Project Response to Review Report of the Data Products Definition Review for the Large Synoptic Survey Telescope Project

Summary of the report findings and general response

The review panel provided the Project with highly valuable recommendations and observations. The recommendations centered on requests for clarifications of the existing requirements and design, providing further details on some data products, and terminology changes to the documentation. These have not significantly affected design, nor MREFC cost or schedule. In particular, no fundamental issues were identified that would preclude start of Construction.

In addition to requests for clarifications, discussion during the review provided additional algorithmic and process/policy options to consider in future value engineering. There was a question regarding the sufficiency of compute and storage capacity provided for user science (Level 3) in the Data Access Centers. We have clarified that the SRD-stated L3 capacities requirement of 10% applies to the total computing and storage budget, not just the portion related to Data Release processing.

Tier 1 Recommendations to be addressed before Final Design Review

While the project will attempt to address all recommendations and observations over time, the current focus of the project and the reason for conducting this review were to address any issues deemed critical for Final Design Review, i.e. the Tier 1 Recommendations. The project response has been added in red, italicized, text to the text of the committee report, attached below. The updated Data Products Definition Document has been submitted for approval to the LSST Change Control Board.

Report for the review of “LSST Data Products Definition Document” and Project Response

Introduction:

The committee appreciates the preparation that went into the talks and documents for this review. Its members thank the LSST Project for the opportunity to participate and believe that the LSST is a tremendously exciting project which will bring great benefit to the whole of the astrophysics community.

Regarding the specific review of document LSE-163, "LSST Data Products Definition Document", we agree that this is an essential and timely document which properly fills the gap between the LSST Science Requirements Document and the downstream detailed implementation documents. LSE-163 will help guide construction of the LSST DM system.

With regard to the specific questions of the charge, the committee as a whole has reviewed the presentation materials and composed their responses:

Q1 - Does the definition of Level 1 Data Products satisfy the data product requirements laid out in the LSST Science Requirements Document?

The plan for the Level 1 Data Products (L1DP) expressed in the Data Products Definition Document (DPDD) meets the requirements expressed in the Science Requirements Document (SRD) for the data release schedule that L1DP are 'generated continuously every observing night, including alerts to objects that have changed brightness or position'.

We note some slight confusion about terminology regarding derived vs. Level 1 ‘raw’ products. For instance, S4.1 states that 'Level 1 data products are a result of difference image analysis' but we think the SSObject and DIAObject tables may be thought of as 'derived data products'.

The plan for the L1DP appears to be mostly consistent with the minimum specifications. They provide extensive information on the 'source' properties such as positions, fluxes and shapes.

The L1DP plans do not adequately address the metadata that will be available for each image and source. For instance, the background level at each source. The issue of deblending L1DP sources needs to be further addressed. For instance, a flag could be introduced indicating that the L1DP source is blended with another source.

The L1DP could include information on the N next-nearest source(s), rather than just three.

The plans for the L1DP software framework appear to be consistent with the DPDD design specification. The proposed software framework is already being used for processing images from other telescopes (e.g. CFHT Megacam and Subaru Hyper-SuprimeCam) and the fact that it is already flexible enough for use on multiple different camera systems bodes well for its application to the LSST L1DP framework.

The plans for the L1DP appear to be consistent with the stretch goals as outlined in the SRD. The planned software system has been demonstrated to be flexible enough to enable new algorithms to be implemented easily (e.g. the flux-dependent PSF width for thick CCDs). We suggest that care be taken to maintain the flexibility of incorporating new algorithms as the software matures - it is too easy to let the code 'harden' as the operations deadline approaches.

The plans for the timely release of L1DP images and catalog data are consistent with the SRD.

The planned L1DP update schedule is consistent with the schedule laid out in the SRD.

The planned L1DP in the DPDD is consistent with the SRD's requirements that the 'fast release of data on likely optical transients will include measurements of position, flux, size and shape, using appropriate weighting functions, for all the objects detected above transSNR signal-to-noise ratio in difference images (design specification: 5). The data stream will also include prior variability information and data from the same night, if available. The prior variability information will at the very least include low-order light-curve moments and probability that the object is variable, and ideally the full light curves in all available bands.'

We note that the SRD used a design specification of transSNR=5 for sources in the difference images but the DPDD usually refers to a specific value of SNR=5. This should be corrected so as not to tie the DPDD to a specific transSNR as the committee had concerns about the false detection rate at SNR=5.

The SRD places limits on the 'minimum number of candidate transients per field of view that the system can report in real time' (transN). Phrased as such it sounds as if the L1DP *must* provide at least transN candidate transients per FOV but we think that the intent is that the L1DP must be *capable* of providing at least transN events per FOV. There was no discussion in the DPDD on whether the system will meet this requirement. On the other hand, we have little doubt that it can meet the SRD minimum spec of transN=1,000 - the document should simply discuss how the project plans to meet the limit.

Our main concern is whether the minimum transN spec is sufficient in the presence of false detections. We naively expect on the order of 100(s) of statistical false detections per FOV and expect there to be many more times systematic false detections (glints, edge effects, etc.) so that a minimum spec of 1,000 transients may be filled only by false detections. A possible solution would be simply to increase the transSNR limit, but the DPDD assumes SNR=5.

The L1DP described in the DPDD appear to be consistent with the SRD's requirement that the transient data stream can be modified with a 'query-like pre-filtering'. The project plans to allow the user to query the LSST database as required. Furthermore, the project plans on providing several pre-defined filters for 'popular transients'.

Q1: Tier 1 -- Recommendations For Immediate Actions:

* Modify the DPDD to use transSNR rather than tying the transient threshold to SNR=5.

We have made it clear that the numerical value (5) derives from the SRD transSNR requirement. Also, to simplify the maintenance of the document, we've defined a \transSNR LaTeX macro that defines the value of transSNR in one place, and is easy to change if transSRR changes.

In general, for readability purposes we strive to have the actual numerical values in the DPDD and similar descriptive documents, rather than symbolic references (though we do note where the numerical values came from). The normative requirements documents (e.g., the SRD, LSR, OSS and DMSR) are all written in SysML, and use its mechanisms to ensure consistency and traceability.

* Add sigma/uncertainty values on trailFlux, trailLength, trailAngle quantities.

The uncertainties (the full covariance matrix) are included in the trailCov column. We have clarified that this is the case.

* Implement the new IAU H/G12 system for asteroid absolute magnitudes and slope parameters.

Agreed. We've replaced the old (H, G) system, with the Muinonen et al. 2010 magnitude-phase system.

* Consider changing the use of the word 'source' to the community-accepted 'detection'.

While we share the committee's concern, we worry that this would require changing a significant number of documents in the relatively short time before Final Design Review, thus drawing resources from other issues that the committee has raised.

We would therefore prefer to consider it after the FDR.

Q1: Tier 2 -- Recommendations For Actions (longer term)

* Users should be able to request and obtain the detection efficiency for an object in an image as a function of (ra, dec, wavelength, rateOfMotion, etc.).

* Introduce a simple means of determining whether a model fit is 'good' e.g. including values for DOF, chi^2, effective number of pixels etc.

* Clarify the reason why DIAObject and SSObject tables are also considered L1DP, even though they are in some sense 'derived data products' from the DIASouce table.

* Consider including more information in the DIASource table --   e.g. # of neighbors within N arcsec, correlation with other surveys, background information, deblending flags. Of course, this must be balanced against data size and search limitations.

* It appears that the plan is to mix observations of SSObjects from other observatories with the observations from LSST. This needs to be considered carefully as mixing the data will make it difficult to inventory the solar system with LSST in an unbiased way. There must be some means of ensuring that the LSST's selection effects can be measured and adding other data into the mix will make it difficult.

* the MOID specification is a pair of numbers but does not specify what object the MOID is measured with respect to. i.e. in some cases, a MOID with respect to centers other than the Sun is useful.

Q1: Tier 3 - Concerns

* The time between visits needs to be assessed for asteroid linking in the presence of false detections. Linking pairs of detections into tracklets on a night and then

linking them between nights into tracks is easy only when the false detection rate is low and the time between visits is small.

* The association of a visit 'source' with an SSObject will be a first. All surveys to-date have only done associations of objects with tracklets. The associations can probably only be made for 'numbered' asteroids for which the astrometric uncertainty is on the order of an arcsec. Is it really worth making the associations per visit when most LSST asteroids will be unknown and unnumbered? i.e. consider the benefits of a small reduction in the unknown transient rate compared to the effort and danger of false linkages in doing so.

* There needs to be consideration or requirements placed on the template used in differencing; e.g. the template should not be too close in time to the visit in order that very slow moving objects can be detected in the DIA.

* Improve wording to the trailed source model fit: all DIASources are fit to all source models - not just the ones that move sufficiently fast.

• Observations

* There is almost no discussion of comets in the DPD. They are considerably more difficult to detect and process than asteroids.

* There needs to be a flag indicating whether trails border a mask, CCD edge, etc.

Q2. Are LSST time domain event alert contents and distribution plans adequate to enable science cases depending on immediate follow-up?

The current specifications for the content of time domain event alerts and the distribution mechanism generally enable science cases envisioned for LSST that rely on immediate follow-up with other instruments.

Alerts contain enough information to significantly reduce the need for additional queries using LSST database before the follow-up can be triggered. Context information that enables linking alerts with all relevant Level 1 and 2 data products has been provided. Plans for a basic event brokering service have been presented to ensure that alerts are available to the world when LSST begins regular survey operations regardless of the availability of an externally supported service. While this service will be limited in capacity and function, it is recognized that the development of a fully featured event broker is outside the project scope.

However, these plans can be further improved and adjusted to support a wider range of science investigations without overtaxing the LSST resources.

Individual snaps within each visit contain information on variability on a time-scale of ~15 seconds. LSST is therefore uniquely positioned to support studies of very fast variability down to an unprecedented sensitivity level. Notable examples are real-time detection and follow-up as well as statistical studies of M dwarf flares and optical flashes associated with gamma-ray bursts. Even more important is the potential for entirely new discoveries in this uncharted territory. In order to maximize the utility of LSST data in this area, we recommend measuring the flux difference and motion between the two snaps of each visit for all variable objects.

Q2: Tier 1 -- Recommendations For Immediate Action

* Alerts should include information about the quality of the model fit, at least for the point source model, to support prioritization of potential follow-up targets by automated event brokers and individual users. This could be automatically included with the full Object record, but it was not clear from the definition of the Object table that the information on the effective number of degrees of freedom is propagated along with the log likelihood values. This issue should be addressed explicitly in the document.

To address this recommendation, we've added columns for χ^2 and the number of data points used in the fit wherever logL was quoted before. This should enable the end-user to compute the number of degrees of freedom, or perform model selection based on any of the numerous criteria described in the literature (BIC, AIC, etc.).

We note that should the fitting algorithm change to require an explicit column for the effective number of degrees of freedom, we will modify the schema accordingly. As the recommendation notes, the key is to provide the users with information necessary to prioritize and make decisions.

Q2: Tier 2 -- RFAs

* Consider the exciting possibility of keeping information on the difference of the two snaps in a single visit -- there will likely be interesting sources that vary in time or space on 15 sec timescales. To the extent possible, this information should be included in Level 1 data products and distributed with real-time alerts. At the least, the measurements that do show a significant difference between the snaps should be kept and distributed in real time.

* Currently there is no estimate of the probability that the difference flux is blended in a given LSST exposure. In addition to random line-of-sight pairs there may be a significant contribution from binary variable stars (see for example Kochanek et al. 2006). Such cases should be considered and processed by the deblender in the production system. Deblended measurements and corresponding flags should be provided in Level 1 data tables and included in real-time alerts.

* The data products specification should include an estimate of the rate of false positives in the alert stream and demonstrate that the data management system has been sized to handle the implied data volume. Also, the peak rates of alerts will be much larger in crowded fields than average rates across the sky. Quantitative estimates of peak rates and their impact on the latency in alert distribution should be collected and factored into the bandwidth budget.

* A detailed association of real-time alerts with external catalogs is better performed in down-stream event brokers. However, some cross-identification information and multi-wavelength measurements should be included with each alert to enable preliminary inferences about the physics of the object without the need to search other databases. Examples include SDSS object id and magnitudes, infrared magnitudes, and nearest radio source.

* Due to a very large expected rate of alerts, only a small fraction of all real-time events will be followed up. This means that follow-up groups need to be very selective and may want to frequently change how they prioritize follow-up targets. If filter code on real-time follow-up is to change or evolve with time, the project should state what sort of flexibility it is willing to support for individual users from the LSST event broker. Can frequent changes of Python and SQL snippets be accommodated by the system?

* Support for external event brokers needs to be specified in more detail and further quantified. How many live connections that transmit all alerts can be sustained on the LSST side? In case the demand exceeds this capacity, how will the decision be made as to which event brokers will be directly connected?

* The area of the sky covered by pixel cutouts in alerts should be reassessed and the binning factor optimized based on user needs. The current specification is that 30x30 pixel postage stamps will be provided. Given a small LSST pixel size, this will not be sufficient in many cases; for example: to evaluate the host galaxy environment for numerous extragalactic transients.

Q2 Tier 3 -- concerns

* While the XML based VOEvent format is currently a standard way of packaging alerts for public distribution, it is very inefficient with respect to storage capacity and network bandwidth. If a suitable, more compact format standard emerges, we recommend considering it.

Q3. Does the definition of Level 2 Data Products satisfy the data product requirements laid out in the LSST Science Requirements Document?

The SRD is very general, and the committee finds that the definition of L2DPs is satisfied in spirit. In detail, we would like a number of things to be fleshed out and clarified.

Accessing and handling masks has been a recurrent issue in the analysis of large surveys. The panel sees here an opportunity for the LSST Project to set a generic standard for masks. This should be done keeping in mind the necessity to include masks in the estimation of correlation functions.

Related to masks, it is important to provide a coverage map, including masks where the photometry is unreliable (missing data, saturated pixels, halo around bright stars, satellite trails, etc.). Such masks are crucial in (i) estimating the robustness of a detection, (ii) estimating surface densities of objects in the sky (the quantity required to measure spatial correlation functions).

Q3 -- Tier 1 -- RFAs Immediate:

* The document should describe how the sky map will be stored and how users will retrieve this. The data products must provide a map of the sky-level and its uncertainty for each object in the source catalog.

A discussion of image characterization data has been added to Section 3. This includes the determination of the sky level, and a commitment to store it. We have also added the entries for estimates of the sky background to the relevant Source and DIASource tables.

Additionally, we have added a more explicit discussion about storing the variance (weight) and mask maps, including a commitment to investigate delivering simplified masks of the covered survey area (e.g., similar to what's commonly done with Mangle today).

* There is no discussion of masks in the document. In the short term, some brief description of the creation, storage, of masks should appear in the Data Products Definition Document.   The associated data products should be outlined.

We agree this was an omission. A discussion of masks has been added to Image Characterization Data subsection in Section 3.

Q3 Tier 2 RFAs:

* The DPDD or a derived document should eventually have a detailed description of how masks will be created, stored and made available.

Q3 -- Concerns and Comments

* We don't think lensing and photometric redshift (photo-z) requirements are addressed in a sufficiently explicit manner in either the SRD or this document. There

are placeholders for model fits and minimal likelihood sampling. These products as explicitly outlined appear to be sufficient to do photometry and support the time domain science, but not necessarily the lensing or photo-zs. (We address this more in response to question 4).

It became clear during discussions that there is much that is planned that is simply not written down in the document. A lot of these ideas will be important for the success of LSST. An example is the longer term plans for algorithm development and use of the
Level 3 system to aid in development and bring the community along. The system as envisioned in fact seems to be more capable and flexible than is indicated. We think generally the authors should write more of these capabilities, flexibility and design ideas into the document at an earlier stage.

Q4. Are the catalog Data Products envisioned for Level 1 and Level 2 appropriately comprehensive to enable a significant fraction of science envisioned for the LSST (the LSST Science Drivers, LSST SRD, Section 2), without further access to pixel data?

From the presentations and discussion, it appears that the Data Products and the associated system are comprehensive enough to facilitate a significant fraction of the LSST science -- but the explicit content in the DPDD gives a more limited picture.

The explicit examples given in the DPDD text do not appear to be sufficient for photo-z and lensing. For example, most modern algorithms for lensing require a full sampling of the likelihood with thousands of points, but the examples given suggest of order 100 likelihood samples would be saved for model fits, which would not be sufficient. In discussions which went beyond the formal review material, it was mentioned that the system could support significant compression of information that would allow storage of more samples or whatever other data is needed for interpretation of the stored results. This compression could be bit compression of the data or the ability for the database to calculate and store sufficient information for a subset of objects which can be interpolated to other objects.

We think the imaging "cutouts" for measured objects are an important data product; in particular the set of cutouts from all single epoch images at the location of a coadd object or some region of sky. In discussions it was indicated that this "product" will most likely manifest as a "service", whereby the users can request the cutout images.

Q4 -- Tier 1 RFAs immediate

* Clarify in the DPDD what the real constraints are on eventually being able to implement ambitious (and not yet fully determined) weak-lensing (and photo-z)

algorithms.   If the real constraint is on *storage* space rather than number of items stored per object, or what is stored, that should be made clear.

*We have added a section named 'Supporting Science Cases Requiring Full Posteriors', where we clarify this point. In particular, we view the current allocations to represent **conservative estimates**, with improvements on order of factor ~4 (at least) being expected once compression techniques are applied to the data (and we give examples of such techniques).*

* The DPDD should be explicit about how the PSF information will be stored and made available to the users.

We have added a section on image characterization data to Section 3 of the document. We are not able to specify the exact format of PSF at this point, given it is likely to depend heavily on the final algorithms; however, we can guarantee that a user will be able to obtain a realization of the PSF at any point In an image.

* Make sure it is clear in the document how signal-to-noise is defined, as the term is used in different ways in different places in the document (i.e. transSNR, SNR, per pixel, per object, per visit, per coadd, etc).

We've scrubbed the document and clarified the definition of SNR where we found it to be unclear.

Q4 -- Tier 2 RFAs

* Make explicit details of the imaging cutout service, and how large numbers of cutouts could be served to a users in an efficient fashion. The cutouts should be full resolution, and sized variably as appropriate for each object (or at least optionally so). The exception would be in the alerts where other constraints may be more important.

Q4 -- Concerns and Observations

* In discussions it became clear that significant computation time will be available per object (of order a few seconds per single epoch image). This leaves open the possibility of running very interesting and computationally intensive algorithms, and thus the possibility for rich data products. This should be made more explicit in the document.

*   It would be good to have the storage requirements on disk detailed for each major data product (e.g. coadds vs. visits vs. catalogs etc) in the DPDD or an easily accessible associated document.

It may be useful to size the cutouts 2^N or 3*2^N to facilitate fast FFTs. Perhaps allowing the user to specify the cutout size would be possible.

Q5. Do planned Level 3 capabilities, in conjunction with Level 1 and Level 2 data products, reasonably enable science cases that greatly benefit from co-location with LSST data and computing, within the guidelines and boundaries defined in the LSST SRD.

Level 3 capabilities can provide the community with an interesting environment to perform analyses involving LSST software and data. While the idea is definitely attractive, the effectiveness of this platform, as currently defined, is unclear.

The current plan is to allocate 10% of the storage and computing resources to users among the community, with access regulated via a 'proposal and allocation committee' mechanism. If the quoted resources have to be distributed among a large number of users (likely dominated by graduate students who often use more resources than actually needed) it is not clear what the computing capabilities will allow each user to do, given the size of the LSST dataset. For a given analysis (especially involving a sizable amount of external data), it is not clear in which instance a user should upload the external data to a Data Access Center vs. downloading LSST data subsets to their local computer.

It became clear in discussions that level 3, while a very capable service, is actually quite small in terms of dedicated resources. The impression of the committee, thought of in terms of expected 2020 computing capabilities, are perhaps comparable to an SDSS skyserver type of service, with some added processing capability. Significant science will be done at the catalog level and on small numbers of images, but major computation (weak lensing reanalysis) will not be possible.

From the DPDD: "As a part of the Level 3 Programming Environment and Framework, the LSST will make available the LSST software stack to users, to aid in the analyses of LSST data." It is not clear whether this "software stack" will allow users to test the LSST source extraction pipeline. It will be of great interest for users to upload simulated images (or images from other surveys) and test the completeness and purity of the source extraction software. This might be useful in various cases: to test the photometry of a faint source blended with a much brighter one, assess the detectability of certain types of variable sources, etc. Such a capability does not exist in the context of the SDSS and has been a limitation for a number of analyses.

Q5 -- Tier 1 RFAs immediate:

* We recommend stating in the DPDD which ‘source extraction processing codes and capabilities’, if any, will be accessible and present in early implementations of Level 3, as well as providing a bit more background on the role of Level 3 within the LSST project.

We’ve clarified this at the beginning of “Level 3 Programming Environment and Framework” section.

In general, LSST will make all its data management code available for reuse at Level 3, allowing for rare exceptions where licensing reasons may prevent us from doing so. It is LSST DM policy to use and deliver all code under GPLv3 license, unless the inability to do so would put the entire project in jeopardy.

Q5 -- Tier 2 RFAs

* Survey science greatly benefits from multiwavelength analyses. We see a need to make non-LSST data available in this environment. For example GALEX, WISE, Planck, NVSS, etc. Data from such all-sky surveys will be needed by most people and should be provided in a common fashion to all interested users. The DPDD should reflect this capability explicitly.

Q5 Tier 3 Concerns and comments

The inclusion of pixel-level data at other wavelength will be welcome.

Question 6:

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

Yes, with qualifications.

Q6 Tier 1 -- RFAs Immediate:

* As mentioned above, the metadata tables at all levels (1, 2, and 3) should be made more explicit, i.e. include (preliminary) metadata table schema listings in the DPDD to a similar extent that the Object and Source/Detection related tables were provided.

We have now defined the baseline metadata schema in a separate document. It is available in web-accessible form at http://ls.st/8g4.

Q6 Tier 2 -- RFAs:

* As mention in 3 and 4 above, the 'exposure level' data, including information on sky and sky noise for each piece of the sky should be made more explicit, and thought should go into how easily this information may be used by scientists.

Q6 Tier 3 -- Concerns and comments

* It would be helpful to describe in the DPDD in somewhat more detail how users can determine code versions and processing metadata (date and time and versions of processing stacks) and how actual code repository access will work.

As highlighted above, the description of the storage and meaning of bit-plane masks (bleed trails, satellite trails, very bright stars, bad pixels, etc) and sky-weight information must eventually be made crystal clear, and a reasonable (efficient) mechanism should be provided for inputting this weight information into, for example, large scale structure correlation results (i.e. two point correlation function of galaxies, fully corrected for masked regions and variable sky depth coverage). It should be possible to determine depth of a coadd at any given point or region on the sky without regard to whether or not an object exists at that point in an efficient way.

LSST has the opportunity to become the de-facto standard for storing and using on-sky position (rather than object) based mask and weight information, if it is clearly defined and easy to use. It is not clear that there is any standard in the field yet with regard to this. One may wish to consult with experts on such mask/weight efforts as 'MANGLE', ‘STOMP’ and with those who have done large scale imaging (not just spectroscopic) galaxy correlation analyses to understand their concerns on incorporating sky weight maps and masks.

* Timekeeping information should be made explicit (i.e. TAI vs. UTC, beginning, end or average time within a visit, etc).

The use of MySQL is a bit of a concern due to well-known technical limitations. The authors in discussions indicated that they had "workarounds" for some of the limitations of this technology. But there are more mature and feature-full open-source alternatives that are standards compliant and do not suffer the same technical limitations (for example: postgres). We understood that the alternatives will have their own trade-offs, strengths and weaknesses. The authors indicated they can "drop in" other databases as needed. We think it would be worthwhile to seriously evaluate alternatives. There was a concern expressed about database technology maturity (i.e. mySQL and its two backend engines myISAM and INNODB), and database backup/recovery plans (or lack thereof) for a system of such large scale.

It was viewed as a strong positive by the committee that LSST is a leading player in the XLDB movement.

* The planned method of replacing some tables on an annual basis (i.e. as false positive detection codes improve) is reasonable, however some effort should be maintained by the project of (at least) keeping a table of matching id numbers between deprecated object tables to more easily support users who are working with older versions of the tables.

* The optimal representation of light-curve metrics is another area where there is not a consensus in the field. The place holder provided is adequate but the project should keep in touch with developments.

Question 7: Do designs for LSST Data Products include adequate provisions for adjustments during construction, commissioning, and operations as needed to ensure the continued scientific adequacy of LSST Data Products?

A: Yes, with minor reservations
With regard to staffing for implementing the software effort, while slightly beyond the scope of this question, the committee believes that while the suggested numbers of FTEs proposed to carry out the work is a reasonable number, there is impedance-matching risk: risk of being able to find, retain and manage the right people with the right balance of science and coding expertise as well as keeping up motivation throughout the development stages.

Because of the long lead time of the LSST, and because of the evolving-state-of-the art of the algorithms and computing hardware and software issues, the project will need to remain flexible, and will need to keep in especially close contact with other active surveys (i.e. HyperSuprimeCam, DES) to pay close attention to what they are discovering. It was excellent to note, for example, that the recently discovered 'PSF-width dependence on brightness' relation for deep well CCDs has already been realized and acknowledged within LSST and ideas are being discussed to address it.

We encourage (and liked what we heard about) processing of publicly available data from other surveys (SDSS, HPC, DES) through the LSST infrastructure as way to keep scientists involved and interested. We encourage regular data challenges (with simulated and real data, if possible) and wide-spread brainstorming (even beyond the LSST collaboration) about key algorithms (deblending, lowest-noise-unbiased-galaxy-model-fitting, weak lensing measures, photo-z, masking-and-weighting, psf-combining), and to keep activity on these algorithms at the forefront of the collaboration’s collective thoughts as first light approaches.

The industry-standard mechanism chosen by LSST, that of having a controlling requirements document, supplemented by a change control policy, is excellent, and we believe essential for a project of LSST's scale.

The tools which LSST has chosen and already established for software management and development (EUPS, SVN/git, Python scripting with C/C++ underneath) give the survey an important foundation for bringing coders and science developers together to work in a uniform, distributed, controllable way.

Examples of where algorithms are rapidly evolving and will require years of scientist expertise include weak lensing, deblending, galaxy flux modeling and photo-z (where priors and likelihood tables require the saving of much information with each object). Optimal solutions to these issues are not known. The current DPDD does offer a path forward, however, there is no guarantee that the ultimate method for obtaining adequate shears, for example, can be accomplished within the scoped resources. The DPDD authors have attempted to scope for a range of possibilities, but for unsolved problems such as lensing and photo-z's, it is impossible to predict the future needs, or what the final data products will look like. For example, it may be that the optimal science-driven computational or storage requirements will be larger than all of the resources available to LSST on a reasonable timescale. In such cases, well-established standards (with flexibility built-in) and communication mechanisms throughout the project will be highly important.

Q7 Tier 1 -- RFAs Immediate:

None

Q7 Tier 2 -- RFA:

* Keep in close touch with the wider astrophysics and engage with the algorithm community on the complex algorithms (i.e. weak-lensing, deblending) needed to analyze LSST data effectively. Maintain flexibility, and remain open to developments.

Q7 Tier 3 -- Concerns

* We encourage reasonable allocation of resources to management of software development personnel.