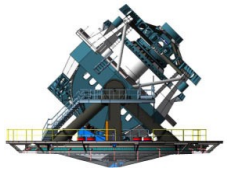
LSST Data Products

Leanne Guy
Data Management Scientist
AURA/LSST

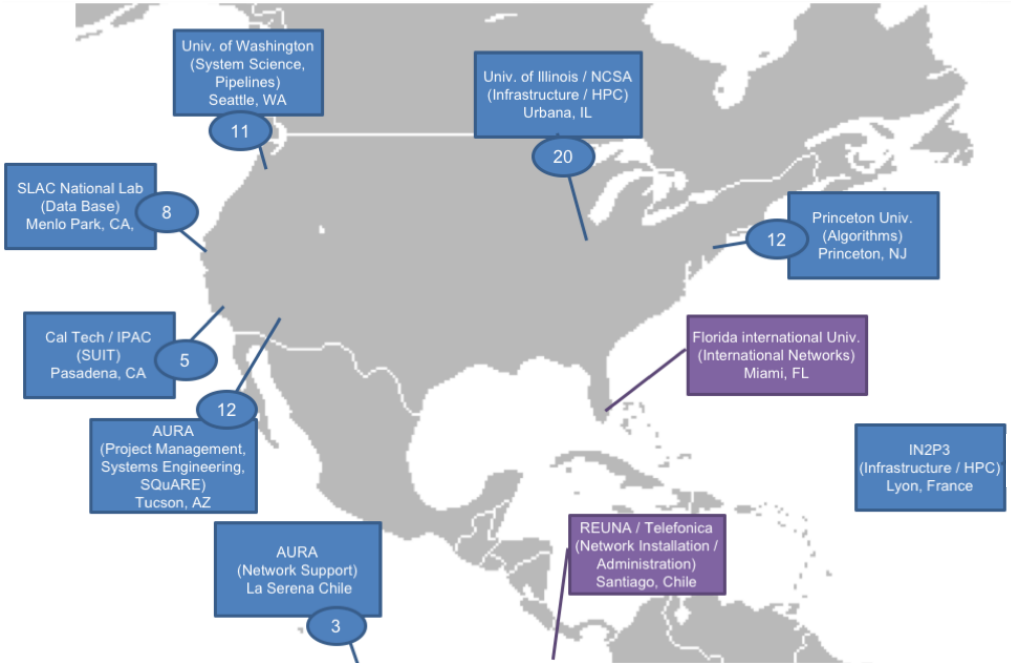
LSST @ Europe 3
Building Science Collaborations

Introduction

- LSST Data Management Scientist
  - Taking over from Mario Jurić
  - Started 14 May 2018
- Based at AURA, LSST HQ Site in Tucson, AZ, USA
- Previously worked on the ESA Gaia mission for over 10 years, on the ATLAS experiment and LHC Computing @ CERN



LSST Data Management



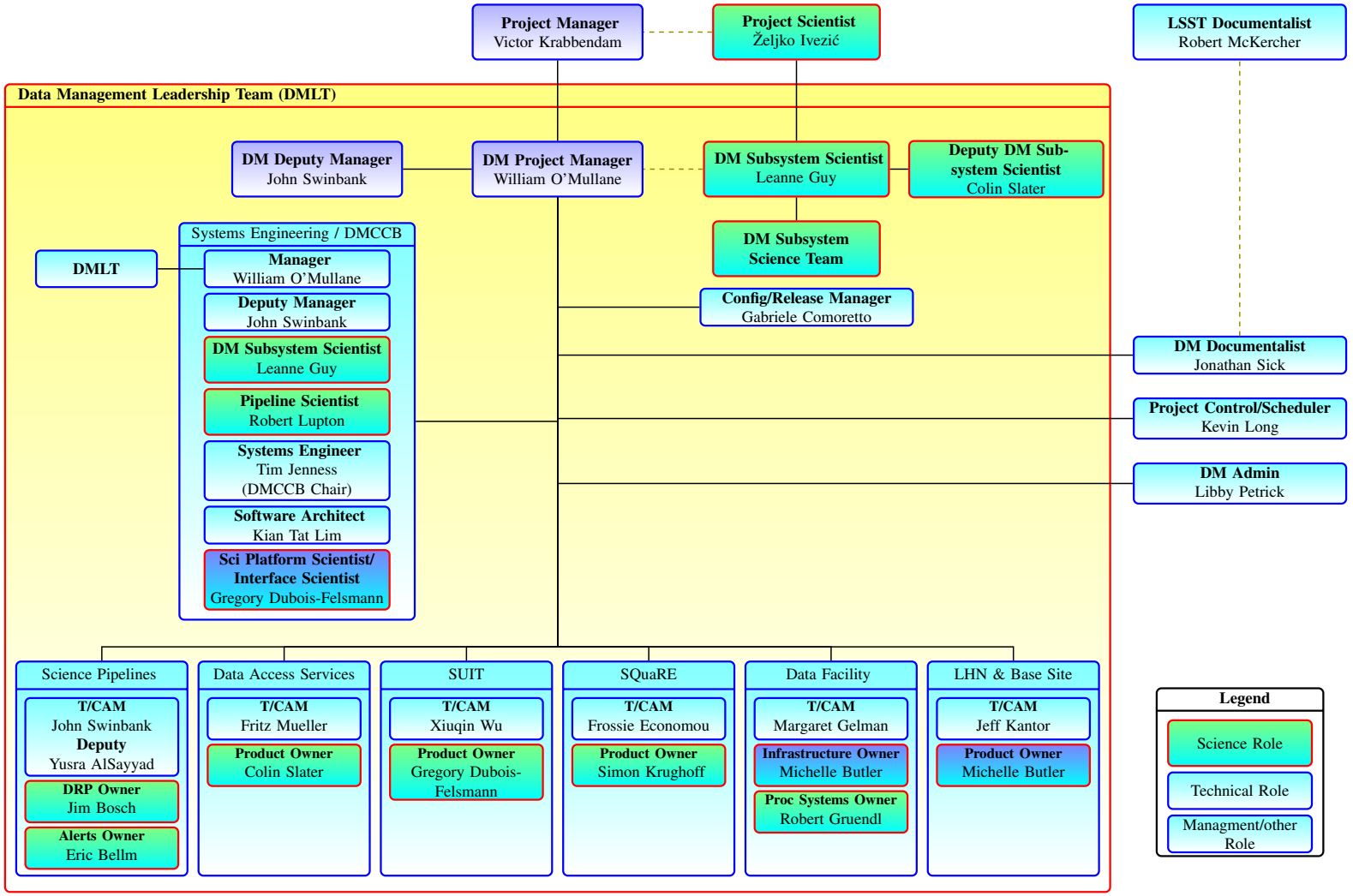
Mission Statement:

"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 LSST deliverable is not the telescope nor the camera, but the final science-ready data products.

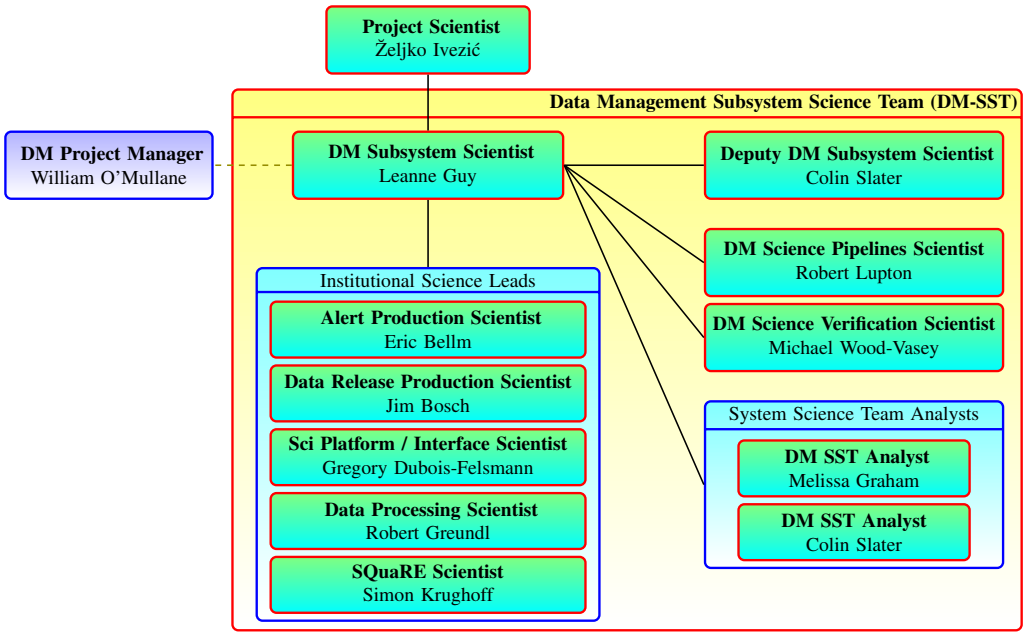
LSST Data Management Organization & Management (LDM-294)

Data Management Organization



LSST Data Management Organization & Management (LDM-294)

DM Subsystem Science



Mandate

" Ensure that the Data Management (DM) initiatives (pipelines and products) provide solutions that meet the overall LSST scientific goals."

Responsibilities

- Work with the science community to understand their needs and how they will be met by the DM System
- Identify and develop new scientific opportunities for the DM System, identify risks and coordinate change
- Lead the Science Validation of the deliverables of the LSST DM System

LSST Data Management Organization & Management (LDM-294)

LSST Data Management

Data Release Data Products
via Annual Data Releases

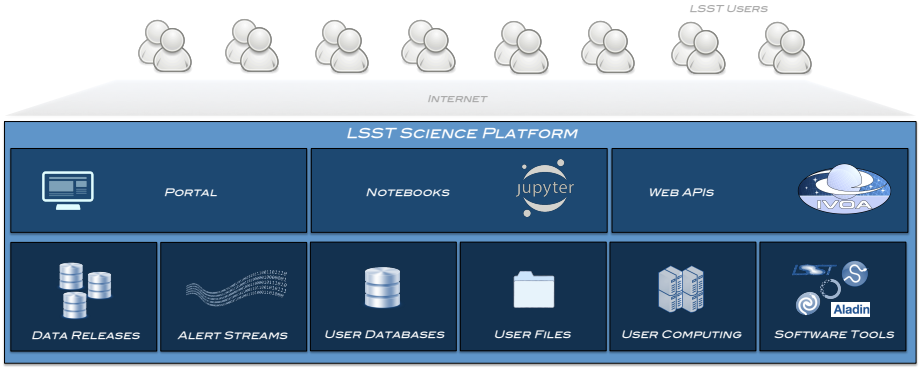
11 Data Releases in 10 years
Final database catalog: 15 PB

20TB raw data/night
(with calibration exposures)

Prompt Data Products
via nightly Alert Streams

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

LSST Science Platform

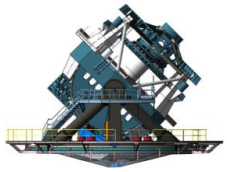


Data Releases
(current & previous)

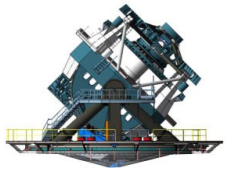
Alerts database

Mini-broker

Data access via Data Access Centres & Services



LSST Key Numbers



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 (20bn galaxies, 17bn stars), ~7 trillion observations ("sources"), and ~30 trillion measurements ("forced sources")
- Produced annually and accessible through online databases.

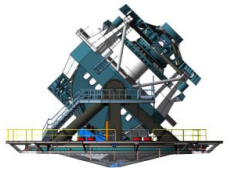
User Generated

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

LPM-231: LSST Data Product Categories & DM Data Products & Key Numbers



LSST Data Product Access

World Public

World Public data can be shared with everyone; data rights not required

Alerts: Alerts are immediately world public and can be shared with everyone

Data Releases: Data Releases are world public **two years** after release.

Education & Public Outreach (EPO): All data products accessed via the EPO platform

Proprietary

Proprietary Data can only be shared with data rights holders

Alerts Database: Archive of all issued alerts

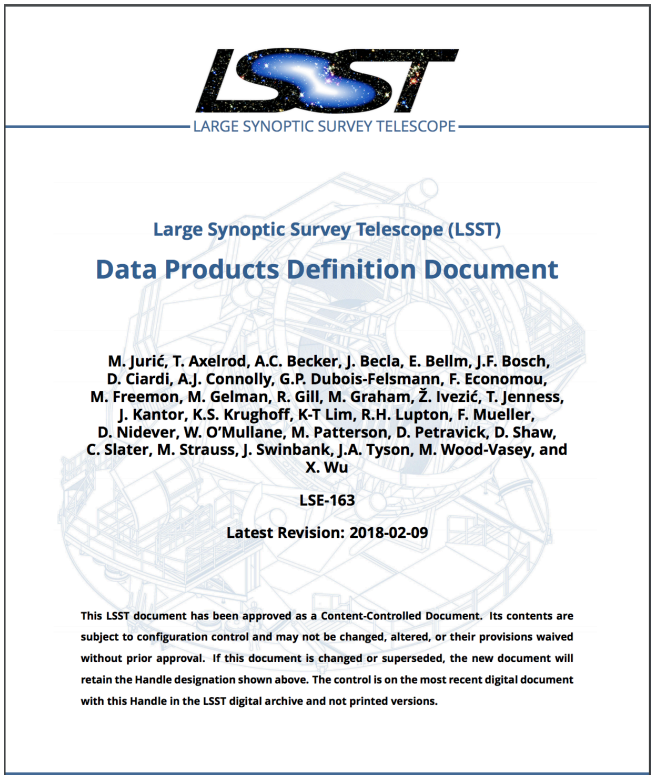
Prompt Products Database (PPDB): Difference Images, Source Catalogs including pre-covery forced photometry

Data Releases: Two year proprietary period for annual 'DRP' Data Releases (images, catalogs)

Science Platform: Computational resources, mini-broker and data portal are accessible only by data rights holders

LSST Data Products Definition Document

High-level description of LSST Data Products & Processing Services


LARGE SYNOPTIC SURVEY TELESCOPE

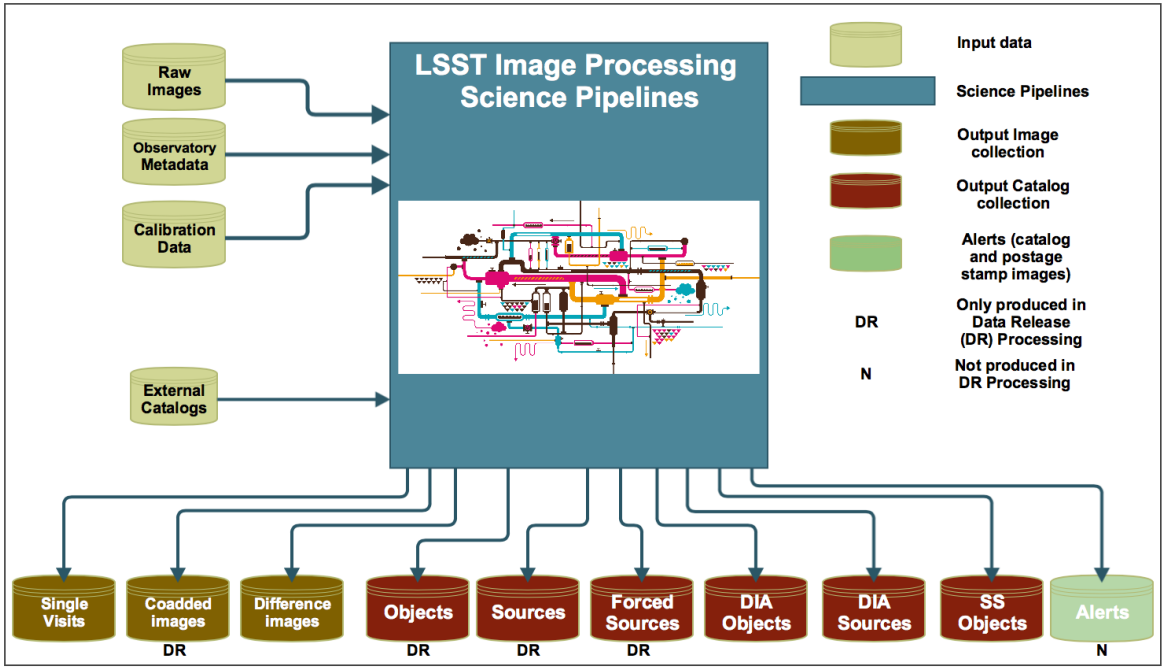
Large Synoptic Survey Telescope (LSST)

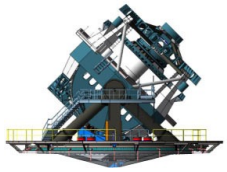
Data Products Definition Document

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. Freemon, M. Gelman, R. Gill, M. Graham, Ž. Ivezić, T. Jenness, 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



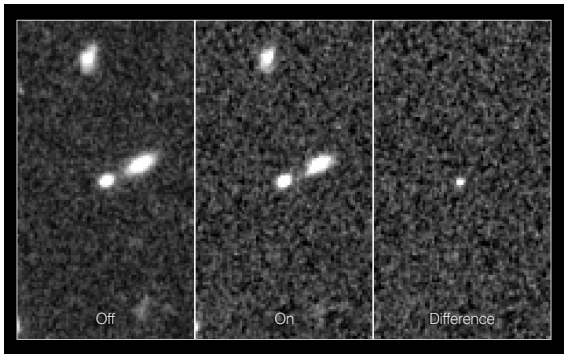


LDM-151: Science Pipelines Design Document & *LSE-163: Data Products Definition Document*

Prompt Data Products

Enables transient & time domain science requiring rapid follow-up

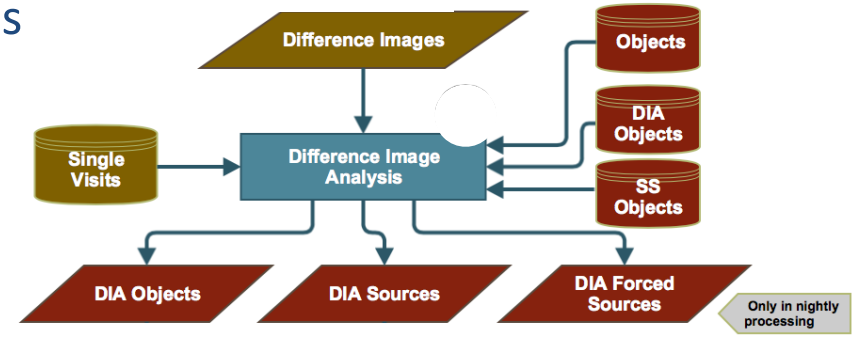
- Transient events: nova, supernova, GRBs
- Nearby Solar System Objects: NEOs, PHAs



Single-epoch data products and alerts

Produced by Difference Imaging Analysis (DIA)

- Sources detected on difference images with S/N > 5 (DIASource)
  - New astrophysical object not previously present (asteroid)
  - Flux changes in existing sources (variable star)



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

LSE-163: Data Products Definition Document

Prompt Products Database

Prompt Products Database: a living database, updated in real-time that contains the objects and sources detected on difference images: **DIA Catalogs**

DIASource *(a table of sources detected on difference image)*

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

DIAObject *(table of associations of DIASources & characteristics)*

- time-averaged coordinates, parallax & proper motion
- fluxes by filter, time-averaged & single-visit
- periodic and non-periodic variability features
- association with nearby objects in the ***Data Release*** catalog

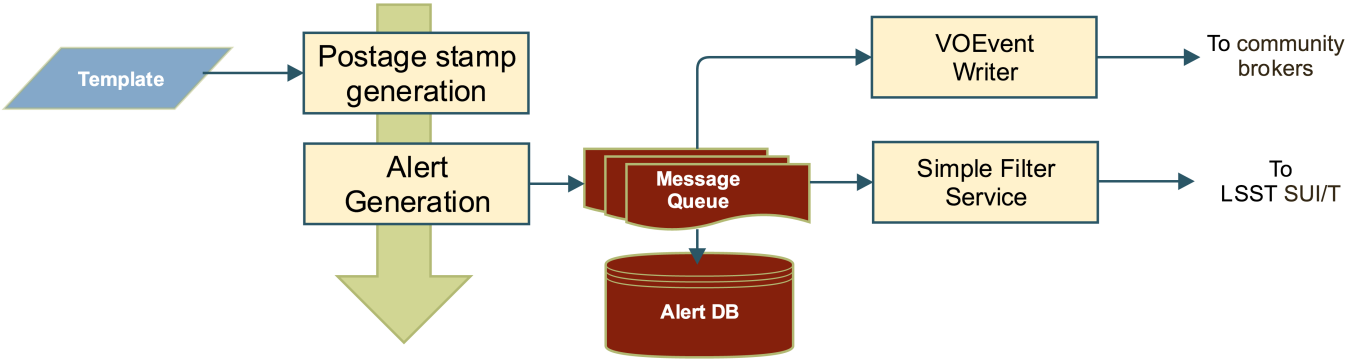
SSObject: *(the catalog of orbits of objects in the Solar System)*

- time-averaged coordinates, parallax & proper motion

LSE-163: Data Products Definition Document

Alert Distribution & Brokers

Alerts will be delivered to a limit set of community brokers that filter and classify events to enable scientific use.



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 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, 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

LDM-151: Science Pipelines Design Document & *LSE-163: Data Products Definition Document*

Alert Packets

Alert Packets: *text file containing the data for one DIA source and associated schema*

- *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)

LSE-163: Data Products Definition Document

Data Release Data Products

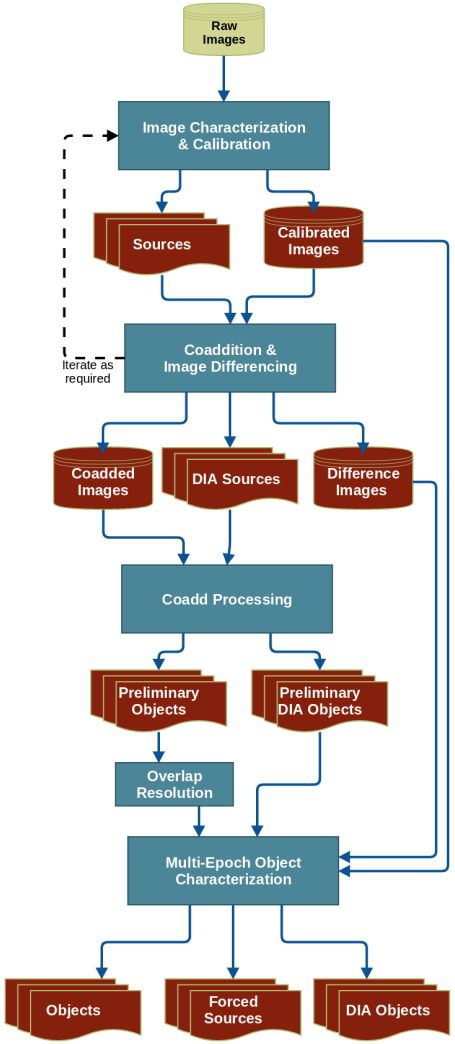
Enables static sky science & time-domain science (not time sensitive)

- Studies of galaxy evolution, or weak lensings
- statistical investigations of variability
- deep-sky and high-Precision astrophysics

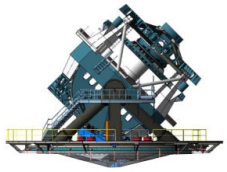
Produced by Direct Image Analysis

- Well calibrated consistently processed catalogs & images combining information from many exposures
- Sources are independently detected and measured in individual visits & coadds, and recorded in the Source table.

Fully reprocessed prompt data products



LDM-151: Science Pipelines Design Document & *LSE-163: Data Products Definition Document*



Data Release Data Products

Data Release products include images and catalogs

Images	**Single-Visit**	fully processed 2 x 15s exposures processed visit images (PVIs)
	Co-add	short-period (yearly & full survey) best seeing & deepest (unless the same) one per filter (ugrizy), and 'multi-color' transient-free template images

Catalogs	**Source (single-epoch)**	*Detections in single-visit images:* positions, aperture fluxes, point source fluxes, de-blending results
	Forced Source (single-epoch)	*In all single-visit images for all sources:* point-source fluxes, forced photometry
	Object (associations of sources representing astrophysical phenomena)	*All sources as well as detections in co-added images:* positions, aperture & point-source fluxes, de-blending results, model fits seeing-independent *galaxy* colours

LDM-151: Science Pipelines Design Document & *LSE-163: Data Products Definition Document*

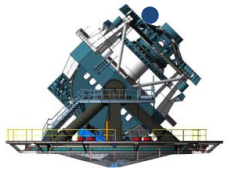
Data Release Availability & Retention

Data Releases

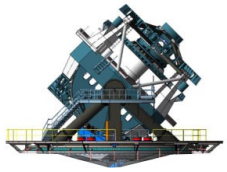
- 11 Data Releases planned over 10 years of LSST operation
  - Two for the first year of operations, one every subsequent year
- Few PB (DR1) to ~70 PB for DR11

Accessibility

- Contents of the most recent and penultimate data releases on fast storage
  - Catalogs loaded into the database.
- Older releases will be archived on mass storage
  - Costly to keep all data releases loaded and accessible at all times.
  - Queries against archived releases *may* not be possible.
  - Available as bulk downloads
- All raw data used to generate any public data product kept & made available
  - raw exposures, calibration frames, telemetry, configuration metadata, etc.



LSE-163: Data Products Definition Document



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
- Detailed variability classification

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 *User-Generated* data product creation

LSE-163: Data Products Definition Document

LSST Special Programs

WFD
Wide-Fast-Deep

DDF
Deep Drilling Field

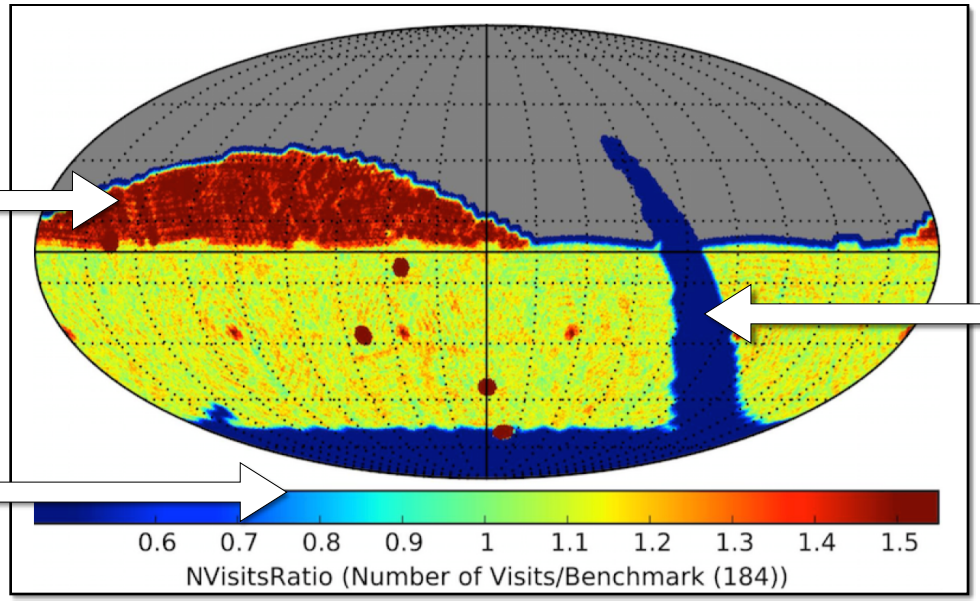
MS
Mini-Survey

What: *anything not "wide-fast-deep" main survey.*
Why: *additional or improved science results.*

Different area, survey strategy, and/or non-standard visit image.
E.g., Deep Drilling Fields, North Ecliptic Spur.

North Ecliptic Spur (solar system)

South Celestial Pole (LMC, SMC)



Galactic Plane (stars and planets)

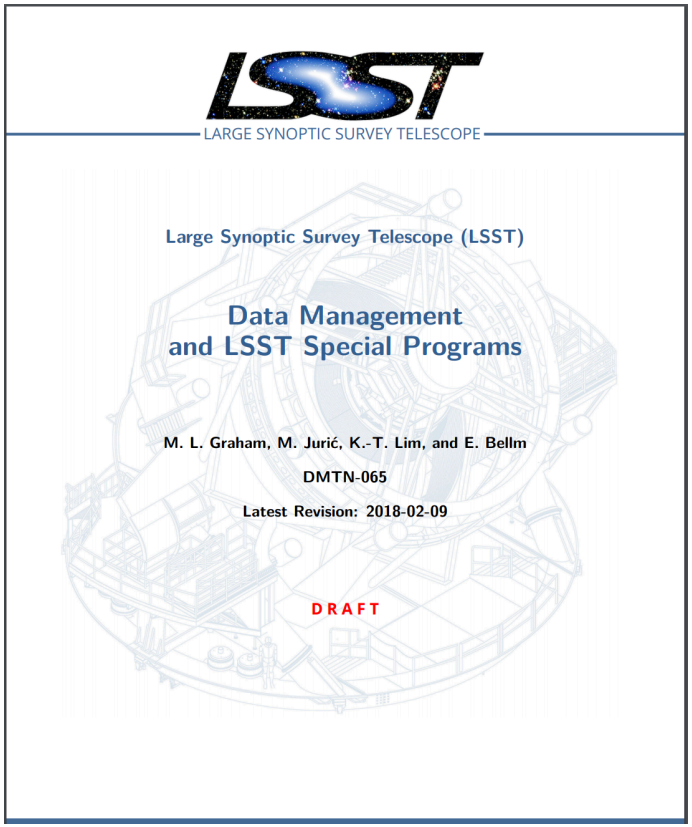
DDF examples

DDF nightly stacks to find high-*z* SN.

Ivezić et al. (2008), Figure 18.

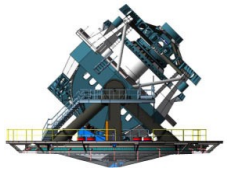
~10% of LSST observing time will be devoted to special programs that extend the areal coverage, depth, and/or sampling cadence to obtain improved coverage of interesting regions

Data Management & Special Programs



LSST Data Management will:

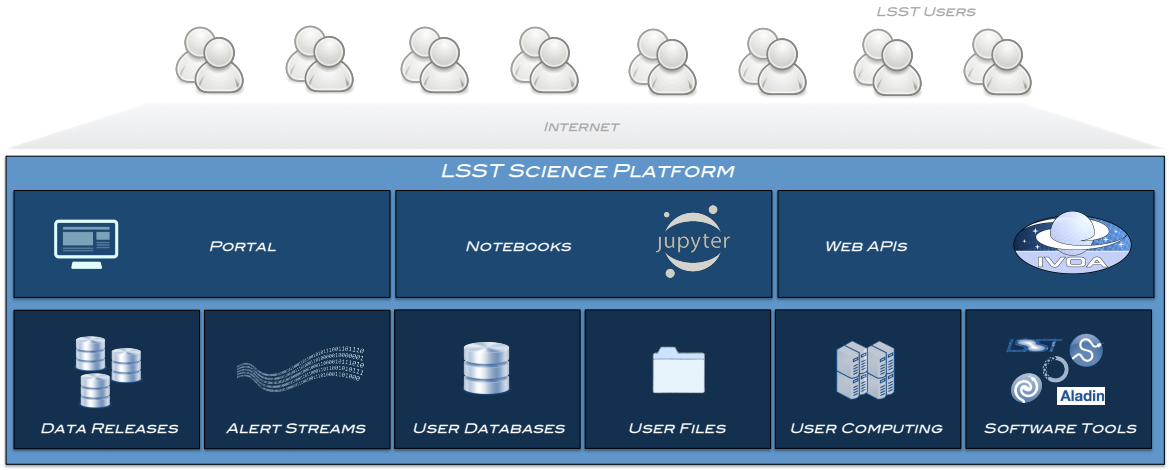
- **not** write unique algorithms for processing special programs data
- reconfigure pipelines to generate separate imaging and catalog products for special programs, where possible
- incorporate special programs data into *Alert* and *Data Release Production*, when scientifically beneficial
- allocate 10% of its resources for processing special programs data
- make the software stack source code available to the community



DMTN-065: Data management and LSST Special Programs

LSST Science Platform: Enabling Science

The *LSST Science Platform* is an integrated web-based service available to LSST data rights holders to to access, visualize and analyse LSST data

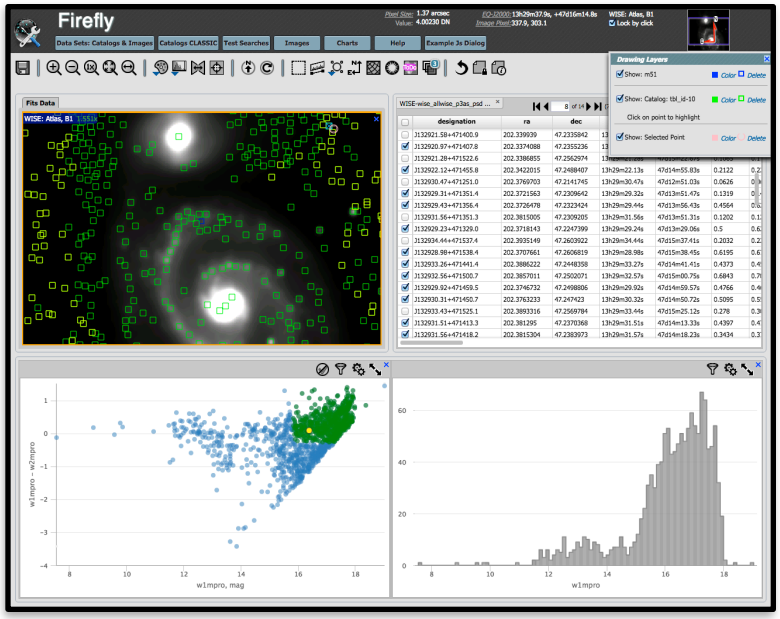


- Integrated service deployed at LSST Data Access Centers (DACs)
- Three primary user-facing “aspects”;
  - Web Portal (novice), JupyterLab (intermediate), Web APIs (expert & remote tools).
- Enables access to LSST Alert Streams and Data Releases
- Supports ‘next-to-the data’ analysis & ***User-Generated*** product creation

LSE-319: LSST Science Platform Vision Document

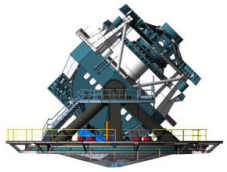
Portal Aspect: A window to the LSST Archive

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



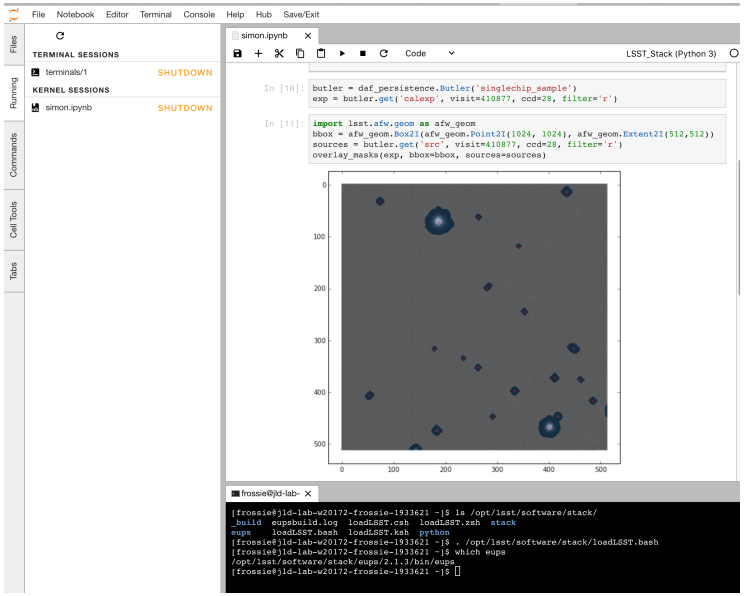
The Firefly Web Science User Interface (Wu et al, 2016; ADASS)

- 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
- Peta-scale capable RDBMS backend



JupyterLab Aspect: Delving deeper

Sophisticated data selection, analysis & creation of User-Generated data products



- Similar to working with Jupyter notebooks
- Computation & analysis on resources at LSST DACs
- Enables science discovery by **'bringing the analysis to the data'**
  - avoid download of 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

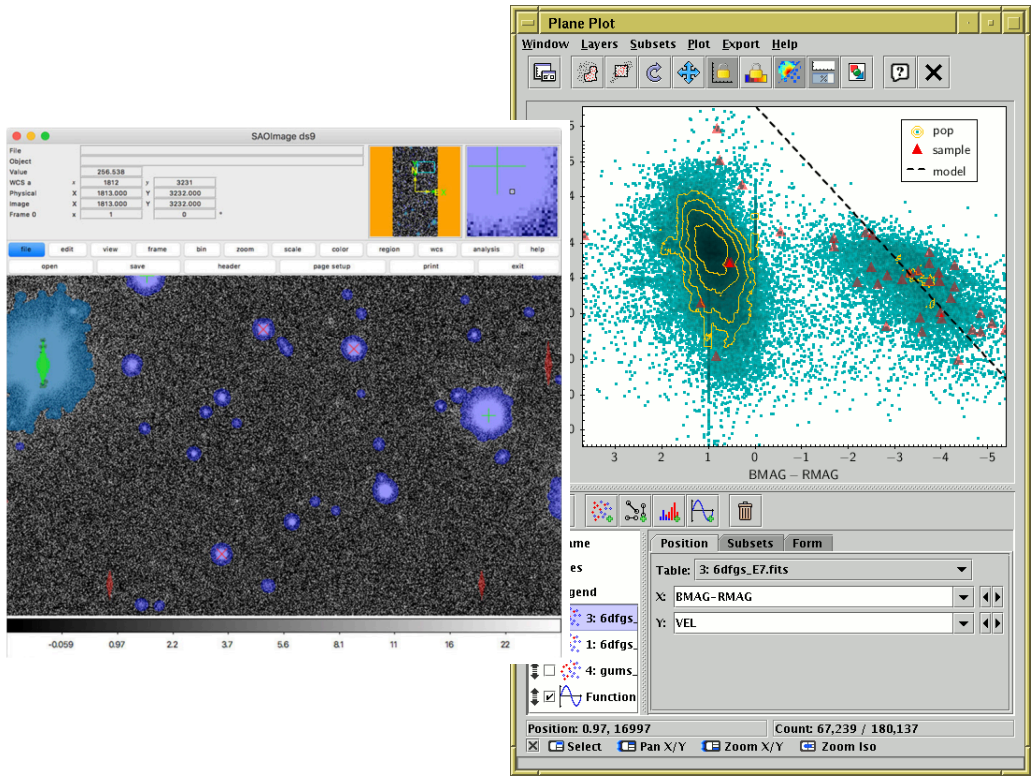
- Data queries will be shareable across the Portal and the JupyterLab aspects
- Submit a complex SQL query in the Notebook, browse & visualize results in the Portal.

LSE-319: LSST Science Platform Vision Document

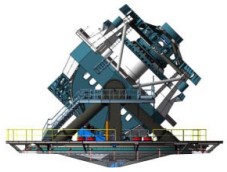
Web API Aspect: Integration with other tools

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 query builder & access results remotely via JupyterLab or TOPCAT



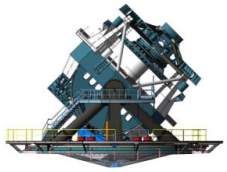
LSE-319: LSST Science Platform Vision Document

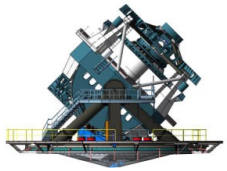
Requirements Documentation Flow-Down

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 *in the DM-SST*

Science Collaboration	SC chairs	DM liaison
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 *(DM-SST Deputy)*
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

DM Community & Communications

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	Latest
Science Public discussions about LSST science. Data Q&A · Survey Strategy · LSST2018 · LSST2017 · LSST2016 · Milky Way (Open)	92	Welcome to community.lsst.org Meta 2 Aug '15
Science Collaborations Science discussions for members of LSST Science Collaborations. PST Telecons	33	Producing and using DIAOjects for analysing DESC/DC2 transient simulation Support imagedifference 4 5h
Announcements Official LSST announcements.	19	New geom package replaces much of lsst.afw.geom DM Notifications 0 2d
Support Community support venue for using the LSST software, services and data. LSST Science Platform	186	OpSim v4 – community access? Simulations 0 3d
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	Science Pipeline release 16.0 - Status and discussion DM Notifications stack-releases 0 3d
		Solar System Science Collaboration (SSSC) May 2018 Update

community.lsst.org

DM Community & Communications

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

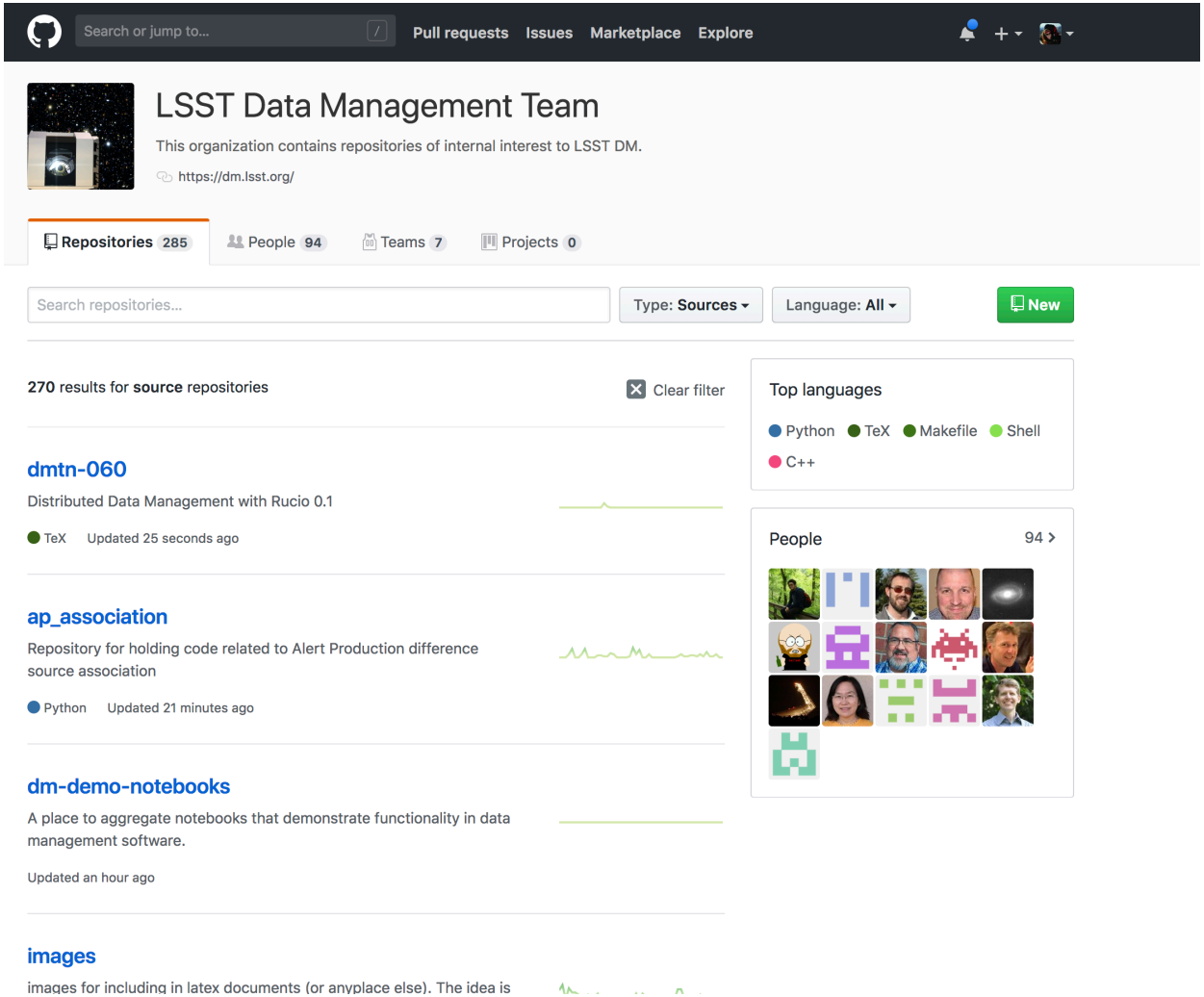
Topic	Users	Replies	Views	Activity
About the Data Q&A category Ask science questions about LSST data products and the Science Platform. Questions will be monitored by the Data Management team and answered in a timely fashion. As often as possible, the answers will point to permanen… read more		0	141	Oct '17
Data Model for variable sources at time of Data Releases? agn	G	4	192	Feb 27
Availability of residual images?		1	116	Feb 23
LSST Filters versus SDSS	Y	5	208	Nov '17
CCD Nonlinearity Near Saturation	P R N	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?.

community.lsst.org

LSST GitHub

All Data Management code and documentation is openly available on GitHub

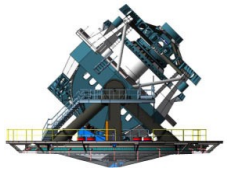


GitHub: lsst-dm

Tutorials

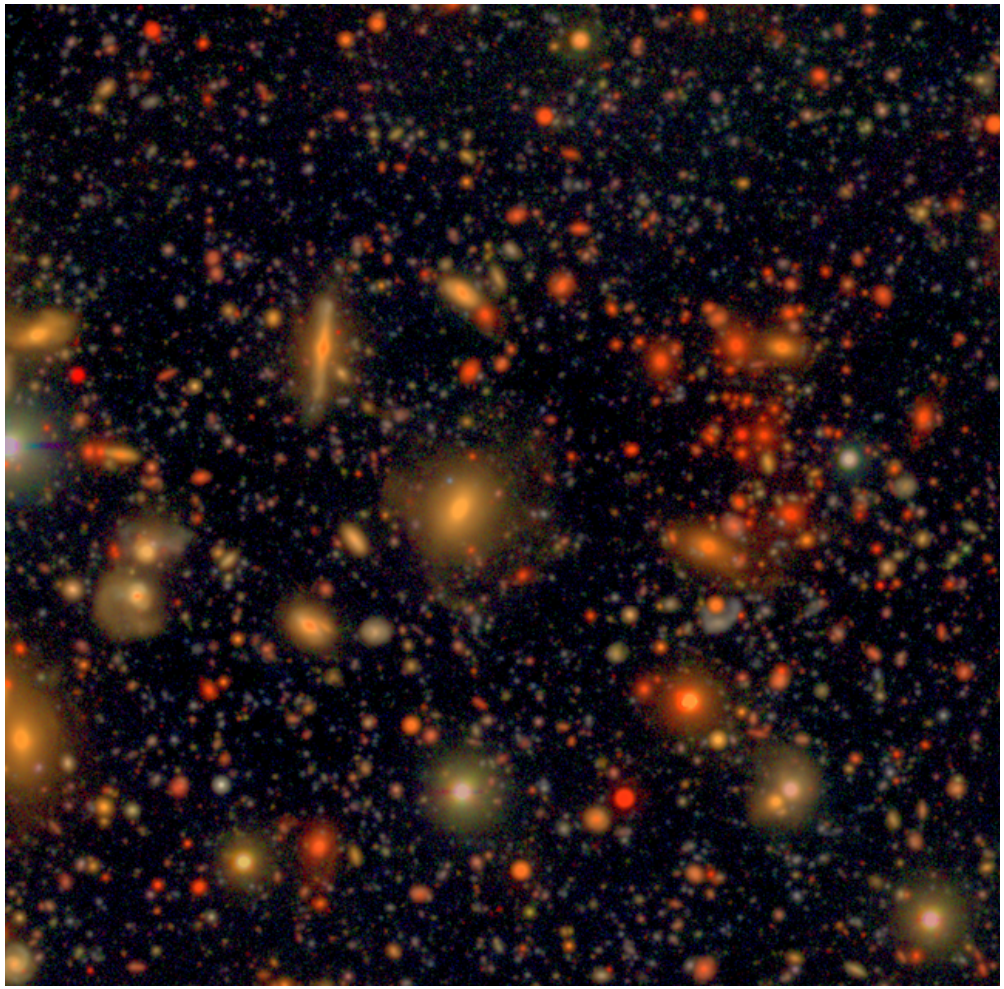
Please provide your GitHib IDs is you want to participate in the tutorials

https://tinyurl.com/lsstdmtutorial



Questions ….

~ 3.5′ SDSS image ≈ 22



HSC image (COSMOS)

g,r(1.5 hrs) ,i(3 hrs) PSF matched co-add (≈ 27.5)

Images: Lupton and HSC collaboration see also Lupton et al. (2004)