LSST:UK Consortium LSST:UK Science Centre

Top Level Phase C Plan for Lasair-LSST

WP3.5(S4)Long Term Plan

Project Acronym	LUSC-C
Project Title	UK Involvement in the Legacy Survey of Space and Time
Document Number	LUSC-C-02

Submission date	29/08/2023
Version	1.0
Status	Final version at start of Phase C
Author(s) inc. institutional affiliation	A.Lawrence, R.Williams, K.Smith, G.Francis, S.Smartt
Reviewer(s)	N/A

Dissemination level	
	Public

Version History

Version	Date	Comments, Changes, Status	Authors, contributors, reviewers
0.1	18/05/23	Initial Draft for comments	Lawrence
0.9	16/08/23	First full draft for team approval	Lawrence, Williams, Smith, Francis, Smartt
1.0	29/8/23	Final version for start of Phase C	Lawrence, Williams, Smith, Francis, Smartt

Top level Phase C plan for Lasair-LSST

Andy Lawrence[1], Stephen Smartt[2], Roy Williams[1], George Beckett[1], Gareth Francis[1], Bob Mann[1], Ken Smith[3], Dave Young[3]

Abstract
We provide a broad-brush top-level plan for the final stages of delivering Lasair, the UK broker for LSST-alerts, as part of Phase C of the LSST:UK project. With a complete working Lasair-ZTF prototype in place, the main purpose of Phase C work will be to port Lasair to Rubin/LSST, optimise performance, put Lasair into operation, complete work with DEV workpackages, and continue to both maintain and further develop the system. In this document, we set out the background, goals, timeline, and development and maintenance methodology.

[1] *University of Edinburgh, United Kingdom*
[2] *University of Oxford, United Kingdom*
[3] *Queen's University Belfast, United Kingdom*

Contents

1. Introduction

1.1 Context

LSST:UK is building an event broker to receive and process the full Rubin alert stream, and to provide an interface to help astronomers get science done with the alert stream in quasi-real time, as well as the accumulating database of alerts. The Rubin project will provide the full alert stream to a relatively small number of approved "community brokers", of which Lasair is one.

Phase A of the LSST:UK project centred on conceptual design, and initial protoyping, using the alert stream from the Zwicky Transient Facility (ZTF). In Phase B we completed a detailed architecture and technology design, carried out experiments on options for the technology stack, tested and improved robustness and scalability, settled on solutions, and implemented these solutions on the UK e-infrastructure (IRIS). Meanwhile we incrementally improved and added to the scientific functionality and the user interface. Through this process interaction with and feedback from an active science user base was crucial. We released a series of successively improved versions of the working prototype, Lasair-ZTF, culminating in Lasair 4.5, considered to be the final version of the ZTF prototype.

As we begin Phase C, with Rubin operations imminent, our prime task is to port the complete working prototype to the Rubin system. Initially this will be tested on a Rubin provided simulated alert stream, but it needs to hit the ground running as Rubin switches on. In addition to beginning operations, we will continue to enhance and improve functionality, as described in later sections, and complete work with key DEV workpackages, making sure that Lasair dovetails optimally with their own work.

1.2 Motivation for this document

The aim of this document is to provide a top level plan for completing the implementation of Lasair, beginning operations, and debating and designing enhancements. The LSST:UK Phase C proposal sets out the broad goals, deliverables and milestones, but in this document we set out how these will be achieved a little more closely. During Phases A and B we developed a successful methodology of six monthly Cycle Plans, coupled with fortnightly progress meetings, and integrated into overall LSST:UK project planning. This document also serves to pull together in one place a short version of material relevant to Lasair that is spread across several other documents and wiki pages.

This document is not intended to be a detailed design or specification document; these will be found elsewhere. It is not intended for external consumption; it is an internal planning document.

1.3 Related documents

This document should be seen in the context of number of key documents in other areas:
Phase B Technical Reports:

- LUSC-B-31 D2.3.2 Design of infrastructure for LASAIR
- LUSC-B-32 D3.2.2 LASAIR user interface
- LUSC-B-39 D3.2.3 Summary documentation for final Lasair system (Mar 2023)

All available at this Phase B wiki page.

LSST:UK Phase C documents and related web pages:

- LSST:UK Phase C wiki page
- The Lasair-ZTF public prototype
- Lasair Phase C planning page including Lasair Phase C deliverables
- The LSST:UK Science Requirements Document
- The LSST:UK Phase C proposal

Rubin-LSST overview publications:

- LSST Overview, Ivezic et al 2019 Ap.J. 873, 111.
- LSST Data Management System, Juric et al 2015, ArXiV 1512.07914.

Rubin-LSST key technical documents:

Note: The URL link can be made by appending `.lsst.io` to the code, for example `DMTN-102.lsst.io`

- DMTN-102 LSST Alerts- Key numbers
- DMTN-118 Review of Timeseries Features
- LDM-151 Data Management Science Pipelines Design
- LDM-542 Science Platform Design
- LDM-554 Data Management LSST Science Platform Design Requirements
- LSE-612 Plans and Policy for LSST Alert Distribution
- LPM-17 LSST System Science Requirements Document
- LSE-163 Data Products Definition Document
- LSE-231 LSST Data Products categories

The **Phase A Technical Reports** are also still relevant:

- LUSC-A-02 Overview of LSST approach to alerts and variable objects
- LUSC-A-03 Ingestion of transient object detections from PanSTARRS1 and ATLAS
- LUSC-A-08 The Lasair-ZTF Transient Broker
- LUSC-A-09 LSST:UK DAC: an initial analysis of requirements from Transient and Variable Science

All available at this Phase A wiki page.

2. Evolution and status of Lasair

The overall plan was to develop a full working prototype, operating on ZTF data, and then to adapt this to become the final Rubin/LSST version. Within this big picture, we have structured Lasair development around a series of "versions". These are not quite "product releases" as one might find with commercial applications, but more like a series of stages or checkpoints in what is really a continuous development process, usually corresponding to formal project Milestones (MS).

Lasair 1.0 was the initial prototype developed during Phase A of LSST:UK. It demonstrated capability to receive and process alerts from ZTF, and provide them to external users.

Lasair 2.0 = MS3 (March 2020) was an improved, stabilised, and documented version of the ZTF prototype, which also included Version 2 of the Sherlock Classifier. This public version was already being productively used by an external science community.

Lasair 3.0 = MS11 (March 2021) was the first version moved to the new UK e-infrastructure (IRIS) and had a set of functionality improvements such as a Python library based API. This version was released privately to beta-testers.

Lasair 3.5 (Jan 2022) was an internal but extremely important stage, in which the system architecture was redesigned, and key new technologies were deployed, aimed at making a scalable system.

Lasair 4.5 = MS25 (Mar 2023) made the technology changes of version 3.5 public, following extensive testing and improvement of system robustness, but also included a completely overhauled user interface, and an extensive and adapatable documentation system. This is considered the final version of the ZTF prototype, and will be maintained while ZTF operates, but will have no further enhancements applied.

Lasair 5.0 = MS23 (June 2023) was an internal but vital release, demonstrating the processing of simulated LSST alerts - the first key stage in porting the system to LSST, and the first Lasair deliverable in Phase C.

3. Lasair in the LSST:UK Phase C plan

Phase C of the LSST:UK project began in April 2023 (Project month 46, or PM46; note that project months are continuous with Phase B) with a planned conclusion in March 2027 (PM93).

3.1 Phase C Objectives

The LSST:UK Phase C programme has five broad objectives, which we could paraphrase briefly as:

- To be ready for Rubin operations
- To support user access to the data
- To complete in-kind contributions
- To maximise UK science return from LSST
- To sustain the prominence of UK researchers in the LSST Science Collaborations

These are all relevant to Lasair except the third; as a Community Broker, Lasair was not formally eligible as an in kind contribution. However, Lasair is a key part of making some UK in-kind contributions - specifically TiDES and Adler - possible, through the provision of infrastructure, processing, and advice.

3.2 Lasair related workpackages

Formal Lasair work is in a single DEV workpackage, WP3.5, with additional support effort available in a DAC workpackage, WP2.5:

WP3.5 Lasair: The UK's community alert broker
Lead: Stephen Smartt (0.1 FTE)
Key funded staff: Ken Smith (1 FTE), Dave Young (0.3 FTE funded by QUB)

WP2.5 DEV Activity Support
Lead: Andy Lawrence (0.1 FTE)
Key funded staff: Roy Williams (0.6 FTE), Gareth Francis (0.4 FTE)

3.3 Lasair related deliverables

Lasair has six formal deliverables, each of which are successive "Versions". As explained above, rather than "product releases" these are intended to be development stages or checkpoints.

D3.5.1 Lasair Version 5 PM46 Apr 2023
D3.5.2 Lasair Version 6 PM58 Apr 2024
D3.5.3 Lasair Version 6.5 PM63 Sep 2024
D3.5.4 Lasair Version 7 PM70 Apr 2025
D3.5.5 Lasair Version 8 PM82 Apr 2026
D3.5.6 Lasair Version 9 PM93 Mar 2027

We set out an initial plan for what the versions will contain in section 8 while noting that this will likely evolve over the next few cycles. Along with the above deliverables, there are associated DAC-related deliverables to support their implementation, to be done in consultation with IRIS staff:

D2.5.1.1 Hosting Platform for Lasair 6 PM57 Mar 2024
D2.5.1.2 Hosting Platform for Lasair 7 PM69 Mar 2025

4. Lasair goals

WP3.5 sets out top-level objectives, which are:

1. **Maintain Lasair-ZTF as long as necessary**
   The ZTF survey continues to be very useful, providing a base of active users who find problems and make requests. All resulting changes to the Lasair-ZTF codebase are copied to the Lasair-LSST codebase.

2. **Optimise pipeline performance**
   The event rates and volumes from LSST will be a factor of at least 30 over ZTF. This performance measurement and optimisation will improve the ingest pipeline, and identify where more compute power is needed.

3. **Provide a platform for external classifiers, providing support where necessary**
   Lasair does not have its own classifiers; rather it is a platform for astronomers to create their own, and the Lasair team will support and encourage them.

4. **Design and build methods and documentation for synergistic use of Lasair and DAC**
   The LSST:UK project has LSST rapid alerts (Lasair), LSST data releases, and many other catalogue resources. The Lasair team will identify science goals that can especially benefit from these synergies and ensure that our infrastructure supports them..

5. **Continue to improve and enhance Sherlock**
   The Sherlock sky context system is a critical part of Lasair, and the team will add new catalogues – especially the southern sky – as well as optimising Sherlock performance.

6. **Design and build components for multimessenger astronomy**
   Lasair is well-placed to find optical counterparts of gravitational-wave, gamma-ray, and neutrino events. The team will select the most likely LSST alerts, both rapidly and automatically, and enhance the Lasair interface to make these available to astronomers.

In more detail, the goals of Lasair are set by a subset of the LSST:UK Science Requirements Document. We reproduce those here, slightly re-phrased and shortened.

- **R2.01 A searchable database containing all the LSST alerts**
- **R2.02 Light-curves:** assimilate all diaSource alerts in diaObjects: providing interactive webpages (linked to database), plots, ability to select ranges, and to submit user added points.
- **R2.03 Postage stamps:** Provide access for all LSST detections and most recent non-detections, with priority for recent detections. Plus multi-colour images from LSST and multiple other surveys.
- **R2.04 Massive catalogue cross-match:** with star, galaxy, AGN, x-ray, radio catalogues, using Sherlock, and machine learning.
- **R2.05 Crossmatch to all previously known transients**: e.g. supernovae, gamma ray-bursts, TDEs, x-ray and radio burst sources
- **R2.06 A database query platform and user-owned storage**: Storage of queries, watchlists, annotations etc. Access by SQL and API.
- **R2.07 Real time cross match** to multiple other time domain surveys at various wavelengths.
- **R2.08 Spectroscopic and/or photometric redshifts:** and hence absolute mags, from other catalogues
- **R2.09 Classification from light curves:** Combine all the above and 24hr-48hr lightcurve information to probabilistically classify all transients as : supernova – kilonova – GRB – Tidal Disruption Event – AGN – XRB – CV – eruption star – microlens – orphan
- **R2.10 Multi-messenger cross-matching:** Probabilistic association with GW and neutrino sources, based on public LIGO/Virgo/Kagra and Ice Cube information, and potentially other sources.
- **R2.11 Provide a stream of transients to external collaborations:** such as 4MOST and SOXS programmes.
- **R2.12 Provide users with a means to upload a "Watchlist":** of up to 10^6 objects, including adjustable search radius, and the means to trigger on magnitude variations.
- **R2.13 Collect a detailed list of additional user requests:** and implement these where possible.

5. Responsibilities in Lasair-related work

We set individual responsibilities for Lasair using a ARIC model, i.e. specifying who is *Accountable*, i.e. to the Consortium, and to STFC, for seeing that the deliverables are achieved; who is *Responsible*, i.e who will in practice undertake the work; who should be *Informed*, i.e. kept abreast of progress, key changes, etc; and who needs to be *Consulted*, i.e. parties affected by changes in Lasair, who should therefore be made aware of such changes, especially ones they may need to have input into. We allocate these roles in three major areas:

Ongoing Requirements Development:
A: Stephen Smartt, Andy Lawrence
R: Stephen Smartt, Andy Lawrence, Roy Williams
I: Bob Mann
C: LSST:UK consortium (via Project Scientist)

Ongoing Technology Development:
A: Andy Lawrence, Stephen Smartt
R: Roy Williams, Gareth Francis, Ken Smith
I: Bob Mann, George Beckett
C: LSST:UK DAC team

Operations and Maintenance:
A: Andy Lawrence, Stephen Smartt
R: Roy Williams, Gareth Francis, Ken Smith, Mark Holliman
I: Bob Mann, George Beckett
C: LSST:UK DAC team

6. Development methodology

The development methodology is unchanged from Phase B, but we summarise it here for completeness. The aim is to keep a balance between long term planning and consistency and short term flexibility and agility. We can think of our development process as divided into timescales.

Four Year Plan. Over the whole lifetime of the project, we will be guided by this document, which in turn is in part an extract and summary of the overall LSST:UK Phase C plan. We will keep in mind the overall goals and deliverables, which will be considered fixed unless there is a good reason to evolve them. The architecture, design, and core functionality of Lasair is now considered complete, and is summarised at Section 7. The overall task of Phase C is therefore preparing for early operations, and developing functionality enhancements.

Six Month Planning Cycles. Every six months we take stock of progress, and take note of any deliverables and milestones due in the coming six months. We then debate and agree goals and tasks to guide our work over the coming six months. We also debate and decide any key *decision points* required in this six month window - for example technology choices which would be hard to row back from later. In the past we have often needed to plan *experiments or tests* to help us make key decisions; most of this is complete, but may occasionally be needed again. Finally, we will make sure that any tasks and goals planned are attached to specific people, and achievable given their available effort during the cycle.

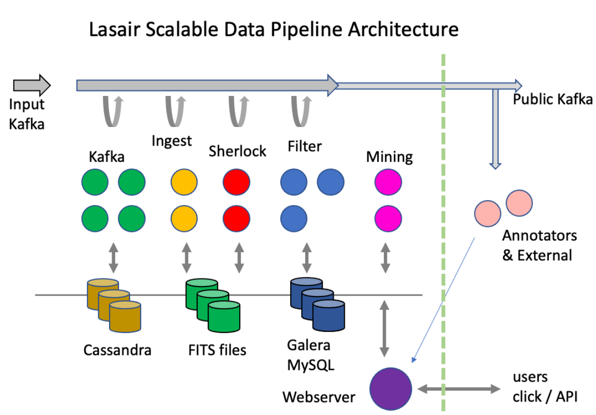
The dates of the planning cycles are included in the schedule section below, and are phased to match key Phase C dates.

Fortnightly Progress Meetings. We will also continue the method we started during Phase A of having regular but short progress meetings. These meetings keep momentum, allow us to react flexibly to developments, help to share and solve problems that come up, and make sure that we do not drift during each planning cycle, by having standing agenda items which help us to keep focus and track progress.

Formal Progress Tracking. We also follow the practices established across the LSST:UK project, by setting up issue tickets in the Zenhub system, and creating "epics" to match our six monthly Cycle planning.

In addition, we rely on some travel between Belfast and Edinburgh to enable in-person technical meetings, on a roughly bi-monthly basis.

7. Summary of Lasair Design



Lasair ingests data with a pipeline of clusters: each cluster does a different job, some more compute/data intensive than others, so it is difficult to know a priori how much resource should be allocated to each. Our design gives flexibility: each cluster can be grown or reduced according to need. Also, there are various persistent data stores, each of which is driven by a resilient cluster that can be grown or reduced according to need. The diagram shows the concept: data enters the Kafka system on the left and progresses to the right. The green cluster reads, processes, and puts different data into the Kafka bus; as soon as that starts the yellow cluster pulls and pushes; eventually the whole pipeline is working. The clusters may also be reading and writing the data stores. We also include the web and annotator nodes in this picture (bottom and right), as well as the mining nodes, although they are not part of the data ingestion pipeline. The web server nodes support users by delivering web pages and responding to API requests. The annotator/classifier nodes may be far from the Lasair computing centre and not controlled by us, but they are in this picture because just like the others, they push data into the data storage and may read from Kafka.

The Kafka system is represented by the green nodes in the diagram as well as the grey arrow at the top. It is responsible for reading and caching the alert packets from the USA, as well as sending them to the compute nodes and receiving their resulting packets.

The Ingest nodes read the original alerts from the Kafka system, and puts the cutout images into the shared filesystem, and the recent lightcurve into the NoSQL (Cassandra) database. It then reformats the alert as JSON – since there is no binary content – then pushes that into the Kafka system.

Each Sherlock node has a SQL database of 5 Tbytes of astronomical sources from 40 catalogues. The sky position of the input alert is used to intelligently decide on the most likely associated source from the catalogues, finding out, for example, if the alert is associated with a known galaxy, or if the alert is a flare from a known CV (cataclysmic variable).

Each filter node computes features of the 30-day light curve that comes with the alert (a year with LSST), as well as matching the alert against user-made watchlists and areas. Records are writen to a local SQL database onboard the node for the object and features, the Sherlock data, the watchlist and area tags. Other tables have already been copied into the local database from the main SQL database. After a batch of perhaps 10,000 alerts are ingested to the local database, it can now execute the user-made queries and push out results via the public Kafka system – or via email if the user has chosen this option.

8. Lasair Version Plan

Based on the Phase C proposal, this is our current expectation of what we will deliver for each Lasair Version in Phase C:

D3.5.1 = Lasair 5 (PM46=Apr 2023) Phase B work completes with a beta release of complete Lasair running with full functionality on LSST simulated or commissioning data.

At the time writing D3.5.1=Lasair 5 has been completed and is under review.

D3.5.2 = Lasair 6 (PM58=Apr 2024) Processing simulated data from science operations. Inclusion of spectroscopic and photometric redshift catalogues. Classification algorithms running in real time, tutorials, documentation, software standards, forced photometry provided, interfaced with 4MOST team.

D3.5.3 = Lasair 6.5 (PM63=Sep 2024) Performance optimisation achieved. Review of first 6 months' operations, integration of other optical survey data (ATLAS, ZTF). Synergies of Lasair with the UK Rubin Science Platform.

D3.5.4 = Lasair 7 (PM70=Apr 2025) Real time processing and services for the LSST alert stream. Inclusion of gravitational wave skymaps – for the LIGO-Virgo-Kagra O5 run (projected Oct 2025). Finalised information flow to and from 4MOST spectroscopic survey.

D3.5.5=Lasair8 (PM82=Apr 2026) Inclusion of LSST DR2 into Sherlock classification (including photo-z values), additions of major new spectroscopic survey data for classification. Test and verification of the 4MOST information flow.

D3.5.6=Lasair 9 (PM93=Mar 2027) Hardened version ready for 8+ years of operations, based on operations during first full year of the final Lasair Version 8, with an LSST all-sky catalogue, updated spectroscopic redshift source catalogue, integration of 4MOST spectra in real time, coordination with O5 gravitational wave sky maps and complete documentation.

9. Key dates and schedule

These are the key dates as we currently understand them, following the Aug 2023 PCW:

```
PM46 Apr 23     C8 start; Phase C begins
PM46 Apr 23     L5.0 release = D3.5.1

PM52 Oct 23     C9 start
PM53 Nov 53     sim alerts flowing from USDF
PM57 Mar 24     hosting platform for L6 = D2.5.1.1

PM58 Apr 24     C10 start
PM58 Apr 24     L6.0 release = D3.5.2
PM62 Aug 24     First photon
PM63 Sep 24     L6.5 release = D3.5.3

PM64 Oct 24     C11 start
PM64 Oct 24     Data Preview 2 (LSST Cam) publ
PM69 Mar 25     hosting platform for L7 = D2.5

PM70 Apr 25     C12 start
PM70 Apr 25     DP1 first real data products
PM70 Apr 25     L7.0 release = D3.5.4
PM74 Aug 25     Scheduled data taking starts

PM76 Oct 25     C13 start
PM75 Mar 26     DP2 useful data products

PM82 Apr 26     C14 start
PM82 Apr 26     L8.0 release = D3.5.5

PM88 Oct 26     C15 start
PM88 Oct 26     DR1 based on first six months
PM93 Mar 27     final month
PM93 Mar 27     L9.0 release = D3.5.6

PM94 Apr 27     Phase D begins
```

Acknowledgments

The authors of this document would like to thank all other members of the group working on building Lasair, as well as the relevant scientists across the LSST:UK consortium.