



LARGE SYNOPTIC SURVEY TELESCOPE

Large Synoptic Survey Telescope (LSST)

Data Management and LSST Special Programs

M. L. Graham and M. Jurić

DMTN-nnn

Latest Revision: 2017-07-31

DRAFT

Abstract

WORKING DRAFT (contains rough notes and partial thoughts) but intended to be an overview of any and all potential requirements on data management (DM) from Special Programs such as the deep drilling fields (DDF) or mini-surveys. We also summarize the plans for how the Special Programs data will be integrated into the wide-fast-deep (WFD) Main Survey (Levels 1 and 2). It is the intent that this document evolve to become an internal catalog of change requests to the DM Level 1 and 2 pipelines and products, and a community resource for future Special Program white papers and preparations for Level 3 pipelines.



Change Record

Version	Date	Description	Owner name
1	2017-04-??	Internal working document.	Melissa Graham

Draft

Contents

1	Introduction	1
1.1	Boundaries of Allowed Special Programs Observations	2
1.2	Interacting with the Science Community	2
1.3	A Summary of Unresolved Concerns	3
2	The Current DM Plans for Special Programs Data Processing	5
2.1	Incorporating SP Data into the WFD Main Survey	6
2.1.1	Level 1 and the Alert Stream	6
2.1.2	MOPS	7
2.1.3	Level 2 Data Release Pipeline	8
2.2	SP Data Processing with Federated and User-Run Level 3	8
2.3	In-Depth Documentation Review	10
3	Questions and Potential Issues	14
3.1	Alert Stream	14
3.2	Databases and Schema	15
3.3	Templates	17
3.4	Snap Handling	18
3.5	Calibrations	18
3.6	Custom Co-Adds	20
3.7	Moving Objects	20

4	Science Cases and their DM Requirements	22
4.1	Nominal Observing Plans for DDF and MS	22
4.2	Science Descriptions (Not Comprehensive)	23
4.3	Special Programs Processing Case Studies	26
4.3.1	A Mini-Survey Searching for TNOs with Shift-and-Stack	26
4.3.2	Searching for Supernovae in Deep Drilling Fields	28
4.3.3	A Twilight Survey of Bright Stars	29
4.3.4	The Galactic Plane Survey for Variable Stars	30
4.3.5	A More Generic Level 3 Shift-and-Stack Case Study, by Mario	31
5	Conclusions	32

Data Management and LSST Special Programs

1 Introduction

The purpose of this study is to:

- (1) summarize the current plans of the DM pipelines with respect to Special Programs,
- (2) summarize the processing that is required to enable science for Special Programs,
- (3) assess whether the DM pipelines meet the expected needs of Special Programs, and
- (4) ensure any necessary changes get written into the requirements, designs, and plans.

This will be accomplished by:

- (1) starting with the DPDD, whitepapers, etc., and compiling relevant information,
- (2) discussing initial issues internally with DM (Mario, the SST), engineering (K-T, Tim), and then
- (3) ensuring the identified needs become specified requirements (i.e., spawn change requests).

*Current Status: This document is a rough draft, with the goal of being exhaustively complete. Many of the identified "**Concerns**" may just be misunderstandings or a lack of knowledge.*

The rest of this introduction contains:

- § 1.1 – defining DM-related boundaries for Special Programs.
- § 1.2 – existing channels for interaction with Science communities.
- § 1.3 – a summary of unresolved **Concerns** that arose in this study.

Later sections attempt to identify potential **Concerns** regarding DM's plans for processing the data from Special Programs with a variety of approaches:

- § 2 – a review of the current DM plans for SP data that are either already explicitly written down in DM documentation, or exist in Mario's thoughts.
- § 3 – a brainstorm of potential issues that might arise, organized by pipeline or product, inspired by internal and external documentation.
- § 4 – a review of science cases for SP and their potential strain on the DM system, including detailed reduction/analysis case studies in § 4.3.

1.1 Boundaries of Allowed Special Programs Observations

In the next couple of years, cadences and pointings for DDF and new mini-surveys will be proposed by the community. The document "Cadence Exploration and Improvement Plan" (by Zeljko, Lynne, Andy) proposes that the next call for DDF White Papers (proposed fields and cadences) be in December 2017, and the call for Mini-Survey White Papers be in October 2018. We will need to define boundaries for these observations from a DM prospective. The relevant Jira issue is <https://jira.lsstcorp.org/browse/LIT-325?filter=-3>.

Start of a list of DM-related considerations for Special Programs data.

Exposure times – (e.g., current shortest is 1 s, with stretch goal of 0.1, but...

Afternoon calibrations – (e.g., must fit into some amount of time left over after the WFD main survey daily calibration frames are obtained)

Time allotted – Nominally, 10% of the total observing time (90% goes to the WFD).

1.2 Interacting with the Science Community

To resolve some open issues we will need to solicit input from the science collaborations (SC; also, remember that the SC Coordinator for DDF is Neil Brandt). It may be better and easier to wait until the next call for community input, and request that a "DM Requirements" section be included in all white papers (as in § 4.3). Also, there is already a community forum on DDF, so far minimally used but ready to be revived when the next calls soliciting white papers are released. <http://community.lsst.org/t/deep-drilling-fields-and-data-management/1115>.

The questions opened up on that forum are:

- (1) What additional processing beyond that currently planned by the DM team (alerting relative to an annually created template) would greatly enhance the DDF science goals?
- (2) Are there DDF or Mini-Surveys specific aspects of the Level 3 system that would add significant value if provided? "Level 3" is the LSST-provided capability that enables non-DM, user-driven, processing of LSST data at the LSST Archive center (or remotely).
- (3) Are there aspects of the Science User Interface and Tools (SUIT) that need to be developed in order to enhance the usefulness of DDF data products.
- (4) To what degree should the DDF or Mini-Survey imaging could/should be incorporated into the main survey's deep stacks and associated data products (as opposed to being processed as separate data products)?

1.3 A Summary of Unresolved Concerns

- Concern I.** Investigate the minimum and maximum exposure times that can be successfully used in difference imaging and thereby contribute to the Alert Stream.
- Concern II.** Confirm whether a latency shorter than `L1PublicT=24` hours for Level 1 data products (LSR-REQ-0104) can be provided to federated and/or user-run Level 3 pipelines for Special Programs.
- Concern III.** Specify the latencies for delivering raw, reduced, and calibrated Special Programs images that are not processed by the Level 1 pipeline.
- Concern IV.** Confirm whether the computational budget for processing the Special Programs data requires more than 10% of the total resources, and on which batch system the bulk of this processing will take place.
- Concern V.** Confirm the expected thermal tolerances of the CCD and the associated minimum and maximum exposure times and exposure cadence that will keep the camera temperature within specifications.
- Concern VI.** Confirm whether processed single visits for Special Programs can stay on disk and accessible for Level 3 pipelines for a longer amount of time, by user request.
- Concern VII.** Query the science community about their science needs, NCSA about their processing capabilities, and DM about incorporating a functionality for the Alert Pipeline to go into a "skip intervening images" mode (or something similar), for < 30 second cadence imaging.
- Concern VIII.** Database schema for DIASource needs to have an element added that contains information about the template that was used to create the difference image, unless it can be shown that this can be handled by provenance (i.e., the code version and time of processing are sufficient to deduce which template was used).
- Concern IX.** Clarify whether `Object.prv_inputID` will identify whether an object is an externally provided coordinate for forced photometry.
- Concern X.** Contact the Science Collaborations for input on whether `float[6x32]` for the characterization parameters for variability is adequate to express the additional timescales and depths covered by all possible Special Programs observing strategies.

- Concern XI.** Contact the Science Collaborations for input on the characterization parameters for extended objects.
- Concern XII.** Confirm whether DM's ISR algorithms will be developed to reduce images with a range of exposure times, dither patterns, etc., and if not, what are the data restrictions for ISR.
- Concern XIII.** Will there be a Science Platform capability set up so that users are notified not just when any data for their Special Program is taken, but when certain sets are finished and ready for processing?
- Concern XIV.** Does the Level 3 pipeline have to include the steps to regenerate the processed exposures, or will DM deliver a service to automatically initiate regeneration by e.g., simply calling for an exposure or difference image?
- Concern XV.** Is it yet known how users will apply for batch processing priority and how they will be notified their jobs have finished, e.g., RSS feed? I could imagine that this is a requirement that would be defined in a future document on the Science Platform deliverables?
- Concern XVI.** Would the internal real/bogus routine be able to run on nightly CoAdd DDF difference images? If so, its outputs might not be reliable because of train/test set mismatch; will the internal LSST routines for real/bogus training be available and adaptable to Level 3 users?
- Concern XVII.** Confirm that resources (packet distribution) for a Level 3 alert stream would be supported and with what latency, whether and how such packets would be identified as non-AP, and that they could be ingested by brokers in real time.
- Concern XVIII.** How long can some Level 3 products remain on disk, such as template images for nightly CoAdd image subtraction?
- Concern XIX.** Confirm that the idea of federated Level 3 codes still exists (e.g., given how the conversation on incorporating a photo- z code ended up in a proposed descope).

2 The Current DM Plans for Special Programs Data Processing

In this section we summarize how DM intends to process the data obtained during Special Programs observations. This high-level "intent" is a combination of specifications that are actually written down in DM-related documents, and unwritten plans that exist in e.g., Mario's Brain. *MLG: Note that Sections 2.1 and 2.2 currently contain a mixture of what is officially written down and what is unofficially planned or expected, usually my interpretation of a conversation with Mario, while Section 2.3 is a direct review of the DM documentation. Once the SST and engineering have converged on a treatment plan for Special Programs data via this work, my plan is to return to the official DM documentation and identify where change requests need to be initiated.*

Overview: Although the data from some Special Programs might be incorporated into the Level 1, 2, and/or MOPS pipelines, every Special Program will probably also need one or more associated Level 3 processing pipelines. These Level 3 processing pipelines will be assembled by the science users, hopefully using existing LSST codes and schemas so that e.g., the Level 3 databases that can be easily joined to the Level 1 and 2 databases. It is clearly specified that DM will not write any specialized algorithms for Special Programs data, but that DM will ensure all pipeline code is extendable to Level 3 and that the relevant software infrastructure exists for science users to create and run their Level 3 pipelines. In addition, LSST will "federate" some Level 3 pipelines, putting them under change control and taking responsibility for their processing (e.g., shift-and-stack for faint moving objects is expected to become a federated Level 3 processing). All user access to processing codes and data products will be through the Science Platform.

Section contents:

Section 2.1 – an overview of whether and how DM will incorporate Special Programs data into the Level 1, 2, and MOPS pipelines and products.

Section 2.2 – an overview of how federated and user-run Level 3 pipelines will process Special Programs data, incorporate WFD data, and/or join to Level 1 and 2 products.

Section 2.3 – a detailed break-down of Special Programs (or Level 3) related content in the three DM documents (DPDD, DMSR, and DMAD), and non-DM LSST documents.

2.1 Incorporating SP Data into the WFD Main Survey

There is not much, if anything, specified regarding guidelines or expectations for incorporating Special Programs data into the Level 1 and Alert Stream, Level 2 DRP, and/or MOPS pipelines that will automatically run on all WFD main survey observations. The concern is that drastic changes in the exposure time or total depth of the observations might negatively impact the science goals that depend on the WFD main survey data products. Here we summarize these concerns.

2.1.1 Level 1 and the Alert Stream

It is beneficial to transient science to include as many LSST images into Level 1 and the Alert Stream as possible. Only products of the Level 1 60-second processing pipeline can contribute to the Alert Stream (e.g., no Level 3 difference imaging pipeline can send its detections to the Alert Stream, although a separate Level 3 alert stream should be possible if the packet/transport mechanism that LSST defines is adopted outside of LSST.). Only images that qualify for Level 1 processing will contribute to the Alert Stream. The template images that are used in the Level 1 difference imaging pipeline will be built from the Level 2 DRP, and so the first factor affecting a SP image's suitability for Level 1 is to be in a region of sky with an existing template. So long as there is a template, when the exposure time is equivalent to a WFD visit image, ~ 30 seconds, treating the image as Level 1 is obviously going to be fine. However, the suitability of shorter and longer exposures is unknown. Here we consider the suitability – or what we need to do in order to determine the suitability – of shorter and longer exposures for the Level 1 and Alert Stream. These considerations are made independently of the science motivations of the Special Programs that proposed the different exposure times.

(1) Shorter Exposures. The minimum supported exposure time is currently specified as 1 second (stretch goal 0.1 seconds), but this is a camera constraint. The minimum exposure time for an image to be successfully reduced with Instrument Signature Removal (ISR) will be a DM constraint, but is currently unspecified. We consider this issue separately in Section 3.5. Assuming that 1 second exposure can be reduced and calibrated, its detected point sources will span a dynamic range of $r \approx 12.9 - 21.0$ magnitudes. A template image built on 15 second exposures will saturate at $r \approx 15.8$, but this still leaves stars between 15.8–21.0 magnitudes to be used in the PSF-matching (and all other filters have a similarly large overlap). However, in order for an image to be successfully PSF-matched to the template, the PSF must be well formed (i.e., no speckle pattern), and have a spatial variance that the pipeline is capable of

modeling (i.e., be smoothly varying on some minimal scale). Another concern is that the Level 1 pipeline would not be able to deliver the required 60 second latency if the camera produces more than 1 image in 30 seconds, which would disqualify a contribution to the Alert Stream.

(2) Longer Exposures. There is no maximum exposure time specified for an LSST image. Given that the template image will be a stack of at least a year or two of data, processing a 5–10 times deeper single image through the difference imaging pipeline should be fine. However, a 2×150 second exposure would saturate at $r \approx 18.3$, and cosmic-ray rejection completeness might suffer (unknown), which could impact the quality of a difference image and the detected sources. With a longer exposure time the 60 second latency requirement could still be met, but any system qualities that vary on short (but > 30 second) timescales could inhibit photometric calibration (e.g., tracking).

In conversation with DM-AP team members (Reiss, Findeisen, Connolly, Bo) there has not yet been a study of the safe range of exposure times that will be allowed to contribute to the Level 1 and Alert Stream. One possibly useful study is Chang et al. (2012), "Atmospheric point spread function interpolation for weak lensing in short exposure imaging data". They show that a 15 second exposure contains PSF variability on short spatial scales across a 1 square degree image which, for extragalactic fields with few stars (i.e., but good for weak lensing), is hard to characterize. They present a new software package to do mitigate the effects. Alternatively, we may need to use software packages PhoSim (Peterson et al. 2015; <http://adsabs.harvard.edu/abs/2015ApJS..218...14P>) or ARROYO (<http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=848256>) to at least simply characterize the PSF stability as a function of exposure time.

Concern I. Investigate the minimum and maximum exposure times that can be successfully used in difference imaging and thereby contribute to the Alert Stream.

2.1.2 MOPS

The WFD main survey observations will be optimized to meet the mandated goals for moving object discoveries, so it is unclear whether the science goals of MOPS can be much improved by incorporating Special Programs data. Since MOPS takes DIASources as input, any SP images that can be run through the Alert Pipeline could be ingested by MOPS. We note that since MOPS does not run during the night, it does not have the same latency requirements, or the associated issues.

2.1.3 Level 2 Data Release Pipeline

Unlike with transient detection and the Level 1 Alert Stream, it is not clear whether it is beneficial to the science goals that depend on the Level 2 DRP to include as many LSST images as possible. We propose that the simplest option is probably that LSST DM deliver a single Level 2 DRP Objects database that is built from a set of images with nearly constant depth and cadence. When Special Programs data brings additional area up to the same level of depth and cadence as the rest of the WFD main survey, or improves the Level 2 DRP in some other way, it can be included in the Level 2 DRP. For example, photometric calibrations may require that some or all of the (shallower) Galactic Plane Special Program survey area be incorporated in order to suppress edge effects and low-order modes in the photometric solutions. In short, DM should probably not promise to incorporate any specific data into the Level 2 DRP images or catalogs.

2.2 SP Data Processing with Federated and User-Run Level 3

Through the Science Platform, member of the science community can assemble Level 3 processing packages using both LSST and custom codes. The DM team is required to write their codes to be extendable to Level 3, and to provide the infrastructure for users to run them on LSST servers, monitor their progress, and analyze their output (including a prioritization system to allocate processing resources). All Level 3 pipelines will have access to processed single visits from both the WFD main survey and Special Programs, as well as difference images, CoAdds, and catalogs from the Level 1 and 2 pipelines. It is expected that some Level 3 pipelines will become "federated", installed and operated (and change controlled) by the LSST Operations team, such as those developed for time domain science that need to run automatically and repeatedly. For both federated and user-run code, the LSST DM team will encourage and facilitate data product databases that are built with the same schema as – and can easily be joined with – the tables of Level 1 and 2.

Level 3 pipelines for Special Programs data for time-domain astronomy will need access to images on a timescale from minutes to weeks after acquisition, as opposed to static-sky Level 3 pipelines that may run on e.g., a yearly schedule like the Level 2 DRP. It is currently unclear what that real-time latency will be, and how processing resources will be allocated; here we break down some of the options.

If the images obtained with a Special Program are incorporated into Level 1 (see Section 2.1), then the Level 3 processing can begin as soon as the Level 1 imaging products are available: raw images, processed single visits (a.k.a. *calexps*), difference images, and their source catalogs. The current maximum time for this is **L1PublicT**, specified by LSE-29 ([2]) to be 24 hours (LSR-REQ-0104) – but this seems far too long, especially for federated Level 3 pipelines, given that they are finished within 60 seconds.

Concern II. Confirm whether a latency shorter than **L1PublicT**=24 hours for Level 1 data products (LSR-REQ-0104) can be provided to federated and/or user-run Level 3 pipelines for Special Programs.

If the images obtained with a Special Program cannot be incorporated into Level 1, then the latency with which they will be made available is unspecified. Ultimately, their delivery will depend why they are not incorporated into Level 1, which will probably fall into two categories: (1) there is no template at their location, or (2) their exposure time prohibits difference imaging. In the former case, it seems likely that single processed images can be provided on a timescale of minutes, but this needs to be specified. In the latter case, it will depend on the exposure time. As mentioned in Section, the minimum exposure time for an image to be successfully reduced with Instrument Signature Removal (ISR) will be a DM constraint, but is currently unspecified, and discussed separately in Section 3.5. It may be that single-visit processing is simply not supported by DM below some exposure time threshold, and so the latency on raw image delivery should be specified. Although it is likely that slightly shorter/-longer exposure images can be reduced and calibrated with the same pipeline as 15 second exposures, even if they are unsuitable for difference imaging, this has not yet been confirmed, and needs to be investigated and specified.

Concern III. Specify the latencies for delivering raw, reduced, and calibrated Special Programs images that are not processed by the Level 1 pipeline.

In the DPDD [14], it specifies that Special Programs data processing will be limited to not more than 10% of the DM computational and storage facilities. During the telecon on Wed Mar 29, KT Lam reviewed the computational budget, but it seemed mainly for WFD with only passing mention of DDF or Special Programs. This 10% of the computational budget seems like an underestimate: although SP data will not be more than 10% of the total data volume, it will be multiply processed: once as WFD Level 1, again perhaps as Level 2, and then possibly in multiple pipelines as Level 3, some of which are extremely computationally intensive (e.g., shift-and-stack). It seems that we could easily find that Special Programs data are processed $\sim 4\times$ as much as WFD. Furthermore, it is unclear whether the location for all this processing

has been identified; Mario asks, for federated real-time Level 3 processing, *"whether this the same batch system we make available to the users for running Level 3 codes or the one that's used to process calibrations. Have these systems been sized?"*.

Concern IV. Confirm whether the computational budget for processing the Special Programs data requires more than 10% of the total resources, and on which batch system the bulk of this processing will take place.

2.3 In-Depth Documentation Review

In this section we attempt to track the flow-down requirements through the official LSST and DM documentation. We find some hardware-related constraints in the OSS that will set the maximum possible divergence of Special Programs raw images from WFD main survey images (i.e., exposure times, rate of filter changes), and these will be relevant to DM processing. We also find that some broad, high-level requirements for Level 3 and Special Programs are written down in the DMSR. The next document down-flow, the DMAD, is focused on the Level 1 and 2 algorithms, and although its Section 7 mentions the DAX and SUIT, it says that they are to be detailed in a different document. The DPDD also does not have much content that directly pertains to Special Programs and Level 3 processing.

Science Requirements Document, SRD, LPM-17, [10]

The SRD makes no direct reference to Special Programs, deep drilling fields, or mini-surveys. It's scope is the science of the WFD only.

LSST System Requirements, SR, LSE-29, [2]

The SR is derived from the SRD. It contains the high-level specifications about fraction of total observing time delegated for Special Programs, minimum exposure times, maximum filter change time, etc., that flow-down to and are repeated in the DMSR and OSS (discussed below).

Observatory System Specifications, OSS, LSE-30, [3]

The OSS contains high-level requirements for the Level 3 data products, and their storage and access, that are similar to those in the DMSR (as they both stem from the SRD and SR). There are two points worth mentioning:

- 3.6.1.4. Minimum exposure time (ID: OSS-REQ-0291): 1 second (stretch 0.1 seconds).
→ There is a note that if the exposure time is shortened from 15 seconds, the spacing be-

tween exposures may need to be lengthened in order to maintain camera thermal stability, and that the thermal stability might also be affected with longer exposure times. However, there doesn't seem to be a specified limit on 'thermal stability'. Knowing the full range of exposure times and image cadences that DM must consider is important for estimating the predicted latencies providing raw, reduced and calibrated images (Section 2.2) and testing the incorporation of non-standard images into the calibration and difference imaging pipelines (Section 2.2). The science community will also need this information when they propose their DDF cadences and write their Special Programs white papers.

Concern V. Confirm the expected thermal tolerances of the CCD and the associated minimum and maximum exposure times and exposure cadence that will keep the camera temperature within specifications.

- 3.6.2.2. Maximum time for operational filter change (ID: OSS-REQ-0293): 120 seconds
 - 3.6.2.4. Minimum filter change count (ID: OSS-REQ-0295): day, 8, night, 4
- The specified maximum filter change time could be a debilitating overhead for, e.g., a deep drilling cadence of consecutive 30 second images wants to rotate through filters. Furthermore, we might expect that in reality the filter change time will be shorter than 120 seconds, which would enable a DDF to sequence filters. Furthermore, while the minimum filter change count is not an issue, I can't find the nightly (or total lifetime) maximum to limit wear and tear on the mechanism (as I've heard about). However, there is already a change request in progress regarding this issue: <https://project.lsst.org/groups/ccb/node/938>.

Data Management Subsystems Requirements, DMSR, LSE-61, [6]

The DMSR contains the high-level requirements for all processing and products. There are two main areas that pertain to Special Programs.

- Section 1.2.2 "Processed Visit Images" (ID: DMS-REQ-0069)
- Specifies that *"Processed science exposures are not archived, and are retained for only a limited time to facilitate down-stream processing. They will be re-generated for users on-demand using the latest processing software and calibrations"*, but also says that *"This aspect of the processing for Special Programs data is specific to each program"*. It is unclear whether processed SP data will remain on disk and accessible for as long as it is required by the science users, or if it will only live a short while.

Concern VI. Confirm whether processed single visits for Special Programs can stay on disk and accessible for Level 3 pipelines for a longer amount of time, by user request.

- Section 2.9, "Level 3 Production"

Specifies services of the DMS for Level 3 processing which are relevant to Special Programs:

2.9.1. Level 3 Data Import (ID: DMS-REQ-0290): to provide the ability to ingest common file formats (e.g., FITS table of external catalog).

2.9.2. Resource Allocation (ID: DMS-REQ-0119): to provide a mechanism to prioritize Level 3 activities and appropriately allocate processing resources.

2.9.3. Data Product Self-Consistency (ID: DMS-REQ-0120): to provide a means for ensuring Level 3 tasks can be carried out on self-consistent inputs.

2.9.4. Provenance (ID: DMS-REQ-0121): to provide a means for recording information about processing performed at DACs (DMS-provided input and user inputs) to help users work in a reproducible way.

2.9.5/6. Software Frameworks (ID: DMS-REQ-0125,DMS-REQ-0128): to provide a means for applying user-provided processing to catalog and image data, and for assessing the completeness of that application.

3.1. Software for Community Re-Use (ID: DMS-REQ-0308): algorithms shall be designed to run on both high-performance platforms and users' desktops.

→ **Questions:** Regarding 2.9.3, I'm not sure what "*self-consistent inputs*" means and so cannot assess any potential issues. Regarding 3.1, I thought that the Science Platform was being developed specifically so that no one needs to process LSST data on their home machines?

Data Management Applications Design, DMAD, LDM-151, [13]

LDM-151 outlines the code and algorithms that will be used in the processing of images in the AP (Level 1), DRP (Level 2), and MOPS. As such, it does not say much about Special Programs. It does say that "*LSST DM is required to facilitate the creation of Level 3 data products by providing suitable APIs, software components, and computing infrastructure, but will not by itself create any Level 3 data products*". The latter means that DM will not create any processing algorithms that are specific to the needs of any Special Programs data. Section 7 briefly describes the Data Access and the Science User Interface and Tool (SUIT), also known as the Science Platform, with a mention that they will be described in full in later documents (probably this one: https://github.com/lsst-dmsst/lsst_science_platform/blob/master/science-platform.pdf). The Data Access layer (DAX) will be for sharing, accessing, searching (etc.) products, and the SUIT/Platform will provide an interactive environment for users, and also serve as portal for Level 3 pipeline development (including installed version of the LSST Software Stack), execution, retrieval, and analysis.

Data Products Definitions Document, DPDD, LSE-163, [14]

The DPDD specifies that the data products for Special Programs will "*be created using the*

same software and hardware as Levels 1 and 2". Like LDM-151, Section 7 is devoted to the data products of Special Programs, explaining how the current plan is to have both the Level 1 and 2 processing pipelines run on SP data in the same way that they run on the WFD, and produce all of the same data products and alerts, when possible. It also states that the Level 3 products for SP data can (should) be kept in separate, but joinable, databases that have identical schema as the WFD databases.

Level 2 Photometric Calibration for the LSST Survey, LSE-180, [12]

LSE-180 is built on OpSim runs that do include some nominal deep drilling fields, but the photometric calibration investigated in this work does not deal with potential issues induced by non-standard visit patterns or exposure times of Special Programs – it's scope the WFD Main Survey.

3 Questions and Potential Issues

In this section we list questions or potential problems identified from reading LSST documents [14, 13], Gregory's AHM 2016 slides¹, or using our imaginations. Another useful resource is the LSST Database Schema Browser: <http://lsst-web.ncsa.illinois.edu/schema/index.php?sVer=baseline>.

For each of these questions or potential issues, a response will be:

blue if the answer is known (even if the answer is a negative),

red if the answer is still to be determined.

If it seems particularly urgent or involved to answer, it will spawn a **Concern**.

3.1 Alert Stream

- Special Programs that use short exposures (e.g., 5 seconds) pose a problem for the Level 1 Alert Pipeline because the higher data acquisition rate (e.g., $\sim 6\times$ higher) would cause a processing back-up: the allocated processing resource is two parallel chains that take 60 seconds each, and so real-time processing only works for 30 second integrations. If there is a demonstrable science need for 60 second latency on each and every one of a consecutive series of short exposures, options for short-term increases in parallel processing power at NCSA will need to be considered (note that this may also be needed for crowded fields with $>10k$ DIASources). However, the main science goal of a short exposure Special Program may simply be to include brighter stars in the photometric catalogs, and not necessarily to monitor variability on short timescales. In this case, there is perhaps no science need to include each and every one of the exposures into the Alert Pipeline, and it might instead be useful enough to just include e.g., every 6th image (which would need only the two parallel processing streams). There is currently no written functionality for this option: i.e., there is no way to send only every Nth image to the pipeline or, more flexibly, to skip over intervening images and only send an image once the processing pipeline is done with the last.

Concern VII. Query the science community about their science needs, NCSA about their processing capabilities, and DM about incorporating a functionality for the Alert Pipeline to go into a "skip intervening images" mode (or something similar), for < 30 second cadence imag-

¹"DM Considerations for Deep Drilling", LSST AHM Aug 2016 presentation by Gregory Dubois-Felsmann <https://zenodo.org/record/61402#.WNVk6hIrIUF>

ing.

- LDM-151, Section 3.3.4 'Alert queuing and persistence' mentions that the *"event message stream and the AlertDB will be synchronized at least once every 24 hours"* and that *"Prior to the start of the subsequent night's observations, the message queue will be flushed and synchronized with the AlertDB. It is possible to persist the message queue on longer timescale but it is a requirement that synchronization be performed within 24 hours of the observations."* Since this is for the main survey in which new alerts are issued for a given field 2–3 times per night, check that this is fine in the case of doing e.g., 200 alert bursts on the same field in a single night from a deep drilling field?

→ **Unanswered.** However, upon review, this requirement seems unrelated to visit locations, and thus not an issue for Special Programs.

- Will near-real-time Level 3 difference-image processing be able to contribute to the same Alert Stream as Level 1? (E.g., transient source detections in nightly stacks of a deep drilling field.)

→ **Answered.** As described in LDM-151, the Alert Stream takes as input only DIAObjects and specifies that *"all alerts should be transmitted within 60s of the closure of the shutter of the final snap within a visit."* Based on this, any Level 3 difference imaging sources would not contribute to the Alert Stream; only imaging that qualifies as Level 1 will contribute to the Alert Stream (but a separate Level 3 alert stream would be possible).

- Alerts on a DIASource will link to the associated DIAObject, but in cases where a Special Program obtains a sequence of images without switching fields, will the DIAObject catalog have had enough time to be updated to include the DIASource from the previous image?

→ **Answered.** Immediate updates of the DIAObject catalog is an expected capability, although it is currently uncertain how it will be accomplished, technically.

3.2 Databases and Schema

- The database schema for DIASource does not appear to have an element that identifies which template image that was used, but this will be needed for both Levels 1 and 3 differencing pipelines and products.

→ **Unaddressed.**

Concern VIII. Database schema for DIASource needs to have an element added that contains information about the template that was used to create the difference image, unless it can

be shown that this can be handled by provenance (i.e., the code version and time of processing are sufficient to deduce which template was used).

- LDM-151 mentions that forced photometry will be done for externally defined targets (Section 3.2.5), and this may be a particular interest to Special Programs, but it is unclear how such targets will be identified or flagged as such in the databases, and unclear whether we need to add an element to the database schema for this. Currently, the Object database contains an element `prv_inputId` which is an integer, and is described as the *“Pointer to prv_InputType. Indicates which input was used to produce a given object.”* Is that all we need?

→ **Unanswered.**

Concern IX. Clarify whether `Object.prv_inputID` will identify whether an object is an externally provided coordinate for forced photometry.

- These are the database elements reserved for variability characterization:

`Object` and `DIAObject.lcPeriodic` = `float[6 x 32]` = Periodic features extracted from light-curves using generalized Lomb-Scargle periodogram

`Object` and `DIAObject.lcNonPeriodic` = `float[6 x 32]` = Non-periodic features extracted from light-curves using generalized Lomb-Scargle periodogram

Will their sizes be adequate to characterize the different kinds of variability that could be measured with Special Programs cadences that are quite different from the main survey (e.g., Miras, flares, and LBV-turned-SNe)? Keep in mind that we just need sufficient information to enable relatively useful filters, from which the downstream broker/user can do additional filtering, and that these elements could be differently populated in the Level 3 databases in order to meet the science goals of a given Special Program. The concern here is just to ensure that they are adequate for the Level 1 database, which might incorporate observations from Special Programs on a variety of timescales.

→ **Unanswered.**

Concern X. Contact the Science Collaborations for input on whether `float[6x32]` for the characterization parameters for variability is adequate to express the additional timescales and depths covered by all possible Special Programs observing strategies.

- Low surface-brightness features and/or faint extended objects are required for some galaxy science, and will show up more in e.g., a DDF than a WFD CoAdd. Will the related shape parameters such as extendedness in the Object database be sufficient to describe the extra level of detail in low-surface brightness features that might come from deeper Special Programs CoAdds? As above, these elements could be differently populated in the Level 3 databases in order to meet the science goals of a given Special Program. Furthermore, it's unlikely that

extremely deep patches will be included in the Level 2 products, but this still should be assessed.

→ **Unassessed.**

Concern XI. Contact the Science Collaborations for input on the characterization parameters for extended objects.

- Images obtained during Special Programs might be used with one or more different types of templates – the same images could be used as multiple surveys. Is it feasible to have separate databases for each?

→ **Answered.** Yes, Level 3 pipelines can create as many separate databases as necessary. There should at least be a separate database for each kind of template image. These databases can have the same base schema, and elements of the schema can be “turned off” if the survey will not use them, to save processing time and disk space.

- Data from Special Programs might be incorporated into the Level 1 and/or 2 databases; if so, how will it be flagged in these databases?

→ **Answered.** In the schema for `DIASource` and `Source` there is an element for `ccdVisitId`, and the database `CcdVisit` has entry for `visitId`, and the database `Visit` has an entry for `programId`, which is currently an integer containing the “*Observing program id (e.g., universal cadence, or one of the deep drilling programs, etc.).*”.

- Will a Level 3 database for a Special Program be able to generate and incorporate forced photometry from WFD images?

→ **Answered.** Yes. All Level 3 pipelines have access to all LSST data.

- Will the Level 1 requirement to generate forced photometry for all new `DIAObjects` within 24 hours be able to run on Level 3 image products of Special Programs data, such as deep nightly stacks of DDF images?

→ **Answered.** No. This forced photometry will run only on other images that are incorporated into Level 1, and this will not include the data products of Level 3 pipelines. Performing forced photometry for new `DIAObjects` on a Level 3 image product would itself be a Level 3 pipeline.

3.3 Templates

- Typically we think of a template as a single deep coadd of image obtained e.g., $\Delta t \geq 1$ year ago. However, the use of a template with constant Δt to the new images (e.g., $\Delta t = 7$ days to

mitigate proper motion induced issues in a difference image for a Galactic plane mini-survey), or the use of a template that is simply a specific other image, might be scientifically necessary. Will this be supported by DM?

→ **Answered.** Yes. Difference imaging with a custom template is Level 3, and the requirement for the generation of custom templates already exists as **DMS-REQ-0032**: The DMS shall provide software to perform image differencing, generating Difference Exposures from the comparison of single exposures and/or coadded images [6].

3.4 Snap Handling

- LDM-151 says (Section 3.2.4.1) that the image differencing pipeline will measure the PSF flux on snap difference images for all DIASources. Could Special Programs data that is only a single exposure, not a visit that is the combination of two snaps, be incorporated into the Level 1 AP?

→ **Unanswered.** But since there is talk anyway of using 1x30 instead of 2x15 snaps, I am not spawning a **Concern** for this question.

3.5 Calibrations

- **Image Reduction** The Special Programs data may have a different exposure time or dither pattern than the WFD survey. The concern here is that if the instrument signature removal (ISR) calibration pipeline, and/or the photometric/astrometric calibration pipeline, have been designed to work for WFD data that problems might be encountered for Special Programs data (e.g., dark frames; brighter-fatter correction; artifacts with an exposure time dependence; charge build-up from no-dither with bright stars). Since there already exists a requirement that DM deliver processed visit images (ID: DMS-REQ-0069) to users (and to internal pipelines), it seems that at least the ISR must be developed for a variety of exposure times, but is this truly the case? Or are there exposure time boundaries beyond which DM will not guarantee delivery of processed images?

→ **Unanswered.**

Concern XII. Confirm whether DM's ISR algorithms will be developed to reduce images with a range of exposure times, dither patterns, etc., and if not, what are the data restrictions for ISR.

- Regarding the above point, LSE-180 offers some relevant information regarding the reduction and calibration of non-WFD images:

→ LSE-180 mentions that it is assumed that all factors affecting the system transmission are stable on 15 second timescales (page 10), but not what the upper limit of that might be.

→ LSE-180 comments on the dither pattern for the WFD survey in that "dither patterns where the overlap is one quarter of the field of view or more produce results meeting the SRD requirements", but this is specific to photometric calibration of the WFD. The LSE-180 also mentions that an inappropriate dither pattern can make it hard to correct for the variation of system bandpass as a function of the focal plane position – but so long as this is solved in the WFD the corrections can be applied to the much smaller amount of data from the Special Programs (unless it is a function of exposure time and the Special Program uses a different exposure time).

• **Image Detection Efficiency** Characterizing the detection efficiencies will be just as important to Special Programs data as the Wide-Fast-Deep survey. Options such as planting fake sources may be both more manageable and more important to some Special Programs science goals. What are the requirements on DM to provide detection efficiencies, and what are the current plans for providing this? (The answer should distinguish between planting fakes in the single images for transient detection efficiencies with difference imaging, *vs.* planting fakes in the CoAdds for point-source limiting magnitudes).

→ The DPDD (LSE-163) does not have any specific data product related to detection efficiencies, but Section 3.2 "Image Characterization Data" does specify that *"Each processed image .. will record information on the pixel variance ... as well as the per-pixel masks ... These will allow the users to determine the validity and usefulness of each pixel in estimating the flux density recorded in that area of the sky"*.

→ The DMSR (LSE-61), Section 1.2.11 "Level 1 Data Quality Report Definition" (ID: DMS-REQ-0097): *"The DMS shall produce a Level 1 Data Quality Report that contains indicators of data quality that result from running the DMS pipelines, including at least ... detection efficiency for point sources vs. mag for each utilized filter."* However, this is a nightly data quality assessment and not a per-image product.

→ The DMAD (LDM-151), Section 5.6.3 "MakeSelectionMaps", states that this calibration step *"is responsible for producing multi-scale maps that describe LSST's depth and efficiency at detecting different classes of object. The details of what metrics will be mapped, the format and scale of the maps (e.g. hierarchical pixelizations vs. polygons), and the way the metrics will be computed are all unknown"*. It also states that this must be extendable to Level 3, but that *"the details of what DM will provide still needs to be clarified to the community"*, and notes that the reprocessing time for fake plants could be prohibitive. (Section 3 "Alert Production" also specifies that in LDM-151 *"we do not address estimation of the selection function for alert generation through*

the injection of simulated sources ... Source detection thresholds can be estimated through the use of sky sources".)

→ Suchyta et al. (2016) presents `balrog`, a fake-embedding software package for detection efficiencies, and applies it in demonstration to DES data (it's not just for point sources, but uses `galSim` to embed shapes built of e.g., Sersic profiles).

→ In a UW DM Brown Bag lunch meeting, it became clear that while there are many opinions on how to plant fake sources, not only are there no actual plans to do this, but there is no funding or FTQs allotted to figure out how the deliverable of detection efficiencies should be achieved.

→ **Unassessed**. But I plan to take this on in the near future with an assessment of our options, their costs, and the scientific payoff, and also this is not a question limited to SP so I have not spawned a **Concern** for this.

- Multiple Special Programs request short-exposure survey, which would provide less/different stars for photometric/astrometric calibration. Would a 1s exposure contain enough stars for calibration?

→ **Answered**. 1 second exposure will contain stars from $12.9 < r < 21.0$, which should be more than enough for a full photometric calibration. However, LSE-180 only considers calibration for the standard WFD data, and so this has not been tested.

3.6 Custom Co-Adds

- Building deep co-adds with no transient contamination (e.g., SN hosts, cosmology uses host parameters correlated with intrinsic SN brightness), or creating deep stacks across filters for faint-object detection at $< 5\sigma$ in a single filter, may be a desired capability for Special Programs data. Will this be possible?

→ **Answered**. Yes. Creating custom co-adds will be a Level 3 capability that will be possible with the Level 2 DRP pipeline algorithms described in LDM-151. This includes multi-filter stacks (DMSR Section 1.3.4. 'Multi-band Coadds' ID:DMS-REQ-0281).

3.7 Moving Objects

- For faint moving objects, users will develop shift-and-stack (SAS) pipelines in Level 3 (MOPS does not SAS). Will they have all the necessary infrastructure for this intensive process? Will they be able to use the LSST real/bogus on their shift-and-stacked frames? Will they be able

to access the difference images over long time scales?

→ **Unanswered.** All of this seems likely, as each of these questions is covered by the DM requirement to *"make all codes extendable to Level 3"*, but ensuring that the answer is yes to each of them will be more involved in reality, and probably happen only with time as the SAS Level 3 pipeline is developed and integrated?

- Can MOPS link newly discovered sources to those discovered earlier on the same night (e.g., in a 4-visits-per-night scenario), or does it have a similar potential problem to DIASource and DIAObject?

→ **Answered.** MOPS does not run during the night in real time. This is not a problem.

4 Science Cases and their DM Requirements

In this section we review science documents from LSST and the community. We compile information about the nominal observing plans for Special Programs that are already written down (Section 4.1), and summarize the science motivations for Special Programs that have so far been proposed (Section 4.2). Potential issues for DM are inferred from these science motivations and have been incorporated into our list of issues in Section 3. In Section 4.3, we explore some Special Programs processing case studies as an additional avenue to identify potential issues.

Resources:

- "LSST: from Science Drivers to Reference Design and Anticipated Data Products" [11]
- The LSST Science Requirements Document [10]
- "General Review of the Proposed DDF and MS", LSST AHM Aug 2016 presentation by Niel Brandt <https://project.lsst.org/meetings/lsst2016/sites/lsst.org.meetings.lsst2016/files/Brandt-DDF-MiniSurveys-01.pdf>
- "Simulations, Metrics and Merit Function for Mini-Surveys and DDF", LSST AHM Aug 2016 presentation by Stephen Ridgway https://project.lsst.org/meetings/lsst2016/sites/lsst.org.meetings.lsst2016/files/Ridgway-SimulationsMetrics_1.pdf
- "LSST's DC [Deep CoAdd] Bias Against Planets and Galactic-Plane Science" by A. Gould, [9] <https://arxiv.org/abs/1304.3455>
- Chapter 10 "Special Surveys" of the Observing Strategy White Paper [17]
- List of LSST Deep Drilling white papers: <https://project.lsst.org/content/whitepapers32012>

4.1 Nominal Observing Plans for DDF and MS

Ivezić et al. (2008, [11]), Section 3.2.1 "Mini-Surveys": describes a nominal DDF data set as ~ 50 consecutive 15 second exposures in each of four filters in one hour per night, once every two nights, for four months. Each observation would have a limit of $r \sim 24.5$; a one-hour nightly stack would have a limit of $r \sim 26.5$; and assuming a 60% completion rate (weather), the four-month ~ 40 hours stacked together with the ~ 180 main survey visits would yield a limit of $r \sim 28$. The LSST Science Requirements Document [10] doesn't contain the terms "deep drilling field" or "mini-survey". Four extragalactic deep drilling fields have already been specified (Table 2). From a scouring of mainly the presentations of Brandt and Ridgway at the

2016 AHM, an *incomplete* list of potential mini-surveys that people are thinking about in Table 2 also.

TABLE 2: Approved DDF and Incomplete List of Potential MS.

Name	Coordinates	Description
DDF Elias S1	00:37:48, -44:00:00	approved, cadence TBD
DDF XMM-LSS	02:22:50, -04:45:00	approved, cadence TBD
DDF Extended Chandra Deep Field-South	03:32:30, -28:06:00	approved, cadence TBD
DDF COSMOS	10:00:24, +02:10:55	approved, cadence TBD
DDF TBD		TBD
North Ecliptic Spur		solar system objects (find and characterize)
Galactic Plane		more intensive stellar surveying
South Equatorial Cap		S/LMC and more Galactic science
Twilight		short exposures (0.1s) for bright stars
Mini-Moons		finding mini-moons
Sweetspot		60 deg from Sun for NEOs on Earth-like orbits
Meter-Sized Impactors		detection a week before impact
GW Optical Counterparts		search and recovery
Old Open Cluster M67	dec +12	compact survey above Galactic plane

4.2 Science Descriptions (Not Comprehensive)

The following is a collection of rough notes on the science goals for some proposed Special Programs, classified by area of astronomy. It is not comprehensive. Any major questions that were inspired by compiling these science goals have been incorporated into Section 3, so no **Concerns** are spawned in this section (although there are some low-level unanswered questions).

Solar System Objects

- Observations in the North Ecliptic Spur region will yield more ≥ 140 m NEOs (Brandt talk).
- Detecting faint SSO down to $r \sim 27$ by applying shift-and-stack processing to a mini-survey (Trans-Neptunian Objects, Trojans, asteroids, long-period comets, dwarf planets). [1]
→ Regarding shift-and-stack, [1] says that “the multi-fit algorithm ... naturally provides a base infrastructure for this process. In particular, the marshaling of the pixels to attempt a given photometric measurement is non-trivial when tens of thousands of images are required. However, the multi-fit middleware is required to do exactly this, so we expect that this issue will be resolved by the time SAS is needed.” Is this true? **Unanswered**, but this does not seem like an urgent concern for DM, so no **Concern** is spawned.
- Mini-moons, temporarily captured satellites of the Earth. (Section 10.2, [17])
- Find meter-sized impactors < 2 weeks before impact. (Section 10.2, [17])

- Sweetspot: twilight fields to find NEOs in Earth-like orbits (never in opposition fields). (Section 10.2, [17])

Stars in the Milky Way and Magellanics

- Deep imaging acquired over a short time span to mitigate proper motion loss (and increase flare detection rates) will identify useful stellar populations. [5].
→ When targeting Galactic plane regions and/or open clusters, typically 1 to 3 filters are needed for detections of e.g., faint stars already detected in z and y but needing i to distinguish from red galaxies [5].
- A nominal deep-drilling observing strategy over the full L/SMC galaxies will characterize stellar variability to $M_V < 6.5$ on timescales from 15s to 3d [18].
→ Special co-adds may be required, e.g., *"To reach variability levels of 0.1 to 0.005 mag will require co-adds depending on the timescale of the particular variables"*. [18].
- A Twilight Short Exposure survey would enable bright stars (with many historical observations) to be put solidly on the same photometric system as the deeper LSST WFD main survey catalog.
- A short-exposure survey of M67, Chapter 10.4 of [17], suggests using *"custom pixel masks to accurately perform photometry on stars as much as 6 magnitudes brighter than the saturation level"*, but is this feasible? **Unanswered**, but this just seems like a description of a potential Level 3 product, so no **Concern** is spawned.

Exoplanets

- Transits. The nominal DDF plan described in [11] would allow for 1% variability detection over hour-long timescales, which is suitable for detecting transits. A DDF field at Galactic latitude 30 degrees would yield 10^6 stars at $r < 21$ that would have $\text{SNR} > 100$ in each single exposure of the sequence. Microlensing events can also be detected with this data set. In both cases, follow-up is required. The Galactic Plane MS is proposed for this. [9] describes how transits can be extract from the same data set.
- Microlensing. Slower than a transit, [9] suggests that ~ 22 mag imaging every 3-4 days (i.e., the WFD nominal cadence) can find microlensing candidates (for follow-up with e.g., LCO). However, more Galactic regions must be included, and this will require image differencing in crowded fields. [9] claims that imaging in the necessary regions of the galaxy has been disfavored by the project on the basis that the eventual deep co-adds would be uselessly confusion limited. (But was this before the Galactic Plane region was defined?)

Supernovae

- Extend the SN sample to $z \sim 1.2$ and providing more densely sampled light curves for cosmological analyses with the nominal DDF plan described in [11], which builds nightly stacks with a limit of $r \sim 26.5$.
 - The optimal exposure time distribution might be 6, 5, 10, 10, 9, 10 in *ugrizy*, which adds up to < 60 minutes and could presumably be done with the standard 30s exposures. [15]
- Generating the best-possible individual SN light curves by building special, deep-as-possible, SN-free host galaxy images and using them as a template. This will be necessary for SNe that appear in the template image (e.g., that last > 1000 days). SN-free images will also be needed to measure correlated properties for cosmology and to do host-galaxy science.
 - The Level 2 DRP codes to create CoAdds should be adequate for this.
 - [7] also mentions *"characterization of ultra-faint SN host galaxies"* in their Galaxies WP.
- A short-exposure survey could include nearby SNeIa on the same photometric system – but again this comes back to the short-exposure questions of whether we can do photometric calibration as well.
- High-cadence observations to detect fast events, particularly young SNe and Gamma-Ray Bursts.

Galaxies

- To build a large collection of low- μ objects: [7] mentions *"identification of nearby isolated low-redshift dwarf galaxies via surface-brightness fluctuations"* and *"characterization of low-surface-brightness extended features around both nearby and distant galaxies"*.
- Characterization of high- z clusters (will depend on ability to deblend extended objects).
- AGN monitoring on a variety of timescales in well-characterized galaxies [7] [8]
 - Study the effects of AGN on e.g., bulge-disk decomposition parameters?

Weak Lensing

- DDF can help with shear systematics and the effects of magnification in the analysis of WFD data (community forum, Jim Bosch)
- *"Will need to process at least some deep drilling fields (high-latitude ones) in the same way we process a full data release production before running the full data release production, so we can use the results to build priors and/or calibrate shear estimates on the wide survey"* (community forum, Jim Bosch)
- *"Will need to process various wide-depth subsets of some deep drilling fields (again, high-latitude ones) using the regular DRP pipeline. We'll definitely want best-seeing, worst-seeing, and probably a couple of independent typical-seeing subsets, but there may be other ways we'd want to subdivide"*

as well." (community forum, Jim Bosch)

- Photo- z are very important to weak lensing [16] and so perhaps the implemented method should be chosen with weak lensing science prioritized.

4.3 Special Programs Processing Case Studies

For further insight to the DM-related needs of potential Special Programs, we can write out all of the data acquisition and processing steps, in order, that some of the proposed Special Programs might use. This kind of thought experiment of describing the reductions and processing could also be a required section of all future white paper proposals. Note that we are not including any analysis in these descriptions. Also, they aren't that fully realized yet, and could use some more thought and input.

Generic steps currently include:

- Step 1. Data Acquisition.
- Step 2. Inclusion in Level 1 AP.
- Step 3. Delivery of LSST Processed Images.
- Step 4. Level 3 Processing.
- Step 5. Level 3 Products.
- (Step 6. Inclusion in Level 2.)

4.3.1 A Mini-Survey Searching for TNOs with Shift-and-Stack

This Special Programs processing summary is based on Becker et al. (2011) white paper to find TNOs with shift-and stack [1]. *MLG: See Section 4.3.5 for a more general processing case-study written by Mario, for a shift-and-stack Level 3 pipeline running regularly on a large amount of data.*

Step 1. Data Acquisition.

The observational sequence is triggered. In a single night, the 9 adjacent fields in a 3x3 grid are observed with 336×15 second r -band exposures. This sequence is always repeated 2-3 nights later. This re-visit sequence is repeated 3 more times: 1.5 months, 3 months, and 13.5 months later. Data obtained in the g -band filter is also acceptable. [1]

Step 2. Inclusion in Level 1 AP.

Each 2×15 second visit is processed as Level 1 and Alerts are released within 60 seconds.

Step 3. Delivery of LSST Processed Images.

The raw, reduced, and calibrated exposures and difference images from the Level 1 pipeline are made available within **L1PublicT** (currently 24 hours, but see Section 2.2). However, this Special Program appears to require a year of dispersed observations before the Level 3 processing would run.

Concern XIII. Will there be a Science Platform capability set up so that users are notified not just when any data for their Special Program is taken, but when certain sets are finished and ready for processing?

Step 4. Level 3 Processing.

The Level 3 pipeline running the shift-and-stack processing will be set up and submitted for batch processing by the user through the Science Platform (this particular Special Program does not appear to need regular or urgent processing, and so its Level 3 code will probably not be federated). Pipeline inputs will be the 336 processed exposures per field per re-visit sequence. The Level 2 difference imaging routine will be used with the same template track/patch for all. Custom code will shift the difference images, then Level 2 routines will stack and do source detection and characterization (including real/bogus?) and generate an object database. Custom code will derive orbital parameters for the detections and add them to the database. The user is notified with processing status updates.

Concern XIV. Does the Level 3 pipeline have to include the steps to regenerate the processed exposures, or will DM deliver a service to automatically initiate regeneration by e.g., simply calling for an exposure or difference image?

Concern XV. Is it yet known how users will apply for batch processing priority and how they will be notified their jobs have finished, e.g., RSS feed? I could imagine that this is a requirement that would be defined in a future document on the Science Platform deliverables?

Step 5. Level 3 Products.

The shifted-and-stacked images. Databases of sources detected by the shift-and-stack process with the desired schema, some of which will overlap with Level 1, 2, and/or MOPS to facilitate table joins, some of which will be generated by custom codes.

Step 6. Inclusion in Level 2.

Given the extreme difference in final depth of these fields, it is unlikely that they will be in-

corporated into the Level 2. However, external Level 3 processing could certainly make deep CoAdds for these fields, join them to the Level 2 DRP catalog, and use them for other science.

4.3.2 Searching for Supernovae in Deep Drilling Fields

Step 1. Data Acquisition.

On a single deep drilling field, the scheduler obtains e.g., 5, 10, 10, 9, and 10 visits with 2×15 second exposures in *grizy* (or similar for the night's filter set) and a small dither pattern between visits.

Step 2. Inclusion in Level 1 AP.

Each 2×15 second visit is processed as Level 1 and Alerts are released within 60 seconds.

Step 3. Delivery of LSST Processed Images.

The raw, reduced, and calibrated exposures and difference images from the Level 1 pipeline are made available within **L1PublicT** (currently 24 hours, but see Section 2.2).

Step 4. Level 3 Processing.

The federated pipeline runs automatically upon delivery of the Level 1 processed images. If a deep template image does not already exist (this would be a Level 3 product), a Level 2 stacking routine will be used to make them. This Level 2 stacking routine is also used to make the nightly deep CoAdd in each filter, and then a Level 1 or 2 routine will generate the difference images and run source detection (and real/bogus?). These sources will be added to Level 3 equivalents of the **DIAObject** and **DIASource** databases. The LSST codes for alert packet and transport are used to distribute the detected objects e.g., to the same brokers that receive the Level 1 Alert Stream.

Concern XVI. Would the internal real/bogus routine be able to run on nightly CoAdd DDF difference images? If so, its outputs might not be reliable because of train/test set mismatch; will the internal LSST routines for real/bogus training be available and adaptable to Level 3 users?

Concern XVII. Confirm that resources (packet distribution) for a Level 3 alert stream would be supported and with what latency, whether and how such packets would be identified as non-AP, and that they could be ingested by brokers in real time.

Step 5. Level 3 Products.

A deep template for this field, the nightly stacked CoAdds and difference images, and analogs of the DIAobject and DIASource databases but built only for this specific Deep Drilling Field.

Concern XVIII. How long can some Level 3 products remain on disk, such as template images for nightly CoAdd image subtraction?

Step 6. Inclusion in Level 2.

At the discretion of DM.

4.3.3 A Twilight Survey of Bright Stars

Step 1. Data Acquisition.

At a specified time (or e.g., 6 degree twilight), the scheduler begins dither pattern of short exposures (1-5 seconds?). Location and exposure times are set by the sky brightness and desired saturation limits. Consider: this survey might obtain a larger fraction of high-airmass (i.e., anti-solar) images.

Step 2. Inclusion in Level 1 AP.

Pending a study of short-exposure suitability for psf-matching and differencing with a deeper template, and pending a decision on whether the AP processing capabilities can be extended to accommodate a faster image-input rate, these data could be incorporated.

Step 3. Delivery of LSST Processed Images.

Pending a study of the ISR pipeline's ability to reduce exposures with a short integration time and a high sky background, the raw, reduced and calibrated exposures (and difference images, if created) are made available within **L1PublicT** (currently 24 hours, but see Section 2.2).

Step 4. Level 3 Processing.

User-generated, user-run scripts might include the following, assuming that the ISR pipeline will produce reduced exposures:

(A.) If difference images are required for science and the Level 1 AP is not run, then custom template images and difference images, and catalogs of difference-image sources, could be created using Level 1 or 2 codes. If the PSF is not well-enough formed, or there is a problem with DCR corrections (e.g., if obtained at high airmass), then some more sophisticated user-written codes will be necessary.

(A2.) If the field includes a nearby supernova, special SN-free template images will be created using Level 1 or 2 image-stacking codes, difference images created with Level 1 DIA codes,

and then photometric calibration done using the results of Level 3 processing.

(B.) A script of Level 2 codes to extract and characterize sources in individual exposures, stack exposures into CoAdds (probably not optimized for depth, but maybe for precision), and extract and characterize sources in the CoAdds will be built and run by users. A specialized deblender for sources that span a large dynamic range might be necessary?

(C.) User-written analysis codes for photometric calibration.

Step 5. Level 3 Products.

(A.) Special template images, difference images, and difference-source catalogs.

(B.) Special CoAdds and source catalogs.

(C.) Catalogs of bright objects in the LSST filter set.

Step 6. Inclusion in Level 2.

Probably not, given the same regions would be surveyed with WFD exposures, these short exposures probably wouldn't help the DRP's objectives.

4.3.4 The Galactic Plane Survey for Variable Stars

Step 1. Data Acquisition.

The schedule incorporates fields in the Galactic Plane, and executes e.g., 2x15 second exposures.

Step 2. Inclusion in Level 1 AP.

These images are incorporated into the Level 1 pipeline. Alerts are produced and DIAObjects is updated.

Step 3. Delivery of LSST Processed Images.

The raw, reduced, and calibrated exposures and difference images from the Level 1 pipeline are made available within **L1PublicT** (currently 24 hours, but see Section 2.2).

Step 4. Level 3 Processing.

User-created, user-run pipelines that incorporate a mix of Level 1 and 2 DM-provided codes and e.g., specialized deblenders, CoAdds, and photometric calibration routines targeted towards specific stellar populations and/or Galactic latitudes, depending on science goals.

Step 5. Level 3 Products.

Possibly a wide variety of difference and CoAdd images and their associated source catalogs.

Step 6. Inclusion in Level 2.

These images are incorporated into the Level 2 DRP to the extent that they assist with the all-sky completeness and photometric calibration; TBD, at the discretion of DM.

4.3.5 A More Generic Level 3 Shift-and-Stack Case Study, by Mario

A Level 3 processing case study for shift-and-stack on a large number images. By Mario.

#1. The scheduler is configured to repeatedly (e.g., 10 times) observe a field during the same night with longer exposure than usual (e.g., 120 sec). [and we should take the actual numbers from the TNO-DDF whitepaper; don't have the internet right now or I would].

#2. The images are processed as regular "Level 1" products within 60 seconds, and transmitted as alerts, with results stored into the regular L1 database. This will happen automatically for all images (perhaps within some range of exposure times?).

#3. The raw images (and all necessary calibrations), calexps, and standard L1 diffims are made available within 10 minutes to the batch system for processing with special programs-specific codes. This is the same batch system we make available to the users for running Level 3 codes [Q for us: is it? or is is the same one that's used to process calibrations? have these systems been sized?]. (**See Concern IV.**)

#3 a). The code running the shift-and-stack processing will be externally developed and delivered, but will be installed and operated (and change controlled!) by the LSST Operations team. That is, we don't expect someone external to the ops team to babysit the code on a nightly basis. In fact, it's the opposite: once the codes are delivered, any changes will go through LSST's software change control process.

Concern XIX. Confirm that the idea of federated Level 3 codes still exists (e.g., given how the conversation on incorporating a photo-*z* code ended up in a proposed descope).

#4. There will be a facility to trigger program-specific processing on the batch system upon the arrival of a new image (above); this processing will then be queued up for execution. We assume that the policy for processing of special programs data may give it preferential treatment relative to general-purpose L3. (**See Concern XIX., above.**)

#5. Once the processing finishes, the results of will be stored to a program-specific database. No alerts (in VOEvent sense) will be issued. We will provide a generic notification facility (perhaps something as simple as an RSS feed) that new data has been made available in a certain database/data store. [This is an example where I'd want to make sure somebody within DM is planning to provide such a facility.]. (**See Concern XV.**)

#6. The outputs stored can be special-program specific (i.e., tables with nearly arbitrary schemas – some columns – like ra/dec for spatial joins – should be present in main tables). The outputs can also contain images (stored in also special-program specific repository), or custom products (treated like opaque files). The visualizations available for these (catalogs, images, arbitrary files) through the Portal will be limited (e.g., generic table visualizations or x-y plots).

#7. When the images are made available to the batch system (step #3), they also become available to *everyone*. I.e., someone else could also run a custom L3 pipeline on these data, feeding their custom L3 database. [this isn't in the requirements right now – right now we say that images will become available in 24hrs. But I'd like to get K-T's reaction to this sort of proposal – this kind of facility would be extremely powerful, and from a technical perspective my hunch is that once you have #3, you get #6 for free as well.] (**See Concern II.**)

5 Conclusions

Conclusions section.

References

- [1] Becker, A.e.a., 2011, Science White Paper for LSST Deep-Drilling Field Observations Opportunities for Solar System Science,
<https://project.lsst.org/content/whitepapers32012>
- [2] **[LSE-29]**, Claver, C.F., The LSST Systems Engineering Integrated Project Team, 2016, *LSST System Requirements*, LSE-29, URL <https://ls.st/LSE-29>
- [3] **[LSE-30]**, Claver, C.F., The LSST Systems Engineering Integrated Project Team, 2016, *LSST System Requirements*, LSE-30, URL <https://ls.st/LSE-30>

- [4] Crots, A.e.a., 2011, Standard Candle Relations and Photo-diversity of Type Ia Supernovae,
<https://project.lsst.org/content/whitepapers32012>
- [5] Dhital, S.e.a., 2011, Science White Paper for LSST Deep-Drilling Field Observations Mapping the Milky Way's Ultracool Dwarfs, Subdwarfs, and White Dwarfs,
<https://project.lsst.org/content/whitepapers32012>
- [6] **[LSE-61]**, Dubois-Felsmann, G., 2016, *LSST Data Management Subsystem Requirements*, LSE-61, URL <https://ls.st/LSE-61>
- [7] Ferguson, H.C.e.a., 2011, Science White Paper for LSST Deep-Drilling Field Observations: LSST Deep Drilling for Galaxies,
<https://project.lsst.org/content/whitepapers32012>
- [8] Gawiser, E.e.a., 2011, Science White Paper for LSST Deep-Drilling Field Observations: Ultra-deep *ugrizy* Imaging to Reduce Main Survey Photo-z Systematics and to Probe Faint Galaxy Clustering, AGN, and Strong Lenses,
<https://project.lsst.org/content/whitepapers32012>
- [9] Gould, A., 2013, ArXiv e-prints, ADS Link
- [10] **[LPM-17]**, Ivezić, Ž., The LSST Science Collaboration, 2011, *LSST Science Requirements Document*, LPM-17, URL <https://ls.st/LPM-17>
- [11] Ivezić, Ž., Tyson, J.A., Abel, B., et al., 2008, ArXiv e-prints, ADS Link
- [12] **[LSE-180]**, Jones, L., 2013, *Level 2 Photometric Calibration for the LSST Survey*, LSE-180, URL <https://ls.st/LSE-180>
- [13] **[LDM-151]**, Jurić, M., Lupton, R., et al., T.A., 2013, *Data Management Applications Design*, LDM-151, URL <https://ls.st/LDM-151>
- [14] **[LSE-163]**, Juric, M., et al., 2017, *LSST Data Products Definition Document*, LSE-163, URL <https://ls.st/LSE-163>
- [15] Kessler, R.e.a., 2011, Science White Paper for LSST Deep-Drilling Field Observations: Supernova Light Curves,
<https://project.lsst.org/content/whitepapers32012>
- [16] Ma, Z.e.a., 2011, Science White Paper for LSST Deep-Drilling Field Observations: Using LSST Deep Drilling Fields to Improve Weak Lensing Measurements,
<https://project.lsst.org/content/whitepapers32012>

- [17] Marshall, P., et al., 2017, Science-Driven Optimization of the LSST Observing Strategy, in preparation,
<https://github.com/LSSTScienceCollaborations/ObservingStrategy>
- [18] Szkody, P.e.a., 2011, Science White Paper for LSST Deep-Drilling Field Observations High Cadence Observations of the Magellanic Clouds and Select Galactic Cluster Fields,
<https://project.lsst.org/content/whitepapers32012>

Draft