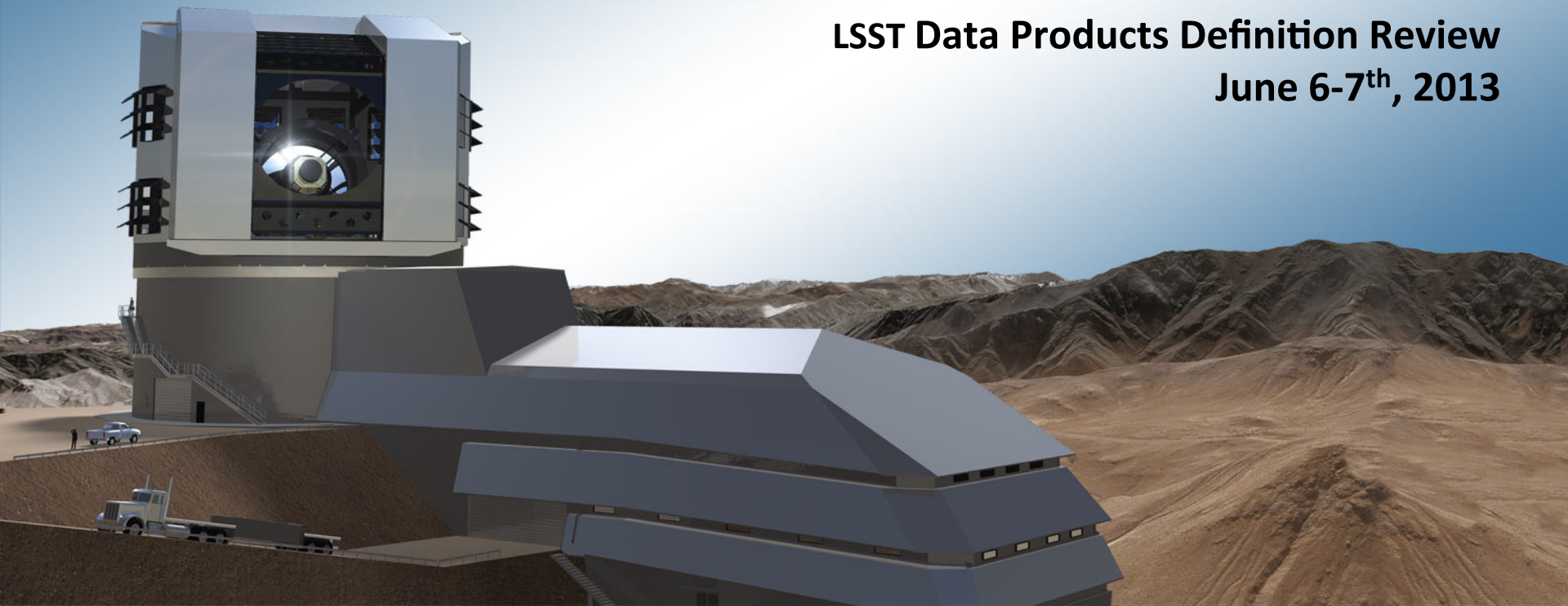
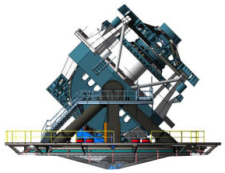
Level 1 Data Products

Mario Juric
LSST Data Management Project Scientist

LSST Data Products Definition Review
June 6-7th, 2013



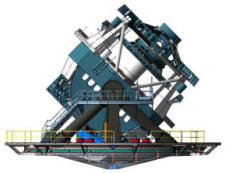


Addressing the Charge

1. *Does the definition of Level 1 Data Products satisfy the data product requirements laid out in the LSST Science Requirements Document?*

2. *Are LSST Transient Alert contents and distribution plans adequate to enable science cases depending on immediate follow-up?*

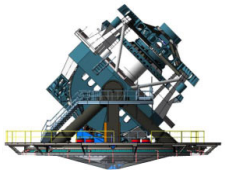
6. *Is the definition of all levels of LSST data products appropriately understood and mature to enable the start of construction of LSST Data Management systems (subject to formal flow-down to specification documents)?*



Presentation Overview

LSST DATA MANAGEMENT

- Definition of, and requirements for, Level 1 data products
- A Walkthrough of Level 1 Processing
- A Walkthrough of Level 1 Catalogs
- Level 1 Image Products
- Transient Alerts and Event Brokers
- Level 1 Data Reprocessing
- Open Issues



Level 1 Data Products Overview

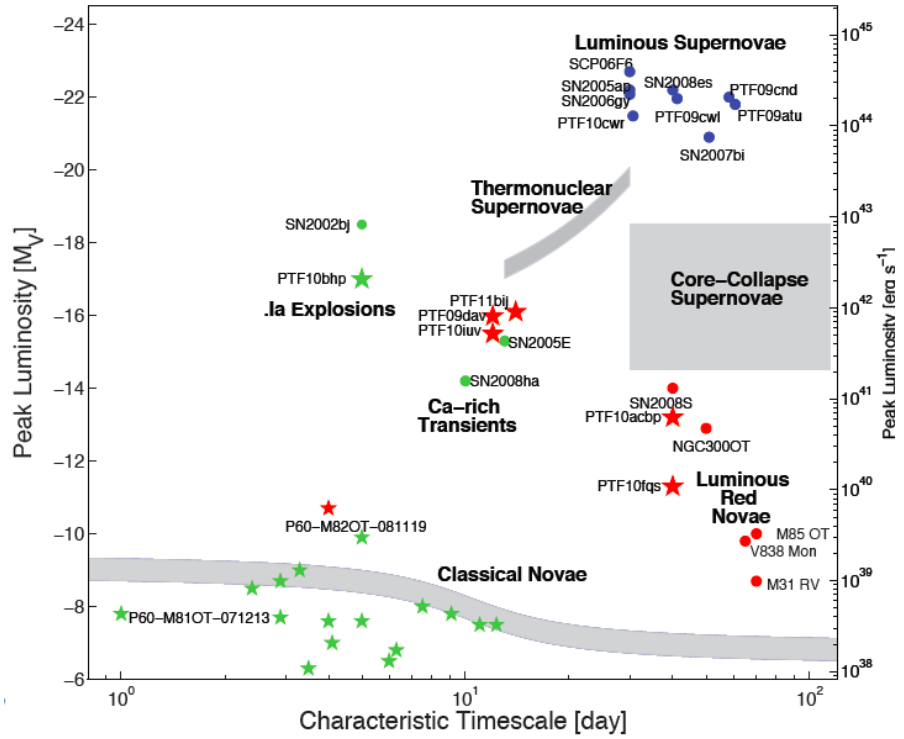
- **Primary purpose:**
  - Satisfy science cases requiring rapid identification and follow-up (transients, fast-moving NEOs, etc.)

Level 1 Data Products Overview

- **Primary purpose:**
  - Satisfy science cases requiring rapid identification and follow-up (transients, fast-moving NEOs, etc.)

- **Transient science**
  - Nova, supernova, GRBs
  - Source characterization
  - Instantaneous discovery

- Nearby Solar System Objects
  - NEOs, PHAs

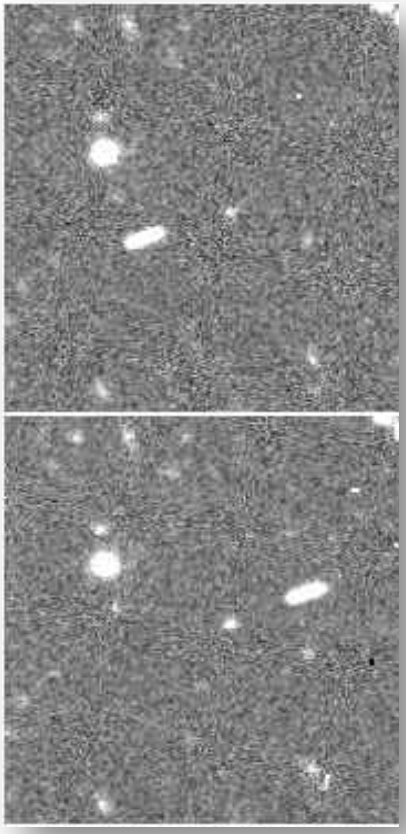


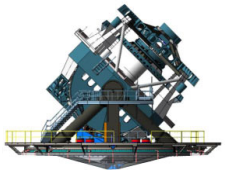
Level 1 Data Products Overview

- **Primary purpose:**
  - Satisfy science cases requiring rapid identification and follow-up (transients, fast-moving NEOs, etc.)

- Transient science
  - Nova, supernova, GRBs
  - Source characterization
  - Instantaneous discovery

- **Nearby Solar System Objects**
  - NEOs, PHAs

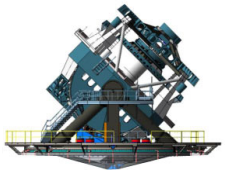




Level 1 Data Products Overview

- **Primary purpose:**
  - Satisfy science cases requiring rapid identification and follow-up (transients, fast-moving NEOs, etc.)

- **Use Cases:**
  - Discover and characterize objects that have changed flux, shape, or moved
  - Identify Solar System objects



Level 1 Data Product Requirements

LSST DATA MANAGEMENT

p. 32, LSST SRD

3.5 Data Processing and Management Requirements

Detailed requirements on data processing and management will be described in the LSST System Requirements Document (for example, specifications for catalog completeness and reliability). Here, only a rough guidance is provided. There will be three main categories of data products:

- **Level 1** data products are generated continuously every observing night, including alerts to objects that have changed brightness or position.

Level 1 Latency

p. 33, LSST SRD

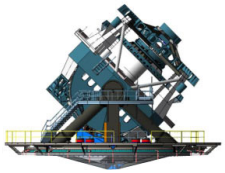
Specification: Images and catalog data will be released to publicly-accessible repositories as soon as possible after they are obtained. This latency, and the exact form of the data to be continuously released, are left unspecified at this time pending further discussion within the project. Data on likely optical transients, however, will be released with a latency of at most OTT1 minutes.

Alert Latency

p. 34, LSST SRD

Quantity	Design Spec	Minimum Spec	Stretch Goal
DRT1 (year)	1.0	2.0	0.5
OTT1 (min)	1.0	2.0	0.5

Table 28: Requirements for the data release cadence and for the transient reporting latency.



Level 1 Alert Requirements

LSST DATA MANAGEMENT

Contents

Throughput

Filtering

p. 34, LSST SRD

The fast release of data on likely optical transients will include measurements of position, flux, size and shape, using appropriate weighting functions, for all the objects detected above transSNR signal-to-noise ratio in difference images (design specification: 5). The data stream will also include prior variability information and data from the same night, if available. The prior variability information will at the very least include low-order light-curve moments and probability that the object is variable, and ideally the full light curves in all available bands.

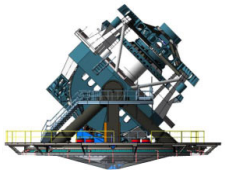
Specification: The system should be capable of reporting such data for at least transN candidate transients per field of view and visit (Table 29).

Quantity	Design Spec	Minimum Spec	Stretch Goal
transN	10^4	10^3	10^5

Table 29: The minimum number of candidate transients per field of view that the system can report in real time.

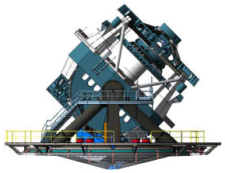
p. 34, LSST SRD

The users will have an option of a query-like pre-filtering of this data stream in order to select likely candidates for specific transient type. Users may also query the LSST science database at any time for additional information that may be useful, such as the properties of static objects that are positionally close to the candidate transients. Several pre-defined filters optimized for traditionally popular transients, such as supernovae and microlensed sources, will also be available, as well as the ability to add new pre-defined filters as the survey continues.



Level 1 Data Products Overview

- **Primary purpose:**
  - Satisfy science cases requiring rapid identification and follow-up (transients, fast-moving NEOs, etc.)
- **Use Cases:**
  - Discover and characterize objects that have changed flux, shape, or moved
  - Identify Solar System objects
- **Requirements:**
  - Measure position, flux, size and shape
  - Publish image and catalog data “as soon as possible”
  - Alert on discoveries within 60 seconds of end of exposure (visit)
  - Allow end-users to filter the alert stream



Level 1 Data Products Overview

DATA MANAGEMENT

- **Primary purpose:**
  - Satisfy science cases requiring rapid identification and follow-up (transients, fast-moving NEOs, etc.)
- **Use Cases:**
  - Discover and characterize objects that have changed flux, shape, or moved
  - Identify Solar System objects
- **Requirements:**
  - Measure position, flux, size and shape
  - Publish image and catalog data “as soon as possible”
  - Alert on discoveries within 60 seconds of end of exposure (visit)
  - Allow end-users to filter the alert stream
- **Method:**
  - Detection: Image differencing
  - Characterization: Measure on difference image, direct image, summarize.

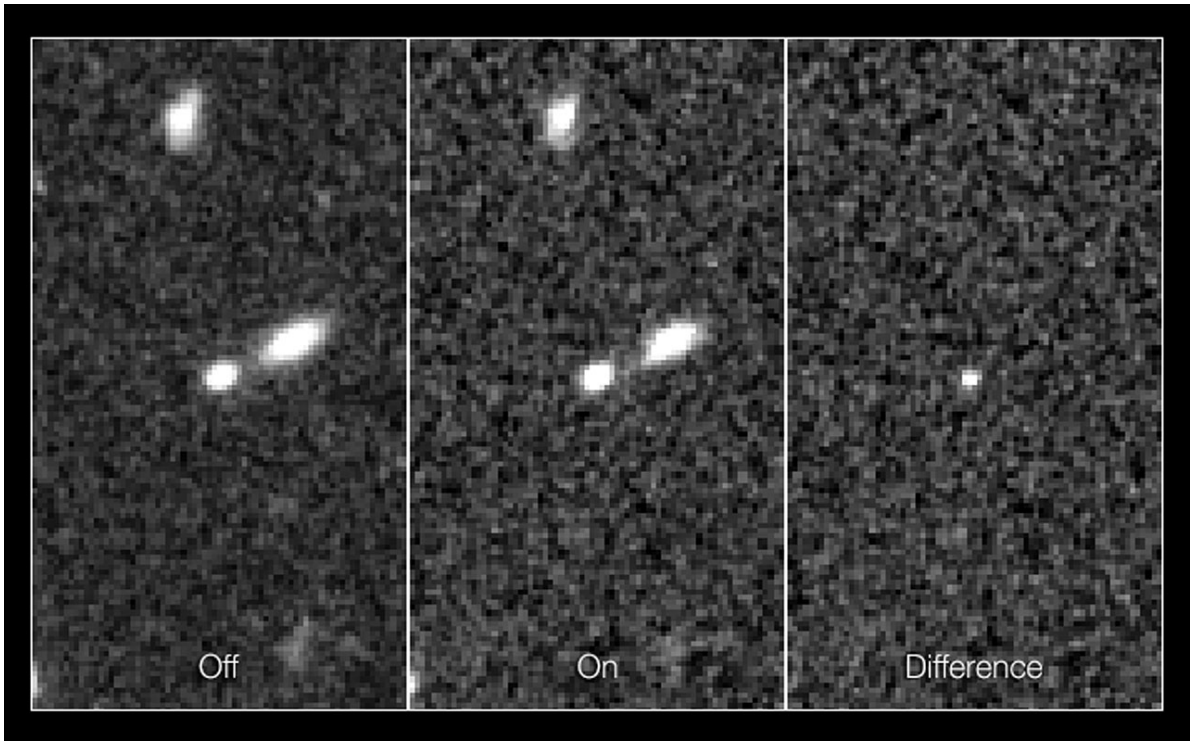
Level 1 Data Processing

Section 4.1 of the DPDD

- Detection performed only on visit differenced against a deeper *template*
- Measurement performed on the *difference image* and *direct image*
- Associated with pre-existing observations and stored in a relational database

- Alerts transmitted to enable rapid follow-up

- Daytime: Identify new Solar System objects by associating discovered sources assuming Sun-centered Keplerian orbits.

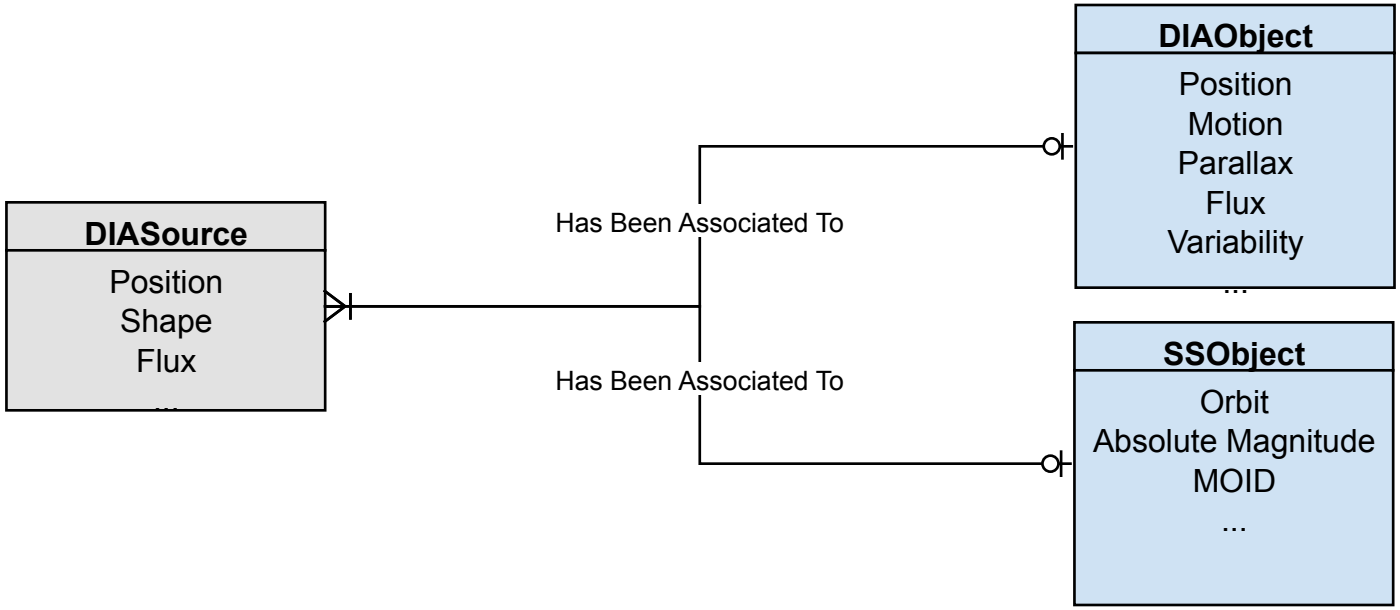


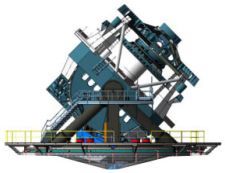
CANDELS (http://www.spacetelescope.org/images/heic1306d/)

Primary Catalog Outputs

Section 4.1 of the DPDD

- Level 1 processing will populate the Level 1 database tables, including:
  - `DIASource`: table of sources detected on difference images
  - `DIAObject`: objects and their characteristics, inferred from DIASource detection
  - `SSObject`: the catalog of orbits of objects in the Solar System



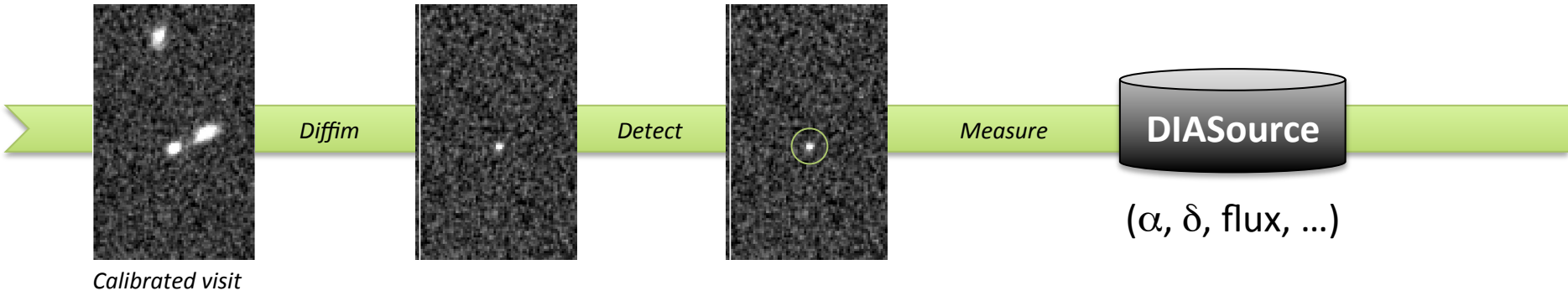


Introducing Some Jargon

- In normal survey mode, LSST will operate by capturing two back-to-back, 15-second exposures for each pointing.
  - The two exposures are referred to as ***snaps***
  - They are combined to a ***visit***, which is the basic input image product for Level 1 and 2 data products
  - The primary purpose of the snaps is to enhance cosmic ray rejection
    - They're not to be confused with 30 to 90 minute ***revisits***, scheduled to support Solar System science.

Level 1 Pipeline: Source Detection

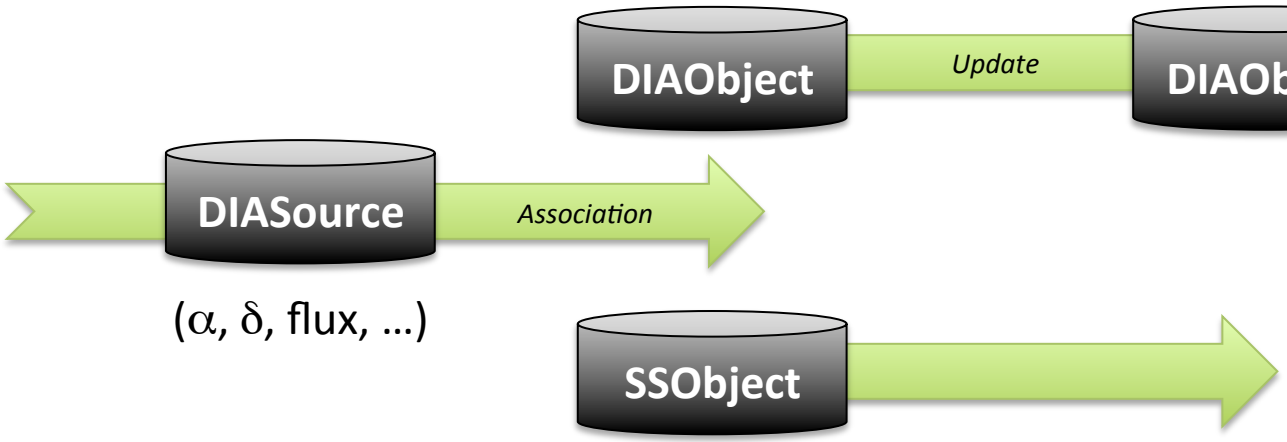
Section 4.2 of the DPDD



- A visit is acquired and reduced (cosmic ray rejection, combining of snaps, bias subtraction, flat fielding, etc.).
- The visit image is then differenced against the appropriate template (for example, a template approximately matching the seeing and airmass of the visit)
- DIASources are detected
- DIASource position, flux, and shape are measured with multiple algorithms
  - The details will be discussed in upcoming slides
- The list is stored to a relational database for Level 1 data products (into the DIASource table)

Level 1 Pipeline: Source Association

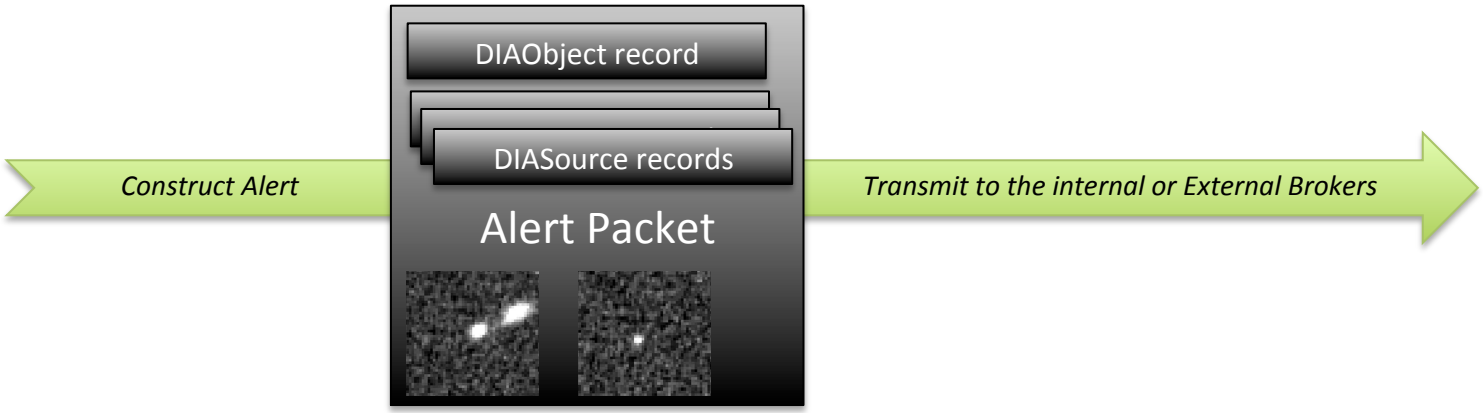
Section 4.2 of the DPDD



- The Level 1 database is searched for a match with an existing DIAObject or an SSObject. New DIAObject is created if no match is found.
  - If the DIASource has been associated with an SSObject, we issue an alert and stop. All further SSObject processing (eg., orbit recomputation) occurs in daytime.
  - Otherwise, the associated DIAObject measurements are updated with new data (centroids, light-curves, etc.).
- The Level 2 database is searched for Objects close to the DIAObject, out to some maximum radius, and their IDs are stored in the DIAObject record.
- Forced photometry is performed for all DIAObjects overlapping the field of view to which a DIASource from this visit has *not* been associated.

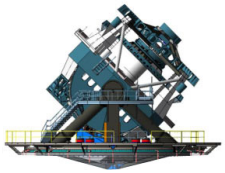
Level 1 Pipeline: Issuing Alerts

Section 4.2 of the DPDD



- An alert packet is constructed that includes all previous DIASource measurements, all information contained in the DIAObject record, image cut-outs, and all metadata necessary to query the Level 1 database for a copy of this alert.

- We plan to use VOEvent as the format for alert transmission
- The packet will be transmitted to VOEvent Brokers
  - We expect the community to provide sophisticated brokers with classification engines, cross-match capabilities to other catalogs, etc.
  - LSST will provide a default, limited, broker (to be discussed later).
- The alerts will be transmitted within 60 seconds of the end of visit, per SRD req.



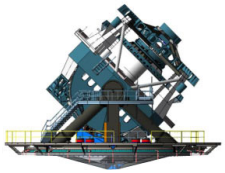
Level 1 Pipeline: Solar System Processing

LSST DATA MANAGEMENT

Section 4.2.2 of the DPDD

- Processing related to Solar System objects is planned to occur in daytime (after the night of observing has ended). Note that trailed objects will already have been alerted on.

- Steps:
  1. The orbits and physical properties of all SSObjects reobserved on the previous night are recomputed. External orbit catalogs (or observations) may be used to improve orbit estimates. Updated data are entered to the SSObjects table.
  2. All DIASources detected on the previous night, that have not been matched at a high confidence level to a known DIAObject, SSObject, or an artifact, are analyzed for potential pairs, forming tracklets.
  3. The collection of tracklets collected over the past 30 days is searched for subsets forming tracks consistent with being on the same Keplerian orbit around the Sun.
  4. For those that are, an orbit is fitted and a new SSObject table entry created. DIASource records are updated to point to the new SSObject record.
  5. Precovery linking is attempted for all SSObjects whose orbits were updated in this process. Where successful, SSObjects (orbits) are recomputed as needed.



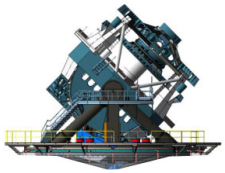
Level 1 Catalogs: DIASources

LSST DATA MANAGEMENT

Table 1, p. 17, of the DPDD

- Includes measures of:
  - Position
  - Shape (adaptive Gaussian moments; Bernstein & Jarvis 2002)
  - Model fits:
    - Point source model
      - Measure of flux and position assuming the object is a stationary point source
    - Trailed source model
      - Measure of flux, position, and direction of motion, assuming an object moves sufficiently fast to appear as a trail in the image. Designed for Solar System objects.
    - Dipole model fit
      - Fit the source with a "dipole" model, a positive next to a negative point source
      - NOTE: Missing from Table 1 of the version of the document being reviewed.
  - Direct image point source photometry

- Some DIASources will be false positives (unrecognized bad columns, diffraction spikes, etc.). We will flag suspect DIASources
- We plan to size the processing for 2-5x the expected true DIASource rate.



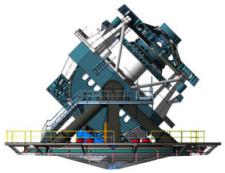
Level 1 Catalogs: DIAObjects

LSST DATA MANAGEMENT

Table 2, p. 21, of the DPDD

- Characterization of the underlying astrophysical objects detected in difference images
  - Computed from associated DIASource records (recomputed as needed)
  - Primary goal: Enable object classification

- Include:
  - Fits of position, parallax, and proper motion
  - Mean point source flux (difference image and direct image)
  - Variability characterization (e.g., Richards et al. 2011 parameters)
  - Pointers to nearby objects in the Level 2 catalog



Level 1 Catalogs: SSObjects

LSST DATA MANAGEMENT

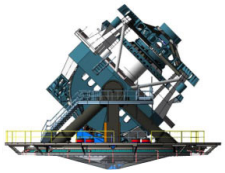
Table 3, p. 22, of the DPDD

- Catalog of Solar System objects
  - Asteroids, comets, KBOs, etc.

- Includes:
  - Orbital elements
  - MOID
  - Estimates of mean absolute magnitude (H, G)
    - Estimates performed in LSST bands

- No taxonomical classification, shape reconstruction
  - Outside the scope of the Project

- Convenience functions to compute the phase angle, reduced, H(α), and absolute, H, asteroid magnitudes will be provided by the Level 1 database.

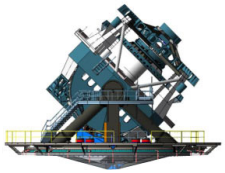


Level 1 Image Data Products

LSST DATA MANAGEMENT

Section 4.4 of the DPDD

- Raw and processed visit images will be made available for download no later than 24 hours from the end of visit acquisition.
- Difference images will be made available for download no later than 24 hours from the end of visit acquisition.
- Note: Raw, processed, and difference images will will remain accessible with low-latency (seconds from request to start of download) for at least 30 days, with slower access afterwards (minutes to hours).
- Templates for difference image analysis will be available for download no later than the first time they're used
- Note: Small cut-outs (aka. postage stamps) will be included with transient alerts (next slide).



Reporting Time-domain Events

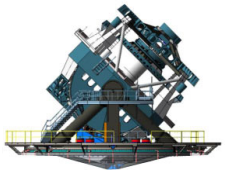
LSST DATA MANAGEMENT

Section 4.5 of the DPDD

- For each detected DIASource, LSST will emit an "Event Alert" within 60 seconds of the end of visit (defined as the end of image readout from the LSST Camera).

- Each alert include the following:
  - Alert and database ID: IDs uniquely identifying this alert.
  - The DIASource record that triggered the alert
  - The entire DIAObject (or SSObject) record
  - All previous DIASource records
  - 30x30 pixel cut-out of the difference image (FITS)
  - 30x30 pixel cut-out of the template image (FITS)

- 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)



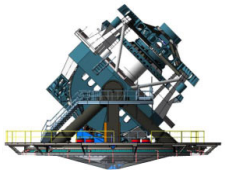
Event Brokers

DATA MANAGEMENT

Section 4.5.2 of the DPDD

- It has been understood for a number of years (dating back to pre-PDR) that it is *outside the scope of the LSST Project to develop and provide advanced event brokering facilities* (for example, brokers employing machine learning classifiers to filter the most interesting events in the event stream)
  - The rationale is that this is an unsolved, open-ended, cutting-edge research problem where no one solution is likely to satisfy all use cases. Developing such techniques does not fit the LSST's role as a *facility*.

- Nevertheless, the Project *does* have a mandate to provide a basic filtering facility:
  - *"The users will have an option of a query-like pre-filtering of this data stream in order to select likely candidates for specific transient type. Users may also query the LSST science database at any time for additional information that may be useful, such as the properties of static objects that are positionally close to the candidate transients. Several pre-defined filters optimized for traditionally popular transients, such as supernovae and microlensed sources, will also be available, as well as the ability to add new pre-defined filters as the survey continues."*
    – LSST SRD, Section 3.5, p. 34



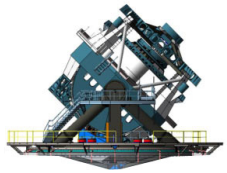
LSST Event Broker

LSST DATA MANAGEMENT

Section 4.5.2 of the DPDD

- To directly serve the end-users, LSST will provide a basic, limited capacity, alert filtering service. This service will run at the LSST U.S. Archive Center. It will let astronomers create simple filters that limit what 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.

- Limitations:
  - No information beyond what is contained in the VOEvent packet will be available to filter on.
  - The number of VOEvents transmitted to each user will be limited (to not less than 20, per visit).
  - The complexity and run time of user defined filters will be limited by available resources; execution latency will not be guaranteed.
  - The total number of simultaneous subscribers will be limited.



LSST Event Broker

Section 4.5.2 of the DPDD

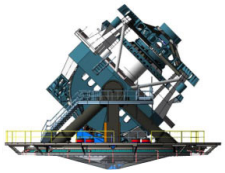
- Such a broker ensures that LSST transients science will not be "dropped to the floor" if there are no community-operated brokers ready in 2021.
- Despite its stated limitations, it is close to current state-of-the-art, and likely to be sufficient for a large fraction of initial of use cases.

- It satisfies the SRD requirements (not just the letter, but the spirit as well).

- Filter example:

```
# 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
```



Level 1 DP Annual Reprocessing

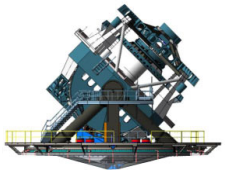
DATA MANAGEMENT

Section 4.3.5 of the DPDD

- Level 1 data products will be reprocessed annually
  - Full reprocessing (improved templates, difference images, new detection, association, etc.)

- Once reprocessing finishes and catches up to current date, the current Level 1 database and software will be switched for the reprocessed database (and software)
  - Alerts will not be reissued

- Side-effects:
  - All IDs will change, source-object associations will change, downstream classifications may change, etc.
    - We are currently assuming that outside event brokers will provide old-new ID matching services
  - Outweighed by the benefits: algorithmic consistency, improved association, improved measurement precision, better artifact rejection/classification, etc.
  - Reflects the true nature of any catalog: our best *interpretation* of the data at hand. Not much different from current practice (eg., SDSS).



(Selected) Open Issues

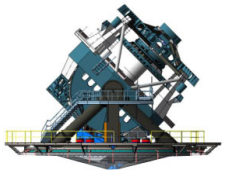
LSST DATA MANAGEMENT

Section 4.6 of the DPDD

- *Should we measure on individual snaps (or their difference)?* Is there a demonstrable science case requiring immediate follow-up that would be triggered by flux changes over a 15 second period?
  - The technical cost is in additional memory (up to 2x) needed to keep both snaps while processing is ongoing

- *Should we associate alerts with external catalogs?* LSST will have a copy of many external catalogs (or parts thereof), and could in principle provide limited positional association. Right now, it's up to the down stream brokers to perform such associations.

- *When should we (if ever) stop performing forced photometry on positions of DIAObjects?* Depending on the rate of false positives, unidentified artifacts, or unrecognized Solar System objects, the number of forced measurements may dramatically grow over time.

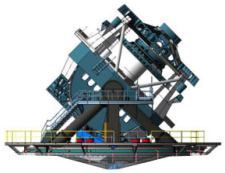


Summary: Addressing the Charge

1. *Does the definition of Level 1 Data Products satisfy the data product requirements laid out in the LSST Science Requirements Document?*

2. *Are LSST Transient Alert contents and distribution plans adequate to enable science cases depending on immediate follow-up?*

6. *Is the definition of all levels of LSST data products appropriately understood and mature to enable the start of construction of LSST Data Management systems (subject to formal flow-down to specification documents)?*



Questions?