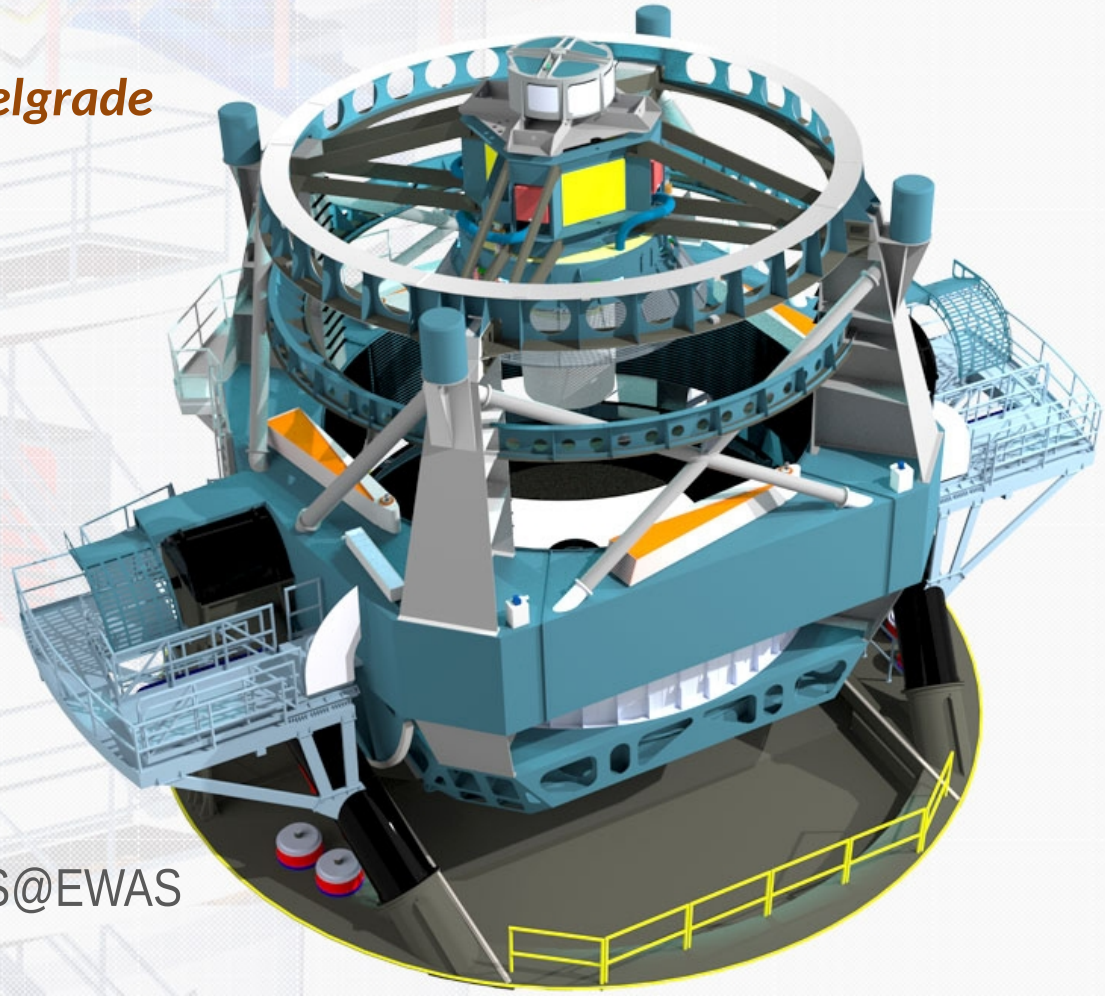
The LSST Data Products and Services

Darko Jevremović
Astronomical Observatory of Belgrade

for the LSST & LSST DM Team



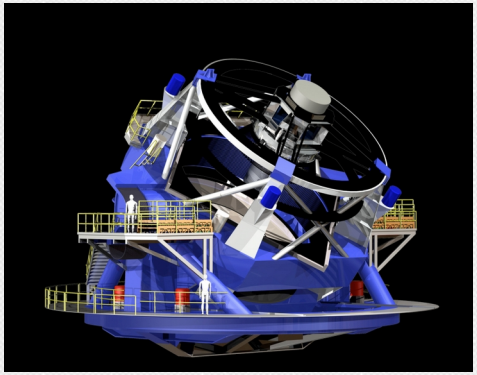
ASTROINFORMATICS@EWAS
Prague, June 30 2017

A Dedicated Survey Telescope

- **A wide (half the sky), deep (24.5/27.5 mag), fast (image the sky once every 3 days) survey telescope. Beginning in 2022, it will repeatedly image the sky for 10 years.**

- The LSST is an integrated survey system. The Observatory, Telescope, Camera and Data Management system are all built to support the LSST survey. There's no PI mode, proposals, or time.

- **The ultimate deliverable of LSST is not the telescope, nor the instruments; it is the fully reduced data. LSST is a *facility* that delivers *data products* and *data access and analysis services*.**



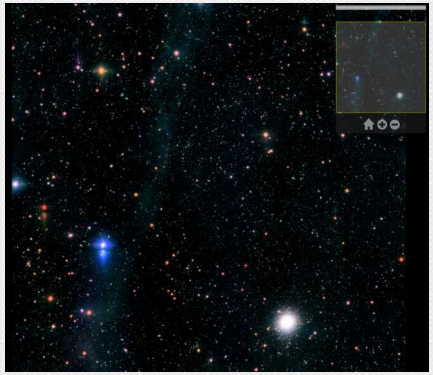


Table 4: Level 2 Catalog `Object` Table

Name	Type	Unit	Description
psRadecTai	double	time	Point source model: Time at which the object was at position `radec`.
psPm	float[2]	mas/yr	Point source model: Proper motion vector.
psParallax	float	mas	Point source model: Parallax.
psFlux	float[ugrizy]	nmgy	Point source model fluxes[58].
psCov	float[66]	various	Point-source model covariance matrix[59].
psLnL	float		Natural *log* likelihood of the observed data given the point source model.
bdRadec	double[2]	degrees	B+D model[60]: (α, δ) position of the object at time `radecTai`, in each band.

Telescope → Images → Catalogs

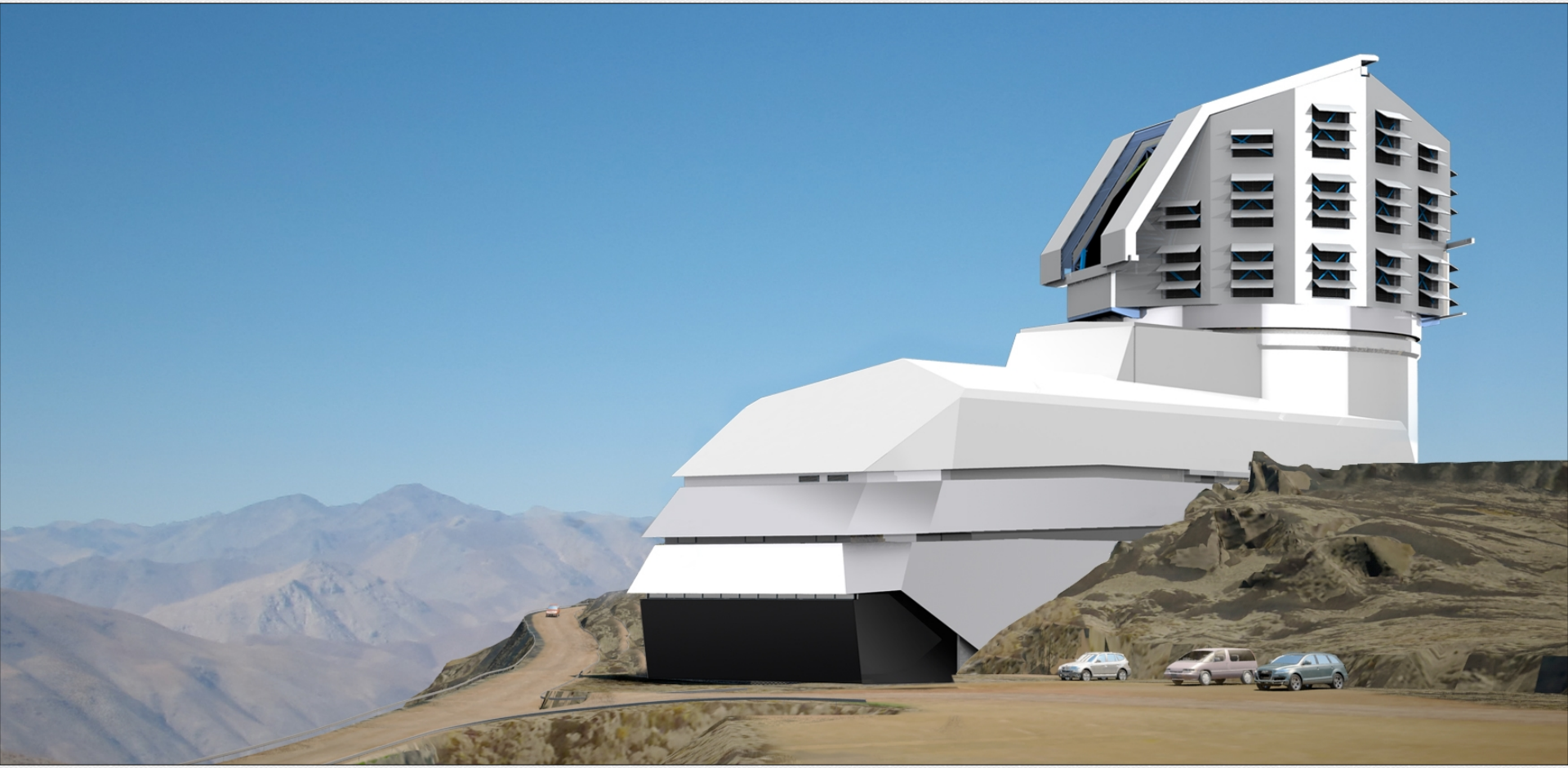
LSST Site (April 14th, 2015 – First Stone ceremony)

The Summit, Wednesday.

The Observatory is Taking Shape

May 15th 2017

LSST Observatory (cca. late ~2018)

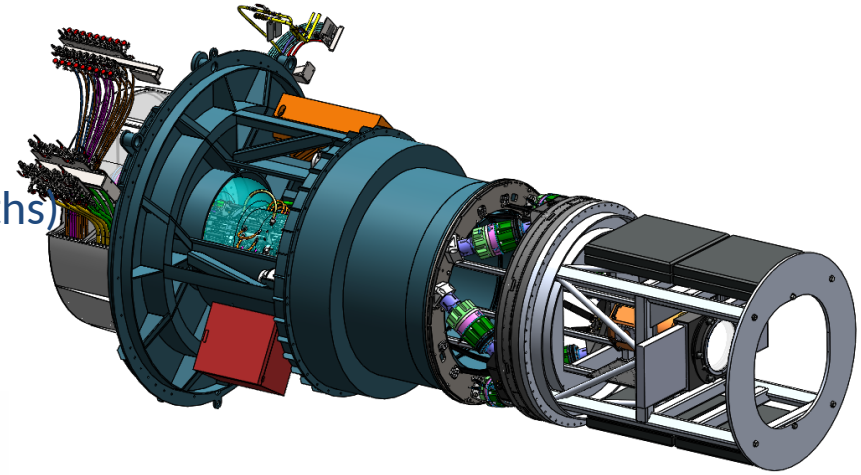


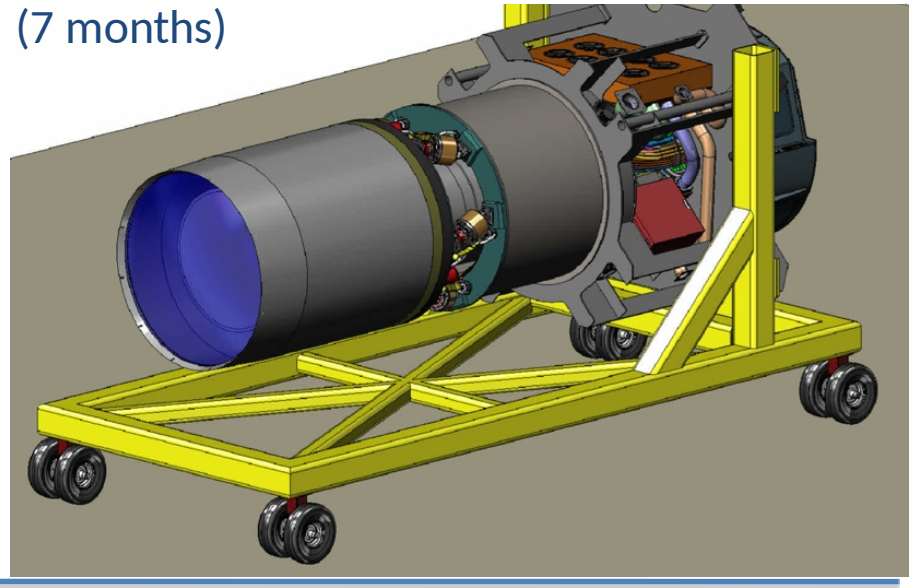
We are <2 years away from the observatory building being close to complete!

The Commissioning Plan is structured in phases

- Phase 0: Pre-Commissioning
  - Test procedure definition
  - Test analysis script development using the LSST code base
  - Procedure & script validation with simulation tools
  - Special hardware/software development
  - Subsystem oversight an monitoring
- Phase I: Early System AI&T with ComCam (6 months)
  - early interface testing
  - procedure verification
  - telescope alignment and AOS control
  - sub-scale DM pipeline testing
  - early science verification tests
- Phase II: Full System AI&T with Science Camera (7 months)
  - camera-telescope integration, alignment & AOS control
  - full-scale DM pipeline testing
  - full-scale science verification test
- Phase III: Science Verification (5 months)
  - mini-Survey 1 focus on Alert Production
  - mini-Survey 2 focus on 10-year depth
  - Data release processing

ComCam + integrating structure





LSSTCam + integrating structure on transport cart

Data Production Milestones

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

LSST From a Scientist's Perspective

- A stream of ~10 million time-domain events per night, detected and transmitted to event distribution networks within 60 seconds of observation.
- A catalog of orbits for ~6 million bodies in the Solar System.

Level 1

- 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.

Level 2

- Services and computing resources at the Data Access Centers to enable analysis and production of added value products.
- Software and APIs enabling development of analysis/added value codes.

Level 3

The Data Products Definition Document

LSST Data Products Definition Document

A document giving a high-level description of LSST data products.

http://ls.st/dpdd

Level 1 Data Products: Section 4.

Level 2 Data Products: Section 5.

Level 3 Data Products: Section 6.

Special Programs DPs: Section 7.

Level 1:

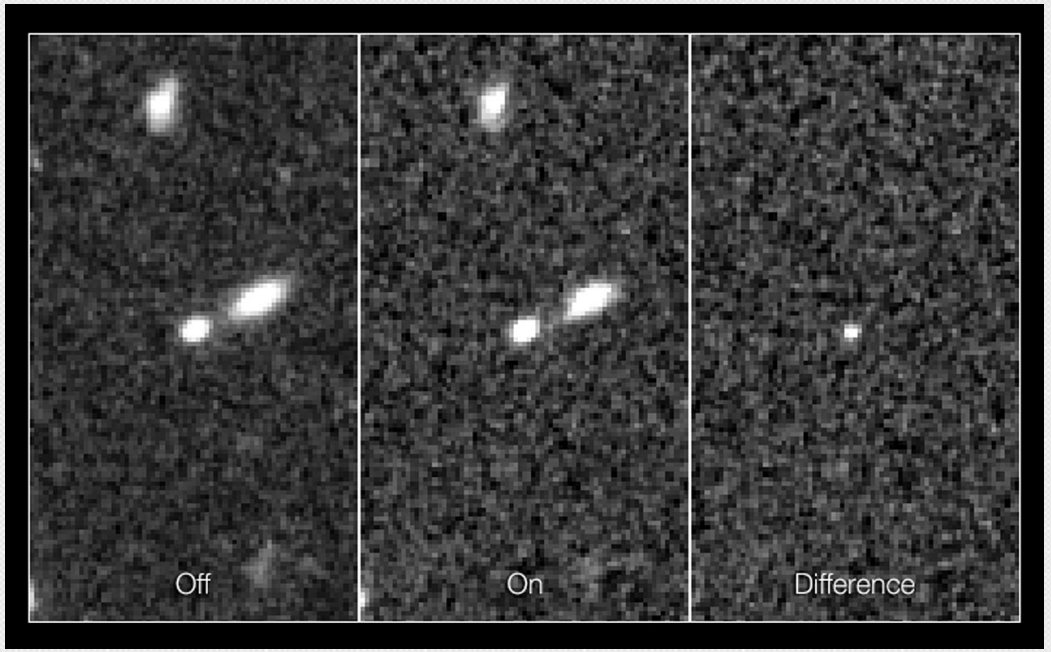
Enabling Discovery and Rapid Follow-up of Time Domain Events

Level 1 Data Products: Time Domain

- **Real-time image differencing as observing unfolds each night**
- Detection performed on image differenced against a deep template
- Measurement performed on the difference image and direct image
- Associated with pre-existing observations and stored in a database

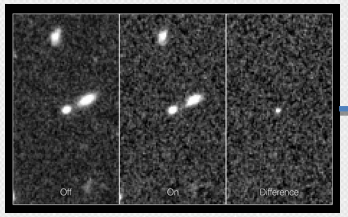
- For every source detected in a difference image, we will emit an “Event Alert” within 60 seconds of observation.

The primary use case is to enable real-time recognition and follow-up of transients of special interest.



CANDELS (http://www.spacetelescope.org/images/heic1306d/)

Level 1 Data Products and Flows



(5 sec)

Difference, detect and measure

(50 sec)

Level 1 Data Store (internal)

Alerts

(60 sec)

Level 1 SQL Database

(24 hrs)

Images

(24 hrs)

Solar System Object Catalog

(24 hrs)

Level 1: Time-Domain Event Alerts

- **Each alert will include the following:**
  - Alert and database ID: IDs uniquely identifying this alert.
  - The photometric, astrometric, and shape characterization of the detected source
  - 30x30 pixel (on average) cut-out of the difference image (FITS)
  - 30x30 pixel (on average) cut-out of the template image (FITS)
  - The time series (up to a year) of all previous detections of this source
  - Various summary statistics ("features") computed of the time series

- **The goal is to transmit nearly everything LSST knows about any given event, enabling downstream classification and decision making *without* the need to call back into LSST databases (thus introducing extra latency)**

- We expect a high rate of alerts, **approaching 10 million per night**.

LSST Alert Filtering Service

- Most end-users will **not** be interested in reception of the full stream, but **only a subset that matches their scientific interest** (e.g., potential SNe candidates, variable stars, or moving objects).

- To support selecting such subsets of alert candidates, **LSST will provide an alert filtering service**. This service will let astronomers create simple *filters* that limit which alerts are ultimately forwarded to them.

- These user defined filters will be possible to specify using an SQL-like declarative language, or short snippets of (likely Python) code (n.b. this is our current thinking, subject to change).

Example of a User-Defined Filter (a sketch!)

```
# Keep only never-before-seen events within two
# effective radii of a galaxy. This is for illustration
# only; the exact methods/members/APIs may change.

def filter(alert):
    if len(alert.sources) > 1:
        return False
    nn = alert.diaobject.nearest_neighbors[0]
    if not nn.flags.GALAXY:
        return False
    return nn.dist < 2. * nn.Re
```

The user will subscribe to the alert stream by specifying a filtering function such as the one shown above. Once specified, only the alerts for which the function returns True will be forwarded to the user's VOEvent client.

Public VOEvent Brokers and Networks

- **We also anticipate that advanced, public, filtering services – VOEvent brokers – will be established by the community.**

- These may provide advanced functionality such as:
  - cross-correlation of LSST alerts with external catalogs and other alert streams,
  - classification engines,
  - more extensive annotation of alerts,
  - coordination of follow-up groups,
  - incorporation of other contextual information needed to decide on whether a transient is worth following up.

- **We are encouraging the community to self-organize and develop such alert brokers and networks.**
  - **US: ANTARES project led by NOAO**
  - **Europe: an opportunity to build on existing archives and expertize!**

Level 1: Solar System Objects

- **Solar System objects will be identified and linked together based on compatibility of their observed positions with motion around the Sun.**
  - Enhanced variant of MOPS algorithm; advanced prototype in hand.

- **Planning to:**
  - Identify and link observations of Solar System objects
  - Measure their orbital elements
  - Measure their photometric properties
  - *Expect to provide orbits for >= 70% of all NEOs brighter than H=22*

- **Availability: within 24 hours of orbit determination**

Level 2:

Enabling Deep Sky and High-Precision Astrophysics

Level 2: Annual Data Releases

- **Well calibrated, consistently processed, catalogs and images**
  - Catalogs of objects, detections, detections in difference images, etc.

- **Made available in *Data Releases***
  - Annually, except for Year 1
    - Two DRs for the first year of data

- **Complete reprocessing of all data, for each release**
  - Every DR will reprocess ***all*** data taken up to the beginning of that DR

- **Projected catalog sizes:**
  - **18 billion objects** (DR1) → **37 billion** (DR11)
  - **750 billion observations** (DR1) → **30 trillion** (DR11)

Level 2: Archive Contents

- Processed visits ("calibrated exposures")
  - Visit images with instrumental signature removed, background, PSF, zero-point and WCS determined
- Coadds
  - Deep coadds across the entire survey footprint (multiple flavors)

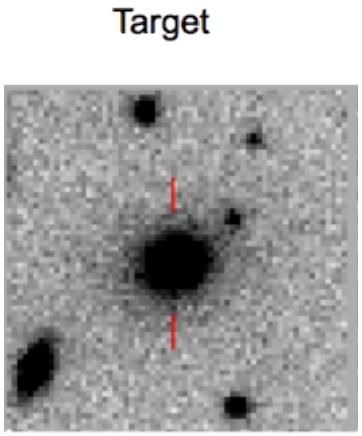
More in DPDD, Section 5.4

- Catalogs of Sources
  - Measurements of sources detected on calibrated exposures
- Catalogs of Objects
  - Characterization of objects detected on multi-epoch data
- Catalogs of ForcedSources
  - Forced photometry performed on all exposures, at locations of all Objects

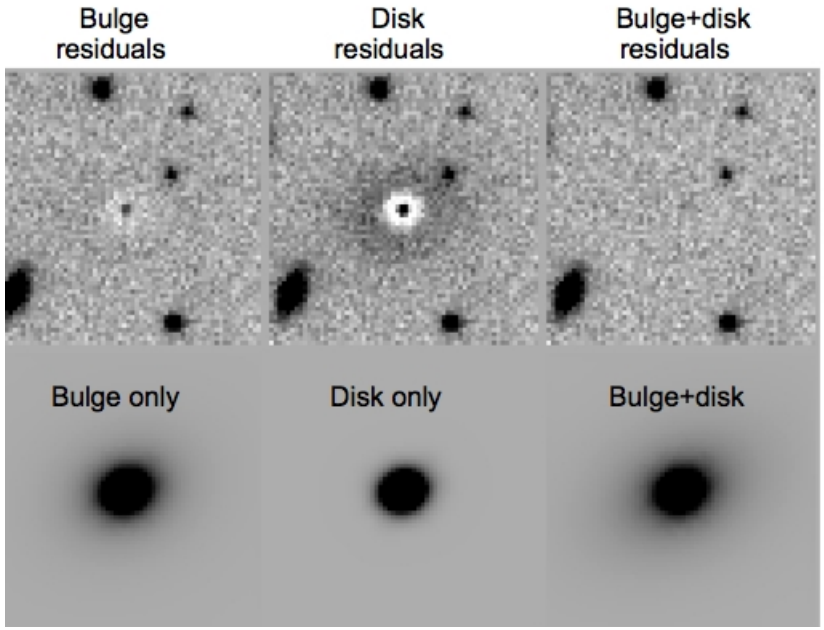
More in DPDD, Section 5.3

LSST Catalog Contents (Level 2)

- **Object characterization (models):**
  - Moving Point Source model
  - Double Sérsic model (bulge+disk)
    - Maximum likelihood peak
    - **Samples of the posterior (hundreds)**
- **Object characterization (non-parametric):**
  - Centroid: (α, δ), per band
  - Adaptive moments and ellipticity measures (per band)
  - Aperture fluxes and Petrosian and Kron fluxes and radii (per band)
- **Colors:**
  - Seeing-independent measure of object color
- **Variability statistics:**
  - Period, low-order light-curve moments, etc.



LSST Science Book, Fig. 9.3



Level 3:

Enabling the Creation of Added-Value Data Products

Level 3: Added Value Data Products

- **Level 3 Data Products: Added-value products created by the community**

- **These may enable science use-cases not fully covered by what we'll generate in Level 1 and 2:**
  - Custom processing of deep drilling fields
  - SNe photometry (e.g. CFHT-LS type forward modeling)
  - Catalogs of SNe light echos
  - Characterization of diffuse structures (e.g., ISM)
  - Extremely crowded field photometry (e.g., globular clusters)
  - Custom measurement algorithms
- **The LSST wants to make it easier for the community to create and distribute Level 3 products**
  - Making the LSST software stack available to end-users
  - Enabling limited end-user analysis and processing at the LSST data center
  - User databases and workspaces ("mydb")
- Level 3 products may be migrated to Level 2 (with owners' permission); this is one of the ways how Level 2 products will evolve.

Enabling the creation of Level 3 Data Products

- We are engineering the LSST software stack to be modular, reusable, documented, supported, and end-user friendly. It will be available under free software or public domain licenses.

- We will enable user computing at the LSST archive, making available to the users ~10% of our storage and computing resources (~50-100 TFLOPS). We will use this to power a JupyterHub-type remote analysis environment and a small HPC-type processing cluster.

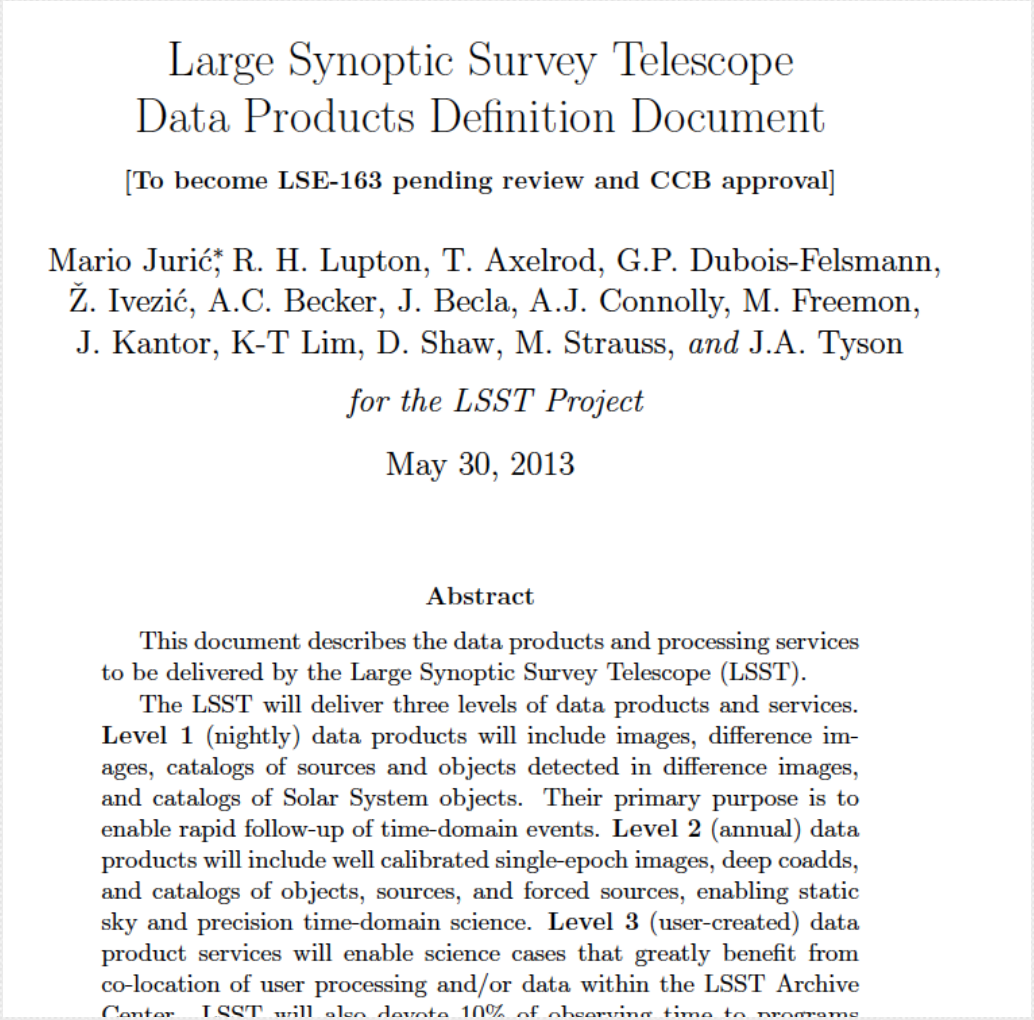
- LSST archive will be located in the National Petascale Computing Facility at National Center for Supercomputing Applications (NCSA). Significant additional supercomputing is expected to be available at the same site (e.g., NPCF currently hosts the Blue Waters supercomputer).

- Rights-holders may build their own computing facilities to support larger-scale processing, reusing our software (pipelines, middleware, databases) to the extent possible.

LSST@EUROPE

- It is time that European LSST data rights holders put act together. There are few hundred of them and it is important not to rediscover the wheel…
- Community networking – either through COST (thinking of successor of BSE) or as INFRAIA starting community
- There are lot of e-infrastructure calls – IN2P3 are targeting e-infrastructure to build the prototype of data access center (Contact Dominique Boutigny/Fabio Hernadez)
- UK – will probably build their own DAC (Edinburgh)
- LSST@EUROPE3 Conference in Lyon June 11-15 2018
- We should also propose session for EWASS2018 – deadline is mid July

Finding Out More: The "DPDD"



Large Synoptic Survey Telescope
Data Products Definition Document

[To become LSE-163 pending review and CCB approval]

Mario Jurić*, R. H. Lupton, T. Axelrod, G.P. Dubois-Felsmann, Ž. Ivezić, A.C. Becker, J. Becla, A.J. Connolly, M. Freemon, J. Kantor, K-T Lim, D. Shaw, M. Strauss, *and* J.A. Tyson

for the LSST Project

May 30, 2013

Abstract

This document describes the data products and processing services to be delivered by the Large Synoptic Survey Telescope (LSST).

The LSST will deliver three levels of data products and services. **Level 1** (nightly) data products will include images, difference images, catalogs of sources and objects detected in difference images, and catalogs of Solar System objects. Their primary purpose is to enable rapid follow-up of time-domain events. **Level 2** (annual) data products will include well calibrated single-epoch images, deep coadds, and catalogs of objects, sources, and forced sources, enabling static sky and precision time-domain science. **Level 3** (user-created) data product services will enable science cases that greatly benefit from co-location of user processing and/or data within the LSST Archive Center. LSST will also devote 10% of observing time to programs

LSST Data Products Definition Document

A document giving a high-level description of LSST data products.

http://ls.st/dpdd

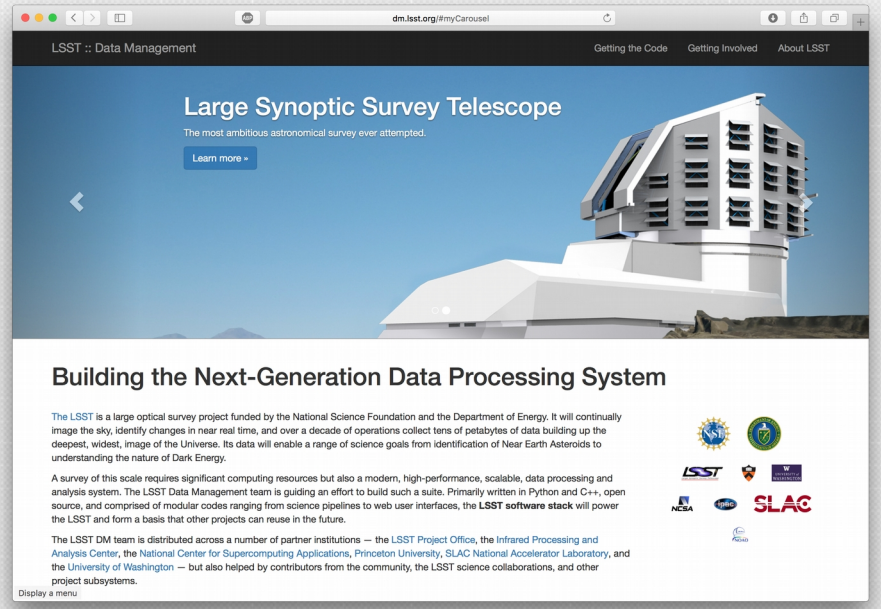
Level 1 Data Products: Section 4.

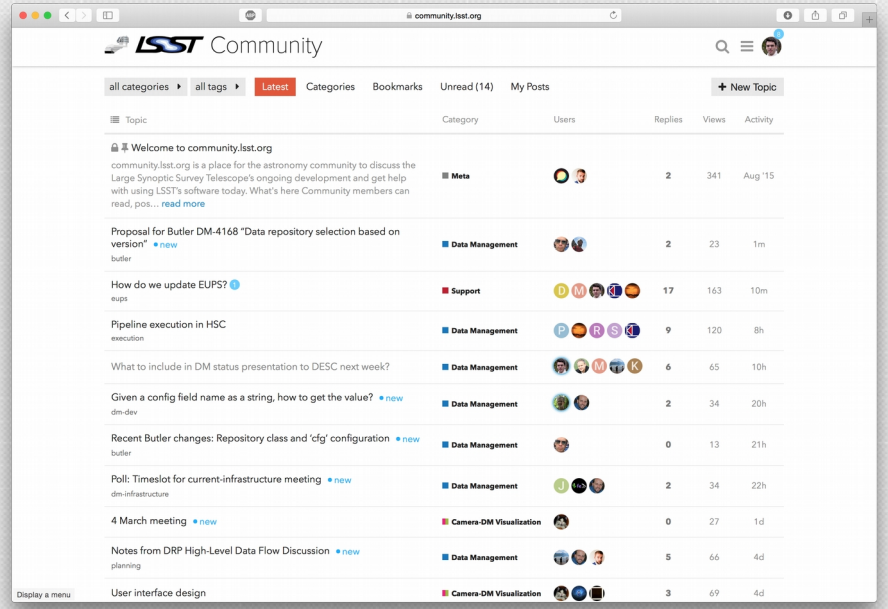
Level 2 Data Products: Section 5.

Level 3 Data Products: Section 6.

Special Programs DPs: Section 7.

Finding out more

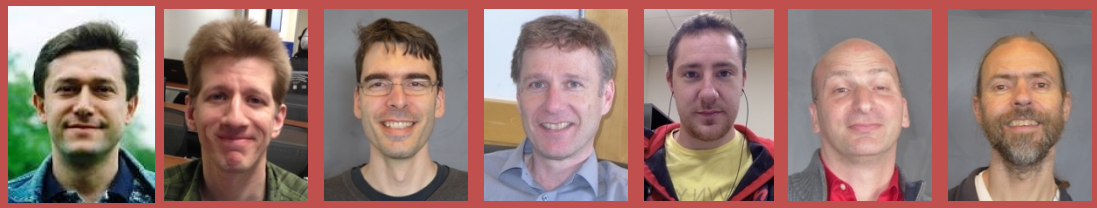




http://dm.lsst.org: *LSST Data Management portal*

http://community.lsst.org: *internal and external project communication*

LSST DM people @ LSST@Europe2 – talk to us!



And many others:

- http://github.com/LSST
- http://pipelines.lsst.io
- http://developer.lsst.io
- https://confluence.lsstcorp.org/display/DM/Data+Management+Home

Prototype LSST Science Pipelines Are Running on HSC Survey …

HSC "ultra deep" gri imaging in COSMOS, with a total of 1.5 hours in g and r and 3 hours in i; (280/550 LSST visits).

The visits were processed, calibrated, registered, added, and the resulting coadds processed using the LSST stack.

These catalogues are being used to carry out first-year HSC science.

Credit: HSC collaboration, Robert Lupton and LSST DM @ Princeton.

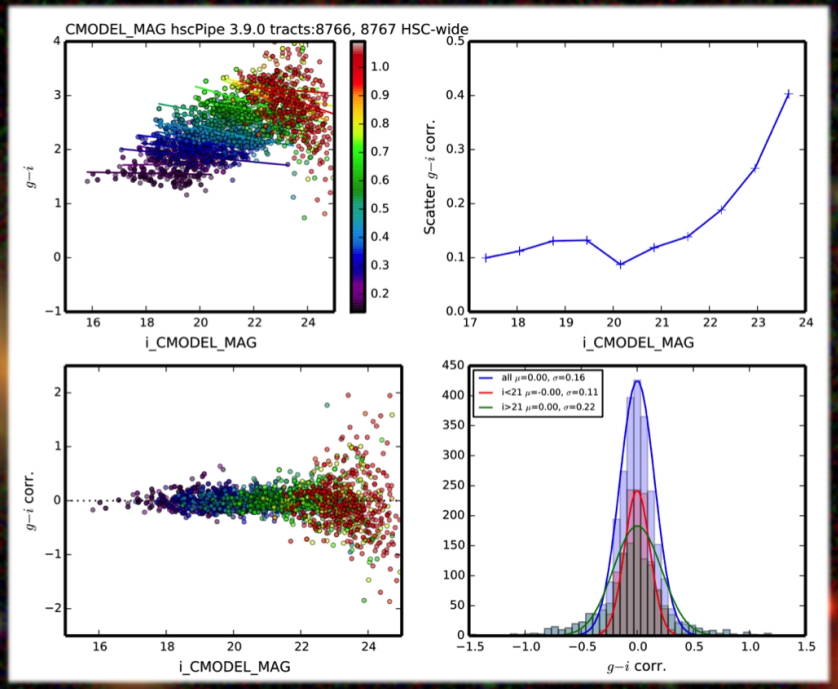
Prototype LSST Science Pipelines Are Running on HSC Survey …

HSC "ultra deep" gri imaging in COSMOS, with a total of 1.5 hours in g and r and 3 hours in i; (280/550 LSST visits).

The visits were processed, calibrated, registered, added, and the resulting coadds processed using the LSST stack.

These catalogues are being used to carry out first-year HSC science.

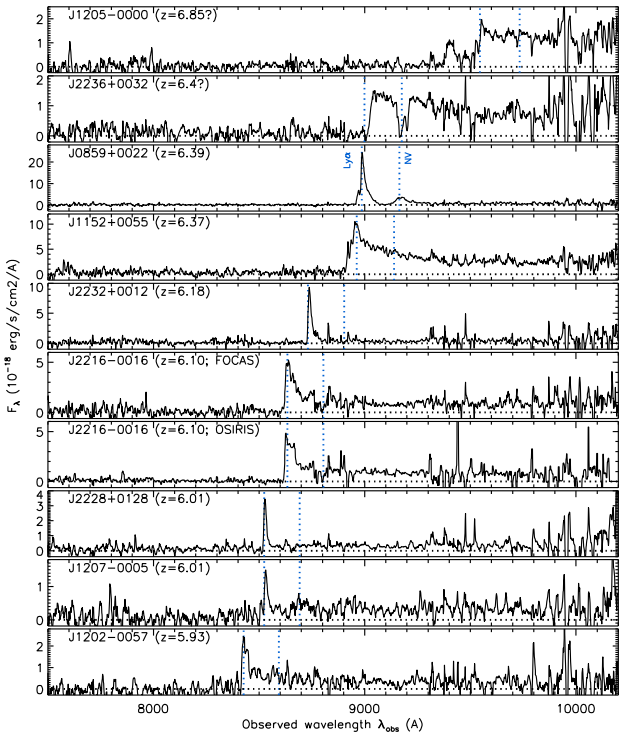
Credit: HSC collaboration, Robert Lupton and LSST DM @ Princeton.



Width of the E-S0 red sequence for SDSS `red mapper' galaxies, measured from the HSC `wide' data. The results are comparable to those from SDSS; the small extra scatter is attributed to problems with deblending data going several magnitudes deeper than SDSS.

Credit: Bob Armstrong, Atsushi Nishizawa, and the HSC collaboration.

… and enabling science.



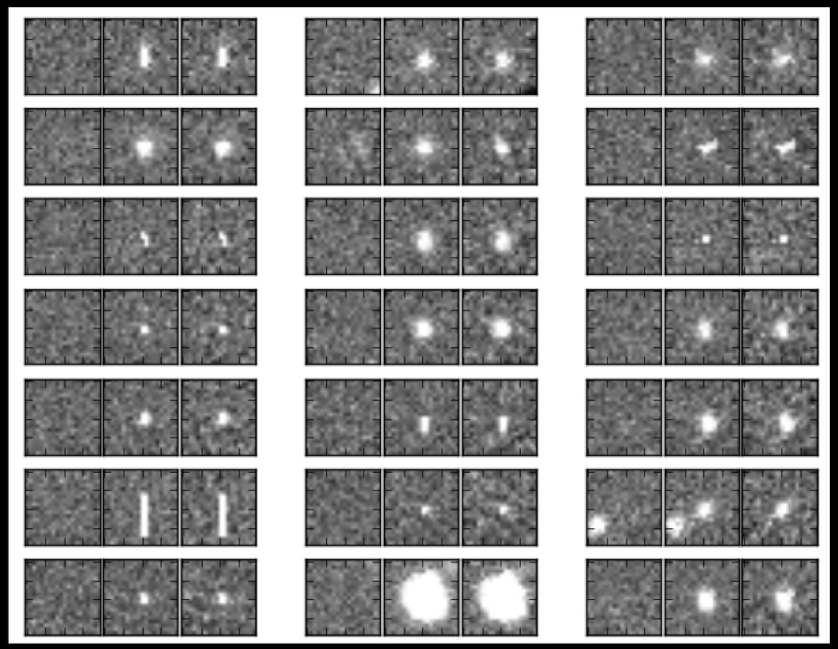
High-redshift quasars selected from the HSC `wide' data (10 minutes per band in gr and 20 minutes in /izy/; equivalent to 30 and 60 LSST visits, respectively).

Followup spectra were taken at Subaru, ***resulting in the discovery of second-highest known QSO redshift****.*

Credit: Yoshii Matsuoka et al. 2016.

HSC has been a good test bed for early LSST technology. Both HSC and LSST benefited.

LSST Time Domain Pipelines: Testing on DECam



Credit: Colin Slater and LSST DM @ U. of Washington
Data courtesy of L. Allen, NOAO

- We've begun running the LSST image differencing pipelines on data acquired by DECam.
- The immediate goal is to characterize LSST's asteroid detection capabilities. This requires low false positive rate. Preliminary runs are showing performance rates known to be clean enough for MOPS.
- Left: A representative set of 21 detections. Only dipoles were identified and rejected; no machine-learning afterburners were applied. Note the low instrumental artifact rates.

Pushing the State of the Art in Image Differencing

- Current transient surveys produce O(30-100k) detections per square degree (at similar depths). The vast majority of these are artifacts and misidentifications, which are then filtered out with machine learning (ML) afterburners.

- Using a similar setup, we find that a large majority of these detections come from misestimation of noise level: PSF matching and convolution induce correlated noise between pixels which is hard to track.
- Covariance matrix is, in principle, N_{pixels} x $N_{pixels,}$ but tracking (smaller) local estimates may be sufficient (following Price & Magnier on PS1). This will ensures that 5σ detections are really 5σ. Our code will do this sort of thing

- This reduces the need for post-fitering with machine learning afterburners. Significant gain in scientific output and system simplicity.

Source Type	Counts per Square Degree
Original Detections	19,475
After Correlated Noise Correction	1,022
Candidate Moving Objects (excludes stars)	344

Expect of order ~100 real objects

Investments in LSST algorithms research and software benefits the broader community.

Credit: Colin Slater and LSST DM @ U. of Washington

Finding Out More: The "DPDD"

Large Synoptic Survey Telescope
Data Products Definition Document

[To become LSE-163 pending review and CCB approval]

Mario Jurić*, R. H. Lupton, T. Axelrod, G.P. Dubois-Felsmann,
Ž. Ivezić, A.C. Becker, J. Becla, A.J. Connolly, M. Freemon,
J. Kantor, K-T Lim, D. Shaw, M. Strauss, *and* J.A. Tyson

for the LSST Project

May 30, 2013

Abstract

This document describes the data products and processing services to be delivered by the Large Synoptic Survey Telescope (LSST).

The LSST will deliver three levels of data products and services. **Level 1** (nightly) data products will include images, difference images, catalogs of sources and objects detected in difference images, and catalogs of Solar System objects. Their primary purpose is to enable rapid follow-up of time-domain events. **Level 2** (annual) data products will include well calibrated single-epoch images, deep coadds, and catalogs of objects, sources, and forced sources, enabling static sky and precision time-domain science. **Level 3** (user-created) data product services will enable science cases that greatly benefit from co-location of user processing and/or data within the LSST Archive Center. LSST will also devote 10% of observing time to programs

LSST Data Products Definition Document

A document giving a high-level description of LSST data products.

http://ls.st/dpdd

Level 1 Data Products: Section 4.

Level 2 Data Products: Section 5.

Level 3 Data Products: Section 6.

Special Programs DPs: Section 7.

Open Data, Open Source: A Community Resource

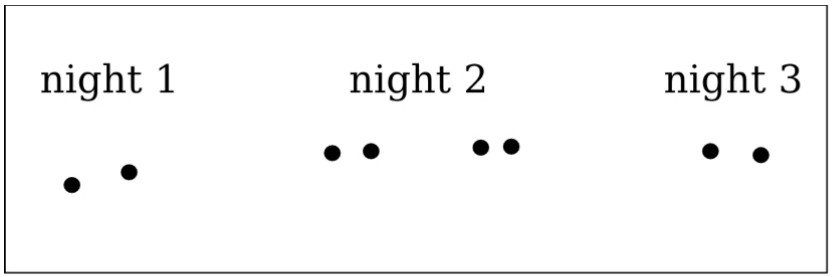
- LSST data, including images and catalogs, will be available with **no proprietary period** to the astronomical community of the **United States, Chile, and International Contributors**

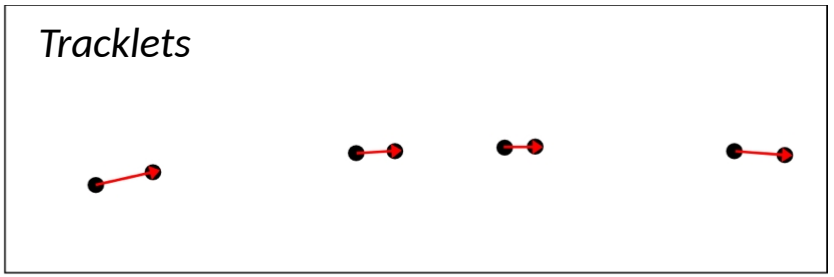
- Alerts to variable sources ("transient alerts") will be **available world-wide within 60 seconds**, using standard protocols

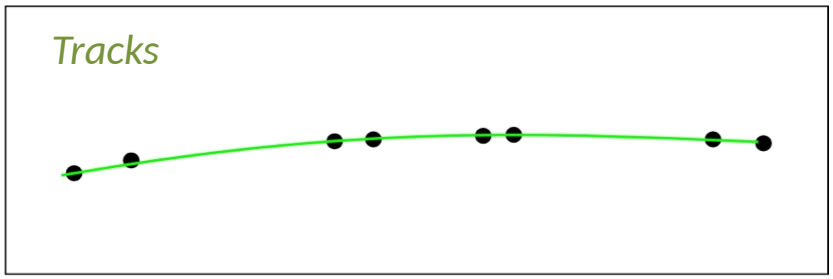
- LSST **data processing stack will be free software** (licensed under the GPL, v3-or-later)

LSST Asteroid Discovery Method: "2+2+2"

Requirement for reportable discovery: <u>at least</u> three pairs taken over three nights in a short (e.g., ~two week) period, fitting a Keplerian orbit (heliocentric).







Initial and Differential Orbit Determination.

Publication and Reporting to MPC.

Terminology:

- ***tracklets***: potential linkages in the same night (linear extrapolation)
- ***tracks***: potential linkages over three nights (quadratic fit)
- ***reportable discovery***: a track that unambiguously fits a Keplerian orbit within the astrometric uncertainties
- ***MOPS***: the software system that links detections into reportable discoveries

Note: some asteroids will also be immediately recognizable due to measurable trailing.

Level 1: Solar System Objects

- **Solar System objects will be identified and linked together based on compatibility of their observed positions with motion around the Sun.**
  - Enhanced variant of MOPS algorithm; advanced prototype in hand.

- **Planning to:**
  - Identify and link observations of Solar System objects
  - Measure their orbital elements
  - Measure their photometric properties
  - Expect to provide orbits for >= 70% of all NEOs brighter than H=22

- **Availability: within 24 hours of orbit determination**