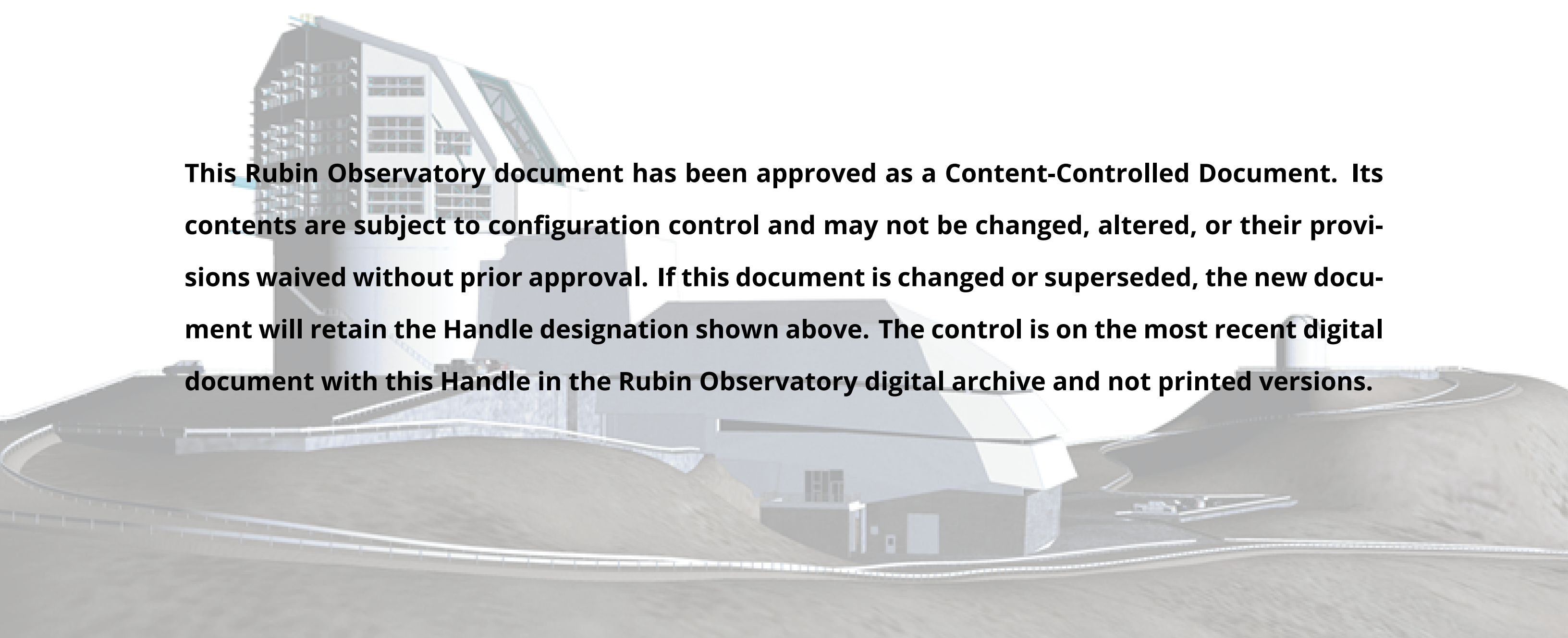
Vera C. Rubin Observatory
Systems Engineering

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, L.P. Guy, 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: 2023-07-10



Abstract

This document describes the data products and processing services to be delivered by the Large Synoptic Survey Telescope (LSST). LSST will deliver three levels of data products and services. **Prompt** data products are computed and released within 24 hours of observation, and 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. **Data Release** data products are computed during annual processing campaigns, and include well-calibrated single-epoch images, deep coadds, and catalogs of objects, sources, and forced sources, enabling static sky and precision time-domain science. The **Science Platform** will allow for the creation of **User Generated** data products and 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 same software and hardware as Prompt and Data Release products. All data products will be made available using user-friendly databases and web services.

Change Record

Version	Date	Description	Owner name
1	2013-10-07	Initial release	Mario Juric
2	2016-09-26	Implementation of LCR-758 Update Data Products Definition Document, LSE-163	Gregory Dubois-Felsmann (LCR), Tim Jenness (document), Robert McKercher (Docushare)
3.0	2017-07-03	Implementation of LCR-962. Reference LSE-61 requirements. Minor clean ups.	Tim Jenness
3.1	2018-01-04	Implementation of LCR-1024. Mention relevant OSS requirements.	Tim Jenness
3.2	2018-02-09	Implementation of LCR-1214. Update Level 1, Level 2 and Level 3 nomenclature as defined in LPM-231. Shrink section Level 3 section and refer to Science Platform.	R. Gill, W. O'Mullane
3.3	2018-10-24	Implementation of LCR-1459. DM-12129: Adoption of nanojansky flux unit. DM-11247: Add additional alert packet contents to DPDD. DM-15566: DPDD incorrectly states "immediate" access to PPDB products. Updated document maintainer, curator and author list, added point estimates for photo-z and a few minor edits.	Ž. Ivezić, Leanne Guy
3.4	2018-07-29	Add DIAForcedSource table. Clarify handling of force photometry, Prompt vs. Data Release products distinction, other minor fixes. (LCR-1916: RFC-517, DM-10540, DM-15671, DM-16688, DM-15784)	C. Slater, E. Bellm
3.5	2020-11-04	Implementation of LCR-2463 (RFC-710, DM-23248): Remove mentions of multifit and multi-epoch fitting. Clarification of DM plans for making and serving coadds. Clarification of current status of computing photometric redshifts.	J. Bosch, L.Guy, E.Bellm
3.5	2020-11-04	Implementation of LCR-22376 (RFC-620, DM-24710) Update of the Solar System Products and Processing baseline	M. Jurić, L.Guy

Continued on next page

Version	Date	Description	Owner name
3.6	2021-12-17	Implementation of LCR-2722 (RFC-695, DM-29731) DIAObjects table updated to reflect the plans for host association; updated to "Rubin" language	M. Graham
3.7	2022-10-18	Implementation of LCR-3293 (RFC-876) Remove CoaddSources and ref to DMS-REQ-0277. Fix ref RDO-013	W. O'Mullane
3.8	2023-07-10	Implementation of LCR-3476 (RFC-921) Add detailed description of plans and data products for shear estimation	J. Bosch

Document curator: Leanne Guy
Document source location: `https://github.com/lsst/LSE-163`

Contents

Data Products Definition Document

Preface

The purpose of this document is to describe the data products produced by the Large Synoptic Survey Telescope (LSST).

To a future LSST user, it should clarify what catalogs, image data, software, and services they can expect from LSST. To LSST builders, it provides direction on how to flow down the LSST System Requirements Document to system design, sizing, budget and schedule as they pertain to the data products.

Though under strict change control, this is a ***living document***. LSST will undergo a period of construction and commissioning lasting no less than seven years, followed by a decade of survey operations. To ensure their continued scientific adequacy, the designs and plans for LSST Data Products will be periodically reviewed and updated.

1 Introduction

1.1 The Vera C. Rubin Observatory

The Rubin Observatory will host the LSST Camera, a large, wide-field ground-based optical telescope system designed to obtain multiple images covering the sky that is visible from Cerro Pachón in Northern Chile. The current baseline design, with an 8.4m (6.7m effective) primary mirror, a 9.6 deg^2 field of view, and a 3.2 Gigapixel camera, will allow about 10,000 square degrees of sky to be covered every night using pairs of 15-second exposures, with typical 5σ depth for point sources of $r \sim 24.5$ (AB). The system is designed to yield high image quality as well as superb astrometric and photometric accuracy. The total survey area will include ~30,000 deg^2 with $\delta < +34.5^{\circ}$, and will be imaged multiple times in six bands, $ugrizy$, covering the wavelength range 320–1050 nm. This is referred to as the Legacy Survey of Space and Time (LSST). For a more detailed, but still concise, summary of Rubin, please see the overview paper [14][1].

The project is scheduled to begin the regular survey operations at the start of next decade. About 90% of the observing time will be devoted to a deep-wide-fast survey mode which will uniformly observe a 18,000 deg^2 region about 1000 times (summed over all six bands) during the anticipated 10 years of operations, and yield a coadded map to $r \sim 27.5$. These data will result in catalogs including over 38 billion stars and galaxies, that will serve the majority of the primary science programs. The remaining 10% of the observing time will be allocated to special projects such as a Very Deep and Fast time domain survey[2].

Rubin will be operated in fully automated survey mode. The images acquired by the LSST Camera will be processed by LSST Data Management software to a) detect and characterize imaged astrophysical sources and b) detect and characterize temporal changes in the LSST-observed universe. The results of that processing will be reduced images, catalogs of detected objects and the measurements of their properties, and prompt alerts to "events" – changes in astrophysical scenery discovered by differencing incoming images against older, deeper, images of the sky in the same direction (*templates*, see §3.4.3). Measurements will be internally and absolutely calibrated.

The *broad*, *high-level*, requirements for LSST Data Products are given by the *LSST Science Re-*

[1] `http://ls.st/2m9`
[2] Informally known as "Deep Drilling Fields".

quirements Document (SRD; LPM-17). This document lays out the *specifics* of what the data products will comprise, how those data will be generated, and when. It serves to inform the flow-down from the LSST SRD through the *LSST System Requirements Document* (the LSR; LSE-29) and the *LSST Observatory System Specifications* (OSS; LSE-30), to the *LSST Data Management System Requirements* (DMSR; LSE-61), the UML model (LDM-133), and the database schema (LDM-153). The OSS explicitly ties this document to the requirements flow down through requirements OSS-REQ-0126 Prompt, OSS-REQ-0133 Data Release, OSS-REQ-0139 User Generated, OSS-REQ-0391 (general conventions), and OSS-REQ-0392 (special programs). Throughout this document margin notes are used to provide linkage to formal LSST requirements and parameters associated with the nearby text.

1.2 General Image Processing Concepts for LSST

A raw image (baselined as a pair of successive 15-second exposures, called snaps), delivered by the LSST camera is processed by the Single Frame Processing Pipeline to produce a single-visit image with, at least conceptually, counts proportional to photon flux entering the telescope pupil (in reality, there are many additional optical, pixel and bandpass effects, including random counting noise and various subtle systematic errors, that are treated during subsequent processing). This single-visit image processed by the pipeline is called a "Processed Visit Image" and its main data structures include counts, their variance and various masks, all defined on per pixel basis. These single-visit images are used downstream to produce coadded and difference images. The rest of the processing is essentially a model-based interpretation of imaging observations that includes numerous astrophysical and other assumptions.

The basic interpretation model assumes a sum of discrete (but possibly overlapping) sources and a relatively smooth background. The background has a different spectral energy distribution than discrete sources, and it can display both spatial gradients and temporal changes. Discrete sources can vary in brightness and position. The motion can be slow or fast (between two successive observations, less or more motion than about the seeing disk size), and naturally separates stars with proper motions and trigonometric parallax from moving objects in the Solar System. Some objects that vary in brightness can be detectable for only a short period of time (e.g., supernovae and other cosmic explosions).

The image interpretation model separates time-independent model component from a temporally changing component ("DC" and "AC", respectively). Discrete DC sources are *not* operationally nor astrophysically associated with discrete AC sources even when they are spatially

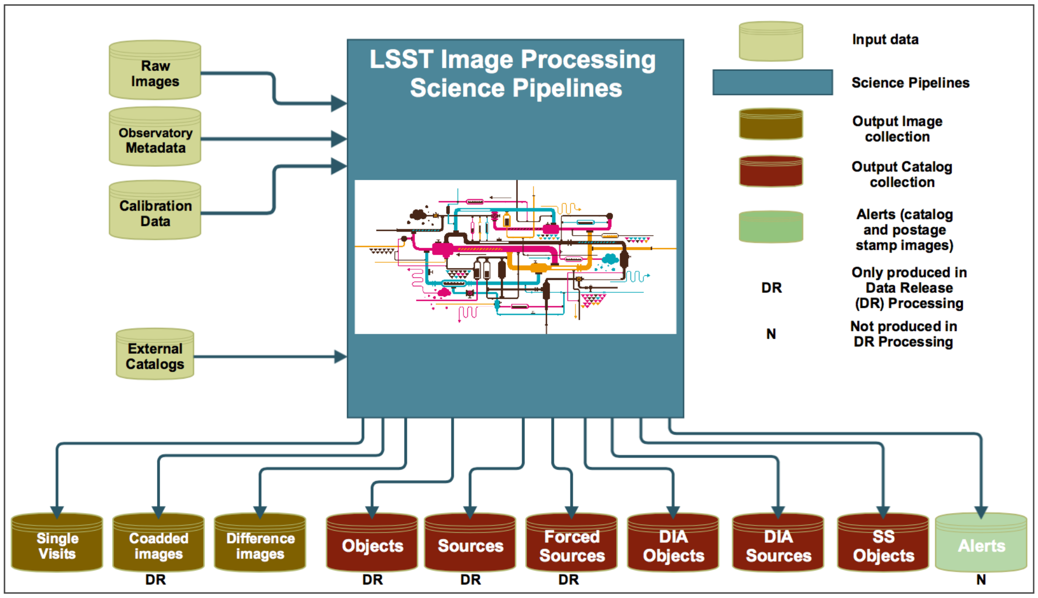


FIGURE 1: Overview of data products produced by LSST Imaging Processing Science Pipelines.

coincident.

Images of discrete objects are modeled using two models (§4.2.1). A two-component galaxy model includes a linear combination of bulge and disk, with their radial intensity variation described using Sersic profiles. Stars are modeled using a moving point source model with its parallax motion superposed on a linear proper motion. This model shares motion parameters across the six bandpasses and assumes constant flux in each band, and thus includes 11 free parameters. Both galaxy and stellar models are fit to all objects, except for fast-moving objects (the Solar System objects), which are treated separately. Discrete objects detected in *difference* images will be modeled using three models: a point source model, a trailed point source model, and a point source dipole model.

1.3 Classes of LSST Data Products

The main LSST data products are illustrated in Figure 1 (see Appendix for a conceptual design of pipelines which will produce these data products). LSST Data Management will perform image analysis on two different timescales, with two corresponding categories of output data products[3]:

[3]Note that prior to 2018 data products were referred to as Level 1, Level 2, and Level 3; this nomenclature was updated in 2018 to Prompt Products, Data Release Products and User Generated Products respectively [LPM-231].

1. **Prompt Processing**, which is performed on a nightly or daily basis and results in **Prompt Data Products**. The goal of this processing is the detection and characterization of astrophysical phenomena that are revealed by their time-dependent nature. The detection of supernovae superimposed on bright extended galaxies is an example of this analysis. Prompt products will include single visit images, difference images, catalogs of sources detected in difference images (`DIASources`), astrophysical objects[4] these are associated to (`DIAObjects`), and Solar System objects (`SSObjects`[5]). The catalogs will be entered into the **Prompt Products database** and made available in near real time. Notifications ("alerts") about new `DIASources` will be issued using community-accepted standards[6] within 60 seconds of observation. Prompt data products are discussed in § 3.

2. **Data Release Processing**, which is performed annually[7] and produces **Data Release** data products. These processing campaigns use all data collected by the survey to date in order to perform a comprehensive analysis, including measurement of both faint static objects through the construction of deep coadds, and light curve characterization through image differencing across all available epochs. The data products produced will include the calibrated single-epoch images, deep coadds, catalogs of characterized `Objects` (detected on deep coadds as well as individual visits[8]), `Sources`[9] (detections and measurements on individual visits), and `ForcedSources` (constrained measurement of flux on individual visits). It will also include difference image analysis tables similar to those produced during Prompt Processing (see §3.3.5). In contrast to the Prompt Products database, which is updated continuously during observing, the Data Release database is static and will not change after release. Data Release data products are discussed in § 4.

The two processing timescales are driven by differing scientific requirements. Changes in flux or position of objects may need to be immediately followed up, lest interesting information be

[4]The LSST has adopted the nomenclature by which single-epoch detections of astrophysical *objects* are called *sources*. The reader is cautioned that this nomenclature is not universal: some surveys call *detections* what LSST calls *sources*, and use the term *sources* for what LSST calls *objects*.

[5]`SSObjects` used to be called "Moving Objects" in previous versions of the LSST Data Products baseline. The name is potentially confusing as high-proper motion stars are moving objects as well. A more accurate distinction is the one between objects *inside* and *outside* of the Solar System.

[6]For example, VOEvent, see `http://ls.st/4tt`

[7]Except for the first two data releases, which will be created six months apart.

[8]The LSST takes two exposures per pointing, nominally 15 seconds in duration each, called *snaps*. For the purpose of data processing, that pair of exposures will typically be coadded and treated as a single exposure, called a *visit*.

[9]When written in bold monospace type (i.e., `\tt`), `Objects` and `Sources` refer to objects and sources detected and measured as a part of Data Release processing.

lost. Thus the primary results of analysis of difference images – discovered and characterized DIASources – generally need to be broadcast as *event alerts* within 60 seconds of end of visit OTT1
acquisition. The analysis of science (direct) images is less time sensitive, and will be done as DMS-REQ-0004
a part of annual data release process.

Recognizing the diversity of astronomical community needs, and the need for specialized processing not part of the automatically generated Prompt and Data Release products, LSST plans to devote 10% of its data management system capabilities to enabling the creation, use, and federation of **User Generated** data products. The **Science Platform** will enable science cases that greatly benefit from co-location of user processing and/or data within the LSST Archive Center. The high-level requirement for this is established in § 3.5 of the LSST SRD. Some details are discussed in § 5 of this document.

Finally, LSST Survey Specifications (§ 3.4 of LSST SRD) prescribe that 90% of LSST observing time be spent in the so-called "universal cadence" mode of surveying the sky. These observations will result in Prompt and Data Release data products discussed above. The remaining 10% of observing time will be devoted to **special programs**, designed to obtain improved coverage of interesting regions of observational parameter space. Examples include very deep ($r \sim 26$, per exposure) 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. The data products for these programs will be generated using the same processing software and hardware and possess the general characteristics of Prompt and Data Release data products, but may be performed on a somewhat different cadence. They will be discussed in § 6.

2 General Considerations

Most LSST data products will consist of images and/or catalogs. The catalogs will be stored and offered to the users as *relational databases* which they will be able to query. This approach was shown to work well by prior large surveys, for example the Sloan Digital Sky Survey (SDSS).

Catalogs may be stored in different databases to meet the operational requirements particular to each dataset. The Prompt Products database will hold catalogs from Prompt Processing, to enable rapid access to recent measurements. Data Release "universal cadence" catalogs will be stored in a Data Release database. The products for special programs may be stored in many different databases, depending on the nature of the program.

Nevertheless, all these databases will follow certain naming and other conventions. We discuss these in the subsections to follow.

2.1 Estimator and Naming Conventions

For all catalogs data, we will employ a convention where estimates of standard errors have the suffix `Err`, while the estimates of inherent widths of distribution (or functions in general) have the suffix `Sigma`[10]. The latter are defined as the square roots of the second moment about the quoted value of the quantity at hand. DMS-REQ-0333

Unless noted otherwise, maximum likelihood values (called likelihood for simplicity) will be quoted for all fitted parameters (measurements). Together with covariances, these let the end-user apply whatever prior they deem appropriate when computing posteriors[11]. Where appropriate, multiple independent samples from the likelihood may be provided to characterize departures from Gaussianity.

We will provide values of log likelihood, the χ^2 for the fitted parameters, and the number of data points used in the fit. Database functions (or precomputed columns) will be provided for frequently used combinations of these quantities (e.g., χ^2/dof). These can be used to assess the model fit quality. Note that, *if the errors of measured quantities are normally distributed,* the likelihood is related to the χ^2 as: DMS-REQ-0331

$$L = \left(\prod_k \frac{1}{\sigma_k \sqrt{2\pi}} \right) \exp\left[-\frac{\chi^2}{2} \right] \tag{1}$$

where the index k runs over all data points included in the fit. For completeness, χ^2 is defined as:

$$\chi^2 = \sum_k \left(\frac{x_k - \bar{x}}{\sigma_k} \right)^2, \tag{2}$$

where $\bar{x}$ is the mean value of x_k.

For fluxes, we recognize that a substantial fraction of astronomers will just want the posteri-

[10] Given N measurements, standard errors scale as $N^{-1/2}$, while widths remain constant.

[11] There's a tacit assumption that a Gaussian is a reasonably good description of the likelihood surface around the ML peak.

ors marginalized over all other parameters, trusting the LSST experts to select an appropriate prior[12]. For example, this is nearly always the case when constructing color-color or color-magnitude diagrams. We will support these use cases by providing additional pre-computed columns, taking care to name them appropriately so as to minimize accidental incorrect usage. For example, a column named `gFlux` may be the expectation value of the g-band flux, while `gFluxML` may represent the maximum likelihood value. DMS-REQ-0331

2.2 Image Characterization Data

Raw images will be processed to remove instrumental signature and characterize their properties, including backgrounds (both due to night sky and astrophysical), the point spread function and its variation, photometric zero-point model, and the world coordinate system (WCS). DMS-REQ-0069 DMS-REQ-0327 DMS-REQ-0070 DMS-REQ-0029 DMS-REQ-0030

That characterization is crucial for deriving LSST catalogs and understanding the images. It will be kept and made available to the users. The exact format used to store these (meta)data will depend on the final adopted algorithm in consultation with the scientific community to ensure the formats in which these data are served are maximally useful. DMS-REQ-0328 DMS-REQ-0072

Each processed image[13], including the coadds, will record information on pixel variance (the "variance plane"), as well as per-pixel masks (the "mask plane"). These will allow the users to determine the validity and usefullness of each pixel in estimating the flux density recorded in that area of the sky.

This information will be per-pixel, and potentially unwieldy to use for certain science cases. We plan to investigate approximate schemes for storing masks based on geometry (e.g., similar to Mangle or STOMP), *in addition* to storing them on a per pixel basis. DMS-REQ-0326

2.3 Fluxes and Magnitudes

DMS-REQ-0043

Because flux measurements on difference images (Prompt data products; § 3) are performed against a template and thus represent a flux difference, the measured flux of a source on the difference image can be negative. The flux can also go negative for faint sources in the presence of noise. Negative fluxes cannot be stored as (Pogson) magnitudes; log of a neg-

[12] It's likely that most cases will require just the expectation value alone.
[13] It is also frequently referred to as *calibrated exposure*, from the Butler product type of `calexp`.

ative number is undefined. We therefore prefer to store fluxes, rather than magnitudes, in database tables[14]. DMS-REQ-0347

LSST fluxes are quoted in units of nanojansky[15] (1 nJy = $10^{?35}$ Wm$^{?2}$Hz$^{?1}$). We will provide columns with flux and flux errors as well as estimates of the relative and absolute photometric calibration errors, and the normalized system response (for details see Section 3.3.4 in the LSST Science Requirements Document).

We acknowledge that the large majority of users will want to work with magnitudes. For convenience, we plan in addition to provide columns with (Pogson) magnitudes and magnitude errors[16], where values with negative flux will evaluate to `NULL`.

2.4 Uniqueness of IDs across database versions

DMS-REQ-0292

To reduce the likelihood for confusion, all IDs shall be unique across databases and database versions, other than those corresponding to uniquely identifiable entities (i.e., IDs of exposures).

For example, DR4 and DR5 (or any other) release will share no identical `Object`, `Source`, `DIAObject` or `DIASource` IDs (see § 3 and 4 for the definitions of `Objects`, `DIAObjects`, etc.).

2.5 Repeatability of Queries

DMS-REQ-0291

We require that queries executed at a known point in time against any LSST-delivered database be repeatable at a later date. This promotes the reproducibility of science derived from LSST data. It is of special importance for Prompt product catalogs (§ 3) that will change on a nightly basis as new time domain data is being processed and added to the catalogs.

The exact implementation of this requirement is left to the LSST Data Management database team. One possibility may be to make the key tables (nearly) append-only, with each row having two timestamps – `createdTai` and `deletedTai`, so that queries may be limited by a `WHERE` clause:

[14] This is a good idea in general. E.g., given multi-epoch observations, one should always be averaging fluxes, rather than magnitudes.

[15] For details, please see LSST DOCUMENT-27758.

[16] These will most likely be implemented as “virtual” or “computed” columns.

```
SELECT * FROM DIASource WHERE 'YYYY-MM-DD-HH-mm-SS'
    BETWEEN createdTAI and deletedTAI
```

or, more generally:

```
SELECT * FROM DIASource WHERE 'data is valid as of YYYY-MM-DD'
```

A perhaps less error-prone alternative, if technically feasible, may be to provide multiple virtual databases that the user would access as:

```
CONNECT lsst-dr5-yyyy-mm-dd SELECT * FROM DIASource
```

The latter method would probably be limited to nightly granularity, unless there's a mechanism to create virtual databases/views on-demand.

3 Prompt Data Products

3.1 Overview

Prompt data products are the result of nightly processing. These data products are single visit images, difference images, and the results of difference image analysis (DIA; §3.2.1). DIA outputs consist of the sources detected in difference images (`DIASources`), astrophysical objects that these are associated to (`DIAObjects`), characterizations of hitherto identified Solar System objects (`SSObject`), and discoveries of new Solar System objects. Various metadata products are also produced during nightly processing and made available to users.

`DIASources` are sources detected on difference images (those with the signal-to-noise ratio $S/N > transSNR$ after correlation with the PSF profile, with $transSNR$ defined in the SRD and presently set to 5). They represent changes in flux with respect to a deep template. Detections with high probability of being instrumental non-astrophysical artifacts may be excluded. Physically, a `DIASource` may be an observation of new astrophysical object that was not present at that position in the template image (for example, an asteroid), or an observation of flux change in an existing source (for example, a variable star). Their flux can be negative (eg., if a source present in the template image reduced its brightness, or moved away). Their shape can be complex (eg., trailed, for a source with proper motion approaching ∼deg/day, or "dipole-like", if an object's observed position exhibits an offset – true or apparent – compared to its position on the template). Some `DIASources` will be caused by background fluctuations; with $transSNR = 5$, we expect about one such false positive per CCD (of the order 200,000 per typical night). The expected number of false positives due to background fluctuations is a very strong function of adopted $transSNR$: a change of $transSNR$ by 0.5 results in a variation of an order of magnitude, and a change of $transSNR$ by unity changes the number of false positives by about two orders of magnitude. transSNR

Clusters of `DIASources` detected on visits taken at different times are associated with either a `DIAObject` or an `SSObject`, to represent the underlying astrophysical phenomenon. The association can be made in two different ways: by assuming the underlying phenomenon is an object within the Solar System moving on an orbit around the Sun[17], or by assuming it to be distant enough to only exhibit small parallactic and proper motion[18]. The latter type of associ- DMS-REQ-0285

[17]We don't plan to fit for motion around other Solar System bodies; eg., identifying new satellites of Jupiter is left to the community.

[18]Where 'small' is small enough to unambiguously positionally associate together individual apparitions of the

ation is performed during difference image analysis right after the image has been acquired. The former is done at daytime by the object linking element of the Solar System Processing Pipeline, unless the `DIASource` is an apparition of an already known `SSObject`. In that case, it will be flagged as such during difference image analysis. DMS-REQ-0089 DMS-REQ-0002

At the end of the difference image analysis of each visit, we will issue time domain event alerts for all newly detected `DIASources`[19].

3.2 Prompt Data Processing

3.2.1 Difference Image Analysis

The following is a high-level description of steps which will occur during regular *nightly* difference image analysis (see Figures 3 and 5):

1. A visit is acquired and reduced to a single *Processed Visit Image* (cosmic ray rejection, instrumental signature removal[20], combining of snaps, etc.). DMS-REQ-0069
2. The Processed Visit Image is differenced against the appropriate template and `DIASources` are detected. If necessary, deblending will be performed at this stage. Both the parent blend and the deblended children will be measured and stored as `DIASources` (see next item), but only the children will be matched against `DIAObjects` and alerted on. Deblended objects will be flagged as such. DMS-REQ-0010
3. The flux and shape[21] of the DIASource are measured on the difference image. PSF photometry is performed on the Processed Visit Image at the position of the `DIASource` to obtain a measure of the total flux. DMS-REQ-0269
4. The Prompt Products database (see §3.3) is searched for a `DIAObject` or an `SSObject` with which to positionally associate the newly discovered `DIASource`[22]. If no match is found, a new `DIAObject` is created and the observed `DIASource` is associated to it. DMS-REQ-0273 DMS-REQ-0271 DMS-REQ-0285

object.

[19]For observations on the Ecliptic near the opposition Solar System objects will dominate the `DIASource` counts and (until they're recognized as such) overwhelm the explosive transient signal. It will therefore be advantageous to quickly identify the majority of Solar System objects early in the survey.

[20]Eg., subtraction of bias and dark frames, flat fielding, bad pixel/column interpolation, etc.

[21]The "shape" in this context consists of weighted 2nd moments, as well as fits to a trailed source model and a dipole model.

[22]The association algorithm will guarantee that a `DIASource` is associated with not more than one existing `DIA-`

5. If the `DIASource` has been associated with an `SSObject` (a known Solar System object), it will be flagged as such and an alert will be issued. Further processing will occur in daytime (see section 3.2.2). DMS-REQ-0274

6. Otherwise, the associated `DIAObject` measurements will be updated with new data collected during previous 12 months. Hence, the computed parameters for `DIAObjects` have a "memory" of past data that does not extend beyond this cutoff[23]. All affected columns will be recomputed, including proper motions, centroids, light curves, etc. DMS-REQ-0319 diaCharacterizationCutoff

7. To aid in rapid prioritization of new `DIAObjects` in the latency interval before the precovery forced photometry is run (see below), a table with the median noise in the difference image per visit is queried. For each visit in the last twelve months at the position of the `DIAObject`, if there are no `DIASource` or `DIAForcedSource` records for that visit, the time of that visit, the filter, and the difference image noise are included in the alert packet. These data allow computation of appropriate upper limits in the difference image light curve. L1PublicT

8. The Data Release database[24] is searched for the three nearest stars and three nearest galaxies to the `DIAObject` in `Objects` out to some maximum radius. The IDs of these nearest-neighbor `Objects` are recorded in the `DIAObject` record and provided in the complete `DIAObject` record issued with the issued event alert (see below). diaNearbyObjMaxStar diaNearbyObjMaxGalaxy diaNearbyObjRadius DMS-REQ-0271

9. An alert is issued that includes: the timestamp of when this database has been queried to issue this alert, the `DIASource` ID, the complete `SSObject` or `DIAObject` record[25], *including all* `DIASources` and `DIAForcedSources` from the last 12 months that are linked with the `SSObject` or `DIAObject`. The science content associated with the Data Release database `Objects` will not be included. See Section 3.5 for a more complete enumeration. DMS-REQ-0002 diaCharacterizationCutoff DMS-REQ-0274

10. For all `DIAObjects` overlapping the field of view, including those that have an associated new `DIASource` from this visit, forced photometry will be performed on the difference image (point source photometry only). Those measurements will be stored as `DIAForcedSources`. No alerts will be issued for these `DIAForcedSources`, but the `DIAForcedSource` measurements will be included in any future alerts triggered by a new `DIASource` at that location. DMS-REQ-0317

`Object` or `SSObject`. The algorithm will take into account the parallax and proper (or Keplerian) motions, as well as the errors in estimated positions of `DIAObject`, `SSObject`, and `DIASource`, to find the maximally likely match. Multiple `DIASources` in the same visit will not be matched to the same `DIAObject`.

[23] This restriction is removed when Prompt processing is rerun during Data Release production, see § 3.3.5.

[24] Data Release database is a database resulting from annual data release processing. See § 4 for details.

[25] We guarantee that a receiver will always be able to regenerate the alert contents at any later date using the included timestamps and metadata.

11. Within 24 hours of discovery, *precovery* PSF forced photometry will be performed on any difference image overlapping the position of new `DIAObjects` taken within the past 30 days, and added to the `DIAForcedSource` table. Alerts will not be issued with precovery photometry information but the resulting `DIAForcedSource` measurements will be included in future alerts from this `DIAObject`. L1PublicT precoveryWindow DMS-REQ-0287

In addition to the processing described above, a smaller sample of sources detected on difference images *below* the nominal $transSNR = 5$ threshold will be measured and stored, in order to enable monitoring of difference image analysis quality. transSNR DMS-REQ-0270

Also, the system will have the ability to measure and alert on a limited[26] number of sources detected below the nominal threshold for which additional criteria are satisfied. For example, a $transSNR$ = 3 source detection near a gravitational keyhole[27] may be highly significant in assessing the danger posed by a potentially hazardous asteroid. The initial set of criteria will be defined by the start of LSST operations.

3.2.2 Solar System Object Processing

The Solar System Processing described in this section occurs in daytime, after a night of observing . Its goal is to link (identify) previously unknown `SSObjects`, given the additional night of observing and report the discoveries to the Minor Planet Center, as well as compute physical (e.g., absolute magnitudes) and other auxiliary properties (e.g., predicted apparent magnitudes and coordinates in various coordinate systems) for known Solar System objects and their LSST observations. The process is graphically illustrated in Figure 6. DMS-REQ-0004 DMS-REQ-0089

The pipeline consists of the following conceptual steps:

1. **Linking:** All `DIASources` detected on the previous night that have *not* been matched at a high confidence level to a known `Object`, `DIAObject`, `SSObject`, or an artifact, are analyzed for potential pairs, forming *tracklets*. The collection of tracklets collected over no fewer than past 14 days[28] is searched for subsets forming *tracks* consistent with being on the

[26] It will be sized for no less than ∼10% of average `DIASource` per visit rate.
[27] A gravitational keyhole is a region of space where Earth's gravity would modify the orbit of a passing asteroid such that the asteroid would collide with the Earth in the future.
[28] The exact time window is largely computationally limited; longer windows increase the discovery rates and are preferable.

same Keplerian orbit around the Sun.

2. **Reporting:** These newly discovered Solar System objects are reported to the MPC, using the standard data-exchange protocols (e.g., the ADES format). The measurements of all `DIASources` detected on the previous night that have been matched at a high level of confidence to a known `SSObject` are also submitted to the MPC.

3. **Catalog Update:** An updated orbit catalog, incorporating previously submitted LSST discoveries as well as discoveries by other contemporaneous programs, is downloaded from the MPC and ingested into the Prompt Products database. `DIASource` records are updated to point to the relevant `SSObject` records. `DIAObjects` "orphaned" by this unlinking are deleted.[29].

4. **Physical Characterization:** The physical properties of all known `SSObjects`, as defined by the orbit catalog, are recomputed. Updated data are entered into the relevant tables. DMS-REQ-0288 DMS-REQ-0273

5. **Precovery:** Precovery linking is attempted for all `SSObjects` whose orbits were updated in this process (or are new). Where successful, newly discovered observations are queued up for submission to the MPC. DMS-REQ-0286

3.3 Prompt Catalogs

The alert processing design relies on the Prompt Products database that contains the objects and sources detected on difference images. At the very least[30], this database will have tables of `DIASources`, `DIAObjects`, and `SSObjects`, populated in the course of nightly and daily difference image and Solar System object processing. As these get updated and added to, their updated contents becomes visible (query-able) to users within 24 hours of the corresponding observation times. In the case of SSObjects, this means 24 hours after orbits have been determined or updated. DMS-REQ-0269 DMS-REQ-0271 DMS-REQ-0273 L1PublicT DMS-REQ-0312

This database is *only loosely coupled to the Data Release database*. All of the coupling is through positional matches between the `DIAObjects` entries in the Prompt Products database and the `Objects` in the Data Release database. There is no direct `DIASource`-to-`Object` match: in general, a time-domain object is not necessarily the same astrophysical object as a static-sky ob-

[29]Some `DIAObjects` may only be left with forced photometry measurements at their location (since all `DIAObjects` are force-photometered on previous and subsequent visits); these will be kept but flagged as such.

[30]It will also contain exposure and visit metadata, SSP-specific tables, etc. These are either standard/uncontroversial, implementation-dependent, or less directly relevant for science and therefore not discussed in this document.

ject, even if the two are positionally coincident (eg. an asteroid overlapping a galaxy). Therefore, adopted data model emphasizes that *having a* `DIASource` *be positionally coincident with an* `Object` *does not imply it is physically related to it*. Absent other information, the least presumptuous data model relationship is one of *positional association*, not *physical identity*.

This may seem odd at first: for example, in a simple case of a variable star, matching individual `DIASources` to `Objects` is exactly what an astronomer would want. That approach, however, fails in the following scenarios:

- *A supernova in a galaxy.* The matched object in the `Object` table will be the galaxy, which is a distinct astrophysical object. We want to keep the information related to the supernova (eg., colors, the light curve) separate from those measurements for the galaxy.
- *An asteroid occulting a star.* If associated with the star on first apparition, the association would need to be dissolved when the source is recognized as an asteroid (perhaps even as early as a day later).
- *A supernova on top of a pair of blended galaxies.* It is not clear in general to which galaxy this `DIASource` would "belong". That in itself is a research question.

`DIASource`-to-`Object` matches can still be emulated via a two-step relation (`DIASource-DIAObject-Object`). For ease of use, views or pre-built table with these matches will be offered to the end-users. DMS-REQ-0324

In the sections to follow, we present the *conceptual schemas* for the most important Prompt Products database tables. These convey *what* data will be recorded in each table, rather than the details of *how*. For example, columns whose type is an array (eg., `radec`) may be expanded to one table column per element of the array (eg., `ra`, `decl`) once this schema is translated to SQL[31]. Secondly, the tables to be presented are largely normalized (i.e., contain no redundant information). For example, since the band of observation can be found by joining a `DIASource` table to the table with exposure metadata, there's no column named `band` in the `DIASource` table. In the as-built database, the views presented to the users will be appropriately denormalized for ease of use.

[31] The SQL realization of this schema is defined in the `cat` package and documented in LDM-153 and can be browsed at `http://ls.st/8g4`

3.3.1 DIASource Table

This is a table of sources detected at $transSNR \geq 5$ on difference images[32] (DIASources). On average, the LSST SRD expects up to ∼10,000 astrophysical DIASources per visit (∼10M per night; 100,000 per deg^2 of the sky per hour). transSNR LSR-REQ-0101 transN

Some $transSNR \geq 5$ sources will not be caused by observed astrophysical phenomena, but by artifacts (bad columns, diffraction spikes, etc.). The difference image analysis software will attempt to identify and flag these as such.

Unless noted otherwise, all DIASource quantities (fluxes, centroids, etc.) are measured on the difference image. DMS-REQ-0269

Table 1: DIASource Table

Name	Type	Unit	Description
diaSourceId	uint64		Unique source identifier
ccdVisitId	uint64		ID of CCD and visit where this source was measured
diaObjectId	uint64		ID of the DIAObject this source was associated with, if any.
ssObjectId	uint64		ID of the SSObject this source has been linked to, if any.
parentDiaSourceId	uint64		ID of the parent DIASource this object has been deblended from, if any.
midPointTai	double	time	Time of mid-exposure for this DIASource[33].
radec	double[2]	degrees	Centroid, (α, δ)[34].
radecCov	float[3]	various	radec covariance matrix.
xy	float[2]	pixels	Column and row of the centroid.
xyCov	float[3]	various	Centroid covariance matrix.

Continued on next page

[32]This requirement is specified in the LSST SRD.
[33]The visit mid-exposure time generally depends on the position of the source relative to the shutter blade motion trajectory.
[34]The astrometric reference frame will be chosen closer to start of operations.

Table 1: `DIASource` Table

Name	Type	Unit	Description
apFlux	float	nJy[35]	Calibrated aperture flux. Note that this actually measures the flux *difference* between the template and the visit image.
apFluxErr	float	nJy	Estimated uncertainty of `apFlux`.
SNR	float		The signal-to-noise ratio at which this source was detected in the difference image.[36]
psFlux	float	nJy	Calibrated flux for point source model. Note this actually measures the flux *difference* between the template and the visit image.
psFluxErr	float	nJy	Estimated uncertainty of `psFlux`.
psRadec	double[2]	degrees	Centroid for point source model.
psCov	float[6]	various	Covariance matrix for point source model parameters.
psLnL	float		Natural log likelihood of the observed data given the point source model.
psChi2	float		χ^2 statistic of the model fit.
psNdata	int		The number of data points (pixels) used to fit the model.
trailFlux	float	nJy	Calibrated flux for a trailed source model[37,38]. Note this actually measures the flux *difference* between the template and the visit image.

Continued on next page

[35]LSST fluxes are reported in nanojansky (1 nJy = 10^{-35} Wm^{-2}Hz^{-1}). See §2.3 and LSST Document-27758 for details.

[36]This is not necessarily the same as apFlux/apFluxErr, as the flux measurement algorithm may be more accurate than the detection algorithm.

[37]A *Trailed Source Model* attempts to fit a (PSF-convolved) model of a point source that was trailed by a certain amount in some direction (taking into account the two-snap nature of the visit, which may lead to a dip in flux around the mid-point of the trail). Roughly, it's a fit to a PSF-convolved line. The primary use case is to characterize fast-moving Solar System objects.

[38]This model does not fit for the *direction* of motion; to recover it, we would need to fit the model to separately to individual snaps of a visit. This adds to system complexity, and is not clearly justified by increased Solar System object linking performance given the added information.

Table 1: `DIASource` Table

Name	Type	Unit	Description
trailRadec	double[2]	degrees	Centroid for trailed source model.
trailLength	float	arcsec	Maximum likelihood fit of trail length[39,40].
trailAngle	float	degrees	Maximum likelihood fit of the angle between the meridian through the centroid and the trail direction (bearing, direction of motion).
trailCov	float[15]	various	Covariance matrix of trailed source model parameters.
trailLnL	float		Natural log likelihood of the observed data given the trailed source model.
trailChi2	float		χ^2 statistic of the model fit.
trailNdata	int		The number of data points (pixels) used to fit the model.
dipMeanFlux	float	nJy	Maximum likelihood value for the mean absolute flux of the two lobes for a dipole model[41].
dipFluxDiff	float	nJy	Maximum likelihood value for the difference of absolute fluxes of the two lobes for a dipole model.
dipRadec	double[2]	degrees	Centroid for dipole model.
dipLength	float	arcsec	Maximum likelihood value for the lobe separation in dipole model.
dipAngle	float	degrees	Maximum likelihood fit of the angle between the meridian through the centroid and the dipole direction (bearing, from negative to positive lobe).

Continued on next page

[39] Note that we'll likely measure trailRow and trailCol, and transform to trailLength/trailAngle (or trailRa/trailDec) for storage in the database. A stretch goal is to retain both.

[40] TBD: Do we need a separate trailCentroid? It's unlikely that we do, but one may wish to prove it.

[41] A *Dipole Model* attempts to fit a (PSF-convolved) model of two point sources, with fluxes of opposite signs, separated by a certain amount in some direction. The primary use case is to characterize moving stars and problems with image differencing (e.g., due to astrometric offsets).

Table 1: `DIASource` Table

Name	Type	Unit	Description
dipCov	float[21]	various	Covariance matrix of dipole model parameters.
dipLnL	float		Natural log likelihood of the observed data given the dipole source model.
dipChi2	float		χ^2 statistic of the model fit.
dipNdata	int		The number of data points (pixels) used to fit the model.
totFlux	float	nJy	Calibrated flux for point source model measured on the visit image centered at the centroid measured on the difference image (forced photometry flux)
totFluxErr	float	nJy	Estimated uncertainty of `totFlux`.
diffFlux	float	nJy	Calibrated flux for point source model centered on `radec` but measured on the difference of snaps comprising this visit[42].
diffFluxErr	float	nJy	Estimated uncertainty of `diffFlux`.
fpBkgd	float	nJy/arcsec^2	Estimated background at the position (centroid) of the object in the template image.
fpBkgdErr	float	nJy/arcsec^2	Estimated uncertainty of `fpBkgd`.
Ixx	float	arcsec^2	Adaptive second moment of the source intensity. See Bernstein & Jarvis [2] for detailed discussion of all adaptive-moment related quantities[43].
Iyy	float	arcsec^2	Adaptive second moment of the source intensity.
Ixy	float	arcsec^2	Adaptive second moment of the source intensity.

Continued on next page

[42] This flux can be used to detect sources changing on timescales comparable to snap exposure length (∼15 sec).
[43] Or `http://ls.st/5f4` for a brief summary.

Table 1: DIASource Table

Name	Type	Unit	Description
Icov	float[6]	arcsec^4	Ixx, Iyy, Ixy covariance matrix.
IxxPSF	float	arcsec^2	Adaptive second moment for the PSF.
IyyPSF	float	arcsec^2	Adaptive second moment for the PSF.
IxyPSF	float	arcsec^2	Adaptive second moment for the PSF.
extendedness	float		A measure of extendedness, computed using a combination of available moments, or from a likelihood ratio of point/trailed source models (exact algorithm TBD). $extendedness = 1$ implies a high degree of confidence that the source is extended. $extendedness = 0$ implies a high degree of confidence that the source is point-like.
spuriousness	float		A measure of spuriousness, computed using information[44] from the source and image characterization, as well as the information on the Telescope and Camera system (e.g., ghost maps, defect maps, etc.).
flags	bit[64]	bit	Various useful flags.

Table 2: DIAForcedSource Table

Name	Type	Unit	Description
diaForcedSourceId	uint64		Unique source identifier
ccdVisitId	uint64		ID of CCD and visit where this source was measured
diaObjectId	uint64		ID of the DIAObject this forced photometry was seeded by.

Continued on next page

[44]The computation of spuriousness will be "prior free" to the extent possible and not use any information about the astrophysical neighborhood of the source, whether it has been previously observed or not, etc. The intent is to avoid introducing a bias against unusual sources or sources discovered in unusual environments.

Table 2: `DIAForcedSource` Table

Name	Type	Unit	Description
midPointTai	double	time	Time of mid-exposure for this `DIAForcedSource`.
psFlux	float	nJy	Calibrated flux for point source model. Note this actually measures the flux *difference* between the template and the visit image.
psFluxErr	float	nJy	Estimated uncertainty of `psFlux`.
totFlux	float	nJy	Calibrated flux for point source model measured on the visit image centered at the `DIAObject` centroid.
totFluxErr	float	nJy	Estimated uncertainty of `totFlux`.

Some fast-moving, trailed, sources may be due to passages of nearby asteroids. Their trails may exhibit significant curvature. While we do not measure the curvature directly, it can be inferred by examining the length of the trail, the trailed model covariance matrices, and the adaptive shape measures. Once curvature is suspected, the users may fit curved trail models to the cutout provided with the alert.

3.3.2 `DIAObject` Table

DMS-REQ-0271
DMS-REQ-0272
diaNearbyMaxObj

Table 3: `DIAObject` Table

Name	Type	Unit	Description
diaObjectId	uint64		Unique identifier.
radec	double[2]	degrees	(α, δ) position of the object at time `radecTai`.
radecCov	float[3]	various	`radec` covariance matrix.
radecTai	double	time	Time at which the object was at a position `radec`.
pm	float[2]	mas/yr	Proper motion vector[45].

Continued on next page

[45] High proper-motion or parallax objects will appear as “dipoles” in difference images. Great care will have to be taken not to misidentify these as subtraction artifacts.

Table 3: `DIAObject` Table

Name	Type	Unit	Description
parallax	float	mas	Trigonometric parallax.
pmParallaxCov	float[6]	various	Proper motion - parallax covariances.
pmParallaxLnL	float		Natural log of the likelihood of the linear proper motion-parallax fit[46].
pmParallaxChi2	float		χ^2 statistic of the model fit.
pmParallaxNdata	int		The number of data points used to fit the model.
psFluxMean	float[ugrizy]	nJy	Weighted mean of point-source model flux, `psFlux`.
psFluxMeanErr	float[ugrizy]	nJy	Standard error of `psFluxMean`.
psFluxSigma	float[ugrizy]	nJy	Standard deviation of the distribution of `psFlux`.
psFluxChi2	float[ugrizy]		χ^2 statistic for the scatter of `psFlux` around `psFluxMean`.
psFluxNdata	int[ugrizy]		The number of data points used to compute `psFluxChi2`.
totFluxMean	float[ugrizy]	nJy	Weighted mean of forced photometry flux, `totFlux`.
totFluxMeanErr	float[ugrizy]	nJy	Standard error of `totFluxMean`.
totFluxSigma	float[ugrizy]	nJy	Standard deviation of the distribution of `totFlux`.
lcPeriodic	float[6×32]		Periodic features extracted from `DIASource` light-curves using generalized Lomb-Scargle periodogram [Table 4, 18][47].
lcNonPeriodic	float[6×20]		Non-periodic features extracted from `DIASource` light-curves [Table 5, 18].
nearbyObj	uint64[6]		Closest `Objects` (3 stars and 3 galaxies) in Data Release database.

Continued on next page

[46] `radec`, `pm`, and `parallax` will all be simultaneously fitted for.

[47] The exact features in use when LSST begins operations are likely to be different compared to the baseline described here. This is to be expected given the rapid pace of research in time domain astronomy. However, the *number* of computed features is unlikely to grow beyond the present estimate.

Table 3: DIAObject Table

Name	Type	Unit	Description
nearbyObjDist	float[6]	arcsec	Distances to nearbyObj.
nearbyObjLnP	float[6]		Natural log of the probability that the observed DIAObject is the same as the nearby Object[48].
nearbyExtObj	unit64[3]		Three extended Objects with lowest separations in Data Release database[49].
nearbyExtObjSep	float[3]		Separations of nearbyExtObj.
nearbyLowzGal	str[1]		External catalog name of the nearest low-redshift galaxy[50].
nearbyLowzGalSep	float[1]		Separation of nearbyLowzGal[51].
flags	bit[64]	bit	Various useful flags.

3.3.3 MPCORB, SSObject and SSSourceTables

The three tables presented below capture the information that will be available in the database for each Solar System Object (SSObject) and for each observation of a Solar System Object (SSSource). **The details of the schema will evolve as the implementation progresses, maintaining the high-level content**. The implementation will be maintained in the catalog repository at `https://github.com/lsst/cat`, as well as linked at `http://ls.st/ssp`. DMS-REQ-0273

Table 4: MPCORB Table

Name	Type	Unit	Description
mpcDesignation	VARCHAR(8)		MPCORB: Number or provisional designation (in packed form)

Continued on next page

[48]This quantity will be computed by marginalizing over the product of position and proper motion error ellipses of the Object and DIAObject, assuming an appropriate prior.
[49]Separations should be calculated with respect to the transient location using the second moments of each Object's luminosity profile, as described in DMTN-151.
[50]External catalog will be, e.g., the NGC/IC, unless the community provides one that they deem to be more scientifically useful, as described in DMTN-151.
[51]Separations will be radial offset in arcseconds unless a community-provided catalog includes galaxy characteristics that would enable an alternative, as described in DMTN-151.

Table 4: MPCORB Table

Name	Type	Unit	Description
mpcNumber	INTEGER		MPC number (if the asteroid has been numbered; NULL otherwise). Provided for convenience.
ssObjectId	BIGINT		LSST unique identifier (if observed by LSST)
mpcH	FLOAT	mag	MPCORB: Absolute magnitude, H
mpcG	FLOAT		MPCORB: Slope parameter, G
epoch	DOUBLE	MJD	MPCORB: Epoch (in MJD, .0 TT)
M	DOUBLE	degrees	MPCORB: Mean anomaly at the epoch, in degrees
peri	DOUBLE	degrees	MPCORB: Argument of perihelion, J2000.0 (degrees)
node	DOUBLE	degrees	MPCORB: Longitude of the ascending node, J2000.0 (degrees)
incl	DOUBLE	degrees	MPCORB: Inclination to the ecliptic, J2000.0 (degrees)
e	DOUBLE		MPCORB: Orbital eccentricity
n	DOUBLE	degrees/day	MPCORB: Mean daily motion (degrees per day)
a	DOUBLE	AU	MPCORB: Semimajor axis (AU)
uncertaintyParameter	VARCHAR(1)		MPCORB: Uncertainty parameter, U
reference	VARCHAR(9)		MPCORB: Reference
nobs	INTEGER		MPCORB: Number of observations
nopp	INTEGER		MPCORB: Number of oppositions
arc	FLOAT	days	MPCORB: Arc (days), for single-opposition objects
arcStart	DATETIME		MPCORB: Year of first observation (for multi-opposition objects)
arcEnd	DATETIME		MPCORB: Year of last observation (for multi-opposition objects)
rms	FLOAT	arcsec	MPCORB: r.m.s residual (″)

Continued on next page

Table 4: `MPCORB` Table

Name	Type	Unit	Description
pertsShort	VARCHAR(3)		MPCORB: Coarse indicator of perturbers (blank if unperturbed one-opposition object)
pertsLong	VARCHAR(3)		MPCORB: Precise indicator of perturbers (blank if unperturbed one-opposition object)
computer	VARCHAR(10)		MPCORB: Computer name
flags	INTEGER		MPCORB: 4-hexdigit flags. See https://minorplanetcenter.net//iau/info/MPOrbitForm for details
fullDesignation	VARCHAR(26)		MPCORB: Readable designation
lastIncludedObservation	FLOAT	MJD	MPCORB: Date of last observation included in orbit solution

The `MPCORB` table will be ingested, at least daily, from the Minor Planet Center. Its exact contents will mirror the columns available in the MPCORB at the time the LSST is operations. While MPCORB does not contain orbital element covariances at this point, it is expected these will become available in time for LSST operations.

Table 5: `SSObject` Table

Name	Type	Unit	Description
ssObjectId	BIGINT		Unique identifier.
discoverySubmissionDate	DOUBLE	date	The date the LSST first linked and submitted the discovery observations to the MPC. May be NULL if not an LSST discovery. The date format will follow general LSST conventions (MJD TAI, at the moment).
firstObservationDate	DOUBLE	date	The time of the first LSST observation of this object (could be precovered)
arc	FLOAT	days	Arc of LSST observations

Continued on next page

Table 5: `SSObject` Table

Name	Type	Unit	Description
numObs	INTEGER		Number of LSST observations of this object
lcPeriodic	BLOB(768)		Periodic light curve features computed on phase/distance-corrected magnitudes (H), 6x32 floats
MOID	FLOAT	AU	Minimum orbit intersection distance to Earth
MOIDTrueAnomaly	FLOAT	deg	True anomaly of the MOID point
MOIDEclipticLongitude	FLOAT	deg	Ecliptic longitude of the MOID point
MOIDDeltaV	FLOAT	km/s	DeltaV at the MOID point
uH	FLOAT	mag	Best fit absolute magnitude (u band)
uG12	FLOAT	mag	Best fit G12 slope parameter (u band)
uHErr	FLOAT	mag	Uncertainty of H (u band)
uG12Err	FLOAT	mag	Uncertainty of G12 (u band)
uH_uG12_Cov	FLOAT	mag^2	H-G12 covariance (u band)
uChi2	FLOAT		Chi^2 statistic of the phase curve fit (u band)
uNdata	INTEGER		The number of data points used to fit the phase curve (u band)
gH	FLOAT	mag	Best fit absolute magnitude (g band)
gG12	FLOAT	mag	Best fit G12 slope parameter (g band)
gHErr	FLOAT	mag	Uncertainty of H (g band)
gG12Err	FLOAT	mag	Uncertainty of G12 (g band)
gH_gG12_Cov	FLOAT	mag^2	H-G12 covariance (g band)
gChi2	FLOAT		Chi^2 statistic of the phase curve fit (g band)
gNdata	INTEGER		The number of data points used to fit the phase curve (g band)
rH	FLOAT	mag	Best fit absolute magnitude (r band)
rG12	FLOAT	mag	Best fit G12 slope parameter (r band)
rHErr	FLOAT	mag	Uncertainty of H (r band)
rG12Err	FLOAT	mag	Uncertainty of G12 (r band)

Continued on next page

Table 5: SSObject Table

Name	Type	Unit	Description
rH_rG12_Cov	FLOAT	mag^2	H-G12 covariance (r band)
rChi2	FLOAT		Chi^2 statistic of the phase curve fit (r band)
rNdata	INTEGER		The number of data points used to fit the phase curve (r band)
iH	FLOAT	mag	Best fit absolute magnitude (i band)
iG12	FLOAT	mag	Best fit G12 slope parameter (i band)
iHErr	FLOAT	mag	Uncertainty of H (i band)
iG12Err	FLOAT	mag	Uncertainty of G12 (i band)
iH_iG12_Cov	FLOAT	mag^2	H-G12 covariance (i band)
iChi2	FLOAT		Chi^2 statistic of the phase curve fit (i band)
iNdata	INTEGER		The number of data points used to fit the phase curve (i band)
zH	FLOAT	mag	Best fit absolute magnitude (z band)
zG12	FLOAT	mag	Best fit G12 slope parameter (z band)
zHErr	FLOAT	mag	Uncertainty of H (z band)
zG12Err	FLOAT	mag	Uncertainty of G12 (z band)
zH_zG12_Cov	FLOAT	mag^2	H-G12 covariance (z band)
zChi2	FLOAT		Chi^2 statistic of the phase curve fit (z band)
zNdata	INTEGER		The number of data points used to fit the phase curve (z band)
yH	FLOAT	mag	Best fit absolute magnitude (y band)
yG12	FLOAT	mag	Best fit G12 slope parameter (y band)
yHErr	FLOAT	mag	Uncertainty of H (y band)
yG12Err	FLOAT	mag	Uncertainty of G12 (y band)
yH_yG12_Cov	FLOAT	mag^2	H-G12 covariance (y band)
yChi2	FLOAT		Chi^2 statistic of the phase curve fit (y band)
yNdata	INTEGER		The number of data points used to fit the phase curve (y band)

Continued on next page

Table 5: `SSObject` Table

Name	Type	Unit	Description
maxExtendedness	FLOAT		maximum 'extendedness' value from the DIASource
minExtendedness	FLOAT		minimum 'extendedness' value from the DIASource
medianExtendedness	FLOAT		median 'extendedness' value from the DIASource
flags	BIGINT		Flags, bitwise OR tbd.

The G_{12} parameter for a large fraction of asteroids may not be well constrained until later in the survey. We may decide not to fit for it at all over the first few DRs and add it later in Operations. Alternatively, we may fit it using strong priors on slopes poorly constrained by the data. The design of the data management system is insensitive to this decision, making it possible to postpone it to Commissioning to ensure it follows the standard community practice at that time.

Per-observation quantities: The LSST database will provide an auxiliary table (`SSSource`) or equivalent functions to compute the phase (Sun-Asteroid-Earth) angle α for every observation, the reduced, $H(\alpha)$, and absolute, H, asteroid magnitudes in LSST bands, as well as any other quantities defined in the realized schema. DMS-REQ-0323

Table 6: `SSSource` Table

Name	Type	Unit	Description
ssObjectId	BIGINT		Unique identifier of the object.
diaSourceId	BIGINT		Unique identifier of the observation
mpcUniqueId	BIGINT		MPC unique identifier of the observation
nearbyObj	BIGINT[6]		Closest Objects (3 stars and 3 galaxies) in Level 2 database.
nearbyObjDist	FLOAT[6]		Distances to nearbyObj

Continued on next page

Table 6: SSSource Table

Name	Type	Unit	Description
nearbyObjLnP	FLOAT[6]		Natural log of the probability that the observed DIAObject is the same as the nearby Object
eclipticLambda	DOUBLE	deg	Ecliptic longitude
eclipticBeta	DOUBLE	deg	Ecliptic latitude
galacticL	DOUBLE	deg	Galactic longitude
galacticB	DOUBLE	deg	Galactic latitute
phaseAngle	FLOAT	deg	Phase angle
heliocentricDist	FLOAT	AU	Heliocentric distance
topocentricDist	FLOAT	AU	Topocentric distace
predictedMagnitude	FLOAT	mag	Predicted magnitude
predictedMagnitudeSigma	FLOAT	mag	Prediction uncertainty (1-sigma)
residualRa	DOUBLE	deg	Residual R.A. vs. ephemeris
residualDec	DOUBLE	deg	Residual Dec vs. ephemeris
predictedRaSigma	FLOAT	deg	Predicted R.A. uncertainty
predictedDecSigma	FLOAT	deg	Predicted Dec uncertainty
predictedRaDecCov	FLOAT	deg^2	Predicted R.A./Dec covariance
heliocentricX	FLOAT	AU	Cartesian heliocentric coordinates (at the emit time)
heliocentricY	FLOAT	AU	
heliocentricZ	FLOAT	AU	
heliocentricVX	FLOAT	AU	Cartesian heliocentric velocities (at the emit time)
heliocentricVY	FLOAT	AU	
heliocentricVZ	FLOAT	AU	
topocentricX	FLOAT	AU	Cartesian topocentric coordinates (at the emit time)
topocentricY	FLOAT	AU	
topocentricZ	FLOAT	AU	
topocentricVX	FLOAT	AU	Cartesian topocentric velocities (at the emit time)
topocentricVY	FLOAT	AU	

Continued on next page

Table 6: SSSource Table

Name	Type	Unit	Description
topocentricVZ	FLOAT	AU	

3.3.4 Precovery Measurements

When a new DIASource is detected, it's useful to perform forced PSF photometry at the location of the new source on images taken prior to discovery. These are colloquially known as *precovery measurements*[52]. Performing precovery in real time over all previously acquired visits is too I/O intensive to be feasible. We therefore plan the following: DMS-REQ-0287 DMS-REQ-0286

1. For all newly discovered objects, perform precovery PSF photometry on visits taken over the previous 30 days[53]. These measurements will be stored in the DIAForcedSource table. precoveryWindow
2. Make available a "precovery service" to request precovery for a limited number of DIASources across all previous visits, and make it available within 24 hours of the request. Web interface and machine-accessible APIs will be provided. DMS-REQ-0341

The former should satisfy the most common use cases (eg., SNe), while the latter will provide an opportunity for more extensive yet timely precovery of targets of special interest.

3.3.5 Image differencing during Data Release Production

In what we've described so far, the Prompt Products database is continually being added to as new images are taken and DIASources identified. Every time a new DIASource is associated to an existing DIAObject, the DIAObject record is updated to incorporate new information brought in by the DIASource. Once discovered and measured, the DIASources would never be re-discovered and re-measured at the pixel level. DMS-REQ-0312

This would be far from optimal. The instrument will be better understood with time. Newer versions of LSST pipelines will improve detection and measurements on older data. Also,

[52]When Solar System objects are concerned, precovery has a slightly different meaning: predicting the positions of newly identified SSObjects on previously acquired visits, and associating with them the DIASources consistent with these predictions.

[53]We will be maintaining a cache of 30 days of processed images to support this feature.

precovery photometry should optimally be performed for *all* objects, and not just a select few. The annual Data Release productions will therefore also include a full reanalysis of the time-domain information in the collected dataset , by performing image differencing on all images collected by the survey . The images will be processed using a single version of the image differencing and measurement software, resulting in a consistent set of `DIASources`, `DIAObjects`, `ForcedSources`, and `SSObjects` as part of a data release. DMS-REQ-0313 DMS-REQ-0325

There will be three main advantages of time-domain data produced during Data Release processing, compared to the Prompt Products database: i) even the oldest data will be processed with the latest software, ii) astrometric and photometric calibration will be better, and iii) there will be no 12-month limit on the width of data window used to computed associated `DIAObject` measurements (proper motions, centroids, light curves, etc.). diaCharacterization-Cutoff

3.4 Prompt Image Products

3.4.1 Visit Images

DMS-REQ-0069

Raw and Processed Visit Images will be made available for download no later than 24 hours from the end of visit acquisition. L1PublicT

The images 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). l1CacheLifetime

3.4.2 Difference Images

DMS-REQ-0010

Complete difference images will be made available for download no later than 24 hours from the end of visit acquisition. L1PublicT

The images 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). l1CacheLifetime

3.4.3 Image Differencing Templates

DMS-REQ-0280

Coadded images will be used as templates for difference image analysis. The coaddition process will take care to remove transient or fast moving objects (eg., asteroids) from the templates. The time range of the epochs included in a template is TBD, and will be chosen to minimize false positives due to high proper-motion stars (favoring shorter ranges) with the need to correct for DCR and maximize depth (favoring longer ranges). The numbers and types of coadds used for templates are also TBD.

3.5 Alerts to `DIASources`

3.5.1 Information Contained in Each Alert

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). These alerts will be issued in `VOEvent` format[54], and should be readable by `VOEvent`-compliant clients. OTT1 DMS-REQ-0002

Each alert (a `VOEvent` packet) will at least include the following:

- *alertID*: An ID uniquely identifying this alert. It can also be used to execute a query against the Prompt Products database as it existed when this alert was issued. DMS-REQ-0274
- Science Data:
  - The `DIASource` record that triggered the alert, as well as the `filterName` and `programId` of the corresponding `Visit`
  - The entire `DIAObject` (or `SSObject`) record. `DIAObject` records include matching `Object` IDs from the latest Data Release, if they exist.
  - Any `DIASource` and `DIAForcedSource` records that exist, and difference image noise estimates where they do not, taken from the previous 12 months. diaCharacterization-Cutoff
- 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)

The variable-size cutouts will be sized so as to encompass the entire footprint of the detected source, but be no smaller than 30×30 pixels. The provided images will comprise of a flux (32 bit float), variance (32 bit float), and mask (at least 16 bit flags) planes, and include metadata necessary for further processing (e.g., WCS, zero-point, PSF, etc.).

[54] Or some other format that is broadly accepted and used by the community at the start of LSST commissioning.

The items above are meant to represent the *information* transmitted with each alert; the content of the alert packet itself will be formatted to confirm to `VOEvent` (or other relevant) standard. Where the existing standard is inadequate for LSST needs, LSST will propose extensions and work with the community to reach a common solution.

With each alert, we attempt to include as much information known to LSST about the `DIASource` as possible, to minimize the need for follow-up database queries. This speeds up classification and decision making at the user end, and relaxes the requirements on the database on the Project end.

3.5.2 Receiving and Filtering the Alerts

Alerts will be transmitted in `VOEvent` format, using standard IVOA protocols (e.g., VOEvent Transport Protocol; VTP[55]). As a very high rate of alerts is expected, approaching ∼10 million per night, we plan for public VOEvent Event Brokers[56] to be the primary end-points of LSST's event streams. End-users will use these brokers to classify and filter events for subsets fitting their science goals. End-users will *not* be able to subscribe to full, unfiltered, alert streams coming directly from LSST[57]. DMS-REQ-0002

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 (at NCSA). It will let astronomers create simple filters that limit what alerts are ultimately forwarded to them[58]. These *user defined filters* will be possible to specify using an SQL-like declarative language, or short snippets of (likely Python) code. For example, here's what a filter may look like: DMS-REQ-0342

```
# 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.
```

[55]VOEvent Transport Protocol is currently an IVOA Note, but we understand work is under way to finalize and bring it up to full IVOA Recommendation status.

[56]These brokers are envisioned to be operated as a public service by third parties who will have signed MOUs with LSST.

[57]This is due to finite network bandwidth available: for example, 100 end-users subscribing to a ∼100 Mbps stream (the peak full stream data rate at end of the first year of operations) would require 10Gbps WAN connection from the archive center, just to serve the alerts.

[58]More specifically, to their VTP clients. Typically, a user will use the Science User Interface (the web portal to LSST Archive Center) to set up the filters, and use their VTP client to receive the filtered `VOEvent` stream.

```
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
```

We emphasize that this LSST-provided capability will be limited, and is *not* intended to satisfy the wide variety of use cases that a full-fledged public Event Broker could. For example, we do not plan to provide any *exclusive* classification to a unique category of object. Following the SRD specification, however, we will provide a limited number of pre-defined filters for a small number of object types of common interest. These will answer non-exclusive questions DMS-REQ-0348
such as "is the light curve consistent with an RR Lyra?", and will have potentially highly overlapping selections, designed to provide good completeness but perhaps only very modest purity. No information beyond what is contained in the VOEvent packet will be available to the pre-defined or user-defined filters (e.g., no cross-matches to other catalogs). The complexity and run time of user defined filters will be limited by available resources. Execution latency will not be guaranteed. The number of VOEvents transmitted to each user per visit will be limited as well (e.g., the equivalent of 20 full-size alert packets per visit per user, dynamically throt- numBrokerAlerts
tled depending on load). Finally, the total number of simultaneous subscribers is likely to be DMS-REQ-0343
limited – in case of overwhelming interest, a TAC-like proposal process may be instituted. numBrokerUsers

4 Data Release Data Products

4.1 Overview

Data Release data products result from direct image[59] analysis. They're designed to enable *static sky* science (eg., studies of galaxy evolution, or weak lensing), and time-domain science that is not time sensitive (eg. statistical investigations of variability). They include image products (reduced single-epoch exposures, called *Processed Visit Images* or, sometimes, *calibrated exposures*, and coadds), and catalog products (tables of objects, sources, their measured properties, and related metadata).

DMS-REQ-0069
DMS-REQ-0279
DMS-REQ-0281
DMS-REQ-0275
DMS-REQ-0267

Similarly to Prompt catalogs of `DIAObjects` and `DIASources`, `Objects` in the Data Release catalog represent the astrophysical phenomena (stars, galaxies, quasars, etc.), while `Sources` represent their single-epoch observations. `Sources` are independently detected and measured in single epoch exposures and recorded in the `Source` table.

DMS-REQ-0276
DMS-REQ-0275
DMS-REQ-0267

The master list of `Objects` in the Data Release will be generated by associating and deblending the list of single-epoch `DIASource` detections and the lists of sources detected on coadds (`Coadds`). We plan to build coadds designed to maximize depth (*"deep coadds"*), although this may still not include some visits with extremely poor seeing.

DMS-REQ-0275
DMS-REQ-0279

Additional coadds may be used but may not be retained (see § 4.4.3 for details). We will provide a facility to generate any coadds we use internally (as well as others we may not use internally) as pipeline tasks in the Science Platform (§ 5).

DMS-REQ-0311
DMS-REQ-0275

The deblender will be run simultaneously on the catalog of peaks[60] detected in the coadds, the `DIAObject` catalog from the Prompt Products database, and one or more external catalogs. It will use the knowledge of peak positions, bands, time, time variability (from Prompt products and the single-epoch `Source` detections), inferred motion, Galactic longitude and latitude, and other available information to produce a master list of deblended `Objects`. Metadata on why and how a particular `Object` was deblended will be kept.

DMS-REQ-0276

[59] As opposed to *difference image*, for Prompt products.

[60] The source detection algorithm we plan to employ finds regions of connected pixels above the nominal S/N threshold in the *PSF-likelihood image* of the visit (or coadd). These regions are called *footprints*. Each footprint may have one or more *peaks*, and it is these peaks that the deblender will use to infer the number and positions of objects blended in each footprint.

The properties of `Objects`, including their exact positions, motions, parallaxes, photometry, and shapes, will be characterized via a broad suite of algorithms (see § 4.2.1, mostly (but not entirely) performed on coadds.

Finally, to enable studies of variability, the fluxes of all `Objects` will be measured on individual visits (using both direct and difference images), with their shape parameters and deblending resolutions kept constant. This process is known as *forced photometry* (see § 4.2.4), and the flux measurements will be stored in the `ForcedSource` table. DMS-REQ-0268

4.2 Data Release Data Processing

Figures 3 and 4 present a high-level overview of the Data Release data processing workflow[61]. Logically[62], the processing begins with single-visit image reduction and source measurement, followed by global astrometric and photometric calibration, coadd creation, detection on coadds, association and deblending, object characterization, and forced photometry measurements.

The following is a high-level description of steps which will occur during regular Data Release data processing (bullets 1 and 2 below map to pipeline 1, *Single Visit Processing*, in Figure 3, bullet 3 is pipeline 2, *Image Coaddition*, bullets 4-6 map to pipeline 3, *Coadded Image Analysis*, and bullets 7-8 map to pipeline 4, *Multi-epoch Object Characterization*):

1. *Single Visit Processing*: Raw exposures are reduced to Processed Visit Images, and `Sources` are independently detected, deblended, and measured on all visits. Their measurements (instrumental fluxes and shapes) are stored in the `Source` table. DMS-REQ-0267

2. *Relative calibration*: The survey is internally calibrated, both photometrically and astrometrically. Relative zero-point and astrometric corrections are computed for every visit. DMS-REQ-0029 Sufficient data is kept to reconstruct the normalized system response function $\phi_b(\lambda)$ (see DMS-REQ-0030 Eq. 5, SRD) at every position in the focal plane at the time of each visit as required by § 3.3.4 of the SRD.

3. *Coadd creation*: Deep, seeing optimized, and short-period per-band coadds are created DMS-REQ-0279 in $ugrizy$ bands, as well as deeper, multi-color, coadds[63]. Transient sources (including DMS-REQ-0337 DMS-REQ-0281

[61]Note that some LSST documents refer to *Data Release Processing*, which includes both Prompt reprocessing (see § 3.3.5), and the Data Release processing described here.

[62]The actual implementation may parallelize these steps as much as possible.

Solar System objects, explosive transients, etc), will be rejected from the coadds. See § 4.4.3 for details.

4. *Source detection*. Sources will be detected on all coadds generated in the previous step. The source detection algorithm will detect regions of connected pixels, known as *footprints*, above the nominal S/N threshold in the *PSF-likelihood image* of the visit. An appropriate algorithm will be run to also detect extended low surface brightness objects (eg., binned detection algorithm from SDSS). Each footprint may have one or more *peaks*, and the collection of these peaks (and their membership in the footprints) are the output of this stage. DMS-REQ-0349

5. *Association and deblending*. The next stage in the pipeline, which we will for simplicity just call *the deblender*, will synthesize a list of unique objects. In doing so it will consider the list of sources detected on `Coadds`, catalogs of `DIASources`, `DIAObjects` and `SSObjects` detected on difference images, and objects from external catalogs[64]. DMS-REQ-0275 DMS-REQ-0034

   The deblender will make use of all information available at this stage, including the knowledge of peak positions, bands, time, time variability (from Prompt products), Galactic longitude and latitude, etc. The output of this stage is a list of uncharacterized `Objects`[65].

6. *Coadd object characterization*. The vast majority of columns in the `Object` table will be measured on one or more coadds. These include positions, shapes, and photometry. For the most part, these are static-sky measurements, but short-period coadds may be used (especially in later data releases) for some measurements with long but finite time scales, such as proper motions. DMS-REQ-0276 DMS-REQ-0275

7. *Forced Photometry*. Source fluxes will be measured on every visit, with the position and motion derived from previous steps held fixed, producing our best estimates of the lightcurve for each object in the survey. The fluxes will be stored in the `ForcedSource` table. Measurements will be performed on both direct images and difference images. DMS-REQ-0268

[63] We'll denote the "band" of the multi-color coadd as 'M'.

[64] Note that `Sources` are not considered when generating the `Object` list (given the large number of visits in each band, the false positives close to the faint end would increase the complexity of association and deblending algorithms). It is possible for intermittent sources that are detected just above the faint detection limit of single visits to be undetected in coaddds, and thus to not have a matching `Object`. To enable easy identification of such `Sources`, the nearest `Object` associated with each `Source`, if any, will be recorded, and some `Object` measurements (e.g. stellar motion parameters) *may* utilize positions from securely-matched `Sources`.

[65] Depending on the exact implementation of the deblender, this stage may also attach significant metadata (eg, deblended footprints and pixel-weight maps) to each deblended `Object` record.

8. *Object Postprocessing*. Measurements from all previous steps are combined in catalog-space algorithms. This includes fitting proper motion and parallax, and characterizing variability. These are also used to populate the `Object` table.

Data Release data processing will include the computation of photometric redshifts for all detected `Objects`. We are still investigating the best approach to take; the DM Science team is working with the community to decide on the most appropriate algorithm and the format for the results. When an algorithm and data product format has been chosen, the exact manner in which it will be incorporated into the Data Release data processing will be described here. DMS-REQ-0046

4.2.1 Object Characterization Measures

Properties of detected objects will be measured as a part of the object characterization step described in the previous section and stored in the `Object` table. These measurements are designed to enable LSST "static sky" science. This section discusses at a high level which properties will be measured and how those measurements will be performed. For a detailed list of quantities being fit/measured, see the table in § 4.3.1.

All measurements discussed in this section deal with properties of *objects*, and will be performed primarily on a suite of multi-epoch coadds. Measurements of sources in individual visits, independent of all others, are described in § 4.2.3, though in rare cases those `Source` measurements and/or the forced measurements described in § 4.2.4 may be used in addition to coadds for `Object` measurements.

The measurements performed on `Objects` include:

- *Point source model fit*. The observed object is modeled as a point source with finite proper motion and parallax and constant flux (allowed to be different in each band). This model is a good description for non-variable stars and other unresolved sources. Its 11 parameters will be constrained using some combination of matched `Source` positions, centroids measured on short-period coadds, and dipole moments measured during difference-image forced photometry[66]. DMS-REQ-0276
- *Bulge-disk model fit*. The object is modeled as a sum of a de Vaucouleurs (Sersic $n = 4$) DMS-REQ-0276

[66] The fitting procedure will account for differential chromatic refraction.

and an exponential (Sersic $n = 1$) component, but with some parameters constrained across the two components (the model described in the Table 7 fixes the ellipticity to be the same and the radii to have a constant ratio while allowing the fluxes to vary independently for example, but it should be emphasized that this just one of several possible models). The object is assumed not to move, and we will probably use the general-purpose centroid for the position rather than include the position as a model parameter. The model will be fit first to the coadds for all bands simultaneously (with only fluxes allowed to vary) and then to each band independently. In all cases the fitting is done via forward modeling – the model is convolved with the PSF before being compared to the image. Our primary goal for this model is to provide robust photometry of most galaxies detected by LSST, with shapes and other morphology information (and particularly bright and/or well-resolved galaxies) secondary goal, and we will tune the model parametrization accordingly.

- *Standard colors*. Colors of the object in "standard seeing" (for example, the third quartile expected survey seeing in the i band, ∼0.9 arcsec) will be measured. These colors are guaranteed to be seeing-insensitive, suitable for estimation of photometric redshifts[67]. DMS-REQ-0276

- *Centroids*. Centroids will be computed independently for each band using an algorithm similar to that employed by SDSS. Information from all[68] epochs will be used to derive the estimate. These centroids will be used for adaptive moment, Petrosian, Kron, standard color, and aperture measurements. DMS-REQ-0276

- *Adaptive moments*. Adaptive moments will be computed using information from all epochs, independently for each band. The moments of the PSF realized at the position of the object will be provided as well. DMS-REQ-0276

- *Petrosian and Kron fluxes*. Petrosian and Kron radii and fluxes will be measured in standard seeing using self-similar elliptical apertures computed from adaptive moments. The apertures will be PSF-corrected and *homogenized*, convolved to a canonical circular PSF[69]. The radii will be computed independently for each band. Fluxes will be computed

[67] The problem of optimal determination of photometric redshift is the subject of intense research. The approach we're taking here is conservative, following contemporary practices. As new insights develop, we will revisit the issue.

[68] Whenever we say *all*, it should be understood that this does not preclude reasonable data quality cuts to exclude data that would otherwise degrade the measurement.

[69] This is an attempt to derive a definition of elliptical apertures that does not depend on seeing. For example, for a large galaxy, the correction to standard seeing will introduce little change to measured ellipticity. Corrected apertures for small galaxies will tend to be circular (due to smearing by the PSF). In the intermediate regime, this method results in derived apertures that are relatively seeing-independent. Note that this is only the case for *apertures*; the measured flux will still be seeing dependent and it is up to the user to take this into account.

in each band, by integrating the light within some multiple of *the radius measured in the canonical band*[70] (most likely the i band). Radii enclosing 50% and 90% of light will be provided. DMS-REQ-0276

- *Aperture surface brightness*. Aperture surface brightness will be computed in a variable number[71] of concentric, logarithmically spaced, PSF-homogenized, elliptical apertures, in standard seeing. DMS-REQ-0276
- *Variability characterization*. Parameters will be provided, designed to characterize periodic and aperiodic variability features [18], in each bandpass. We caution that the exact features in use when LSST begins operations are likely to be different compared to the baseline described here; this is to be expected given the rapid pace of research in time domain astronomy. However, their *number* is unlikely to grow beyond the present estimate. DMS-REQ-0276

4.2.2 Supporting Science Cases Requiring Full Posteriors

Science cases sensitive to systematics, departures of likelihood from Gaussianity, or requiring user-specified priors, demand knowledge of the shape of the likelihood function beyond a simple Gaussian approximation around the ML value. Photometric redshift estimation is the primary example where knowledge of the full posterior is likely to be needed for LSST science cases. DMS-REQ-0046 DMS-REQ-0276

We currently plan to provide this information by providing parametric estimates of the likelihood function (for the photometric redshifts). As will be shown in Table 7, the current allocation is ~100 parameters for describing the photo-z (photometric redshift) likelihood distributions, per object. DMS-REQ-0333

The methods of storing likelihood functions (or samples thereof) will continue to be developed and optimized throughout Construction and Commissioning. The key limitation, on the amount of data needed to be stored, can be overcome by compression techniques. For example, simply noticing that not more than ~0.5% accuracy is needed for sample values allows one to increase the number of samples by a factor of 4. More advanced techniques, such as

[70] The shape of the aperture in all bands will be set by the profile of the galaxy in the canonical band alone. This procedure ensures that the color measured by comparing the flux in different bands is measured through a consistent aperture. See `http://www.sdss.org/dr7/algorithms/photometry.html` for details.

[71] The number will depend on the size of the source.

PCA analysis of the likelihoods across the entire catalog, may allow us to store even more, providing a better estimate of the shape of the likelihood function. In that sense, what is presented in Table 7 should be thought of as a ***conservative estimate***, which we plan to improve upon as development continues in Construction.

4.2.3 Source Characterization

Sources will be detected on individual visits as well as the coadds. Sources detected on coadds will primarily serve as inputs to the construction of the master object list as described in § 4.2, and may support other LSST science cases as seen fit by the users (for example, searches for objects whose shapes vary over time). DMS-REQ-0267

The following `Source` properties are planned to be measured:

- *Centroids*. Centroids are currently planned to be computed using an algorithm similar to that employed by SDSS, which uses an approximation to the PSF model as a weight function. These centroids will be used for PSF photometry, adaptive moments, and aperture magnitude measurements. DMS-REQ-0267
- *Point source photometry*. Fluxes are measured using the PSF model as a weight function, with aperture corrections applied to account for flux beyond the PSF model's extend and put this measurement on the same system as other flux measurements. DMS-REQ-0267
- *Adaptive moments*. Adaptive moments will be computed. The moments of the PSF realized at the position of the object will be provided as well. DMS-REQ-0267
- *Aperture surface brightness*. Aperture surface brightness will be computed in a variable number[72] of concentric, logarithmically spaced, PSF-homogenized, elliptical apertures. DMS-REQ-0267

Note that we do *not* plan to fit extended source Bulge+Disk models to individual `Sources`, nor measure per-visit Petrosian or Kron fluxes. These are object properties that are not expected to vary in time[73], and will be better characterized by coadd-based measurements (in the `Object` table). For example, although a simple extendedness characterization is present in the

[72] The number will depend on the size of the source.

[73] Objects that *do* change shape with time would, obviously, be of particular interest. Aperture fluxes provided in the `Source` table should suffice to detect these. Further per-visit shape characterization can be performed using the Science Platform.

Source table, star-galaxy separation (an estimate of the probability that a source is resolved, given the PSF) will be better characterized by a combination of `Object` measurements.

4.2.4 Forced Photometry

DMS-REQ-0268
DMS-REQ-0287

Forced Photometry is the measurement of flux in individual visits, given a fixed position, shape, and the deblending parameters of an object. It enables the study of time variability of an object's flux, irrespective of whether the flux in any given individual visit is above or below the single-visit detection threshold.

Forced photometry will be performed on all visits, for all `Objects`, using both direct images and difference images. The measured fluxes will be stored in the `ForcedSources` table. Due to space constraints, we only plan to measure the PSF flux.

4.2.5 Crowded Field Photometry

A fraction of LSST imaging will cover areas of high object (mostly stellar) density. These include the Galactic plane, the Large and Small Magellanic Clouds, and a number of globular clusters (among others).

LSST image processing and measurement software, although primarily designed to operate in non-crowded regions, is expected to perform well in areas of crowding. The current LSST applications development plan envisions making the deblender aware of Galactic longitude and latitude, and permitting it to use that information as a prior when deciding how to deblend objects. While not guaranteed to reach the accuracy or completeness of purpose-built crowded field photometry codes, we expect this approach will yield acceptable results even in areas of moderately high crowding.

Note that this discussion only pertains to processing of *direct images*. Crowding is not expected to significantly impact the quality of data products derived from *difference images* (i.e., Prompt products).

4.2.6 Shear Measurement

Cosmology is one of the major pillars of LSST science, and a substantial fraction of the survey's cosmological constraining power will come from weak gravitational lensing, which involves estimating the shear field applied to a population of galaxies by foreground matter.

For our pixel-level measurements of galaxy shapes to be usable for shear estimation, the responses of the algorithms we apply to shear must be carefully controlled and/or measured. While the state of the art for these algorithms continues to evolve, our plan for at least DR1 and DR2 is to use the `Metadetection` algorithm of Sheldon et al. [21, 20]. Additional algorithms may be run as well, as long as their processing time and catalog storage costs are negligible (which is actually the case for some promising but still unproven methods, e.g. Bernstein et al. [3]).

This is a variant of the `Metacalibration` [11, 19] approach, in which the response to shear is measured by applying the same algorithms to sheared versions of the original (coadd) images and their PSF models, and then using finite differences to compute the derivative of those measurements with respect to shear. `Metadetection` requires the algorithms run on these images to include detection and deblending, and like all variants of `Metacalibration`, it requires all algorithms run on the modified images to be simple enough for finite differences to work well. In practice this means measurements must be as close as possible to linear in the pixel values. This is incompatible with our usual desire for algorithms to produce catalogs that map as well as possible to astrophysical reality, especially since the way in which the coadd images are modified represents a slight degradation in signal-to-noise and image quality.

As a result, the measurements from `Metadetection` are best structured as a separate `ShearObject` table representing a completely distinct set of detections and measurements, with many fewer columns than the main `Object` table and $3-5\times$ as many rows (for the modified versions of the original coadds). A match table from the main `Object` table to `ShearObject` will not be provided, as actually using any such matches would void the carefully measured responses in `ShearObject`, rendering it useless.

`Metadetection` measurements will be performed on small coadd cells, over which the PSF can be considered spatially constant and the set of input visits is constant. This will also be true of some regular `Object` measurements, but the cell-based nature of those measurements will be less important; some key `Metadetection` measurements are actually reported per cell rather than per Object.

4.3 The Data Release Catalogs

This section presents the contents of key Data Release catalog tables. As was the case for Prompt Products (see § 3.3), here we present the *conceptual schemas* for the most important Data Release tables (the `Object`, `Source`, and `ForcedSource` tables). The tables themselves are defined in LDM-153[74].

These convey *what* data will be recorded in each table, rather than the details of *how*. For example, columns whose type is an array (eg., `radec`) may be expanded to one table column per element of the array (eg., `ra`, `decl`) once this schema is translated to SQL. Secondly, the tables to be presented are normalized (i.e., contain no redundant information). For example, since the band of observation can be found by joining a `Source` table to the table with exposure metadata, there's no column named `band` in the `Source` table. In the as-built database, the views presented to the users will be appropriately denormalized for ease of use. DMS-REQ-0332

4.3.1 The `Object` Table

DMS-REQ-0275

Table 7: Data Release Catalog `Object` Table

Name	Type	Unit	Description
objectId	uint64		Unique object identifier
parentObjectId	uint64		ID of the parent `Object` this object has been deblended from, if any.
radec	double[6][2]	degrees	Position of the object (centroid), computed independently in each band. The centroid will be computed using an algorithm similar to that employed by SDSS.
radecErr	double[6][2]	arcsec	Uncertainty of `radec`.
psRadecTai	double	time	Point source model: Time at which the object was at position `psRadec`.

Continued on next page

[74]The SQL definition itself can be found in the `cat` package.

Table 7: Data Release Catalog `Object` Table

Name	Type	Unit	Description
psRadec	double[2]	degrees	Point source model: (α, δ) position of the object at time `psRadecTai`.
psPm	float[2]	mas/yr	Point source model: Proper motion vector.
psParallax	float	mas	Point source model: Parallax.
psFlux	float[ugrizy]	nJy	Point source model fluxes[75].
psCov	float[66]	various	Point-source model covariance matrix[76].
psLnL	float		Natural log likelihood of the observed data given the point source model.
psChi2	float		χ^2 statistic of the model fit.
psNdata	int		The number of data points (pixels) used to fit the model.
bdEllip	float[2][ugrizyM]		B+D model: Ellipticity (e_1, e_2) of the object.
bdFluxB	float[2][ugrizy]	nJy	B+D model[77]: (α, δ) position of the object, in each band.: Integrated flux of the de Vaucouleurs component, in both the multi-band fit and the per-band fit.
bdFluxD	float[2][ugrizy]	nJy	B+D model: Integrated flux of the exponential component, in both the multi-band fit and the per-band fit.
bdReB	float[ugrizyM]	arcsec	B+D model: Effective radius of the de Vaucouleurs profile component.
bdReD	float[ugrizyM]	arcsec	B+D model: Effective radius of the exponential profile component.

Continued on next page

[75] Point source model assumes that fluxes are constant in each band. If the object is variable, `psFlux` will effectively be some estimate of the average flux.

[76] Not all elements of the covariance matrix need to be stored with same precision. While the variances will be stored as 32 bit floats (~ seven significant digits), the covariances may be stored to ~ three significant digits (~1%).

[77] Though we refer to this model as "Bulge plus Disk", we caution the reader that the decomposition, while physically motivated, should not be taken too literally.

Table 7: Data Release Catalog `Object` Table

Name	Type	Unit	Description
bdCovM	float[136]		B+D model covariance matrix for the multi-band fit[78].
bdCovM	float[21][ugrizy]		B+D model covariance matrices for the independent per-band fits.
bdLnL	float[ugrizyM]		Natural log likelihood of the observed data given the bulge+disk model.
bdChi2	float[ugrizyM]		χ^2 statistic of the model fit.
bdNdata	int[ugrizyM]		The number of data points (pixels) used to fit the model.
stdColor	float[5]	mag	Color of the object measured in "standard seeing". While the exact algorithm is yet to be determined, this color is guaranteed to be seeing-independent and suitable for photo-z determinations.
stdColorErr	float[5]	mag	Uncertainty of `stdColor`.
Ixx	float	arcsec^2	Adaptive second moment of the source intensity. See Bernstein & Jarvis [2] for detailed discussion of all adaptive-moment related quantities[79].
Iyy	float	arcsec^2	Adaptive second moment of the source intensity.
Ixy	float	arcsec^2	Adaptive second moment of the source intensity.
Icov	float[6]	arcsec^4	`Ixx`, `Iyy`, `Ixy` covariance matrix.
IxxPSF	float	arcsec^2	Adaptive second moment for the PSF.
IyyPSF	float	arcsec^2	Adaptive second moment for the PSF.
IxyPSF	float	arcsec^2	Adaptive second moment for the PSF.
m4	float[ugrizy]		Fourth order adaptive moment.

Continued on next page

[78] See `psCov` for notes on precision of variances/covariances.
[79] Or `http://ls.st/5f4` for a brief summary.

Table 7: Data Release Catalog `Object` Table

Name	Type	Unit	Description
petroRad	float[ugrizy]	arcsec	Petrosian radius, computed using elliptical apertures defined by the adaptive moments.
petroRadErr	float[ugrizy]	arcsec	Uncertainty of `petroRad`
petroBand	int8		The band of the canonical `petroRad`
petroFlux	float[ugrizy]	nJy	Petrosian flux within a defined multiple of the canonical `petroRad`
petroFluxErr	float[ugrizy]	nJy	Uncertainty in `petroFlux`
petroRad50	float[ugrizy]	arcsec	Radius containing 50% of Petrosian flux.
petroRad50Err	float[ugrizy]	arcsec	Uncertainty of `petroRad50`.
petroRad90	float[ugrizy]	arcsec	Radius containing 90% of Petrosian flux.
petroRad90Err	float[ugrizy]	arcsec	Uncertainty of `petroRad90`.
kronRad	float[ugrizy]	arcsec	Kron radius (computed using elliptical apertures defined by the adaptive moments)
kronRadErr	float[ugrizy]	arcsec	Uncertainty of `kronRad`
kronBand	int8		The band of the canonical `kronRad`
kronFlux	float[ugrizy]	nJy	Kron flux within a defined multiple of the canonical `kronRad`
kronFluxErr	float[ugrizy]	nJy	Uncertainty in `kronFlux`
kronRad50	float[ugrizy]	arcsec	Radius containing 50% of Kron flux.
kronRad50Err	float[ugrizy]	arcsec	Uncertainty of `kronRad50`.
kronRad90	float[ugrizy]	arcsec	Radius containing 90% of Kron flux.
kronRad90Err	float[ugrizy]	arcsec	Uncertainty of `kronRad90`.
apNann	int8		Number of elliptical annuli (see below).
apMeanSb	float[6][apNann]	nJy/arcsec^2	Mean surface brightness within an annulus[80].
apMeanSbSigma	float[6][apNann]	nJy/arcsec^2	Standard deviation of apMeanSb.

Continued on next page

[80] A database function will be provided to compute the area of each annulus, to enable the computation of aperture flux.

Table 7: Data Release Catalog `Object` Table

Name	Type	Unit	Description
extendedness	float		A measure of extendedness, computed using a combination of available moments, or from a likelihood ratio of point/B+D source models (exact algorithm TBD). $extendedness = 1$ implies a high degree of confidence that the source is extended. $extendedness = 0$ implies a high degree of confidence that the source is point-like.
lcPeriodic	float[6×32]		Periodic features extracted from difference image-based light-curves using generalized Lomb-Scargle periodogram [Table 4, 18].
lcNonPeriodic	float[6×20]		Non-periodic features extracted from difference image-based light-curves [Table 5, 18].
photoZ	float[2×95]		Photometric redshift likelihood samples – pairs of $(z, \log L)$ – computed using a to-be-determined published and widely accepted algorithm at the time of LSST Commissioning.
photoZ_pest	float[10]		Point estimates for photometric redshift[81], computed using photoZ.
flags	bit[128]	bit	Various useful flags.

4.3.2 `Source` Table

`Source` measurements are performed independently on individual visits. They're designed to enable relative astrometric and photometric calibration, variability studies of high signal-to-noise objects, and studies of high SNR objects that vary in position and/or shape (eg., comets). DMS-REQ-0267

[81] TBD but likely candidates are the mode, mean, standard deviation, skewness, kurtosis, and 1%, 5%, 25%, 50%, 75%, and 99% points from cumulative distribution.

Table 8: Data Release Catalog `Source` Table

Name	Type	Unit	Description
sourceId	uint64		Unique source identifier[82]
ccdVisitId	uint64		ID of CCD and visit where this source was measured
objectId	uint64		ID of the `Object` this source was associated with, if any.
ssObjectId	uint64		ID of the `SSObject` this source has been linked to, if any.
parentSourceId	uint64		ID of the parent `Source` this source has been deblended from, if any.
xy	float[2]	pixels	Position of the object (centroid), computed using an algorithm similar to that used by SDSS.
xyCov	float[3]		Covariance matrix for `xy`.
radec	double[2]	arcsec	Calibrated (α, δ) of the source, transformed from `xy`.
radecCov	float[3]	arcsec	Covariance matrix for `radec`.
apFlux	float	nJy	Calibrated aperture flux.
apFluxErr	float	nJy	Estimated uncertainty of `apFlux`.
sky	float	nJy/arcsec^2	Estimated background (sky) surface brightness at the position (centroid) of the source.
skyErr	float	nJy/arcsec^2	Estimated uncertainty of `sky`.
psFlux	float	nJy	Calibrated point source model flux.
psCov	float[6]	various	Point-source model covariance matrix, including dependence on the centroid[83].
psLnL	float		Natural log likelihood of the observed data given the point source model.
psChi2	float		χ^2 statistic of the model fit.

Continued on next page

[82]It would be optimal if the source ID is globally unique across all releases. Whether that's realized will depend on technological and space constraints.

[83]Not all elements of the covariance matrix will be stored with same precision. While the variances will be stored as 32 bit floats (~ seven significant digits), the covariances may be stored to ~ three significant digits (~1%).

Table 8: Data Release Catalog Source Table

Name	Type	Unit	Description
psNdata	int		The number of data points (pixels) used to fit the model.
Ixx	float	arcsec^2	Adaptive second moment of the source intensity. See Bernstein & Jarvis [2] for detailed discussion of all adaptive-moment related quantities[84].
Iyy	float	arcsec^2	Adaptive second moment of the source intensity.
Ixy	float	arcsec^2	Adaptive second moment of the source intensity.
Icov	float[6]	arcsec^4	Ixx, Iyy, Ixy covariance matrix.
IxxPSF	float	arcsec^2	Adaptive second moment for the PSF.
IyyPSF	float	arcsec^2	Adaptive second moment for the PSF.
IxyPSF	float	arcsec^2	Adaptive second moment for the PSF.
apNann	int8		Number of elliptical annuli (see below).
apMeanSb	float[apNann]	nJy	Mean surface brightness within an annulus.
apMeanSbSigma	float[apNann]	nJy	Standard deviation of apMeanSb.
extendedness	float		A measure of extendedness, computed using a combination of available moments (exact algorithm TBD). $extendedness = 1$ implies a high degree of confidence that the source is extended. $extendedness = 0$ implies a high degree of confidence that the source is point-like.
flags	bit[64]	bit	Various useful flags.

4.3.3 ForcedSource Table

DMS-REQ-0268

[84] Or http://ls.st/5f4 for a brief summary.

Table 9: Data Release Catalog `ForcedSource` Table

Name	Type	Unit	Description
objectId	uint64		Unique object identifier
ccdVisitId	uint64		ID of CCD and visit where this source was measured
psFlux	float	nJy	Point source model flux on direct image, if performed.
psFluxErr	float	nJy	Point source model flux error, stored to 1% precision.
psDiffFlux	float	nJy	Point source model flux on difference image, if performed.
psDiffFluxErr	float	nJy	Point source model flux error, stored to 1% precision.
flags	bit[8]	bit	Various useful flags.

4.3.4 `ShearObject` Table

Table 10: Data Release Catalog `ShearObject` Table

Name	Type	Unit	Description
shearObjectId	uint64		Unique object identifier.
cellId	uint64		Unique identifier for the coadd cell in which this object was detected and measured.
metaStep	int8		Enumeration value for which of the `Metacalibration` modifications was applied to images prior to all measurements.
radec	double[2]	degrees	Position of the object (centroid), measured jointly across all bands.
maskFractionObj	float		Weighted fraction of this object's pixels that were masked, over all visit images input to the coadd.

Continued on next page

Table 10: Data Release Catalog `ShearObject` Table

Name	Type	Unit	Description
maskFractionCell	float		Weighted fraction of the coadd cell's that were masked, over all visit images input to the coadd.[85].
nEpochCell	int32[ugrizy]		Number of input visits that contributed to this coadd cell.
g1	float		Ellipticity, e.g. from image moments, measured consistently over the subset of bands [86] that are considered to be sufficiently well-characterized (survey-wide) for lensing.
g2	float		Ellipticity; see g1.
gCov	float[3]		Covariance matrix for (`g1`, `g2`).
T	float	arcsec^2	Moments trace (Ixx + Ixy) or other size estimated consistently with `g1` and `g2`.
SNR	float		Signal-to-noise ratio measured consistently with `g1` and `g2`.
TErr	float	arcsec^2	Uncertainty on `T`.
g1PSFMeta	float		Ellipticity measured on the metacalibration-modified PSF models.[87]
g2PSFMeta	float		Ellipticity measured on the metacalibration-modified PSF models.
TPSFMeta	float	arcsec^2	Moments trace (Ixx + Ixy) or other size measured on the metacalibration-modified PSF models.
g1PSFOrig	float		Ellipticity measured on the original PSF models.
g2PSFOrig	float		Ellipticity measured on the original PSF models.

Continued on next page

[85] This and `nEpochCell` are the same for all rows with the same `cellId`, but are expected to be access per-object; they may or may not be normalized into a separate table in the physical schema.

[86] Probably riz

[87] This and other PSF quantities are not constant over a cell only because the PSF may depend on the SED of the object.

Table 10: Data Release Catalog ShearObject Table

Name	Type	Unit	Description
TPSFOrig	float	$arcsec^2$	Moments trace (Ixx + Ixy) or other size measured on the original PSF models.
stdFlux	float[ugrizy]	nJy	PSF-corrected aperture fluxes, measured using an approach similar to the "standard color" in the main Object table.[88]
stdFluxErr	float[ugrizy]	nJy	Uncertainty on stdFlux.
stdFluxT	float[ugrizy]	$arcsec^2$	Moments trace (Ixx + Ixy) computed with the same weights as stdFlux.
stdFluxTErr	float[ugrizy]	$arcsec^2$	Uncertainty on stdFluxT.
flags	bit[8]	bit	Various useful flags.

4.4 Data Release Image Products

4.4.1 Visit Images DMS-REQ-0065

Raw exposures, including individual snaps, and Processed Visit Images will be made available for download as FITS files. They will be downloadable both through a human-friendly Science User Interface, as well as using machine-friendly APIs.

Required calibration data, processing metadata, and all necessary image processing software DMS-REQ-0130 will be provided to enable the user to generate bitwise identical processed images from raw DMS-REQ-0298 images[89]. DMS-REQ-0308

4.4.2 Calibration Data

All calibration frames (darks, flats, biases, fringe, etc.) will be preserved and made available DMS-REQ-0130 for download as FITS files. DMS-REQ-0065 DMS-REQ-0298

[88]This is not necessarily expected to provide good estimates of total flux, but *some* estimate of total flux is useful, so this is still reported as fluxes, as opposed to the colors in Object's stdColor.
[89]Assuming identically performing software and hardware configuration.

All auxiliary telescope data, both raw (images with spectra) and processed (calibrated spectra, derived atmosphere models), will be preserved and made available for download.

4.4.3 Coadded Images

Much of Data Release processing relies on coadded images. These will include at least:

- A set of *deep coadds*. One deep coadd will be created for each of the $ugrizy$ bands. DMS-REQ-0279 These coadds will be optimized for a reasonable combination of depth (i.e., employ no DMS-REQ-0281 PSF matching) and resolution (i.e., visits with significantly degraded seeing may be omitted). Transient sources (including Solar System objects, explosive transients, etc), will be removed. Care will be taken to preserve the astrophysical backgrounds[90].

  The PSF model on these coadds will be propagated from the PSF models of their constituent visit-level images, and is thus expected to be discontinuous where the set of input images varies (e.g. due to CCD boundaries or rejected pixels). To mitigate this effect, we are considering both building per-object postage-stamp coadds or by discretizing the inclusion or rejection of an epoch into the coadd over small predefined regions (and hence ensuring PSF discontinuities appear only at the well-defined boundaries between these regions).

  The six per-band coadds will be kept indefinitely and made available to the users. DMS-REQ-0334 *Their primary purpose (after data release production is complete) is to enable the end-users to apply alternative object characterization algorithms, perform studies of diffuse structures, and for visualization*.

- A set of *template coadds*. These will be used for image differencing in that data release and in the following season's Prompt Processing (approaches for image differencing prior to DR1 are still under investigation). They will be retained and made available at least during the period in which they are used by Prompt Processing. See also § 3.4.3.

- RGB color images derived from coadds (utilizing any three of *ugrizy*), intended for visual inspection and EPO . These may be stored using lossy compression (e.g. JPEG) and are DMS-REQ-0103 not intended to be used as an input to algorithmic code

DM will also provide software and configuration that will allow users to create additional coadds using the Science Platform or other resources. These will include at least PSF-matched DMS-REQ-0311 DMS-REQ-0298

[90] For example, using "background matching" techniques [15]

coadds, multi-band detection coadds, yearly or other short-period coadds, and direct coadds built from images with only the visits with the best seeing. . The exact configurations for all coadds actually used in data release processing will be made available, even if the images themselves are not retained. DMS-REQ-0330 DMS-REQ-0281 DMS-REQ-0335 DMS-REQ-0330

The exact details of which coadds to build, and which ones to keep, can change during Construction without affecting the processing system design as the most expensive operations (raw image input and warping) are constant in the number of coadds produced. The data management system design *is* sensitive to the total *number and size* of coadds to be *kept* – these are the relevant constraining variables.

We reiterate that **not all coadds will be kept and served to the public**[91], though sufficient metadata will be provided to users to recreate them on their own. Some coadds may be entirely "virtual": for example, the PSF-matched coadds could be implemented as ad-hoc convolutions of postage stamps when the colors are measured. DMS-REQ-0311

We will retain smaller sections of all generated coadds, to support quality assessment and targeted science. Retained sections may be positioned to cover areas of the sky of special interest such as overlaps with other surveys, nearby galaxies, large clusters, etc. DMS-REQ-0338

4.5 Data Release Availability and Retention Policies

Over 10 years of operations, Rubin will produce eleven data releases: two for the first year of survey operations, and one every subsequent year. Each data release will include reprocessing of all data from the start of the survey, up to the cutoff date for that release.

The contents of data releases are expected to range from a few PB (DR1) to ~70 PB for DR11 (this includes the raw images, retained coadds, and catalogs). Given that scale, it is not feasible to keep all data releases loaded and accessible at all times.

Instead, ***only the contents of the most recent data release, and the penultimate data release will be kept on fast storage and with catalogs loaded into the database***. Statistics collected by prior surveys (eg., SDSS) show that users nearly always prefer accessing the most recent data release, but sometimes may use the penultimate one (this is especially true just after the DMS-REQ-0313

[91] The coadds are a major cost driver for storage. Rubin Data Management system is currently sized to keep and serve seven coadds, *ugrizyM*, over the full footprint of the survey.

publication of a new data release). Older releases are used rarely.

To assist with data quality monitoring and assessment ***small, overlapping, samples of data from older releases will be kept loaded in the database***. The sample size is expected to be on order of ~1-5% of the data release data, with larger samples kept early on in the survey. The goal is to allow one to test how the reported characterization of the same data varies from release to release. DMS-REQ-0339

Older releases will be archived to mass storage (tape). The users ***will not be able to perform database queries against archived releases***. They will be made available as bulk downloads in some common format (for example, FITS binary tables). Database software and data loading scripts will be provided for users who wish to set up a running copy of an older (or current) data release database on their systems. DMS-REQ-0077 DMS-REQ-0300

All raw data used to generate any public data product (raw exposures, calibration frames, telemetry, configuration metadata, etc.) will be kept and made available for download. DMS-REQ-0309 DMS-REQ-0346

5 User Generated Data Products and the Science Platform

The Science Platform is envisioned to enable science cases that would greatly benefit from co-location of user processing and/or data within the Archive Center. The high-level requirement for this is established in § 3.5 of the SRD. The science platform vision is described in in LSE-319 with more formal requirements listed in LDM-554 and the design described in LDM-542. The reader is referred to these documents for more information on the platform. As a non exhaustive summary authenticated and authorized users will have access to the platform granting them: OSS-REQ-0140 DMS-REQ-0125 DMS-REQ-0128 DMS-REQ-0033 DMS-REQ-0032 DMS-REQ-0308 DMS-REQ-0119

1. Access to the Data Products as described in this document
2. processing resources for image reprocessing, etc. DMS-REQ-0127
3. storage resources to hold processed image outputs, query outputs or uploaded catalogs, etc. DMS-REQ-0290 DMS-REQ-0123
4. access control to the storage resources including groups DMS-REQ-0124 DMS-REQ-0340
5. the LSST tools and processing framework including the pipeline code `https://pipelines.lsst.io/`
6. a python based interface as provided by JupyterHub (or equivalent)
7. a fairly standard job control environment (like Condor).

The platform is intended, when put to fullest use, to allow users to run their own productions-like runs on bulk image and/or catalog data, with mechanisms for creating and tracking large groups of jobs in a batch system.

The storage resources for User Generated data products come out of the SRD requirement for 10% of Rubin data management capabilities to be devoted to user processing. In general, they are likely to be controlled by some form of a "resource allocation committee". Users will probably have some small baseline automatic allocation, beyond which a SAC proposal is needed. The SAC may take into account scientific merit, length of time for which the storage is requested, and openness of the data to others, in setting its priorities. DMS-REQ-0119

The details of access rights and transitive rights for user generated products are dealt with in RDO-013.

5.1 Migration of User Generated data products to Data Release products

User Generated data products that are found to be generally useful can be migrated to Data Release products. This is a fairly complex process that ultimately involves the project taking responsibility for supporting and running Rubin-style code that implements the algorithm necessary to produce the data product (it's not just relabeling an existing User Generated data product as a Data Release data product). The project will provide necessary support for such migrations.

- For the migration to be considered, the creator of the L3DP will need to agree to make their data product public to the entire Rubin data-rights community, along with supporting documentation and code. The code at first need not be in the Rubin framework or even in a Rubin-supported language.
- If the original proponent wrote her/his code in the C++/Python Rubin stack environment (the "Science Platform"), it will be easier to migrate it to a Data Release product (though, obviously, using the same languages/frameworks does not guarantee that the code is of production quality).
- If the original code was written in another language or another data processing framework, the project may consider rewriting it to required Rubin standards.
- Taking on a new Data Release data product means that the project is committing to code maintenance, data quality review, space allocation, and continuing production of the new Data Release data product through DR11.

6 Data Products for Special Programs

LSR-REQ-0075

The LSST Survey Specifications (SRD, § 3.4) specify that 90% of LSST observing time will be spent executing the so-called "universal cadence". These observations will result in Prompt and Data Release data products described earlier in this document.

The remaining 10% of observing time will be devoted to special programs, obtaining improved coverage of interesting regions of observational parameter space. Examples include very deep ($r \sim 26$, per exposure) 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.

The details of these special programs or micro surveys are not yet defined[92]. Consequently, the specifics of their data products are left undefined at this time. Instead, we just specify the *constraints* on these data products, given the adopted Prompt/Data Release/User Generated architecture. It is understood that no special program will be selected that does not fit these constraints[93]. This allows us to size and construct the data management system, without knowing the exact definition of these programs this far in advance.

Processing for special programs will make use of the same software stack and computing capabilities as the processing for universal cadence. The programs are expected to use no more than ~10% of computational and storage capacity of the Rubin data processing cluster. When special products include time domain event alerts, their processing shall generally be subject to the same latency requirements as Prompt data products.

DMS-REQ-0320

DMS-REQ-0321
DMS-REQ-0344
OTT1

For simplicity of use and consistency, the data products for special programs will be stored in databases separate from the "main" (Prompt and Data Release) databases. The system will, however, allow for simple federation with Prompt/Data Release/User Generated data products (i.e., cross-queries and joins).

DMS-REQ-0322

As a concrete example, a data product complement for a "deep drilling" field designed for supernova discovery and characterization may consist of: i) alerts to events discovered by differencing the science images against a special deep drilling template, ii) a Prompt -like database

[92] The initial complement is expected to be defined and selected no later than Science Verification.
[93] Or will come with additional, external, funding, capabilities, and/or expertise.

iii) one or more "nightly co-adds" (co-adds built using the data from the entire night), produced and made available within ~24 hours, and iv) special deep templates, built using the best recently acquired seeing data, produced on a fortnightly cadence. L1PublicT

Note that the data rights and access rules apply just as they would for Prompt/Data Release/User Generated. For example, while generated event alerts (if any) will be accessible world-wide, the image and catalog products will be restricted to users with Rubin data rights.

A Appendix: Conceptual Pipeline Design

A high-level conceptual overview of the LSST image processing science pipelines is illustrated in Figure 2. The pipeline definitions presented here are driven by their inputs, outputs and processing steps; they do *not* describe exact boundaries in the actual implementation code, execution, or development responsibilities within the Project. Processing from pipelines marked with 1, 2, and 5-8 is executed every day when new data are taken to produce Prompt Data Products. Annual Data Release processing includes pipelines 1-6 and 8 (everything except Alert production). These main conceptual steps in LSST image processing include the following pipelines (enumeration in this list corresponds to enumeration in Figure 2 but note that these steps can be interleaved in the actual processing flow):

1. *Single Visit Processing* pipeline (Figure 3) produces calibrated and characterized single-visit images from raw snaps. The main processing steps include instrumental signature removal, background estimation, source detection, deblending and measurements, point spread function estimation, and astrometric and photometric calibration.

2. *Image Coaddition* pipeline (Figure 4) produces coadded images of different flavors (optimized for depth, seeing, etc.) from an ensemble of single-visit images.

3. *Coadded Image Analysis* pipeline (Figure 4) defines the `Object` list and performs (mostly) static-sky measurements on coadded images.

4. *Multi-epoch Object Characterization* pipeline (Figure 4) performs Forced Photometry on single-visit images (both direct and differences) at the positions of all `Objects`, and then runs a series of primarily catalog-space algorithms that combine information from all previous stages; this includes measuring proper motions and parallax from a combination of `Source` and `Object` measurements, summarizing variability from forced photometry, and computing photometric redshifts.

5. *Image Differencing* pipeline (Figure 5) produces difference images from a single-visit and coadded (template) images.

6. *Difference Image Analysis* pipeline (Figure 5) updates `DIAObject` and `SSObject` lists with new `DIASources` detected on processed difference image, fits a library of image models to Footprints of these `DIASources`, and for all `DIAObjects` overlapping the difference image it performs Forced Photometry and recomputes summary quantities. During

nightly Prompt processing, this pipeline also performs Forced Photometry for all new `DIAObjects` on difference images from the last 30 days. precoveryWindow

7. *Alert Generation and Distribution* pipeline (Figure 5) uses updated `DIAObjects` and `DIASources` to generate and distribute `Alerts` (which also include postage stamp images of the `DIASource` in difference image and coadded template image).

8. *Solar System Processing* pipeline (SSP, Figure 6)) combines all un-associated `DIASources` into plausible linkages, reports the discoveries to the Minor Planet Center, and computes the physical characteristics of observed objects. The three main pipeline stages include associating new `DIASources` with known `SSObjects` (attribution), discovering new `SSObjects` (linking), and the computation of physical characteristics and auxiliary quantities useful for Solar System science.

Further details about the pipeline design and implementation are available from the LSST document LDM-151.

B References

[1] **[LDM-153]**, Becla, J., 2013, *LSST Database Baseline Schema*, LDM-153, URL `https://ls.st/LDM-153`

[2] Bernstein, G.M., Jarvis, M., 2002, AJ, 123, 583 (`arXiv:astro-ph/0107431`), doi:10.1086/338085, ADS Link

[3] Bernstein, G.M., Armstrong, R., Krawiec, C., March, M.C., 2016, MNRAS, 459, 4467 (`arXiv:1508.05655`), doi:10.1093/mnras/stw879, ADS Link

[4] **[RDO-013]**, Blum, R., the Rubin Operations Team, 2020, *Vera C. Rubin Observatory Data Policy*, RDO-013, URL `https://ls.st/RDO-013`

[5] **[LSE-29]**, Claver, C.F., The LSST Systems Engineering Integrated Project Team, 2017, *LSST System Requirements (LSR)*, LSE-29, URL `https://ls.st/LSE-29`

[6] **[LSE-30]**, Claver, C.F., The LSST Systems Engineering Integrated Project Team, 2018, *Observatory System Specifications (OSS)*, LSE-30, URL `https://ls.st/LSE-30`

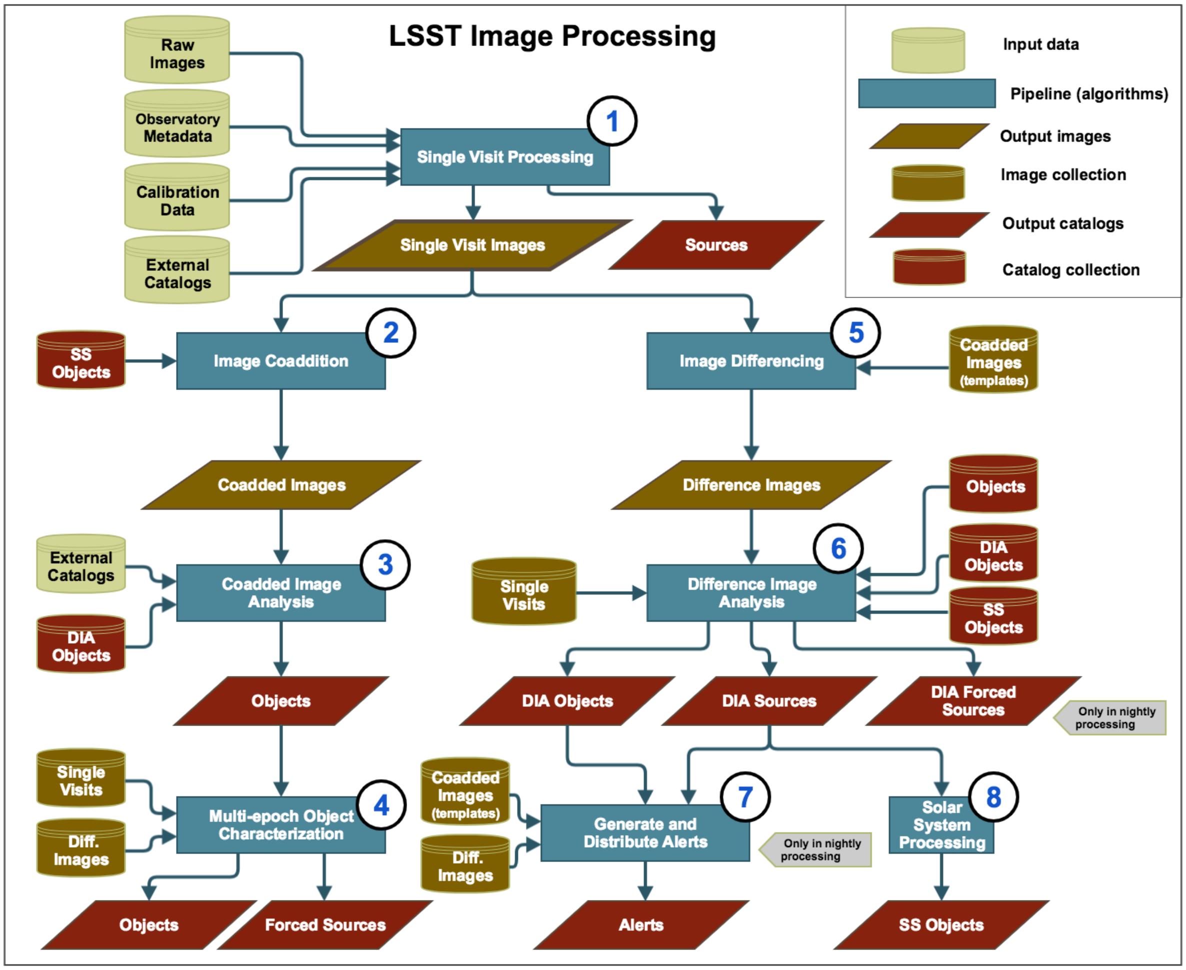


FIGURE 2: Illustration of the conceptual design of LSST science pipelines for imaging processing.

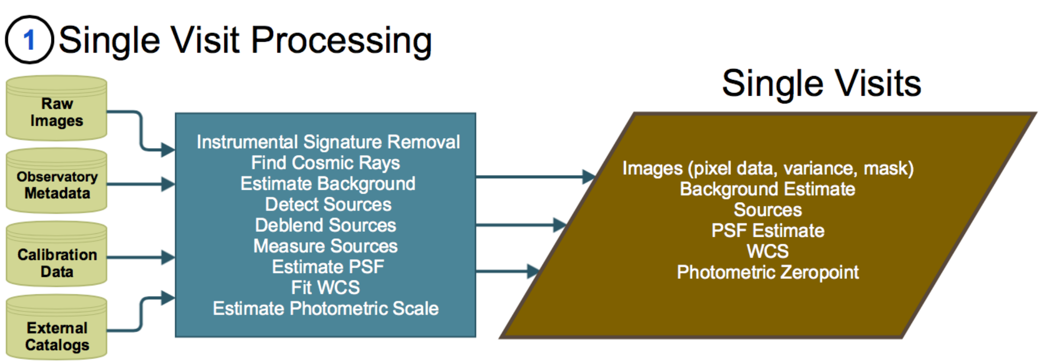


FIGURE 3: Illustration of the conceptual algorithm design for Single Visit Processing pipeline.

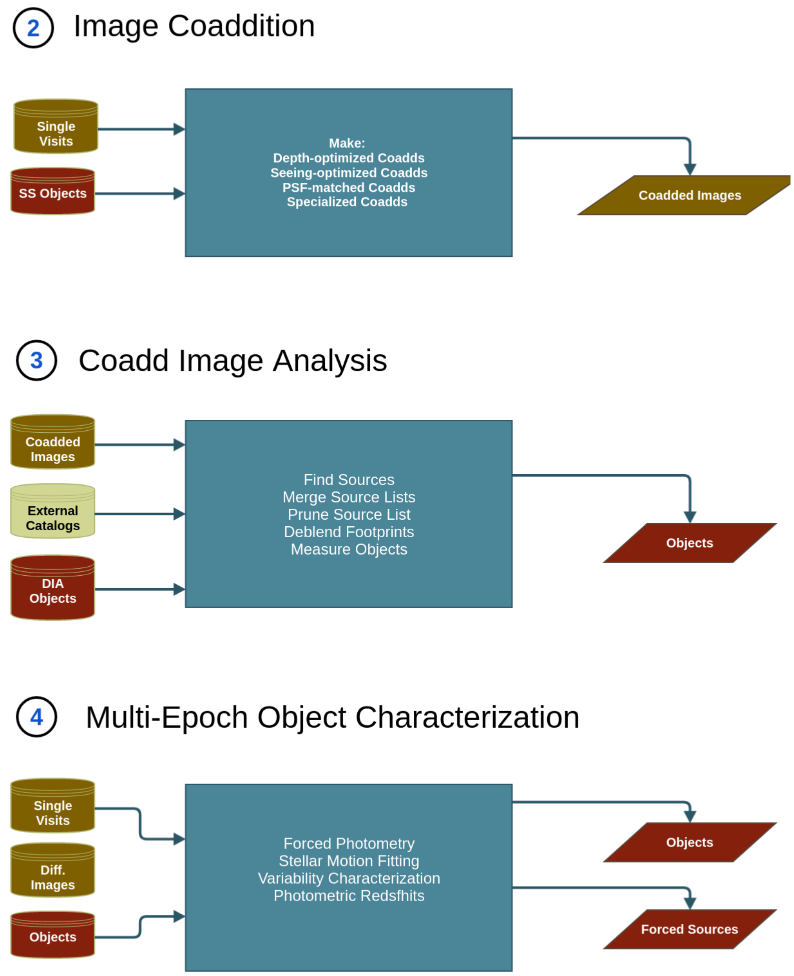


FIGURE 4: Illustration of the conceptual algorithm design for Image Coaddition, Coadded Image Analysis, and Multi-epoch Object Characterization pipelines.

[7] **[LSE-61]**, Dubois-Felsmann, G., Jenness, T., 2019, *Data Management System (DMS) Requirements*, LSE-61, URL `https://lse-61.lsst.io/`,
Vera C. Rubin Observatory

[8] **[LPM-231]**, Dubois-Felsmann, G., Ivezic, Z., Juric, M., 2018, *LSST Data Product Categories*, LPM-231, URL `https://lpm-231.lsst.io/`,
Vera C. Rubin Observatory

[9] **[LDM-554]**, Dubois-Felsmann, G., Ciardi, D., Mueller, F., Economou, F., 2019, *Data Management LSST Science Platform Requirements*, LDM-554, URL `https://ldm-554.lsst.io/`,
Vera C. Rubin Observatory Data Management Controlled Document

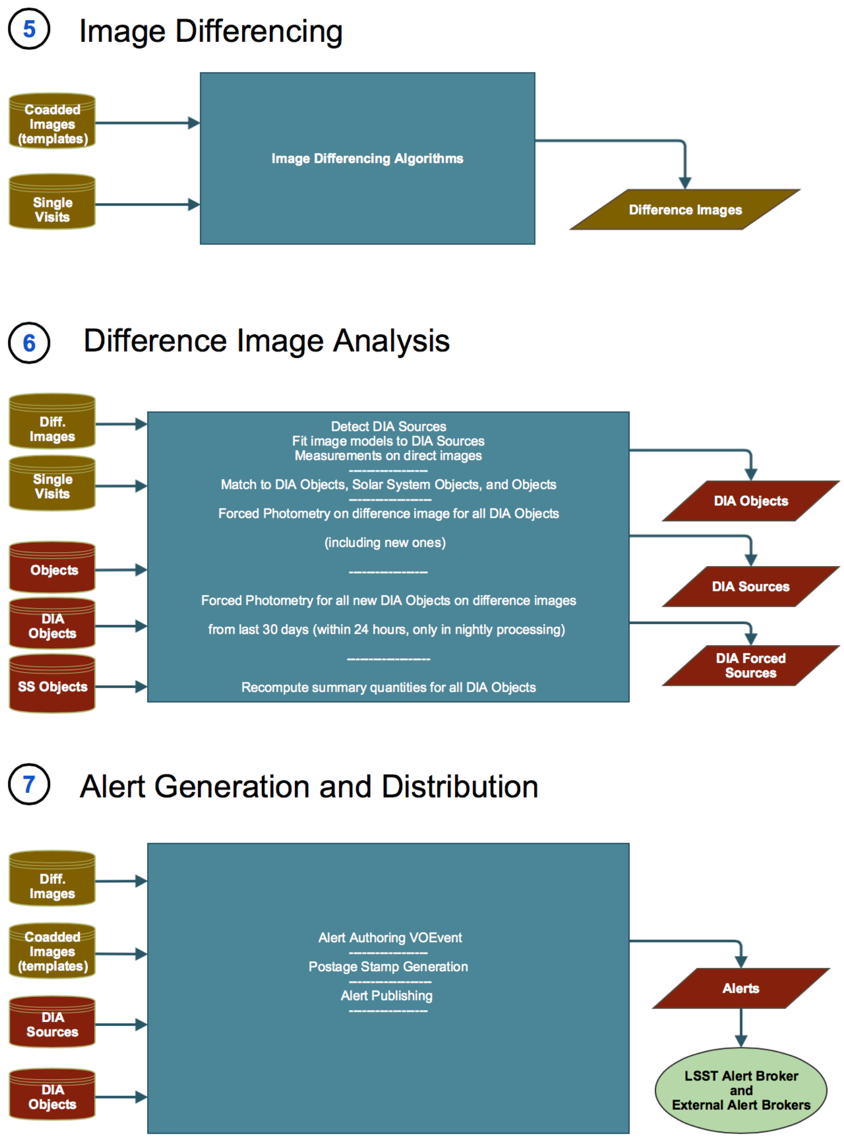


FIGURE 5: Illustration of the conceptual algorithm design for Image Differencing, Difference Image Analysis, and Alert Generation and Distribution pipelines.

[10] **[LDM-542]**, Dubois-Felsmann, G., Economou, F., Lim, K.T., et al., 2019, *Science Platform Design*, LDM-542, URL `https://ldm-542.lsst.io/`,
Vera C. Rubin Observatory Data Management Controlled Document

[11] Huff, E., Mandelbaum, R., 2017, ArXiv e-prints (`arXiv:1702.02600`), ADS Link

[12] **[DOCUMENT-27758]**, Ivezić, Ž., the LSST Project Science Team, 2018, *On the Choice of LSST Flux Units*, DOCUMENT-27758, URL `https://pstn-001.lsst.io/`,
Vera C. Rubin Observatory

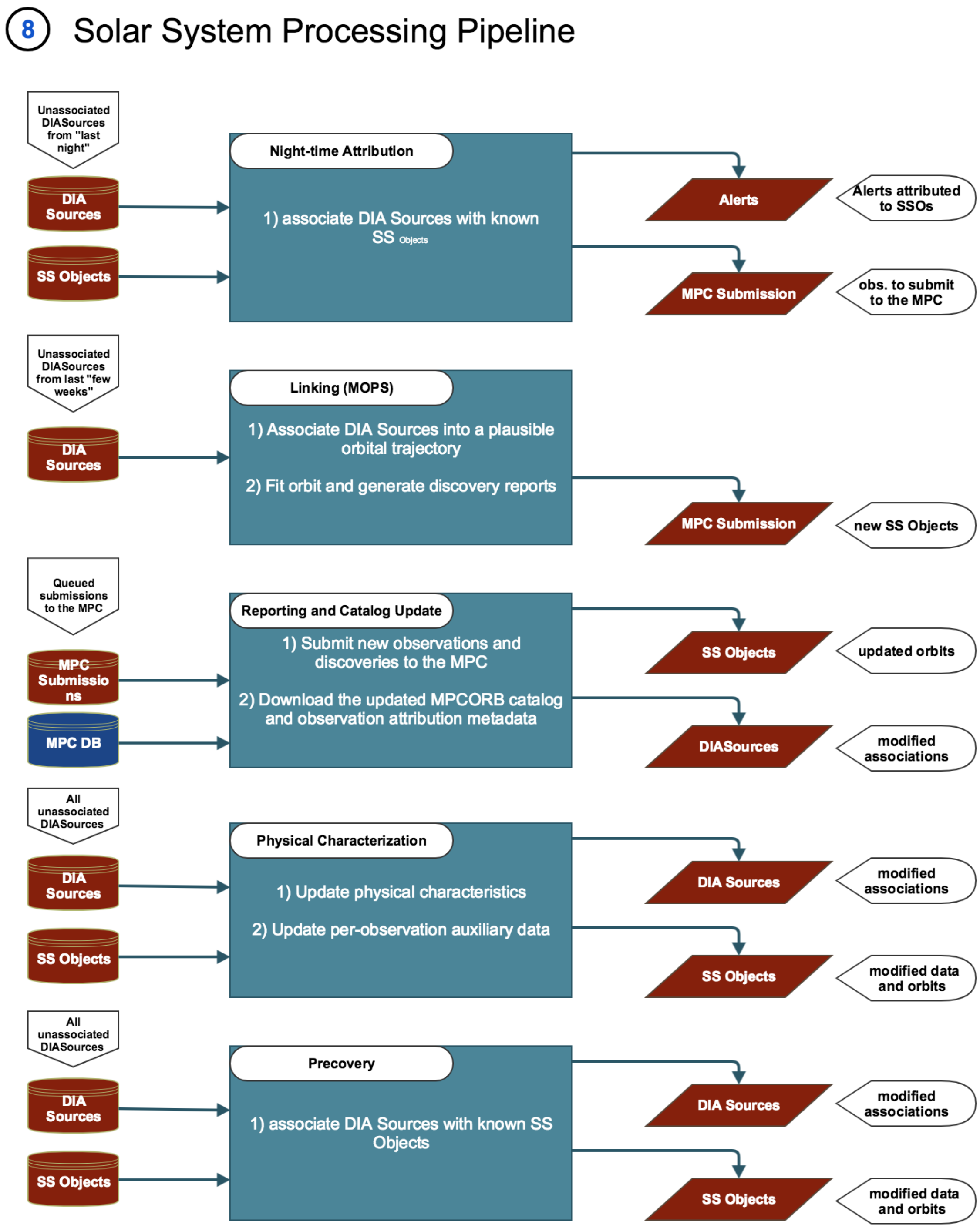


FIGURE 6: Illustration of the conceptual algorithm design for the Solar System Processing pipeline. The elements in blue denote ingestion of data from external sources.

[13] **[LPM-17]**, Ivezić, Ž., The LSST Science Collaboration, 2018, *LSST Science Requirements Document*, LPM-17, URL `https://ls.st/LPM-17`

[14] Ivezić, Ž., Kahn, S.M., Tyson, J.A., et al., 2019, ApJ, 873, 111 (arXiv:0805.2366), doi:10.3847/1538-4357/ab042c, ADS Link

[15] **[DMTN-035]**, Juric, M., Becker, A., Shaw, R., Krughoff, K.S., Kantor, J., 2013, *Winter 2013 LSST DM Data Challenge Release Notes*, DMTN-035, URL https://dmtn-035.lsst.io/, Vera C. Rubin Observatory Data Management Technical Note

[16] **[LDM-133]**, Jurić, M., Lim, K.T., Kantor, J., 2013, *Data Management UML Domain Model*, LDM-133, URL https://ls.st/LDM-133

[17] **[LSE-319]**, Jurić, M., Ciardi, D., Dubois-Felsmann, G., Guy, L., 2019, *LSST Science Platform Vision Document*, LSE-319, URL https://lse-319.lsst.io/, Vera C. Rubin Observatory

[18] Richards, J.W., Starr, D.L., Butler, N.R., et al., 2011, ApJ, 733, 10 (arXiv:1101.1959), doi:10.1088/0004-637X/733/1/10, ADS Link

[19] Sheldon, E.S., Huff, E.M., 2017, ApJ, 841, 24 (arXiv:1702.02601), doi:10.3847/1538-4357/aa704b, ADS Link

[20] Sheldon, E.S., Becker, M.R., MacCrann, N., Jarvis, M., 2020, ApJ, 902, 138 (arXiv:1911.02505), doi:10.3847/1538-4357/abb595, ADS Link

[21] Sheldon, E.S., Becker, M.R., Jarvis, M., Armstrong, R., The LSST Dark Energy Science Collaboration, 2023, arXiv e-prints, arXiv:2303.03947 (arXiv:2303.03947), doi:10.48550/arXiv.2303.03947, ADS Link

[22] **[LDM-151]**, Swinbank, J., Axelrod, T., Becker, A., et al., 2020, *Data Management Science Pipelines Design*, LDM-151, URL https://ldm-151.lsst.io/, Vera C. Rubin Observatory Data Management Controlled Document

C Acronyms

Acronym	Description
AC	Alternating Current
B	Byte (8 bit)
CCD	Charge-Coupled Device
DC	Data Center

DCR	Differential Chromatic Refraction
DIA	Difference Image Analysis
DM	Data Management
DMS	Data Management Subsystem
DMS-REQ	Data Management System Requirements prefix
DMSR	DM System Requirements; LSE-61
DMTN	DM Technical Note
DPDD	Data Product Definition Document
DR	Data Release
DR1	Data Release 1
DR11	Data Release 11
DR2	Data Release 2
EPO	Education and Public Outreach
FDR	Final Design Review
FITS	Flexible Image Transport System
IVOA	International Virtual-Observatory Alliance
LCR	LSST Change Request
LDM	LSST Data Management (Document Handle)
LPM	LSST Project Management (Document Handle)
LSE	LSST Systems Engineering (Document Handle)
LSR	LSST System Requirements; LSE-29
LSST	Legacy Survey of Space and Time (formerly Large Synoptic Survey Telescope)
ML	Machine Learning
MPC	Minor Planet Center
NCSA	National Center for Supercomputing Applications
NGC	New General Catalogue
OSS	Observatory System Specifications; LSE-30
PB	PetaByte
PCA	Principal Component Analysis
PPDB	Prompt Products DataBase
PSF	Point Spread Function
RDO	Rubin Directors Office
RFC	Request For Comment

RGB	Red Giant Branch
SAC	Science Advisory Committee
SDSS	Sloan Digital Sky Survey
SE	System Engineering
SED	Spectral Energy Distribution
SNR	Signal to Noise Ratio
SQL	Structured Query Language
SRD	LSST Science Requirements; LPM-17
SS	Subsystem Scientist
SSP	Solar System Processing
TAC	Time Allocation Committee
TBD	To Be Defined (Determined)
UDF	User Defined Function
UML	unified modeling language
WAN	Wide Area Network
WCS	World Coordinate System
photo-z	photometric redshift