catalogs



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

DIASource

- Signal to noise ratio of the detection
- Coordinates and association with DIAObject or SSOObject
- Time of mid-exposure at location on CCD
- Flux in n the difference and visit images (PSF, aperture)
- Shape parameters (trails, dipoles, FWHM, extendedness)
- Parent/child de-blending flags

DIAObject

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

Non-exhaustive



Alert Packet Contents

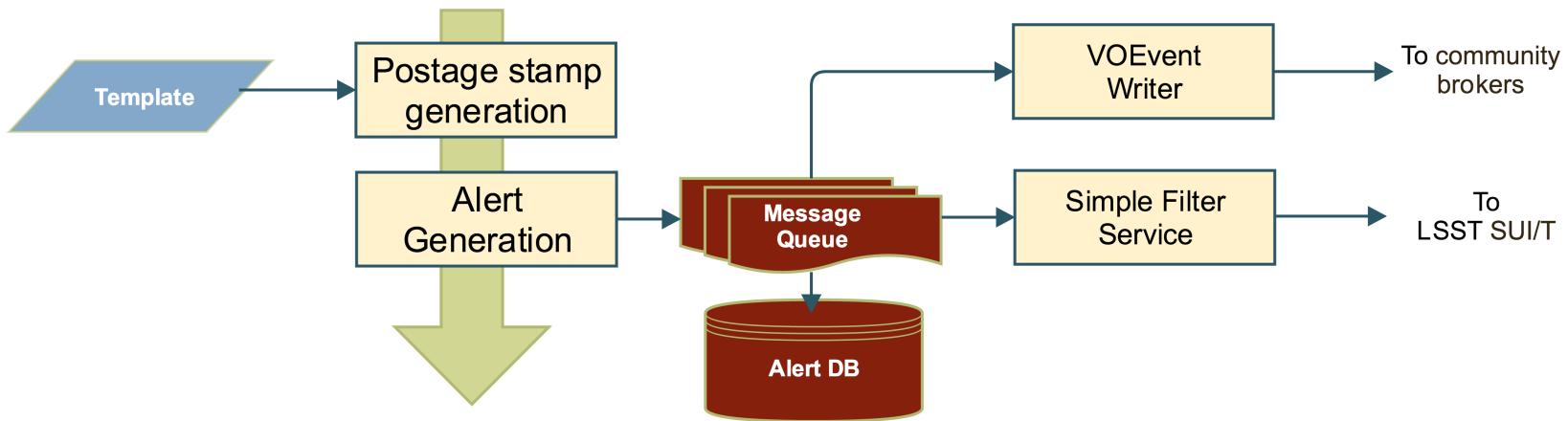


Each alert (a VO event packet) will contain at least the following:

- *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.
- *Level 1 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



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 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



Data Release Data Products



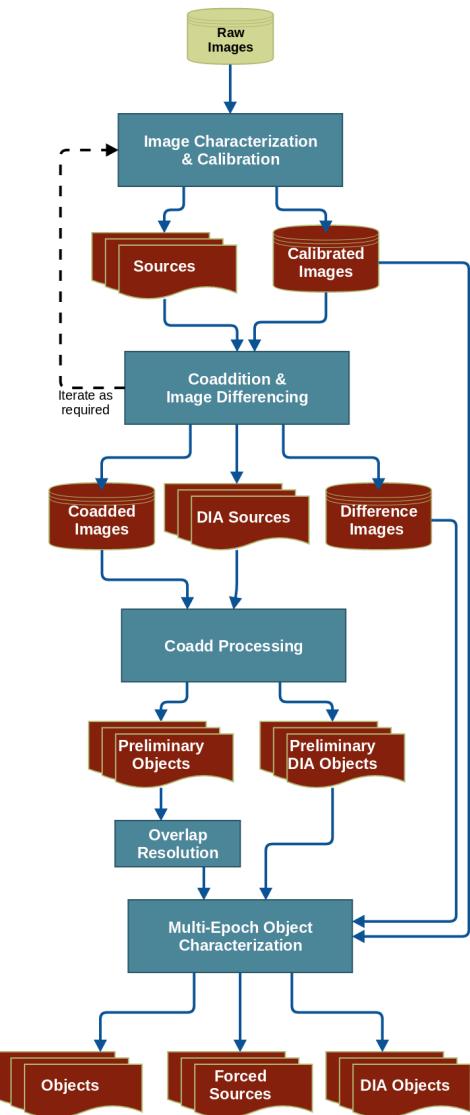
Enable static sky science (eg., studies of galaxy evolution, or weak lensing), and time-domain science that is not time sensitive (eg. statistical investigations of variability).



Data Release Data Products



- Well calibrated, consistently processed catalogs and images
 - Catalogs of objects, detections, detections in difference images, etc.
 - Combine information from many exposures
- Made available via an annual ***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 (LSP)
 - Remote access APIs, VO Protocols (TAP)
- Projected catalog sizes are:
 - 18 billion objects (DR1) -> 37 billion (DR11)
 - 750 billion observations (DR1) -> 30 trillion (DR11)
 - Few PB (DR1) -> 70 PB (DR11)





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
- Source (and ForcedSource)
 - locations and times of all detections
 - aperture and point source flux; PSFmoments
 - 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 & non-periodic features
 - Proper motions and parallax



LSST Database Browser



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



LSST Database Schema Browser alpha

Schema versions available for browsing: [baseline](#) | DC3a | PT1_1 | PT1_2 | ImSim | S12_sdss | S12_lsstsim (underlined showed)

User defined functions documentation: version 0.1, version 0.2, version 0.3 (default on lsst10)

Table List		Details for table <i>DiSource</i>					
ApertureBins		Table to store 'difference image sources'; - sources detected at SNR >=5 on difference images.					
CcdVisit							
CcdVisitMetadata							
DiaForcedSource							
DiaObject							
DiaObject_To_Object_Match							
DiSource							
ForcedSource							
LeapSeconds							
Object							
Object_APMean							
Object_Extra							
Object_NonPeriodic							
Object_Periodic							
prv_cnf_Node							
prv_cnf_Pipeline							
prv_cnf_Pipeline_Tasks							
prv_cnf_Task							
prv_cnf_Task_Columns							
prv_cnf_Task_Files							
prv_cnf_Task_KVParams							
prv_DataBlock							
prv_Node							
prv_Pipeline							
prv_ProcHistory							
prv_RowIdToDataBlock							
prv_Task							
prv_TaskExecution							
prv_TaskExecutionToInputDataBlock							
prv_TaskExecutionToOutputDataBlock							
RawAmpExposure							
RawAmpExposureMetadata							
RawCcdExposure							
RawCcdExposureMetadata							
RawExposure							
sdqa_ImageStatus							
sdqa_Metric							
sdqa_Rating_CcdVisit							
sdqa_Rating_ForAmpVisit							
sdqa_Threshold							
Source							
Source_APMean							

Table List

Details for table *DiSource*

Table to store 'difference image sources'; - sources detected at SNR >=5 on difference images.

name	type	not null	unit	ucd	description
diaSourceId	BIGINT	y		meta.id;obs.image	Unique id.
ccdVisitId	BIGINT	y		meta.id;obs.image	Id of the ccdVisit where this diaSource was measured. Note that we are allowing a diaSource to belong to multiple amplifiers, but it may not span multiple ccds.
diaObjectId	BIGINT			meta.id;src	Id of the diaObject this source was associated with, if any. If not, it is set to NULL (each diaSource will be associated with either a diaObject or ssObject).
ssObjectId	BIGINT			meta.id;src	Id of the ssObject this source was associated with, if any. If not, it is set to NULL (each diaSource will be associated with either a diaObject or ssObject).
parentDiaSourceId	BIGINT			meta.id;src	Id of the parent diaSource this diaSource has been deblended from, if any.
prv_procOrder	INT	y			Position of this diaSource in the processing order relative to other diaSources within a given diaObjectId or ssObjectId.
ssObjectReassocTime	DATETIME				Time when this diaSource was reassigned from diaObject to ssObject (if such reassociation happens, otherwise NULL).
midPointTai	DOUBLE	y	d	time.epoch	Effective mid-exposure time for this diaSource.
ra	DOUBLE	y	deg	pos.eq.ra	RA-coordinate of the center of this diaSource.
raSigma	FLOAT	y	deg	stat.error;pos.eq.ra	Uncertainty of ra.
decl	DOUBLE	y	deg	pos.eq.dec	Decl-coordinate of the center of this diaSource.
declSigma	FLOAT	y	deg	stat.error;pos.eq.dec	Uncertainty of decl.
ra_decl_Cov	FLOAT	y	deg^2	stat.covariance	Covariance between ra and decl.
x	FLOAT	y	pixel	pos.cartesian.x	x position computed by a centroiding algorithm.
z	FLOAT				



User Generated Data Products



User-Generated products: 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.



LSST Special Programs



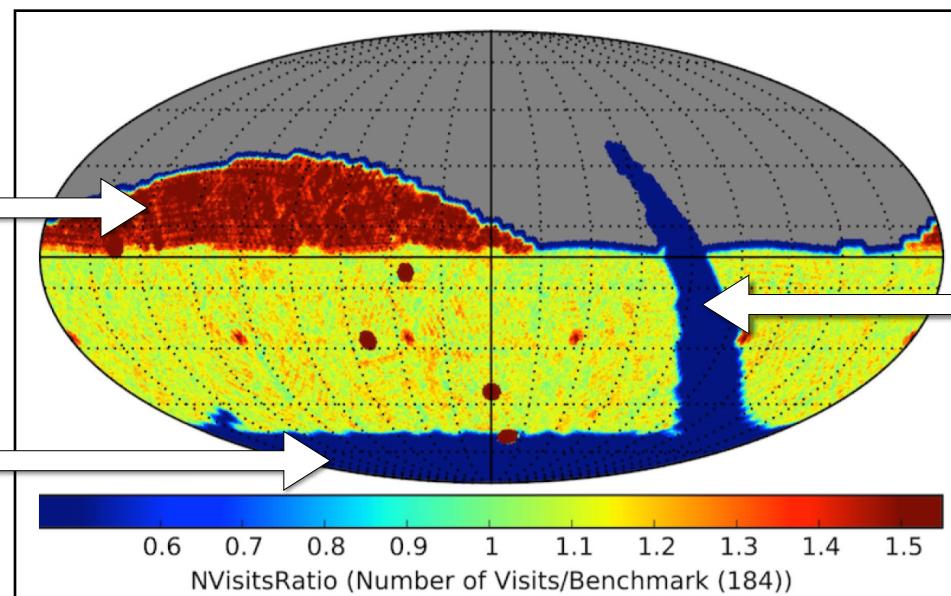
WFD
Wide-Fast-Deep

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 special programs 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 re-processing special programs data
- Allocate 10% of its resources for processing special programs data
- Incorporate special programs data into Alert and Data Release Production, when scientifically beneficial
- Reconfigure pipelines to generate separate imaging and catalog products for special programs, where possible
- Make the software stack source code available to the community



LSST Data Products Overview



CATEGORY
CADENCE

IMAGES

CATALOGS

ALERTS

NIGHTLY

PROMPT
PRODUCTS

ANNUALLY

DATA RELEASE
PRODUCTS

Raw Science Images
Calibrated Science Images
Difference Science Images

Co-add Science Images
Calibration Images
RGB JPEG Images

PROMPT PRODUCTS DB

Catalog of DIA sources
Catalog of DIA Objects
Orbit Catalog
Precovery photometry (30 days)

Catalog of sources from
calibrated science images
Object catalog based on best
measured properties

Reprocessed DIA sources

Transient Alert (<60s)

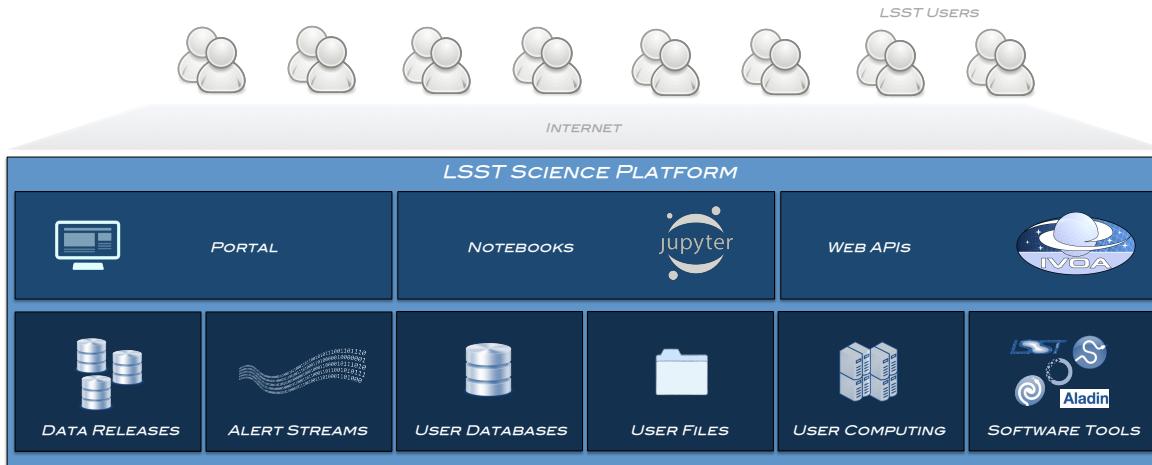
Reprocessed DIASources
Alert statistics and
summaries based on
reprocessing



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.



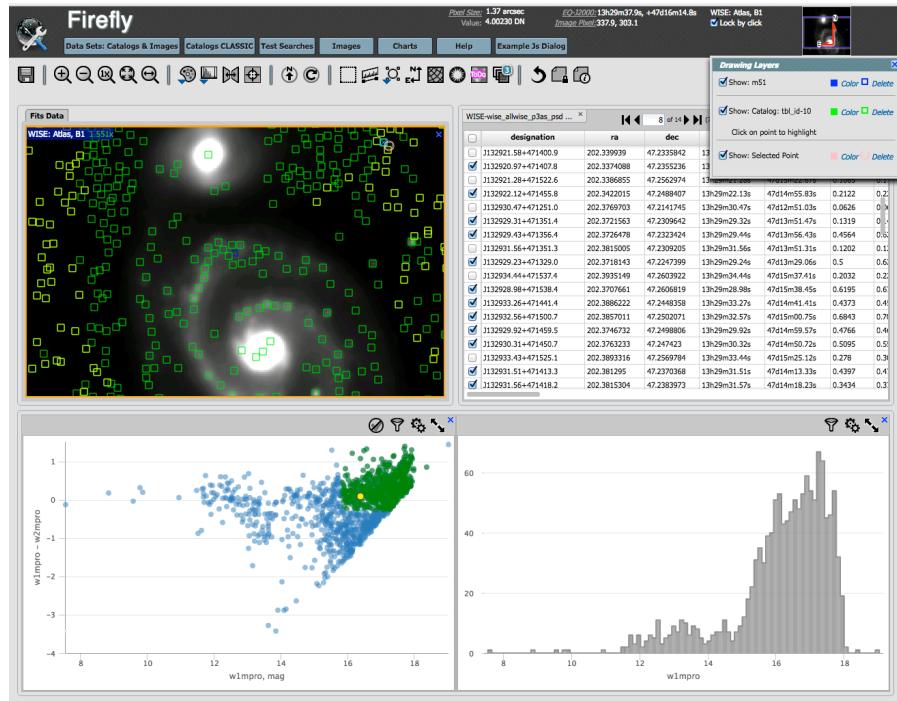
- Set of integrated web applications & services deployed at LSST Data Access Centers (DACs) including software tools and computational resources.
- Underlying DAC services exposed through three primary user-facing “aspects”; the Web Portal (novice), the JupyterLab (intermediate), and the Web APIs (expert 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 utilise community-accepted standards.



Portal Aspect: Delving Deeper



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



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
- Peta-scale capable RDBMS backend



JupyterLab Aspect: Delving deeper



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

The screenshot shows the LSST Science Platform JupyterLab interface. On the left, a sidebar lists 'TERMINAL SESSIONS' (terminals/1) and 'KERNEL SESSIONS' (simon.ipynb). The main area has two code cells and a plot. Cell 10:

```
butler = daf_persistence.Butler('singlechip_sample')
exp = butler.get('calexp', visit=410877, ccd=28, filter='r')
```

 Cell 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)
```

 Below the cells is a scatter plot of astronomical objects. A terminal session at the bottom shows command-line steps to build software and run eups.

```
[frossie@jld-lab ~]
[frossie@jld-lab-w20172-frossie-1933621 ~]$ ls /opt/lsst/software/stack/
_build
_eupsbuild.log
loadLSST.csh
loadLSST.zsh
stack
eups
[frossie@jld-lab-w20172-frossie-1933621 ~]$ . /opt/lsst/software/stack/loadLSST.bash
[frossie@jld-lab-w20172-frossie-1933621 ~]$ which eups
/opt/lsst/software/stack/eups/2.1.3/bin/eups
[frossie@jld-lab-w20172-frossie-1933621 ~]$
```

- 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

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.

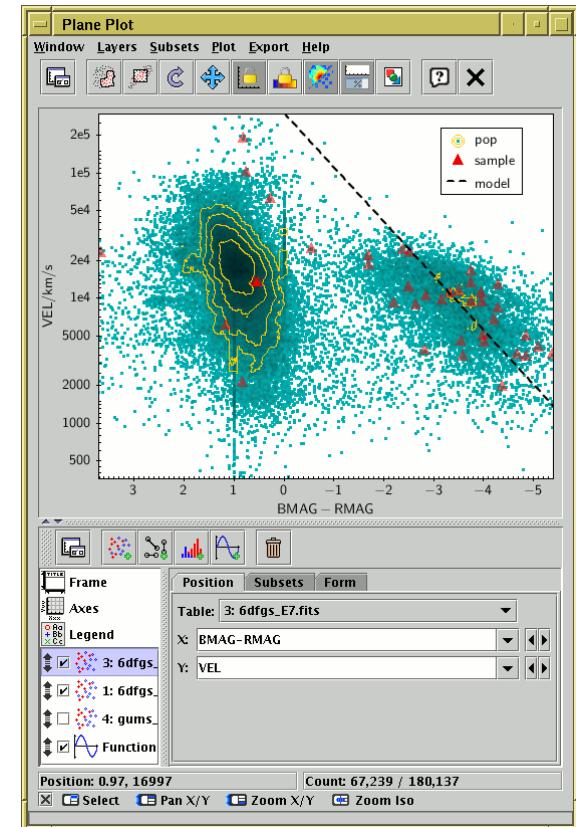
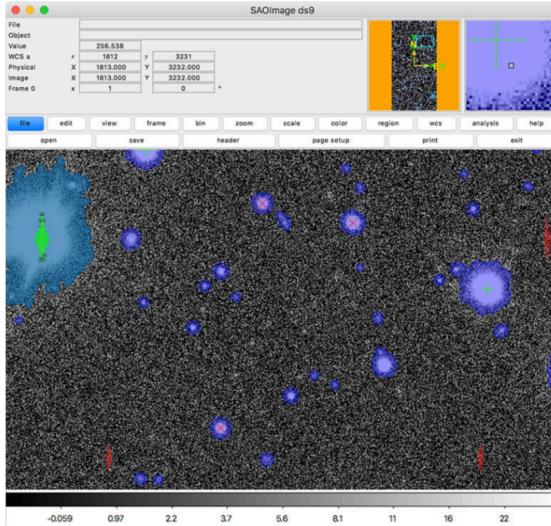


Web API Aspect: Integration with existing tools



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 and TAP (for catalogs) and SIAP (for images) will be supported.



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



Data Release Availability & Retention



- Over 10 years of operation, LSST will produce 11 data releases, two for the first year of operations, one every subsequent year
 - Few PB (DR1) to ~70 PB for DR11
- Costly to keep all DRs loaded and accessible at all times.
 - Only the contents of the most recent and penultimate data releases will be kept on fast storage and with catalogs loaded into the database.
 - Most users will work with the most recent releases
- Older releases will be archived on mass storage
 - Queries against archived releases *may* not be possible.
 - Available as bulk downloads
- All raw data used to generate any public data product (raw exposures, calibration frames, telemetry, configuration metadata, etc.) will be kept and made available



Requirements Documentation Flow-Down



Requirements flow down

LSST Science Requirements Document (SRD) [LPM-17](#)
LSST DM Subsystems Requirements (DMSR) [LSE-61](#)
LSST DM Science Pipelines Design (DMSP) [LDM-151](#)

LSST Data Products Definitions Document (DPDD) [LSE-163](#)
LSST Data Product Categories (DPC) [LPM-291](#)
LSST Science Platform Vision Document (LSP) [LSE-319](#)



LSST Data Management Communications



Every Science Collaboration has a designated liaison '*Point of Contact*' in the DM-SST

Science Collaboration	SC Point of Contact	DM-SST Point of Contact
Solar System	Meg Schwamb, David Trilling	Mario Jurić
Galaxies	Michael Cooper, Brant Robertson	Robert Lupton
Stars, Milky Way & Local Volume	John Bochanski, John Gizis, Nitya Jacob Kallivayalil	Colin Slater (<i>DM-SST Deputy</i>)
Dark Energy	Phil Marshall, Eric Gawiser	Robert Lupton
AGN	Niel Brandt	Željko Ivezić
Transient & Variable Stars	Federica Bianco, Rachel Street	Melissa Graham, Eric Bellm
Strong Lensing	Charles Keeton	Jim Bosch
Informatics & Statistics	Tom Loredo, Chad Schafer	Leanne Guy

LSST Science Collaborations



DM Community & Communications



Every Science Collaboration has a designated liaison '*Point of Contact*' in the DM-SST

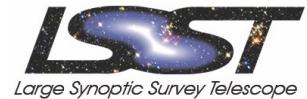
Science Collaboration	SC Point of Contact	DM-SST Point of Contact
Solar System	Meg Schwamb, David Trilling	Mario Jurić *
Galaxies	Michael Cooper, Brant Robertson	Robert Lupton
Stars, Milky Way & Local Volume	John Bochanski, John Gizis, Nitya Jacob Kallivayalil	Colin Slater (<i>DM-SST Deputy</i>)
Dark Energy	Phil Marshall, Eric Gawiser	Robert Lupton
AGN	Niel Brandt	Željko Ivezić
Transient & Variable Stars	Federica Bianco, Rachel Street	Melissa Graham, Eric Bellm
Strong Lensing	Charles Keeton	Jim Bosch
Informatics & Statistics	Tom Loredo, Chad Schafer	Leanne Guy

* Monday only

Present at LSST@Europe 3

LSST Science Collaborations

DM Community & Communications



LSST Community

community.lsst.org



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.

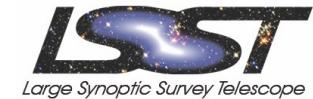
Please take a moment to review our [community guidelines](#).

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

Category	Topics
Science Public discussions about LSST science. Data Q&A Survey Strategy LSST2018 LSST2017 LSST2016 Milky Way (Open)	92
Science Collaborations Science discussions for members of LSST Science Collaborations. PST Telecons	33
Announcements Official LSST announcements.	19
Support Community support venue for using the LSST software, services and data. LSST Science Platform	186
Data Management Discussions with LSST Data Management developers about LSST Stack development. DM Notifications DM Team DM System Specifications DM Meetings DM IN2P3 DM RFD	696

Latest	
Welcome to community.lsst.org Meta	2 Aug '15
Producing and using DIAObjects for analysing DESC/DC2 transient simulation Support imagedifference	4 5h
New geom package replaces much of lsst.afw.geom DM Notifications	0 2d
OpSim v4 - community access? Simulations	0 3d
Science Pipeline release 16.0 - Status and discussion DM Notifications stack-releases	0 3d
Solar System Science Collaboration (SSSC) May 2018 Update	0

DM Community & Communications



 LSST Community

community.lsst.org



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](#).

[Science](#) ▶ [Data Q&A](#) ▶ [all tags](#) ▶ [Latest](#) Bookmarks My Posts [+ New Topic](#) [O](#)

Topic	Users	Replies	Views	Activity
¶ About the Data Q&A category Ask science questions about LSST data products and the Science Platform. Questions will be monitored by the Data Management team and answered in a timely fashion. As often as possible, the answers will point to permanent... read more		0	141	Oct '17
<input checked="" type="checkbox"/> Data Model for variable sources at time of Data Releases? agn		4	192	Feb 27
<input checked="" type="checkbox"/> Availability of residual images?		1	116	Feb 23
<input checked="" type="checkbox"/> LSST Filters versus SDSS		5	208	Nov '17
CCD Nonlinearity Near Saturation		12	288	Oct '17
How will the difference imaging pipeline respond to marginally resolved sources? difference-imaging		2	216	Oct '17

There are no more Data Q&A topics. Why not create a topic?



LSST GitHub



All Data Management code and documentation is openly available on GitHub

Search or jump to... Pull requests Issues Marketplace Explore

LSST Data Management Team
This organization contains repositories of internal interest to LSST DM.
<https://dm.lsst.org/>

Repositories 285 People 94 Teams 7 Projects 0

Search repositories... Type: Sources Language: All New

270 results for source repositories Clear filter

dmtn-060
Distributed Data Management with Rucio 0.1
TeX Updated 25 seconds ago

ap_association
Repository for holding code related to Alert Production difference source association
Python Updated 21 minutes ago

dm-demo-notebooks
A place to aggregate notebooks that demonstrate functionality in data management software.
Updated an hour ago

images
images for including in latex documents (or anywhere else). The idea is

Top languages
Python TeX Makefile Shell C++

People 94 >

