LSST was designed to deliver in four key science areas

- Time domain science
  - Nova, supernova, GRBs
  - Source characterization
  - Instantaneous discovery
- Census of the Solar System
  - MBAs, NEOs, Comets
  - KBOs, Oort Cloud
- Mapping the Milky Way
  - Tidal streams
  - Galactic structure
- Dark energy and dark matter
  - Strong Lensing
  - Weak Lensing
  - Constraining the nature of dark energy

LSST was designed to deliver in four key science areas

- Time domain science
  - Nova, supernova, GRBs
  - Source characterization
  - Instantaneous discovery
- Census of the Solar System
  - MBAs, NEOs, Comets
  - KBOs, Oort Cloud
- Mapping the Milky Way
  - Tidal streams
  - Galactic structure
- Dark energy and dark matter
  - Strong Lensing
  - Weak Lensing
  - Constraining the nature of dark energy

These science cases flow down into science requirements for a monolithic, 10-year imaging survey (Ivezic et al 2008, arXiv:0805.2366)

One system can deliver on this science

The LSST system will include:

(i) an 8.4m (6.7m effective aperture) optical telescope with a 3.5-degree diameter field-of-view, a 3.2 billion pixel camera, and 6 broad-band, optical filters

(ii) a data facility that will process, archive, and distribute survey images, associated transient alerts, and calibrated catalogs, as well as calibration and other metadata.

We will deploy this system for a 10 year, time domain survey covering > 18,000 deg^2

LSST Science Requirements

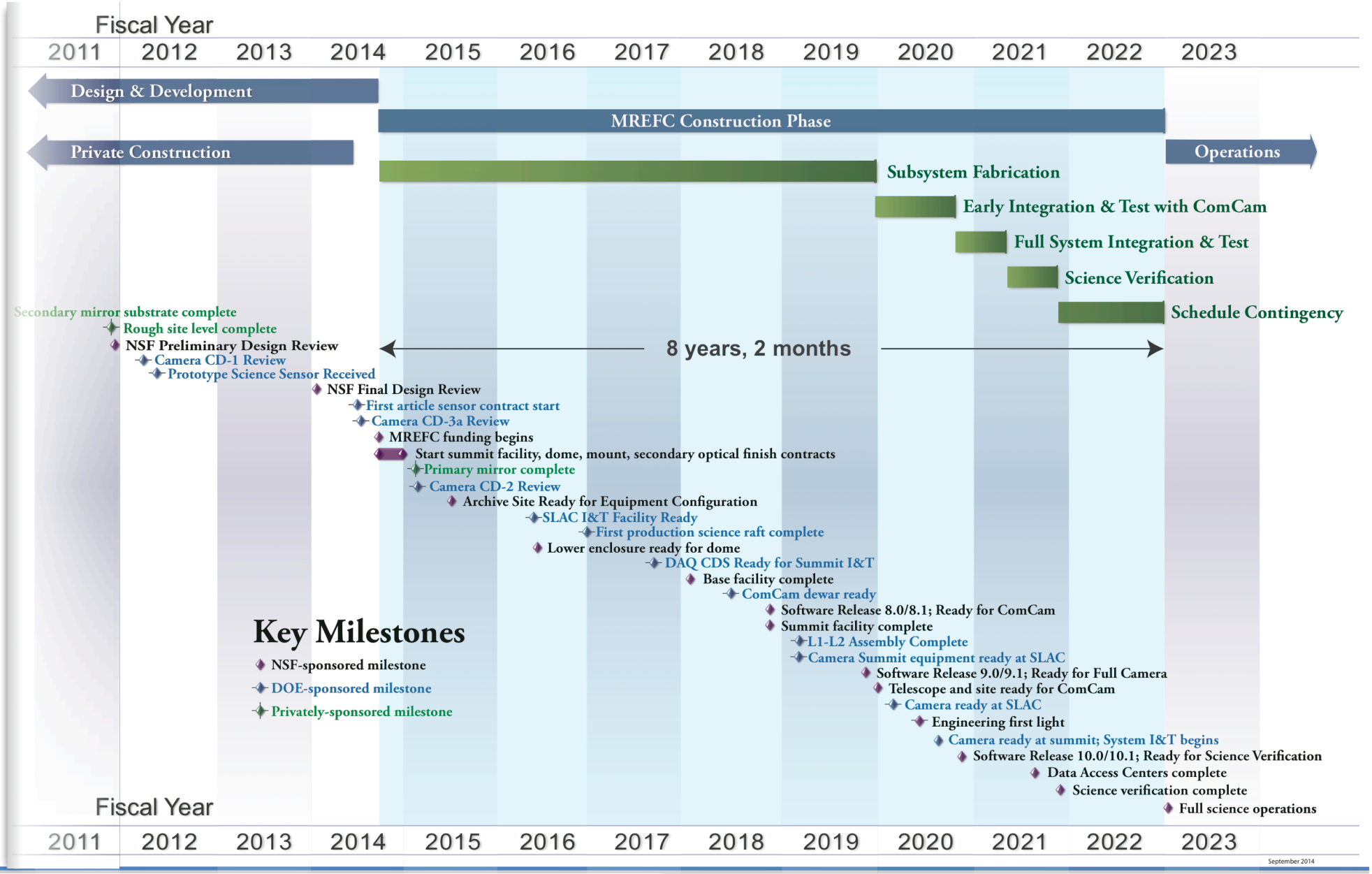
Survey Property	Performance
Main Survey Area	18000 sq. deg.
Total visits per sky patch	825
Filter set	6 filters (ugrizy) from 320 to 1050nm
Single visit	2 x 15s exposures, 2s readout
Single Visit Limiting Magnitude	u = 23.5; g = 24.8; r = 24.4; I = 23.9; z = 23.3; y = 22.1
Photometric calibration	2% absolute, 0.5% repeatability & colors
Median delivered image quality	~ 0.7 arcsec. FWHM
Transient processing latency	60 sec after last visit exposure
Data release	Full reprocessing of survey data annually

LSST is an interagency construction project

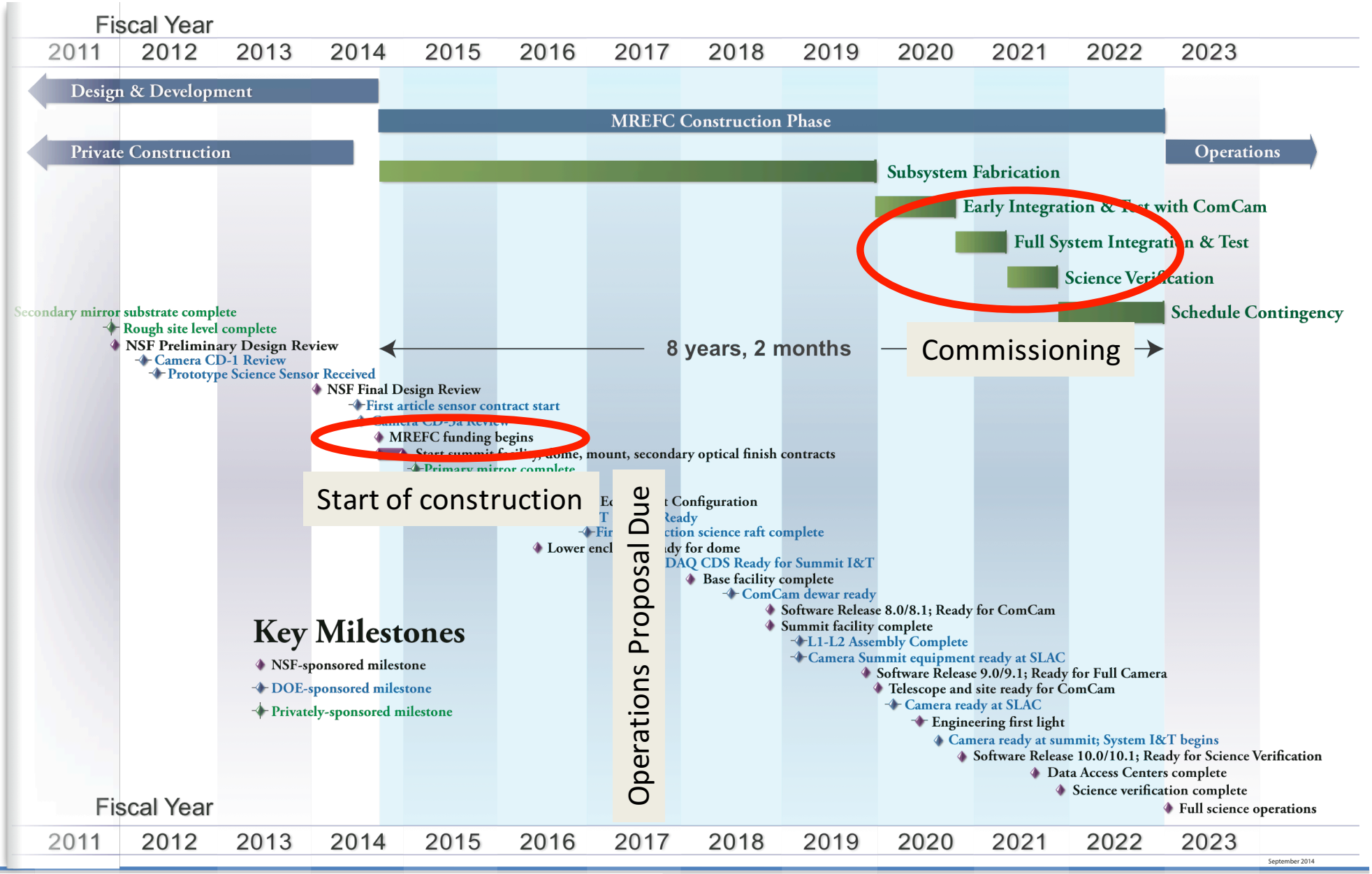
- **The National Science Foundation:**
  - Telescope and site facility construction, data management system, project management office, systems engineering, EPO
  - Major Research Equipment and Facility Construction (MREFC). Total not to exceed cost is $473M.
  - Under Cooperative Agreement with the Association of Universities for Research in Astronomy (AURA)
- **The Department of Energy:**
  - Camera fabrication.
  - Major Item of Equipment (MIE), through the Office of High Energy Physics in the Office of Science. Total projected cost is $168M.
  - SLAC National Accelerator Laboratory is the lead DOE lab.

The Project also received private funds (~$50M) that supported the primary/tertiary mirror, secondary mirror blank, preliminary site preparation, early sensor studies. The responsible organization is the LSST Corporation.

Project Schedule



Project Schedule



Survey Dataset Basics

Scale of the Dataset

Final Image Collection – All DRs	515 PB	All Data Releases Includes Virtual Data (475 PB)
Final Image Collection – DR11	114 PB	Data Release 11 (Year 10) Includes DR11 Virtual Data (88 PB) and all raw images from all years
Final Database	16 PB	Data Release 11 (Year 10) Includes Data, Indexes, and DB Swap
Final Disk Storage	375 PB	Archive and Base Sites
Final Tape Storage	121 PB	Single Site, Single Copy Only
Peak Number of Nodes	1750	Archive and Base Sites Compute and Database Nodes
Number of Alerts Generated	28 billion	Life of survey

- Virtual Data is data that is dynamically recreated on-demand from provenance information

LSST's Required Management of this Dataset

- **Archive Raw Data:** Receive the incoming stream of images that the Camera system generates to **archive the raw images**.

- **Process to Data Products:** **Detect and alert on transient events** within one minute of visit acquisition. Approximately once per year **create and archive a Data Release**, a static self-consistent collection of data products generated from all survey data taken from the date of survey initiation to the cutoff date for the Data Release.

- **Publish:** **Make all LSST data available** through an interface that uses community-accepted standards, and **facilitate user data analysis and production of user-defined data products** at Data Access Centers (DACs) and external sites.

Guiding Principles for Data Products

There are virtually infinite options on what quantities one can measure in images and store in catalogs.

1. **"Maximize science enabled by the catalogs"**
   - Working with images takes time and resources; a large fraction of LSST science cases should be enabled by just the catalog.
2. **"Provide simple but useful, commonly used, external or derived, quantities"**
   - Example: E(B-V) values for each object.
   - Example: Photo-z using well known, published, algorithms.
3. **Minimize information loss**
   - Provide (as much as possible) estimates of likelihood surfaces, not just single point estimators
4. **Provide and document the transformation (the software)**
   - Measurements are becoming increasingly complex and systematics limited; need to be maximally transparent about *how* they're done

Products to be Delivered by the LSST Project

Nightly (Level 1)

- 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.

Annual DRs (Level 2)

- A catalog of ~37 billion objects (20B galaxies, 17B stars), ~7 trillion single-epoch detections ("sources"), and ~30 trillion forced sources, produced annually, accessible through online databases.
- Deep co-added images.

Added Value (Level 3)

- Services and computing resources at the Data Access Centers to enable user-specified custom processing and analysis.
- Software and APIs enabling development of analysis codes.

Products to be Delivered by the LSST Project

Nightly (Level 1)

- 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.

See the Data Products Definition Document, http://ls.st/dpdd for full details

Annual DRs (Level 2)

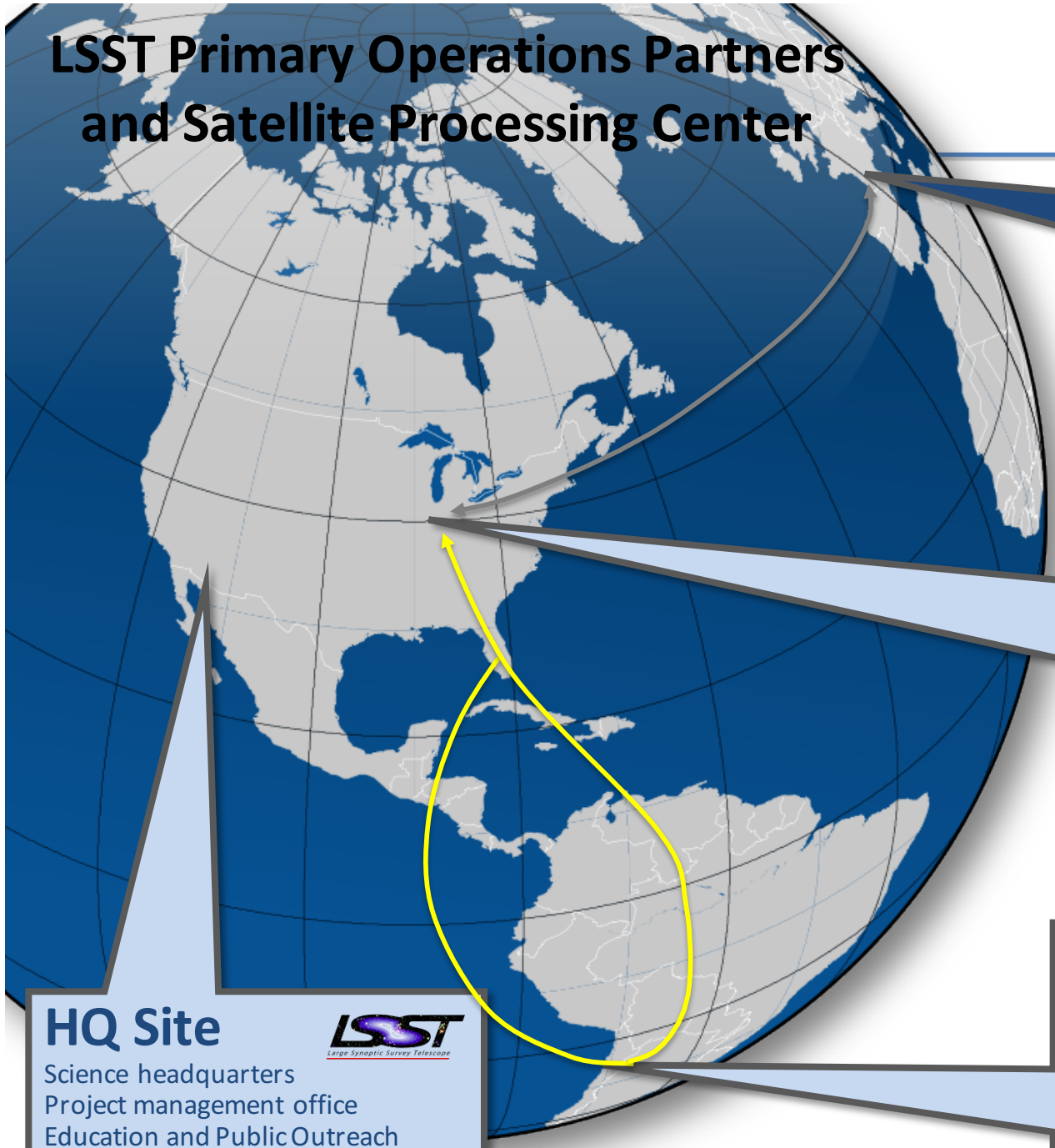
- A catal[illegible]n single-epoch detections ("sources"), and ~30 trillion forced sources, produced annually, accessible through online databases.
- Deep co-added images.

Added Value (Level 3)

- Services and computing resources at the Data Access Centers to enable user-specified custom processing and analysis.
- Software and APIs enabling development of analysis codes.

A data management system is under construction that will deliver these products and services to the scientific community

LSST Primary Operations Partners and Satellite Processing Center



French satellite center

(CC-IN2P3, Lyon, France)

Data Release Production (50%)
French DAC

Archive Site

Archive Center

Alert Production
Data Release Production (50%)
EPO Infrastructure
Long-term Storage (copy 2)

Data Access Center

Data Access and User Services

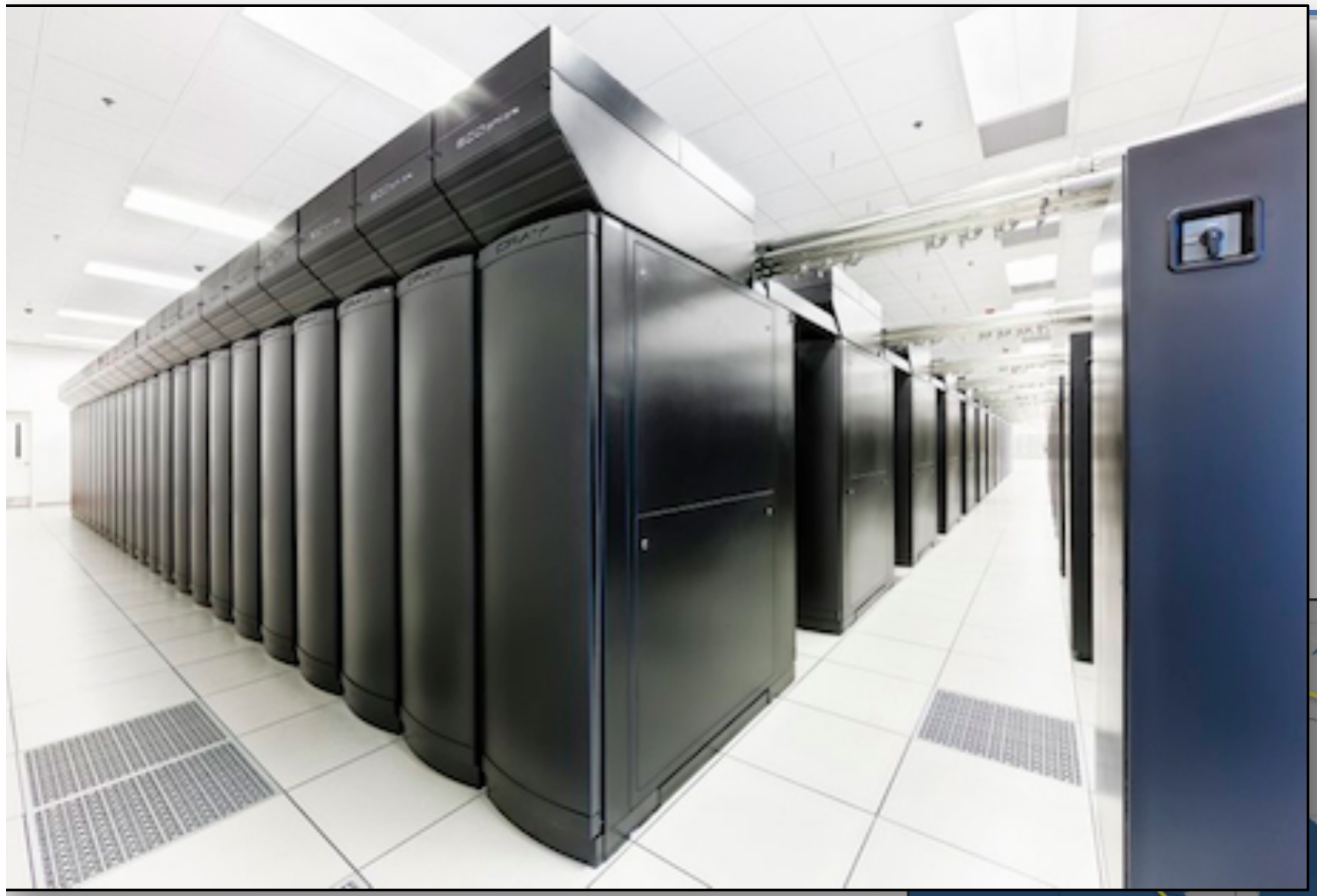
Summit and Base Sites

Telescope and Camera
Data Acquisition
Crosstalk Correction
Long-term storage (copy 1)
Chilean Data Access Center

HQ Site

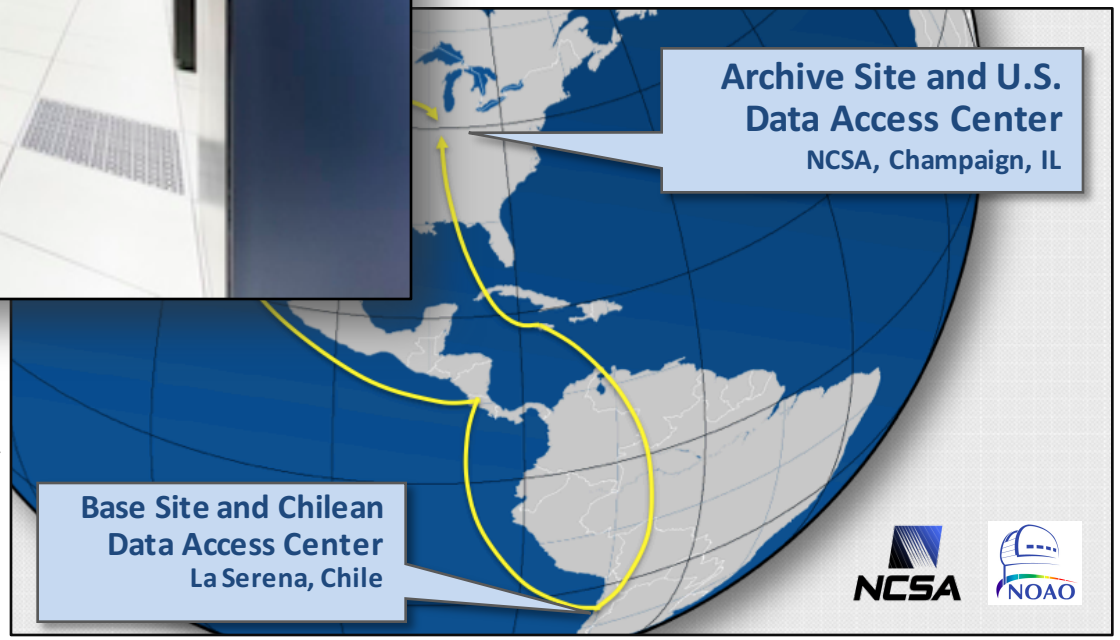
Science headquarters
Project management office
Education and Public Outreach

Infrastructure: Petascale Computing, Gbit Networks



The computing cluster at the **LSST Archive** (at NCSA) will run the processing pipelines.

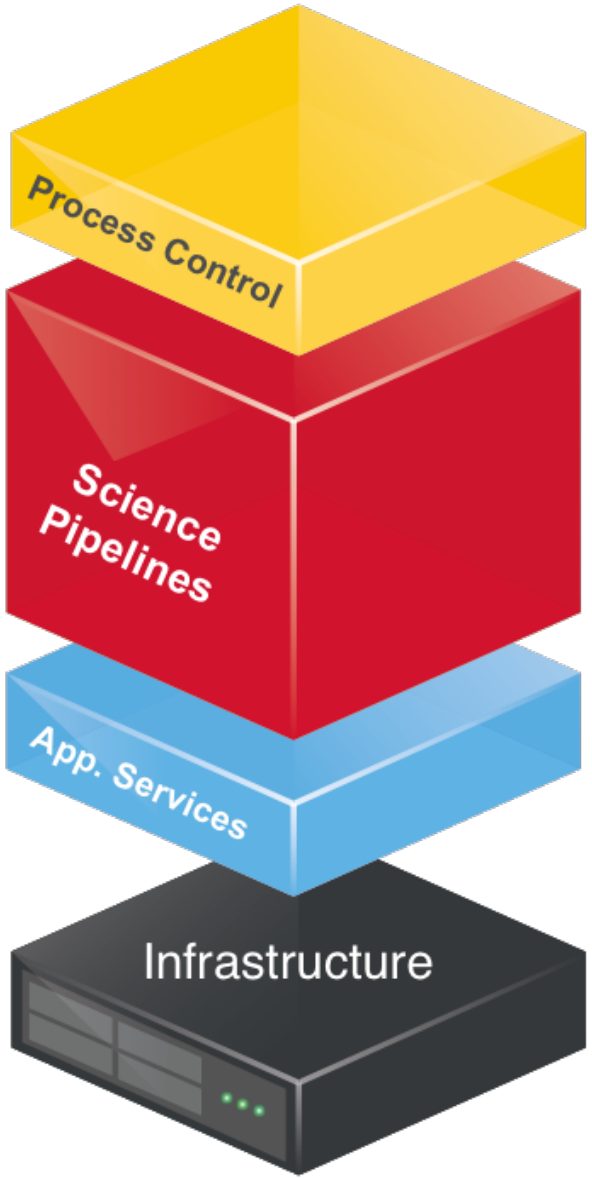
- Single-user, single-application, dedicated data center
- Process images in real-time to detect changes in the sky
- Produce annual data releases



Long Haul Networks to transport data from Chile to the U.S.

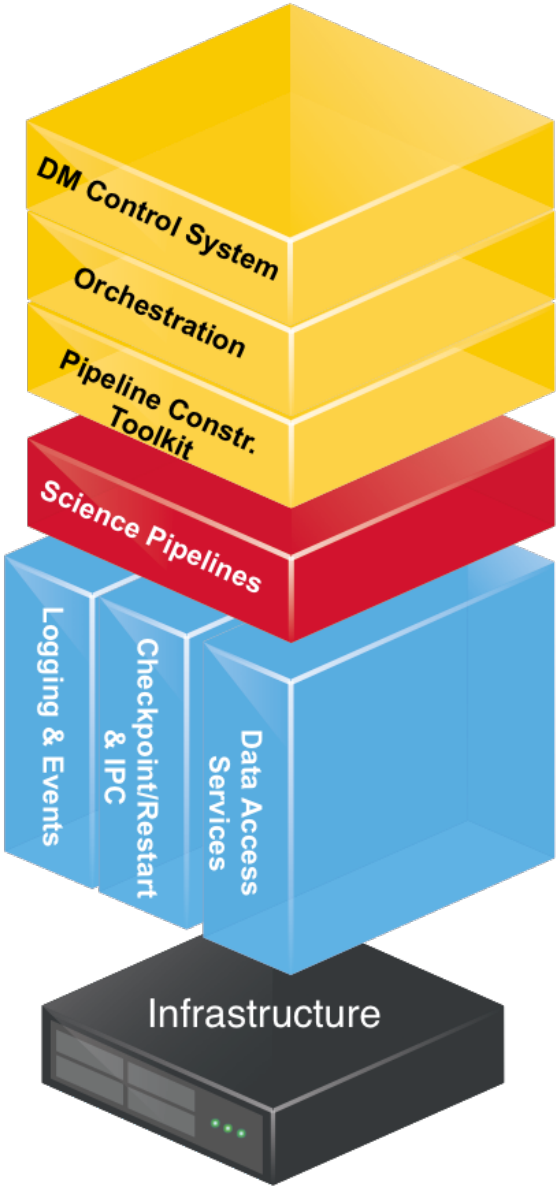
- 200 Gbps from Summit to La Serena (new fiber)
- 2x100 Gbit (minimum) for La Serena to Champaign, IL (protected, existing fiber)

Science Pipelines: Scientific Core of LSST DM



- *Science Pipelines* carry core scientific algorithms that process or analyze raw LSST data to generate output Data Products
- N.b. also referred to as "Applications"
- Variety of processing
  - Image processing
  - Measurement of source properties
  - Associating sources across space and time, e.g. for tracking solar system objects

Middleware Layer: Isolating Hardware, Orchestrating Software



Enabling execution of science pipelines on hundreds of thousands of cores.

- Frameworks to construct pipelines out of basic algorithmic components
- Orchestration of execution on thousands of cores
- Control and monitoring of the whole DM System

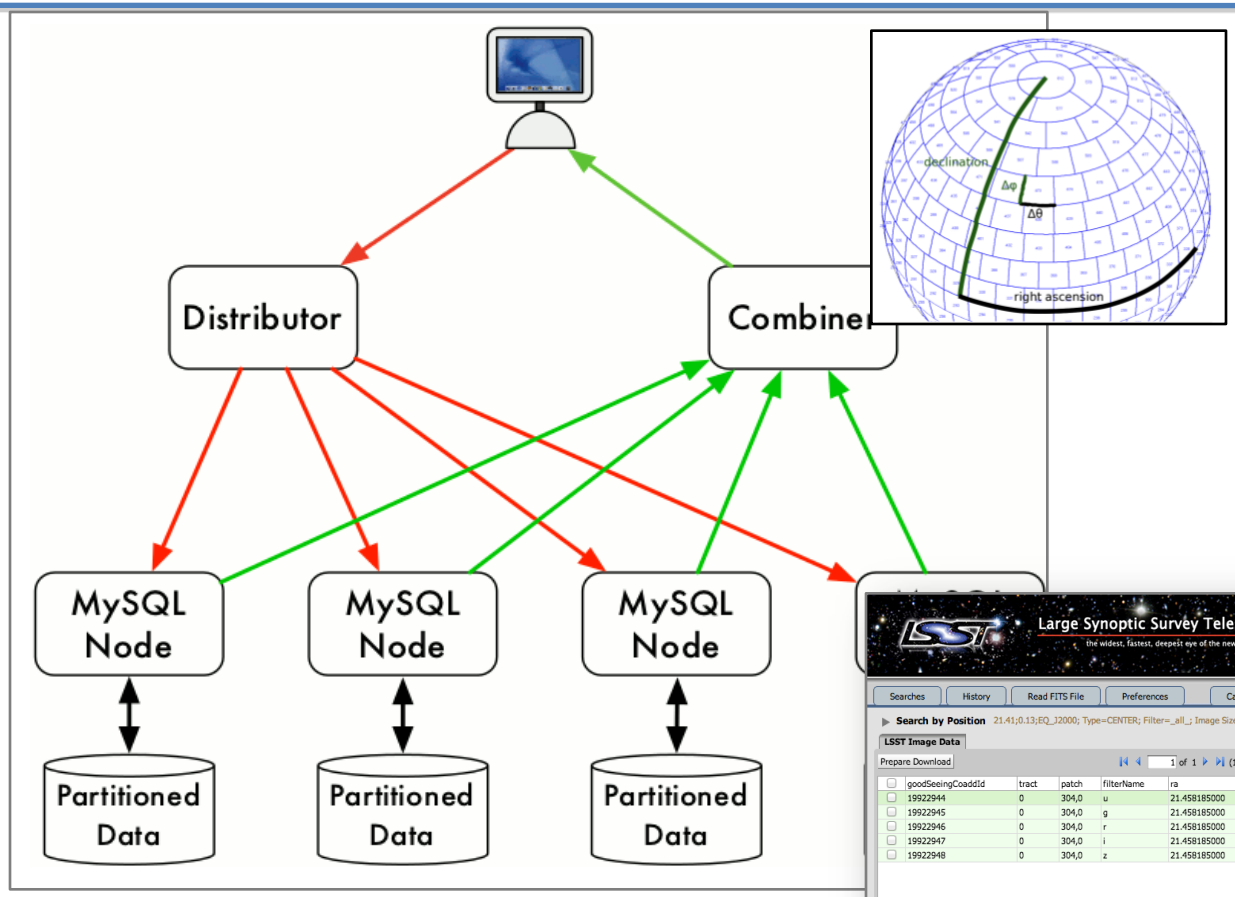
NCSA

Isolating the science pipelines from details of underlying hardware

- Services used by applications to access/produce data and communicate
- "Common denominator" interfaces handle changing underlying technologies

SLAC

Database and Science UI: Delivering to Users



Massively parallel, distributed, fault-tolerant ***relational database.***

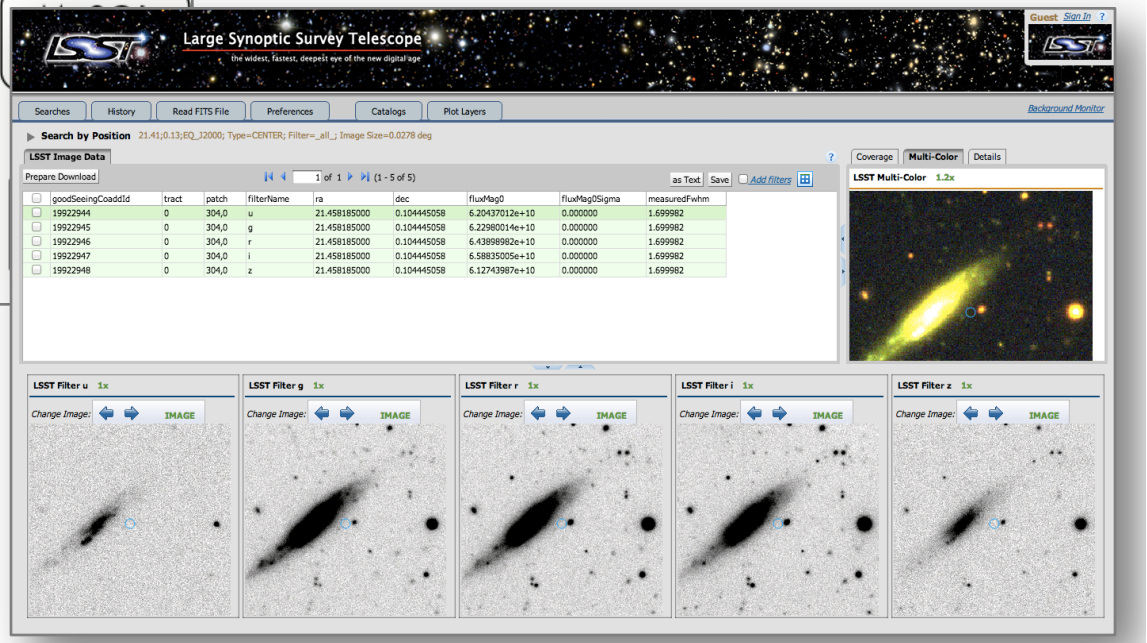
- To be built on existing, robust, well-understood, technologies (MySQL and xrootd)
- Commodity hardware, open source
- Advanced prototype in existence (qserv)

SLAC

Science User Interface *to enable the access to and analysis of LSST data*

- Web and machine interfaces to LSST databases
- Visualization and analysis capabilities

ipac



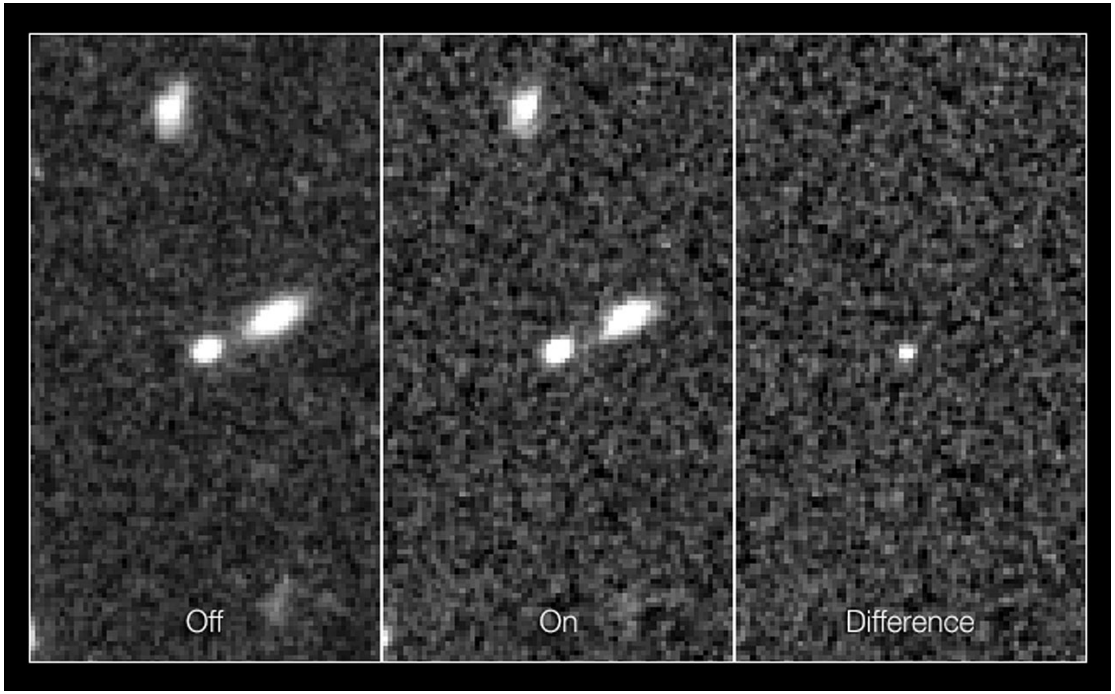
Nightly (Level 1) Data Products

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

- **Result: Alerts transmitted within 60 seconds of observing, enabling rapid follow-up**



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

Level 1: 60 second Transients Alerts

- **LSST computing is sized for 10M alerts/night (average), 10k/visit (average), 40k/visit (peak)**
  - Dedicated networking for moving data from Chile to the US
  - Dedicated image processing clusters
  - New image differencing pipelines with improved algorithms
- **Includes anything that changed wrt. deep template (variables, explosive transients, asteroids, etc.)**
- **Will measure and transmit with each alert:**
  - position
  - Flux (psf and aperture), size, and shape
  - light curves in all bands (up to a ~year; stretch: all)
  - variability characterization (eg., low-order light-curve moments, probability the object is variable)
  - cut-outs centered on the object (template, difference image)

LSST Limited Alert Filtering

- Based on models of the rates of observed asteroids as well as stellar variability (dominant sources of event alerts), we expect a high rate of alerts, approaching 10 million per night.
- A typical LSST user will not have sufficient bandwidth to receive this full stream. Size and bandwidth in the current LSST baseline will allow for transmission of ~three copies of the full data stream to public VOEvent brokers.
- Most end-users will be interested in only a subset that matches their scientific interest (e.g., SNe candidates, variable stars, or moving objects).
- To support selecting subsets of alerts, LSST will provide a basic alert filtering service. This service will let astronomers create simple *filters* that limit which alerts are ultimately forwarded to them.
- This is the Level 1 analog of querying the database in Level 2, a service we also provide. The (significantly smaller) returned subset will then be transmitted to the end-user for analysis.

Limitations of the LSST Alert Filtering Service

- The complexity and run time of user defined filters will be limited by available resources and may be throttled depending on load.
- The number of VOEvents transmitted to each user will be limited and dynamically throttled depending on load.
  - E.g., with a maximum of ~20 events per visit per user (20k/night), we can serve about ~500 simultaneous users at any one time utilizing total bandwidth equivalent to one full stream.
- No information beyond what is contained in the VOEvent packet will be available to user-defined filters (eg., no cross-matches to other catalogs, or other alert streams).
- We will not provide any astrophysical classification (eg., “is the light curve consistent with an RR Lyra?”, or “a Type Ia SN?”).

Public VOEvent Alert Brokers and Networks

- We anticipate that advanced, public, filtering services – VOEvent brokers – will be established by the community (e.g. the ANTARES project)

- 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, and (more generally) incorporation of other contextual information needed to decide on whether a transient is worth following up.

- Because of their advanced functionality, we expect these will be preferred by the end-users, compared to the more limited LSST filtering service.

- However, if such public brokering facilities fail to materialize, we expect the filtering service provided by LSST will be sufficient to enable initial Level 1 science.

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

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

Annual (Level 2) Data Products:

Enabling Deep Sky and High-Precision Astrophysics

Annual Data Releases

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

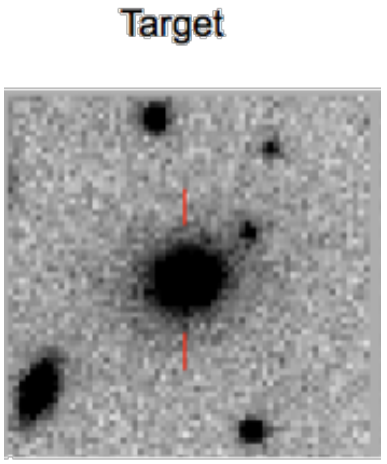
- **Made available in *Data Releases***
  - Annually, except in Year 1
    - Two DRs in the first year
  - **No proprietary period (US/Chile/International Contributors)**

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

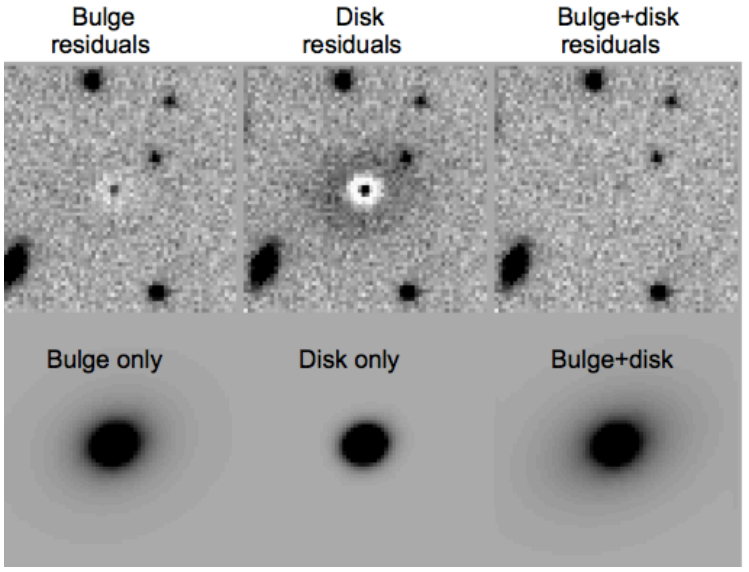
- **Accessing the catalogs**
  - QSERV Database and Science User Interface
  - Remote access APIs, VO protocols (e.g., Table Access Protocol)

Annual Data Release Catalogs

- **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
- **Metadata:** Deblend status, flags, etc.



LSST Science Book, Fig. 9.3



Data Release Processing (DRP)

DRP begins with processing, detection, and measurement on single frames, generating *Sources*.

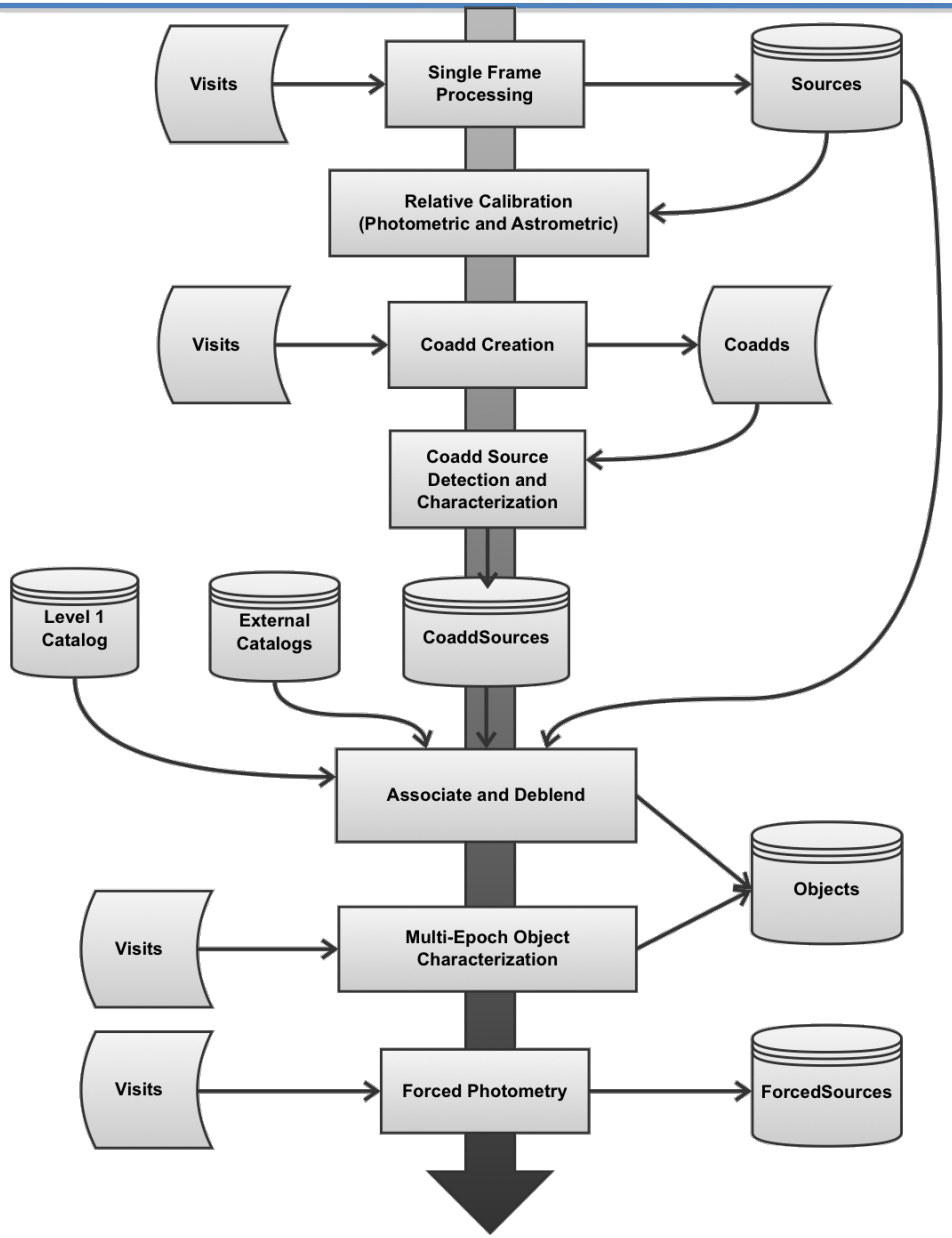
These are used to photometrically and astrometrically calibrate the survey.

A series of *coadds* is built next, where *Objects* are detected.

Detections on co-adds are *deblended* and *associated* to form a master object list.

The objects are simultaneously characterized in all observed epochs (*MultiFit*).

Time variability is characterized by independent measurement of *Forced Sources* in individual epochs.



Level 3 Services:

Enabling User-generated Data Products and Processing

Level 3: Enabling User-created Data Products

- The Project will provide the "Level 3 Service" to enable the community to create new "Level 3 Data Products" using LSST's software, services, or computing resources
- Level 3 Services will include:
  - User databases and workspaces ("mydb")
  - Making the LSST software stack available to end-users
  - Enabling user computing at the LSST data center: processing that will greatly benefit from co-location with the LSST data. Particularly important for science cases involving the images
- Example use cases:
  - Reprocessing images to search for SNe light echos
  - Characterization of diffuse structures (e.g., ISM)
  - Extremely crowded field photometry (e.g., globular clusters)
  - Implementing new measurement algorithms