Overview of LSST approach to Transients

WP2(Data Access Centre)
Task 2.3 (Support of Phase A Dev work)

Project Acronym LUSC-A

Project Title UK Involvement in the Large Synoptic Survey Telescope

Document Number LUSC-A-02 <Contact George.beckett@ed.ac.uk for number>

Submission date	25/06/2019
Version	1.0
Status	Complete
Author(s) inc. institutional affiliation	A.Lawrence, D.Morris, S.Voutsinas (Edin) S.Smartt (QUB)
Reviewer(s)	Bob Mann (Edin)

Dissemination level	
Confidential	*The prime purpose is for informing other members of the consortium. In due course it may be suitable to be public.*

Version History

Version	Date	Comments, Changes, Status	Authors, contributors, reviewers
0.1	23/2/2017	Initial draft	A.Lawrence
0.5	28/2/2017	Full draft, after informal review by LSST project members. Circulated to many other LSST:UK scientists, for informing Phase B proposal development	Lawrence, Morris, Voutsinas, Smartt
1.0	25/6/2019	Final version, at end of Phase A.	Lawrence, Morris, Voutsinas, Smartt

Overview of LSST approach to alerts and variable objects

Andy Lawrence[1], Dave Morris[1], Stelios Voutsinas[1], Stephen Smartt[2], Roy Williams[1]

Abstract
In order to provide a context for LSST:UK DAC planning, we summarise the expected LSST arrangements and performance for alert detection, processing, and distribution; and for providing information from which to construct light curves of variable sources. Science requirements and DAC plans are discussed in separate documents, but we provide here a short summary of implications for the UK DAC and Broker.

[1] *Institute for Astronomy, University of Edinburgh, ROE, United Kingdom*
[2] *Astrophysics Research Centre, Queen's University Belfast, United Kingdom*

Contents

1. Introduction

1.1 Context

Most of the information described in this document is available in various LSST documents, but is here summarised as briefly as possible, and set in a context of UK DAC and DEV plans. Since the first full draft (v0.5 of Nov 2017) LSST:UK has developed a prototype event broker (Lasair) and developed other variability analysis capability plans as part of the Phase B proposal. This version (v1.0 of June 2019) is being completed at the end of Phase A and will help to inform Phase B planning. Apart from updating, the main development since v0.5 is some pruning, and the addition of material describing more general variability-related data processing and products, as opposed to only alert production.

1.2 Motivation

Probably the most important feature of LSST is that it surveys the sky repeatedly, which has several scientific advantages. Over time, stacking of data leads to very deep images; the light curves of variable objects can be constructed; and transient objects can be detected in a timely manner. Transients are of increasing importance in recent astronomy, with important science coming from supernovae, tidal disruption events, gamma-ray bursts, and gravitational wave events. Some of the key issues include:

- Discovery of transients.
- Creation of an alert stream,and its transmission.
- Astrophysical characterisation of transients, and how to filter the stream.
- Interaction of the LSST system with third party brokers.
- How users (astronomers and robots) will consume and act on the filtered stream.
- How post-hoc data mining of an alert database will be enabled.
- How to characterise the variability properties of objects in the LSST database.
- How to construct light curves of objects on various timescales.
- How to trigger on interesting behaviour of objects that don't appear in the alert stream.
- How to enable data-mining on variability properties.

1.3 Definitions and key points

Throughout this document, *DAC* refers to the anticipated UK Data Access Centre, which will receive data from the LSST project and serve it to registered UK users. Likewise *DEV* refers to science software developed in the UK which interacts with and adds value to the DAC infrastructure (i.e. what the LSST project refers to as "user generated products").

By *transient* we mean photometric transients - LSST will also of course be important for the discovery and characterisation of moving objects such as Near Earth Asteroids, but we are not concerned with these here. The term "transient" has a well understood astrophysical meaning - roughly, something that comes out of nowhere - but there is of course a grey area between true transients and extreme variability of known objects.

By *alert*, the LSST project means anything that appears in the nightly *difference images* at 5σ or more. The number of alerts will be far greater than the number of transients, including very large numbers of variable objects. Conversely, depending on the epoch of the template image and the new image, the alerts may or may not include what could be called "slow transients" - for example a quasar undergoing an outburst over several years.

Study of variable objects will need access to both the alert stream and standard data release products (see section 4)

1.4 Related documents

This UK-DAC document sets out an overview of the current plans for how the LSST transient system will work, and more generally how variability information can be derived. It relates to several other UK DAC Phase A documents:

- LUSC-A-03 summarises the QUB experience in processing the PanSTARRS and ATLAS alert stream
- LUSC-A-08 Describes the design of the Lasair event broker
- LUSC-A-09 analyses time domain science requirements and their implications
- The LSST:UK Science Requirements Document

The key related LSST project documents are:

- LSST Overview, Ivezic et al 2019 Ap.J. 873, 111.
- Observing Strategy white paper, arXiv:1708.04058
- LSE-163 Data Products Definition Document
- LSE-231 LSST Data Products categories
- LDM-151 Data Management Science Pipelines Design
- LPM-17 LSST System Science Requirements Document
- LSE-612 Plans and Policy for LSST Alert Distribution
- DMTN-102 LSST Alerts- Key numbers

2. LSST observing pattern

As of June 2019, the Observing Strategy, including main survey cadence, and the nature and number of special surveys, is still under active debate, involving both the project and the community. (See the Observing Strategy White Paper). What we describe here is the *baseline design*, and mostly refers to the main survey.

The *Main Survey* will cover 18,000 sq.deg and take 90% of the observing time. *Special Surveys* take up the remaining 10%. There will be at least four and likely more extragalactic "deep drilling fields" but also other specialised surveys.

Each *exposure* is 15 seconds and covers 9.6 sq.deg. Each *visit* has two such 15 second exposures back-to-back. Each visit will be in a single filter.

The *revisit* distribution in time is not yet fixed. However, it is expected that there will be a *fast revisit* within 15-60 minutes. This is driven by the need to separate Main Belt Asteroids from Near Earth Asteroids, but could also be important for some transient science. This initial revisit could be in the same filter, or a different one. This seems also to be not yet fixed; there are different scientific pros and cons.

The remaining *revisit distribution* in time and filter is not yet fixed, but the basline assumption is that there will be a *Universal Cadence* that keeps everybody equally happy (or unhappy). As a working assumption, one can assume a uniform distribution in epoch while a given field is observable (roughly half the year), with time divided favouring r and i more often than the other filters. The reality will be more subtle, but that's good enough to scope the system.

Given two visits per night and field size of 9.6 sq.deg., in principle, 20,000 sq.deg. can be painted in every 3 nights, which is pretty much all the sky visible from the LSST latitude at any one time of year. So a given field can be revisited *in some filter or other* every few nights.

The *baseline survey design* covers 18,000 sq.deg. over 10 years with a *median number of visits* $N_{vis} = 825$, distributed as 56,80,184,160,160 visits in u, g, r, i and z respectively. This means that *on average* each field is visited every 9 days in some filter or other. (Remember that two of the visits are the same night). The average time before a repeat in r is 40 days, and the average time before a repeat in u is 130 days.

Note however that there will be *overlaps* between fields, so a minority of the sky will will be seen more often.

The observing pattern for deep drilling fields - exposure time per visit, number of filters per visit, time between revisits - is not yet fixed. However, if 10% of the time goes into special surveys, this would be of the order 70 single-filter-field visits per night with the standard 2×15sec exposures. So for example, if say three deep fields are being monitored at any one time, they could have full six-band coverage at 15 times the standard exposure, every 4 nights.

3. Expected alert and transient rates and timescales

The number of expected alerts is of course very uncertain. The project technical note DMTN-102 estimates the rate of alerts for various astrophysical categories, and their dependence on Galactic latitude, and sets a requirement for the data management system to cope with:

- An average of 10,000 alerts/visit, 10 million/night
- Instantaneous peaks of 40,000 alerts/visit

The expected data packet size is 82kB, so that the alert stream is expected to average 782GB/night, and to accumulate 2.2PB over 10 years.

The expected alerts are dominated by variable stars, followed by moving objects. The number of alerts caused by Supernovae and AGN is far smaller, around 300/visit. As supernovae evolve over months, they will be seen repeatedly in the alert stream. The number of new SNe is probably about 30/day. Even more towards the rare but high value end, there might be a few gravitational wave events per year.

The first challenge is then simply coping with the full alert rate, but the more difficult and interesting challenge is finding the needles in the haystack. Downstream systems have to be able to filter the stream, and classify new alerts as quickly as possible.

Differing kinds of astronomical object also have quite different *timescales*. A gamma-ray burst counterpart may last a fraction of second; a flaring M star may flare and fade within a few hours; a bulge microlensing event may last a few days; a Tidal Disruption Event or a Supernova may rise within days but fade over months. Different science areas will therefore have differing requirements on how fast to *act on alerts*. Because at least some transients will be urgent, the LSST goal is to *issue alerts within 60 seconds*.

It is obviously crucial for each science area to develop its own predictions for rates, magnitude dependence, colour evolution, timescales etc. The definitive statement to date is the 2009 Science Book, but we expect that these predictions will be improved, and the LSST:UK SWGs will want to develop their own viewpoint. These parameters will feed into both the final observing strategy design, and into the design of Brokers. It is also worth noting that LSST is venturing into unknown parameter space. This almost certainly means we should be ready to adjust our systems during the commissioning phase.

4. LSST data processing

LSST data processing is divided into prompt products (nightly), data release products (annual), and user generated products (community software, using project APIs and sitting next to the data). These categories were previously referred to as Levels 1,2,3, and that terminology is still in many documents. The overall concept is illustrated in Fig. 1. The different levels are only loosely coupled. Rather than being distinct products from a single pipeline, they are better seen as distinct pipelines that occasionally interact. After some initial processing at the Summit Facility, the data is passed on the Base Facility at La Serena and on to the Archive Center at NCSA and/or the satellite Archive Center in Lyons, and finally to one or more Data Access Centres (DACs).

Sources and Objects. In LSST terminology, a "source" is a thing found in the processing of a single frame, whereas an "object" is a persisting astronomical entity. This is the inverse of the useage in UK survey astronomy for many years, so watch out for that.

4.1 Prompt processing

The exposure images are combined and corrected to make a single *visit image*. Then:

- The visit image is differenced with a *template* image and processed to find *DIA sources*.
- The DIA Sources are associated with persistent *DIA Objects*.
- The sources and objects accumulate in a *Prompt Products Database*.
- All 5σ DIA source are issued as alerts

The goal is to issue alerts within 60 seconds, although this is still under debate. Within 24 hours, precovery forced photometry for each DIA source is produced, going back 30 days. Note also that past DIA Sources are kept for 12 months. The accumulating prompt database will be queryable; this may sometimes be important for transient classification and filtering. The visit images and difference images will be available within 24 hours. Cut out images will be provided with the alert events. The alerts will also come with a *real/bogus score* to assist in filtering out junk.

4.2 Data Release processing

Further processing is carried out to make annual data releases.

- The visit images are calibrated and processed to make *Visit source catalogs*
- The visit images are co-added per-band and processed to make *Co-add source catalogs*
- Sources are merged to make *Objects*
- Objects are *characterised* in various standard ways, such as proper motions, galaxy profile fits, and some simple variability characterisation

The visit images and their source tables will be permanently available, as will annual deep co-adds. All sorts of intermediate co-adds will be produced but not kept, as they can be re-created if required. There is no commitment for example to monthly stacks for making deep light curves, and *light curves* are not a standard data product.

4.3 User generated software and data products

From the UK point of view, this will mean processing using resources at the planned UK DAC. From the transients and variables point of view, it could include:

- Our planned UK alert broker
- Classifying, filtering, and feeding through to follow-up projects such as PESSTO and 4MOST
- Second stage filtering and triggering use both the alerts and the data release products - a "variability broker"

5. Alert data flow

Information about each DIA source/alert is packaged as a standard data packet and output to chosen brokers as an event stream using a standard protocol. This might be VO Event/VTP but will more likely be AVRO/Kafka. Filtered streams emanating from brokers for public consumption are more likely to be VO Event/VTP. The data fields included in each packet are still being determined, but there will certainly be a unique identifier, and two associated postage stamp images, from the difference image and the template image.

5.1 The Primary Stream

LSST plans to provide the primary stream to a small number ($\sim$ 4-7) of third party brokers. Selection of these brokers is through a relatively formal proposal process that is currently underway. The project may also provide a *mini-broker* directly to end-users - a very simple filtering service, with some kind of SQL-like language.

5.2 Brokers

A *Broker* just means a software service that intermediates between the user and the data, in just the same way that consumers deal with Insurance Brokers rather than directly with Insurance Companies. These will be run by Third Parties that have a formal agreement with LSST. (There could be independent tertiary brokers who do whatever they like with the streams emerging from these secondary Brokers). Prototype event brokers are already in performance, mostly using the public ZTF stream. Brokers could fulfill several functions. Not every broker will do all of these.

- Passing on the events to astronomical users
- Providing some kind of user interface (either app-like, or scriptable).
- Classifying, filtering, and cross-matching the event stream.
- Providing an automated interface to follow-up facilities.
- Ingesting the events into a queryable database (including the added value information from classifying and cross-matching etc).

5.3 Processing capability

To satisfy all of typical users, power users, and external projects, the UK broker will want to offer more than just a queryable interface. This is likely to mean:

(i) The ability to filter and re-broadcast the stream in real time.

(ii) A scriptable interface including libraries of routines for accessing the data, and backed up by significant processing power and storage available to users. This is certain to mean a Jupyter notebook hub, but may well go beyond this.

(iii) Access through standard protocols, such as TAP and SIA, so that users can process data wit h external tools such as Topcat, Aladin, and DS9.

(iv) An API to allow power-users and external projects to use a customized version of our software stack for their own purposes.

6. More general variability processing

Much variability science will not be achievable through the alert stream alone, and so will require the data release products. However, as noted above, the standard data products will not include intermediate stacks, or light curves. Note that use of intermediate stacks is not just about fainter variables, but also about improved accuracy, for example to spot trends. Note also that while the Visit Images are available within 24 hours, they are not calibrated until the annual data release process. To cater for UK science needs, the DAC should ideally provide:

- The ability to produce light curves for a given ObjectID, from the Visit Image catalogs.
- The ability to make intermediate stacks on demand
- The ability to calibrate real-time Visit Images
- Libraries of routines to help automated massive data mining, such as hunting for slow-risers, or making outburst-forecasts

Acknowledgments

We are grateful to Eric Bellm for various clarifications. Several other LSST:UK members have indirectly contributed in many conversations, especially Roy Williams.

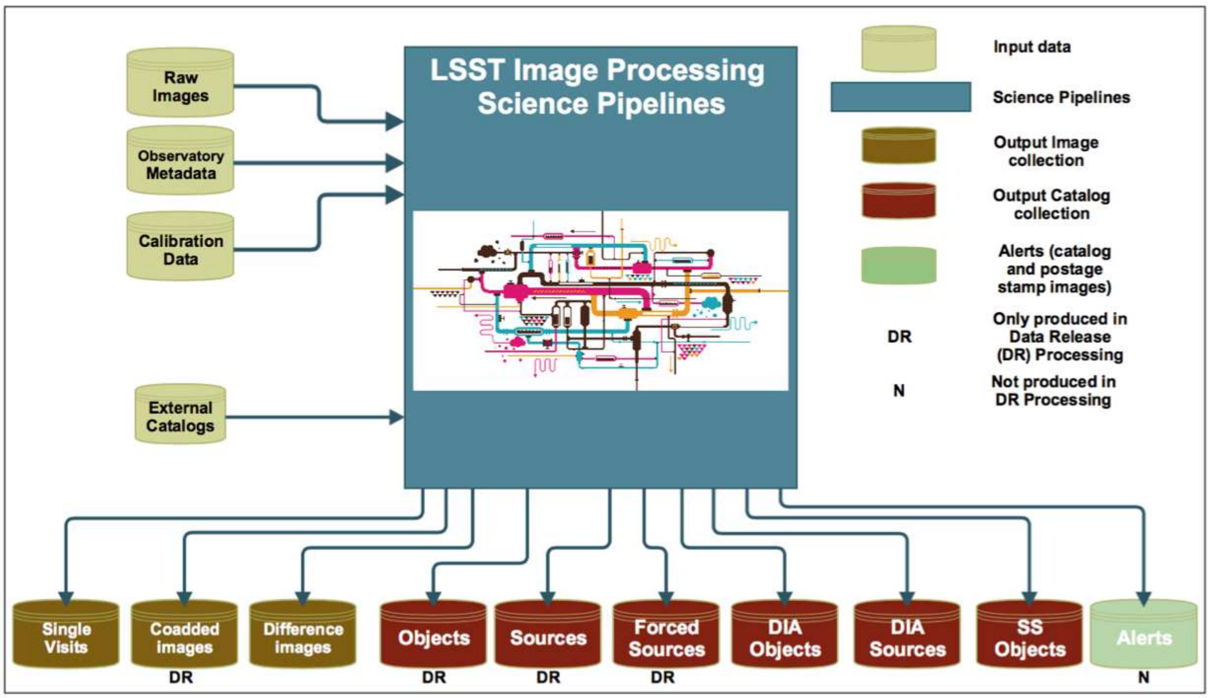

Figure 1. Overview of LSST data flow and products, taken the Data Products Definition document.