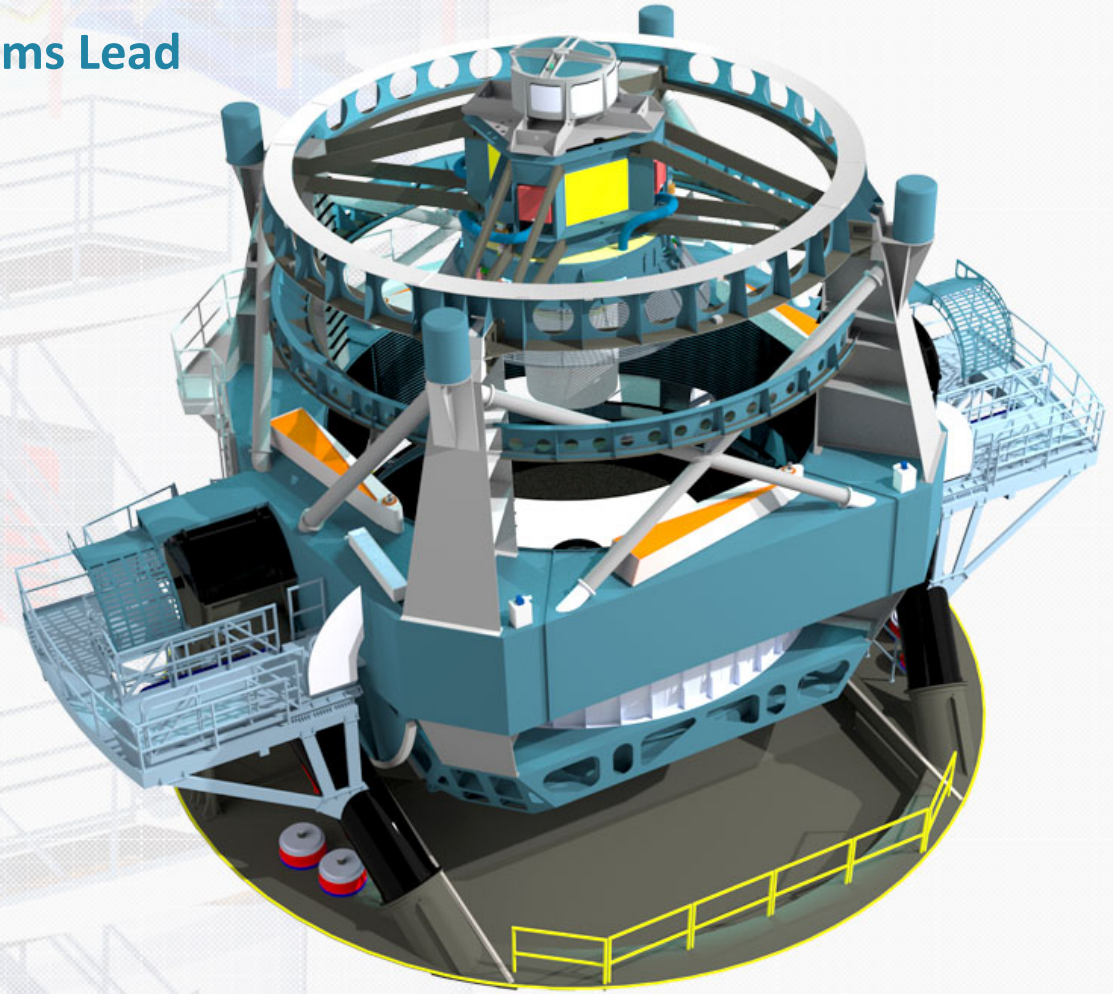
LSST Data Management plans: Pipeline outputs and Level 2 vs. Level 3

~~Mario Juric~~ Robert Lupton

LSST DM ~~Project Scientist~~ Algorithms Lead



LSST SAC

Data Product Requirements: Level 1,2,3

p. 32, LSST SRD, http://ls.st/srd
Details in the DPDD

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.

. . . Nightly

- **Level 2** data products will be made available as annual Data Releases and will include images and measurements of positions, fluxes, and shapes, as well as variability information such as orbital parameters for moving objects and an appropriate compact description of light curves.

. . . Annual

- **Level 3** data products will be created by the community, including project teams, using suitable Applications Programming Interfaces (APIs) that will be provided by the LSST Data Management System. The Data Management System will also provide at least 10% of its total capability for user-dedicated processing and user-dedicated storage. The key aspect of these capabilities is that they will reside "next to" the LSST data, avoiding the latency associated with downloads. They will also allow the science teams to use the database infrastructure to store their results.

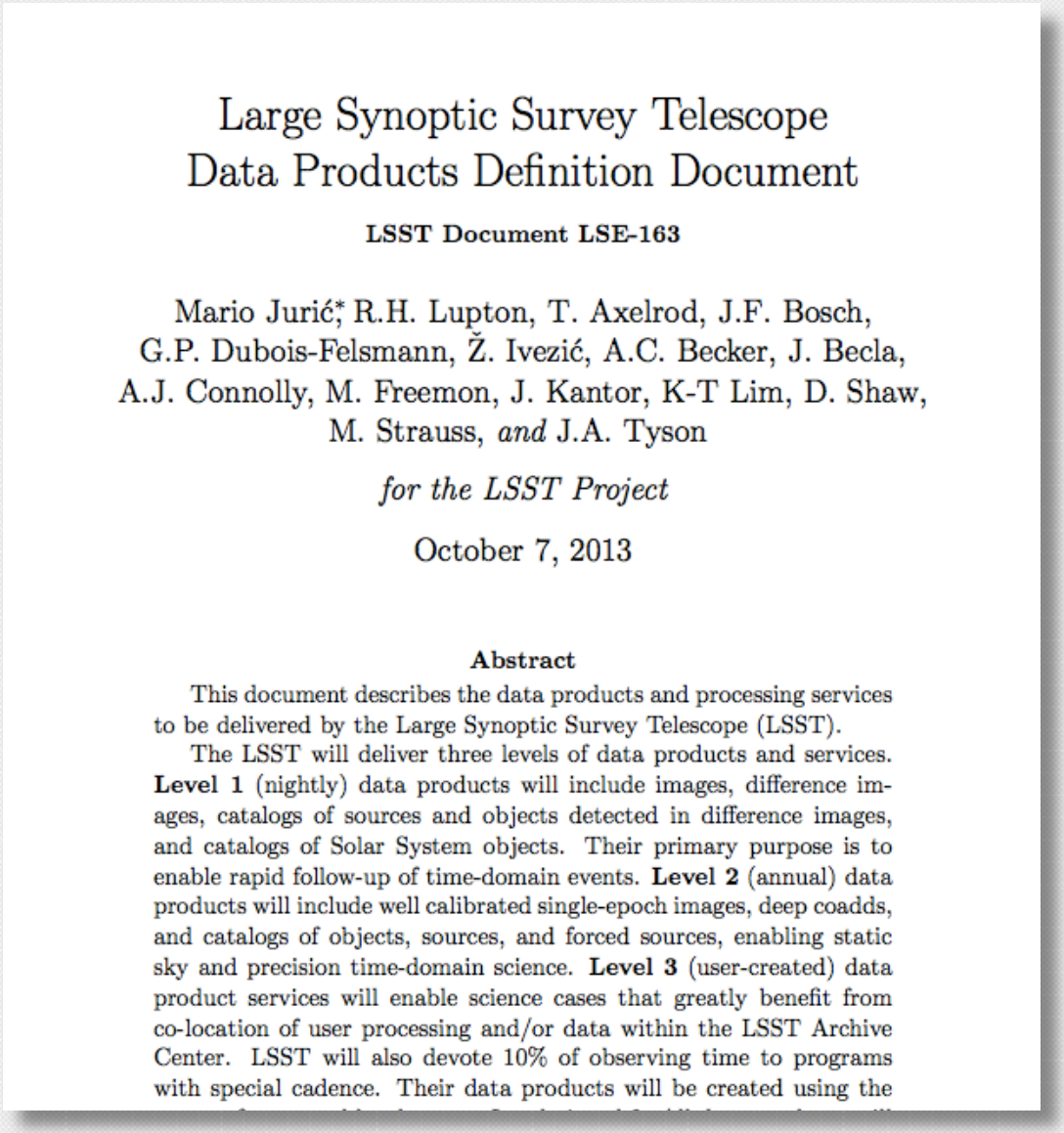
. . . User-driven

Data Product Requirements: Special Programs

p. 26, LSST SRD

As a result of these studies, the adopted baseline design (see Appendix A) assumes a nominal 10-year duration with about 90% of the observing time allocated for the main LSST survey. The same assumption was adopted here to derive the requirements described below. Only visits that satisfy the requirements listed in the previous section are counted towards the specifications listed in this Section. For example, if the photometric accuracy falls below requirements due to complex atmospheric cloud structure, or due to extraneous noise sources inside the system, the data will not be counted. The remaining 10% of observing time will be used to obtain improved coverage of parameter space such as very deep ($r \sim 26$) observations, observations with very short revisit times (~1 minute), and observations of "special" regions such as the Ecliptic, Galactic plane, and the Large and Small Magellanic Clouds. A third type of survey, micro-surveys, that would use about 1% of the time, may also be considered.

LSST Data Products Definition Document



Large Synoptic Survey Telescope
Data Products Definition Document

LSST Document LSE-163

Mario Jurić*, R.H. Lupton, T. Axelrod, J.F. Bosch,
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

October 7, 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 with special cadence. Their data products will be created using the

Reviewed in June 2013,
Yanny et al. committee

LSST Data Products Definition Document

A readable description of LSST data products. Used to communicate with the science community, and to support the formal requirements flow-down. Describes the processing as well as the data products

Level 1 Data Products: Section 4.

Level 2 Data Products: Section 5.

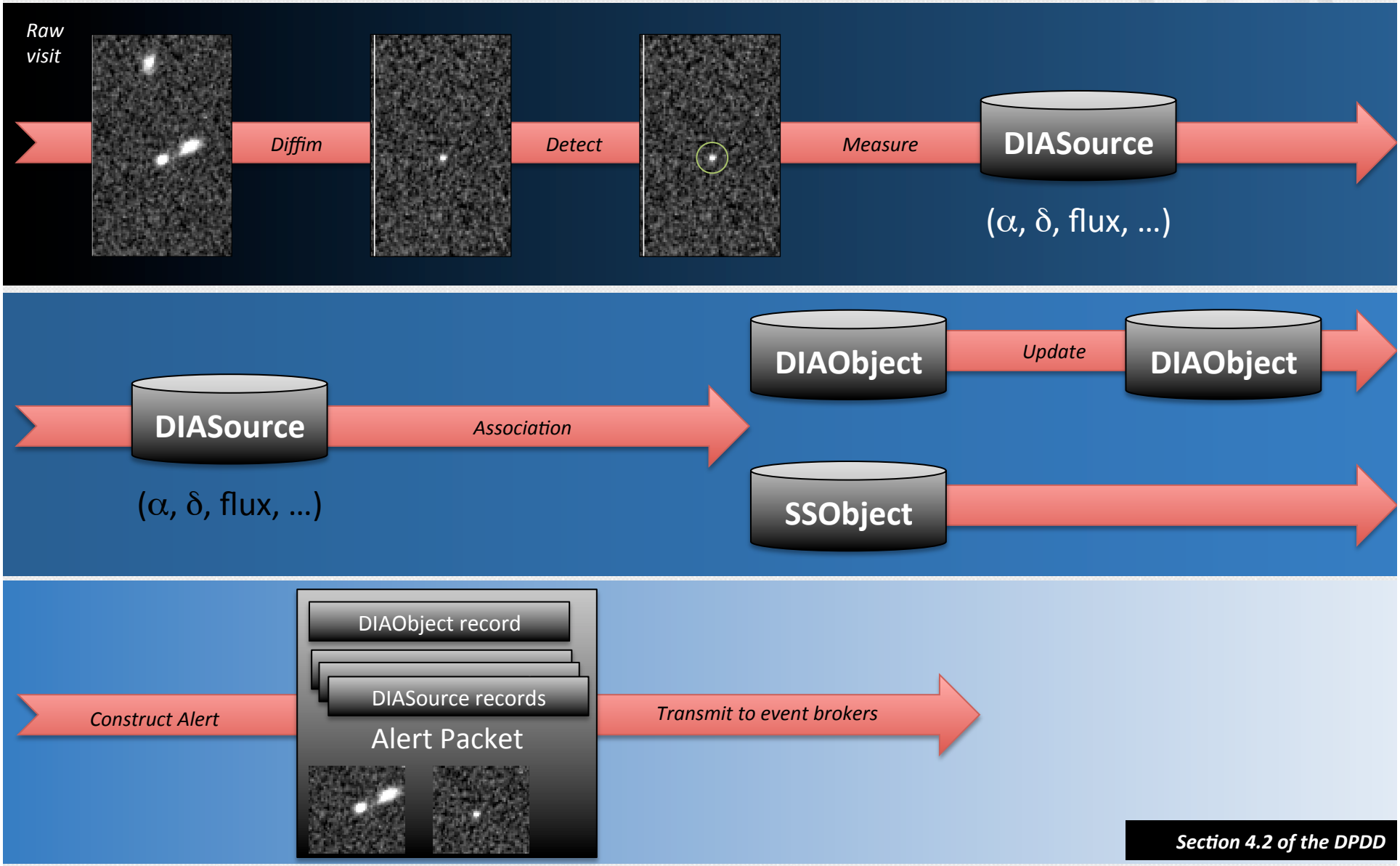
Level 3 Data Products: Section 6.

Special Programs DPs: Section 7.

Level 1 Data Products

- **Processing to enable rapid detection and follow-up of time-domain events**
- Real-time image differencing as observing unfolds each night
- Measurement of position, brightness and shape for each detection

Level 1 (Alert Production): Production Outline



Level 1: 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

- **Will measure and *transmit with each alert*:**
  - position
  - flux, 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)

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**

Level 2 Data Products

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

Level 2 Catalog Guiding Principles

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

- Defining principles for the Level 2 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.

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

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

Level 2 Pipelines: Logical Flow

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

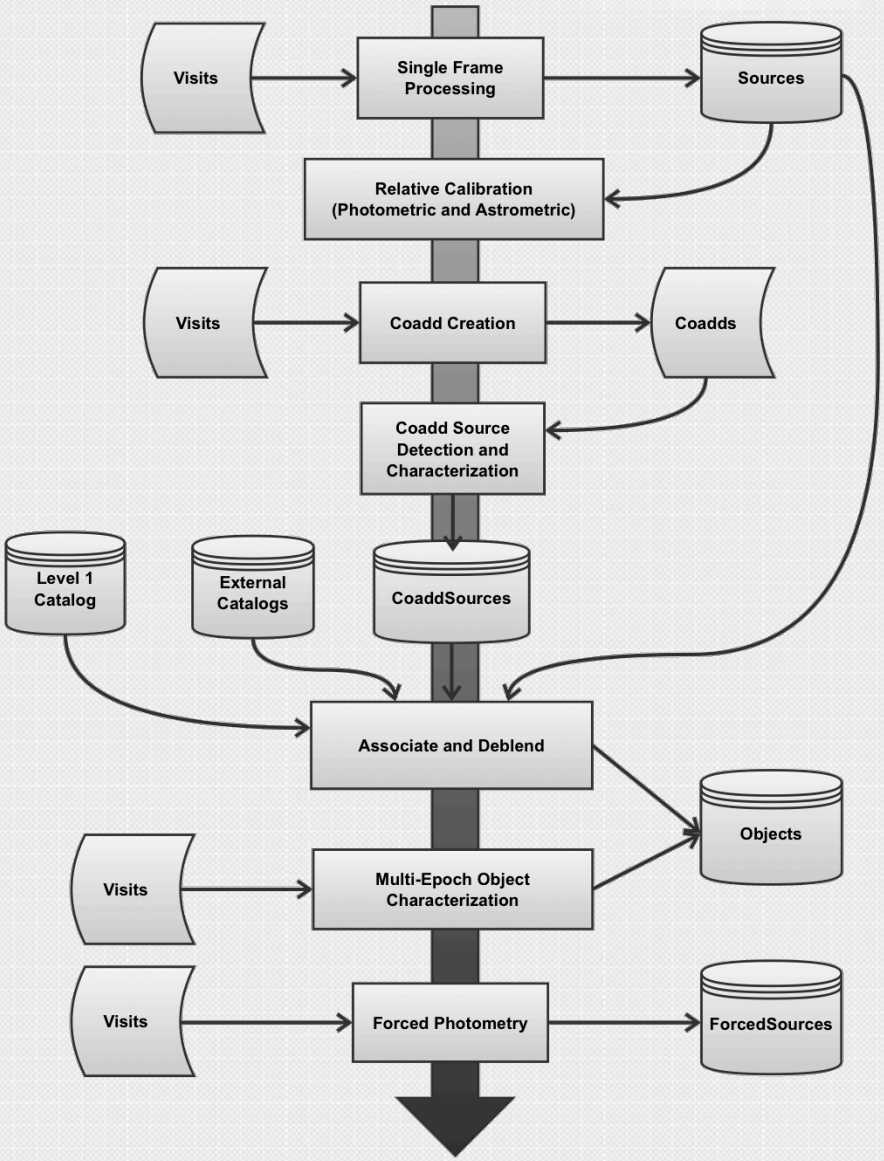
These are used to photometrically and astrometrically calibrate the survey.

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.



Catalogs: Measurements on Detected Sources

- **Object characterization (models):**
  - Point source photometry
- **Object characterization (model-free):**
  - Centroid: (α, δ)
  - Adaptive moments and ellipticity measures
  - Aperture photometry
- **Metadata:** Deblend status, flags, etc.

Catalogs: Measurements on Detected Objects

- **Object characterization (models):**
  - Moving Point Source model
  - Double Sérsic model
    - Maximum likelihood peak
    - Samples of the posterior (hundreds)
- **Object characterization (model-free):**
  - 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.

Level 3: *Enabling* User-created Data Products

- Products created by the community using LSST's software, services, and/or computing resources.

- **For use-cases not fully enabled by Level 1 and 2 processing:**
  - Reprocessing images to search for SNe light echos
  - Characterization of diffuse structures (*e.g.*, ISM)
  - Extremely crowded field photometry (*e.g.*, globular clusters)
  - Custom measurement algorithms

- **Enabling Level 3:**
  - 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

Implementation Strategy:

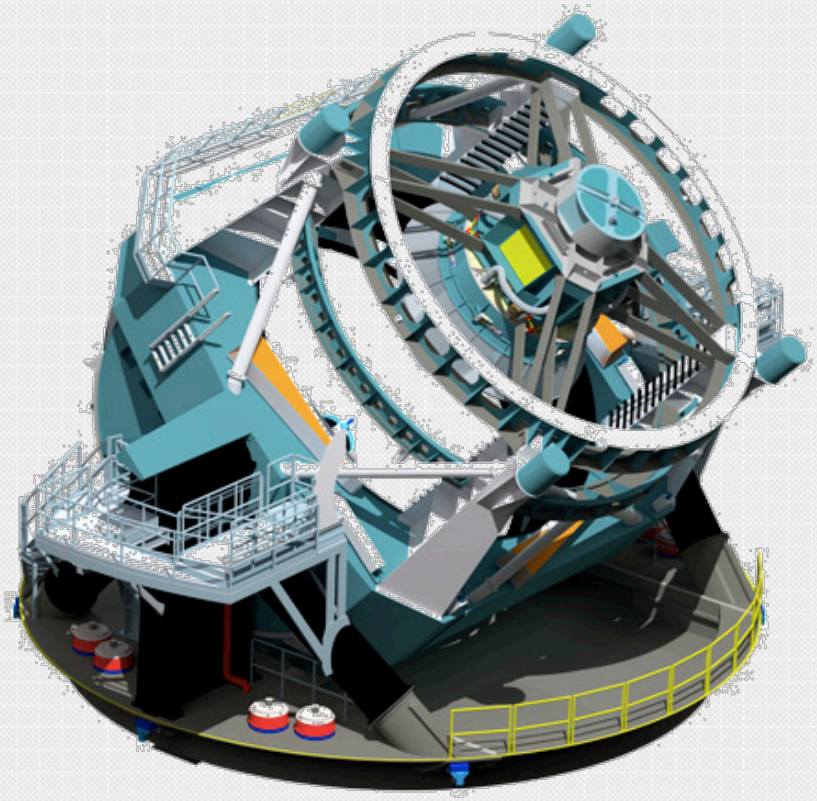
- **By adopting an appropriate architecture, *the support for Level 3 is a natural outgrowth of the data management system*, distinguished from Level 2 largely by access and resource rights.**
  - Level 3 pipelines are built by the users, using the same components that Level 2 applications use
  - The difference is in where these run, and how much resources they're permitted to consume
- Benefits:
  - Reduction in cost of delivery of Level 3 capability
  - Reduction in complexity and cost of migration of Level 3 codes to Level 2
  - Increased return on investment in LSST software

Special Programs

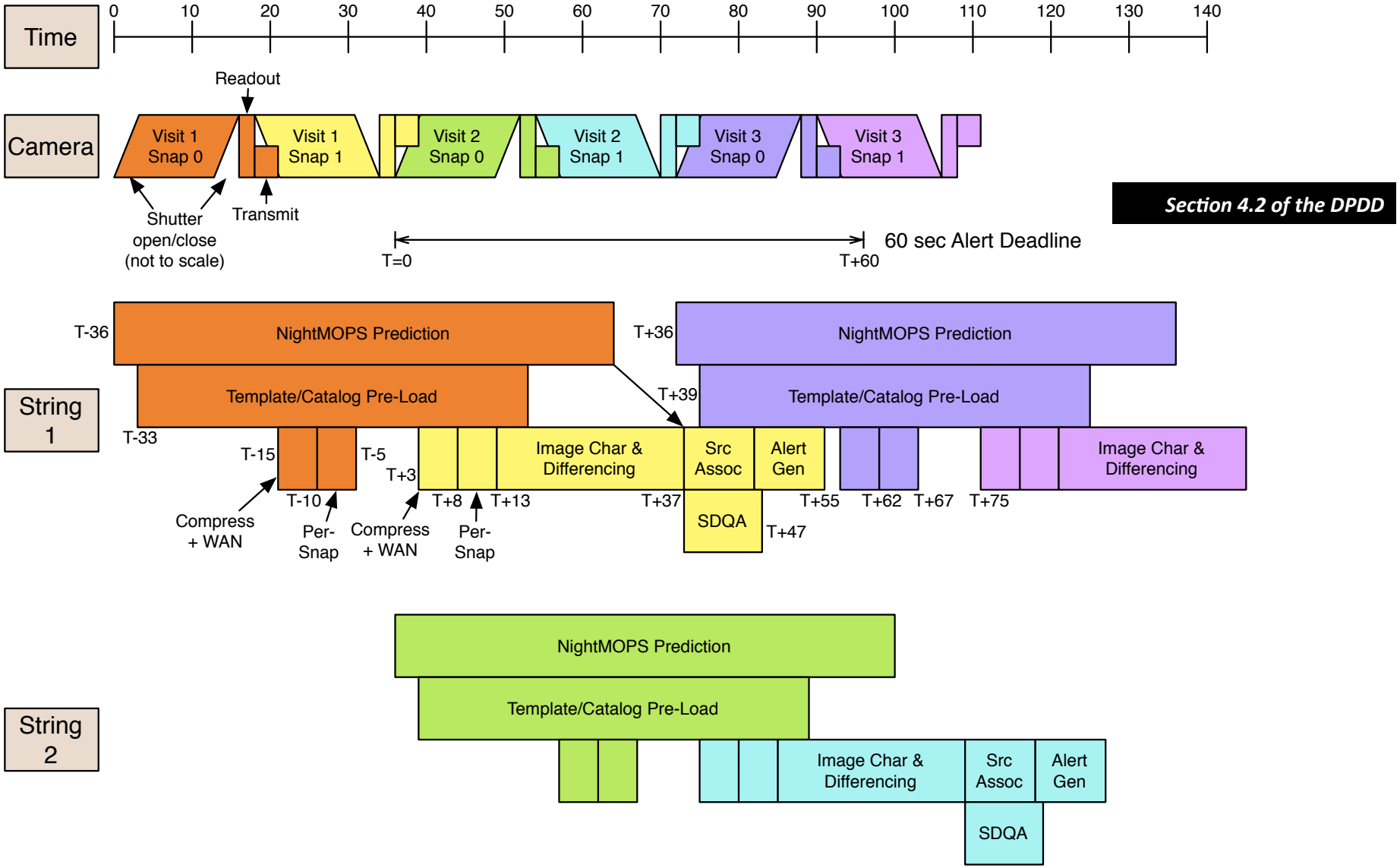
- 10% of LSST observing time will be devoted to special programs, to obtain improved coverage of interesting regions of observational parameter space.
  - Deep observations, observations with short revisit times, and observations of the Ecliptic, Galactic plane, and the Large and Small Magellanic Clouds, etc.

- The details of these special programs and their data products are not yet defined. We instead specify the data product *constraints*, given the adopted Level 1/2/3 architecture.
  - No special program will be selected that does not fit these constraints.
  - All currently proposed programs (e.g., the so-called "deep drilling" fields), satisfy the imposed constraints.

Questions?



Level 1 Processing: Timing Diagram



Level 1 Catalogs: DIASources (see DPDD)

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 object moves sufficiently fast to 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
- Some DIASources will be false positives. We will flag suspect DIASources

Level 1 Catalogs: DIAObjects

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

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

- 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 Pipeline: Solar System Processing

Details in talk by Jones

- 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.