



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: 2017-05-05

**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 (SP) such as the deep drilling fields (DDF) or mini-surveys (MS). 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. **Preliminary action items are listed in Section 1.1.**

MLG: The very high level requirements for Level 3 and Special Programs are written down in the DMSR, but they're broad. The next document down-flow, the LDM-151, is focused on the Level 1 and 2 algorithms; Section 7 mentions the DAX and SUIT, but says they detailed in a different document (?).



## Change Record

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

Draft

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Current Action Items and Open Questions . . . . .	1
1.2	Interacting with the Science Community . . . . .	3
<b>2</b>	<b>The Current DM Plans for Special Programs</b>	<b>4</b>
2.1	Incorporating SP Data into the WFD Main Survey . . . . .	4
2.2	Documentation Review . . . . .	6
<b>3</b>	<b>Questions and Potential Issues</b>	<b>9</b>
3.1	Alert Stream . . . . .	9
3.2	Databases and Schema . . . . .	10
3.3	Templates . . . . .	12
3.4	Snap Handling . . . . .	12
3.5	Calibrations . . . . .	13
3.6	Custom Co-Adds . . . . .	14
3.7	Moving Objects . . . . .	15
<b>4</b>	<b>Science Cases and their DM Requirements</b>	<b>16</b>
4.1	Nominal Observing Plans for DDF and MS . . . . .	16
4.2	Science Descriptions (Not Comprehensive) . . . . .	17
<b>5</b>	<b>Special Programs Processing Summaries</b>	<b>20</b>
5.1	A Mini-Survey Searching for TNOs with Shift-and-Stack . . . . .	21



5.2	Searching for Supernovae in Deep Drilling Fields . . . . .	22
5.3	A Twilight Survey of Bright Stars . . . . .	22
5.4	The Galactic Plane Survey for Variable Stars . . . . .	22
<b>6</b>	<b>Conclusions</b>	<b>22</b>

Draft

# Data Management and LSST Special Programs

## 1 Introduction

The current status of this work is that it is an internal, ongoing study with several interim preliminary action items or outstanding questions so far (Section 1.1).

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, K-T, Tim), and then
- (3) ensure the identified needs become specified requirements (i.e., action items Section 1.1).

*Mario: "Examples include making sure there's a requirement that the system is able to use an arbitrary template for some (deep drilling) field and a plan to deliver that functionality; or that the requirements define how a user-supplied code will be able to run in the batch system to process the last-night's data with shift-and-stack code for KBOs."*

### 1.1 Current Action Items and Open Questions

Action items for Data Management can be done internally as needed – some are just questions of things to investigate. Items that involve gather feedback from the Science Collaborations should probably wait and happen at the same time as the next calls for community input (see Section 1.2).

#### **Data Management:**

- (1) Confirm whether processed single visits for Special Programs can stay on disk and accessible for Level 3 pipelines for a longer amount of time than the WFD images, by user request. (Section 2.2)

- (2) Confirm whether the computational/storage budget requires more than 10% of the total resources for Special Programs and Level 3 processing. (Section 2.2)
- (3) 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. (Section 2.2)
- (4) Find out whether we can expect a filter change time shorter than 120 seconds (the specified maximum), which would enable a DDF to sequence filters. (Section 2.2)
- (5) Confirm whether alternative data products can contribute to the Alert Stream, or if only the Level 1 processing system will create alerts. (Section 3.1)
- (6) Update the database schema for DIASource to add an element that contains information about which template was used to create the difference image (e.g., tractId or patchId). (Section 3.2)
- (7) Clarify whether database schema key Object.prv\_inputID will identify whether an object is an externally provided coordinate for forced photometry. (Section 3.2)
- (8) Find out whether a SP visit with a single exposure (no snaps) disqualifies it from incorporation into Level 1. (Section 3.4)
- (9) Confirm that the ISR algorithms will cover a range of exposure times and dither patterns. (Section 3.5)
- (10) Implanting fake sources remains TBD; will DM provide a code base to do this in Level 3? (Section 3.5)

### Science Collaborations:

- (1) SC should be solicited to provide input about all special program science goals that require 60 second alert latency from non-standard exposure times. (Section 3.1)
- (2) SC should be solicited to provide input on whether the space allotted for object characterization parameters in the database schema such as variability (e.g., lcPeriodic = float[6 x 32]) or morphology (e.g., extendedness = float[1]) are adequate to describe objects with Special Programs observations, which may include a wider range of cadences and depths than the WFD observations. Two examples include: (1) Does the variability parameter lcPeriodic have enough space to adequately characterize high and low cadence observations with a variable dynamic range from a Twilight Survey? (2) Do the morphology parameters have enough elements to adequately represent both bright spiral arms and ultra-low-surface brightness features like tidal tails? DM will first compile and provide a list of planned parameters from the schema. (Section 3.2)
- (3) SC should be solicited to provide input on whether there are science use-cases for spe-

cialized, auto-generated template images e.g., with constant  $\Delta t$ , or image-image differencing (instead of image-template). DM should then assess whether this will require additional database elements. (Section 3.3)

## 1.2 Interacting with the Science Community

To resolve some open issues we will need to solicit input from the science collaborations (SC). However, it may be better and easier to wait until the next call for community input, and package these requests up in a broader request that a kind of "DM Requirements" section be included in all white papers. 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.

Recall that there is already a community forum on this topic, so far minimally used but ready to be revised in the future when the next calls soliciting white papers are released.

<http://community.lsst.org/t/deep-drilling-fields-and-data-management/1115>.

Remember also that the SC Coordinator for DDF: Neil Brandt.

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)?



## 2 The Current DM Plans for Special Programs

In this section we review what is written down in LSST documents with regards to the processing of Special Programs data. We start with an overview, below, and then in Section 2.1 we summarize how Special Programs data will be incorporated into the Level 1, 2, and MOPS pipelines. In Section 2.2 we break it down into specific issues that arise from the three Data Management documents (DPDD, DMSR, and DMAD), and non-DM LSST documents.

**Overview:** Although the data from some special programs might be incorporated into the Level 1, 2, and/or MOPS pipelines (see Section 2.1 below), every special program will also (probably) need one or more associated Level 3 processing pipelines, and databases that include only the products from that Special Program (joinable to Level 1 and 2 databases). 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 an effort will be made to ensure the DM codes are extendable to Level 3. DM will also ensure that the relevant software infrastructure exists for science users, and for LSST to automatically run user-provided Level 3 pipelines (e.g., difference imaging on deep drilling nightly stacks). To enable science goals, some Level 3 pipelines can be run by LSST with a projected latency of (at most) tens of minutes (as opposed to the 60 second guaranteed for WFD). Level 3 pipelines for stacks and co-adds that do not have the same real-time urgency can simply be run by users through the Science Platform.

### 2.1 Incorporating SP Data into the WFD Main Survey

I haven't found much written down in terms of guidelines or expectations for incorporating Special Programs data into the Level 1 and Alert Stream, Level 2 DRP, and/or MOPS pipelines – especially when these data deviate from the standard 30 second visit – but it seems feasible. Here we make these considerations.

#### **Incorporating SP data into Level 1 and the Alert Stream.**

It is reasonable to attempt to include as many LSST images into Level 1 and the Alert Stream as possible. When the exposure time is equivalent to a WFD visit image,  $\sim 30$  seconds, this is obviously going to work fine. When the exposure time is shorter than a WFD visit image, this might be OK as well. The minimum supported exposure time is currently specified as 1

second, which spans 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 (all other filters have a similarly large overlap). **Except: The PSF must be well formed ( $> 5$  seconds?) and a Level 1 template image must exist, for the image to be incorporated into Level 1.** The exposure times of a special program image might be longer than the WFD visit image, but given that the template will typically be a stack of a year or two of data (at least), and that it's hard to imagine a reason for even doing 150 second exposures (given that the saturation would be  $r \approx 18.3$  and the cosmic rays would intensify), it seems unlikely that any images would be rejected from Level 1 for having too long an exposure time. **Except: Are there any system qualities that vary on short timescales which would prohibit the calibration of long images, such as tracking?** This means that the most likely limiting factor of including any/all special programs data into the Level 1 and Alert Stream will be a processing back-up caused by the higher cadence of short exposures, or a lack of template images for special programs that go outside the WFD search area. (Also, I checked with David Reiss and he is not aware of anyone considering problems arising from differencing shorter/longer exposure times.)

### **Incorporating SP data into MOPS.**

Needs and capabilities are currently undefined, but it is expected that MOPS will identify and run on and and all Special Programs data that it can, and that the MOPS pipeline builders will define that boundary.

### **Incorporating SP data into the Level 2 DRP.**

All Special Programs data will NOT be automatically included in Level 2. It will be included only where DM determines that it benefits the main science goals of Level 2 DRP. For example, photometric calibrations may require that some or all of the (shallower) Galactic Plane survey area be incorporated in order to suppress edge effects and low-order modes in the photometric solutions. For another example, the uniform-depth requirement of weak lensing studies may require that the image products of Level 2 DRP be created without the DDF. However, identical Level 3 data pipelines and products may be used for regions with deeper coverage and the two databases can be joined, such that science users can easily access both when needed.

## 2.2 Documentation Review

### DMSR, LSE-61

Data Management Subsystems Requirements [6] contains the high-level requirements for all processing and products. These have flowed down from the SRD (LPM-17), the SR (LSE-29), and the OSS (LSE-30), which are discussed in the next section.

- The only part of the DMSR that directly references Special Programs (or DDF or mini-surveys) is Section 1.2.2 "Processed Visit Images" (ID: DMS-REQ-0069), which specifies that "Processed science exposures are not archived, and are retained for only a limited time to facilitate downstream 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 when required by the science users, or if it will only live a short while.

#### **Action 1. 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 high-level 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. → I thought that the Science Platform was being developed specifically so that no one needs to process LSST data on their home machines?

**DMAD, LDM-151**

LSST Data Management Applications Design [13] 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 meaning 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. 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.

**DPDD, LSE-163**

Data Products Definitions Document [14] specifies that the data products for Special Programs will "be created using the same software and hardware as Levels 1 and 2". 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/desirable. The Level 3 products for SP data can/will be kept in separate (joinable) databases that have identical schema as the WFD databases.

- It is specified that SP data processing will be limited to not use 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.

→ 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). On top of this, add the Level 3 processing that will be applied to WFD data, and we could easily find that the Special Programs data are processed  $\sim 4\times$  as much as WFD.

**Action II. Confirm whether the computational/storage budget requires more than 10% of the total resources for Special Programs and Level 3 processing.**

**SRD, LPM-17**

The Science Requirements Document [10] makes no reference to special programs, deep

drilling fields, or mini-surveys; it's scope is the science of the WFD only.

## LSST SR, LSE-29

The System Requirements document [2] is derived from the LSST Science Requirements Document (SRD, [10]). 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 the DMSR and OSS (detailed and discussed above and below).

## OSS, LSE-30

The Observatory System Specifications (OSS) document [3] 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 LPM-17 and LSE-29). 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 between exposures may need to be lengthened in order to maintain camera thermal stability, and that thermal stability might also be affected with longer exposure times – but there doesn't seem to be an official limit on 'thermal stability'.
  - **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
  - The specified maximum filter change time is 120 seconds, but this could be a debilitating overhead for, e.g., a deep drilling cadence of consecutive 30 second images that rotates through filters.
  - **Find out whether we can expect a filter change time shorter than 120 seconds (the specified maximum), which would enable a DDF to sequence filters.**
- 3.6.2.4. Minimum filter change count (ID: OSS-REQ-0295): day, 8, night, 4
  - There is no related specification limiting the maximum number of total filter changes to avoid wear-and-tear on the mechanism?

## LSE-180

Level 2 Photometric Calibration for the LSST Survey [12] 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 is that of the WFD Main Survey.

### 3 Questions and Potential Issues

In this section we list questions or potential problems identified only from reading LSST documents [14, 13], or issues raised also in Gregory's AHM 2016 slides<sup>1</sup>. 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**, and

**bold if an action item is required**.

#### 3.1 Alert Stream

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

→ **As it is 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.**

→ **Confirm whether alternative products can contribute to the Alert Stream.**

- When we attempt to include short exposures (e.g., 5 seconds) in Level 1 then this higher data acquisition rate (e.g.,  $\sim 6\times$ ) might cause a processing back-up at NCSA. Is there a science need for 60 second latency on a consecutive time-series of short exposures? Or would e.g., 60 seconds latency on every 6<sup>th</sup> image, be preferred?

→ **Query the community about all special program science goals that require 60 second alert latency from non-standard exposure times.**

- 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?

→ **Immediate updates of the DIAObject catalog is an expected capability, although it is cur-**

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

rently uncertain how it will be accomplished, technically.

- 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.

→ Upon review, this requirement seems unrelated to visit locations, and thus not an issue for special programs.

## 3.2 Databases and Schema

- The database schema for DIASource needs to have an element added to identify the template image that was used; this is relevant for the Main Survey WFD Level 1 and also all Level 3 differencing products.

→ **Database schema for DIASource needs to have an element added that contains information about the template that was used to create the difference image.**

- 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.

→ This is definitely a promised Level 1 product, but it is unclear whether we need to add an element to the database schema. 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?

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

- Will the following element sizes be enough to characterize the different kinds of variability that could be measured with cadences that are quite different from the main survey (i.e., longer or shorter cadenced special programs)?



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

→ Although these elements could be differently populated in the Level 3 databases in order to meet the science goals of a given special program, we need to ensure that they are adequate for the Level 1 data which might incorporate many timescales.

→ **Contact the Science Collaborations for input on whether float[6 x 32] for the characterization parameters for variability is adequate to express the additional timescales and depths afforded by a special programs observing strategy.**

- Low surface-brightness features and/or objects are required for some galaxy science, and will show up more in e.g., a DDF than the WFD stack. Will the related shape parameters such as extendedness that will be included in the Object database be sufficient to describe the extra level of detail in low-surface brightness features afforded by e.g., DDF stacks?

→ 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. It's unlikely that extremely deep patches will be included in the Level 2 products, but this needs assessment.

→ **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?

→ 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?

→ 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?

→ 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?

→ This seems unlikely to be included as Level 1 (and unnecessary). 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  $\Delta t$  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?

→ 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, but is there a lower limit to exposure time for this? E.g., would not apply to a 1s twilight survey?

→ Have not yet tried to answer this question, but I wonder if for Special Programs that have single-image visits instead of two snaps, whether the lack of snaps will disqualify them from Level 1 inclusion?

→ Does a SP visit with a single exposure (no snaps) disqualify it from incorporation into Level 1?

### 3.5 Calibrations

- 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 will be characterized for a variety of exposure times – but will it for sure?.

→ Note that 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).

→ **Confirm whether it is implicit in the LDM-151 that all the ISR algorithms will automatically investigate and derive corrections for a range of exposure times, dither patterns, etc. The feeling is that no, very short exposures will not be supported, but Robert will know.**

- 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).

→ **There remain no concrete plans for developing the tools for detection efficiencies in Level 1, and furthermore currently no time in the Level 1 developers' schedule to explore or attempt 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?

→ **A 1 second exposure will contain stars from  $12.9 < r < 21.0$ , which should be more than enough for a full photometric calibration.**

→ Note that LSE-180 only considers calibration for the standard WFD data.

### 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.

→ 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?

→ In particular, people may want to SAS on difference images, which would require DM making these available for some time.

→ **Confirm that science users could write a Level 3 SAS with the existing planned algorithms? E.g., see Section 5.**

- 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?

→ 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.

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](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

[11], Section 3.2.1 "Mini-Surveys": describes a nominal DDF data set as  $\sim 50$  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  $\sim 40$  hours stacked together with the  $\sim 180$  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"; I assume that the SRD is only for WFD 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 and potential data processing/product issues for some proposed special programs, classified by field. It is not comprehensive. The questions for DM that were inspired by considering the proposed special programs are still mingled in amongst these notes, but have also been incorporated into Section 3, and so the answers and solutions to any questions or problems are all in Section 3 also.

### Solar System Objects

- Strategy would be to observe 3x3 grid of 9 fields centered on coordinate of an ecliptic latitude equal to 0 and a longitude at a conjunction of Trojan clouds of Jupiter and Neptune, 4 groups of 2 visits on a 2–3 day revisit schedule, with the 4 groups targeted at e.g., opposition, all done in *r*-band (*g* suitable too). (Is this within the North Ecliptic Spur?) [1].
- North Ecliptic Spur region will yield more  $\geq 140$  m NEOs (Brandt talk).
- Are the three SSO special programs described in Chapter 10.2 of [17] mandated by NASA?
- SSO science goals require enabling the discovery of faint SSO via shift-and-stack of images in specific fields where SSO discovery is maximized [1].
- Does MOPS processing do shift-and-stack, and/or is the MOPS database schema ready to incorporate a shift-and-stack output?
- 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." True?

- To detect Earth impactors *"hours before impact"* using the four visits per night will require real-time linking of new DIASources – or maybe the MOPS handles this differently?

### Stars in the Milky Way and Magellanics

- Special programs will provide images deeper than the main survey in order to identify useful stellar populations; getting the images over a short time span mitigates proper motion losses and increase flare detection rates [5].
- Target Galactic plane regions and/or open clusters, typically 1 to 3 filters needed for detections of e.g., faint stars already detected in  $z$  and  $y$  but needing  $i$  to distinguish from red galaxies [5].
- Characterize stellar variability over the full L/SMC galaxies 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].
- Light curve characterization parameters in the databases should include one for very short events like flares; ensure there's adequate space in the database for multiple variability parameterizations.
- How might LC-characterization parameters fail on e.g., an LBV-turned-SN?
- A Twilight Short Exposure survey would need good on-the-fly sky subtraction of a rapidly changing background; ensure plans are adequate?
- 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"*; feasible?

### 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  $\sim 22$  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. I don't think DM could fix this fact, and deep co-adds (and all DRP) could possibly just be skipped for these regions. Otherwise, the already-planned AP should work fine for Galactic plane survey.

### Supernovae:

- The nominal DDF plan described in [11] would allow for nightly stacks with a limit of  $r \sim 26.5$ , extending the SN sample to  $z \sim 1.2$  and providing more densely sampled light curves for cosmological analyses.
- [4] doesn't say much specific.
- [15] describes their optimal exposure time distribution as 6, 5, 10, 10, 9, 10 in *ugrizy*, which adds up to  $< 60$  minutes and could presumably be done with the standard 30s exposures.
- Template generation: e.g., supernovae will want a DDF template from  $> 30$  days ago (minimum for SNeIa);  $> 100$  days is better (SNIId), and  $> 200$  days safest (SLSNe, SNIIn). So, last DRP will be fine. But some last longer, thousands of days, and so users will want to know the date of the template.
- Building special, deep-as-possible, SN-free host galaxy images in order to measure correlated properties and make corrections to the LC, and do host science. The regular DRP will be inadequate for this (this could be a Level 3 tool to stack requested epochs only?). With short runs on a DDF this might be impossible, would have to revisit. [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 can we calibrate as well with the limited set of stars?

### Galaxies:

- [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*" – building a large collection of low- $\mu$  objects is a science driver but do these even get in the Object database (**check on the 'extendedness'-related parameters,**



**e.g. GALFIT modeling? to what  $\mu$ ?**, or is this assumed entirely Level 3?

- It seems that e.g., characterization of high- $z$  clusters will depend on ability to deblend extended objects.
- AGN monitoring in well-characterized galaxies [7], and for a long-term DDF on many timescales [8]
- Effect 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 [16] and so perhaps the implemented method should be chosen with weak lensing science prioritized

## 5 Special Programs Processing Summaries

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 analysis that would be typical in some kinds of observing proposal should also be a required section of all future white paper proposals.

Generic steps should include:

Step 1. Observations.

Step 2. Inclusion in Level 1?

Step 3. Rapid single-image processing and product delivery.

Step 4. Level 3 processing and product delivery.

Step 5. Inclusion in Level 2?

## 5.1 A Mini-Survey Searching for TNOs with Shift-and-Stack

(1.) The observational sequence is triggered. In a single night, the 9 adjacent fields in a 3x3 grid are observed with  $336 \times 15$  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]

(2.) The individual 15 second images are processed as Level 1 and Alerts are released within 60 seconds. (This is not relevant or useful to the science goals of the Special Program).

(3.) The raw images (and all necessary calibrations), calexp, and standard L1 diffims are made available within 10 minutes (or 24 hours?) to the batch system for processing with special programs-specific codes. **Confirm that 10 minutes is a requirement and/or a plan, and 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?**

(4a.) 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. 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.

(4b.) 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.]. *MLG: This kind of notification does not appear to be covered by any item in the DMSR, but it seems more appropriate to be described in the Science Platform deliverables?*

(4c.) 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).

(4d.) 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.]

(5.) Given the extreme difference in final depth of these fields, it is unlikely that they will be incorporated into the Level 2. However, external Level 3 processing could certainly make deep CoAdds for these fields and use them for other science.

## 5.2 Searching for Supernovae in Deep Drilling Fields

(1) The scheduler

## 5.3 A Twilight Survey of Bright Stars

## 5.4 The Galactic Plane Survey for Variable Stars

# 6 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] Crofts, 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>