

Update on LSST:UK DAC Team Activities and Plans

Thursday 24th January 2019

George Beckett (george.beckett@ed.ac.uk) – LSST:UK Project Manager /
Technical Lead

Overview

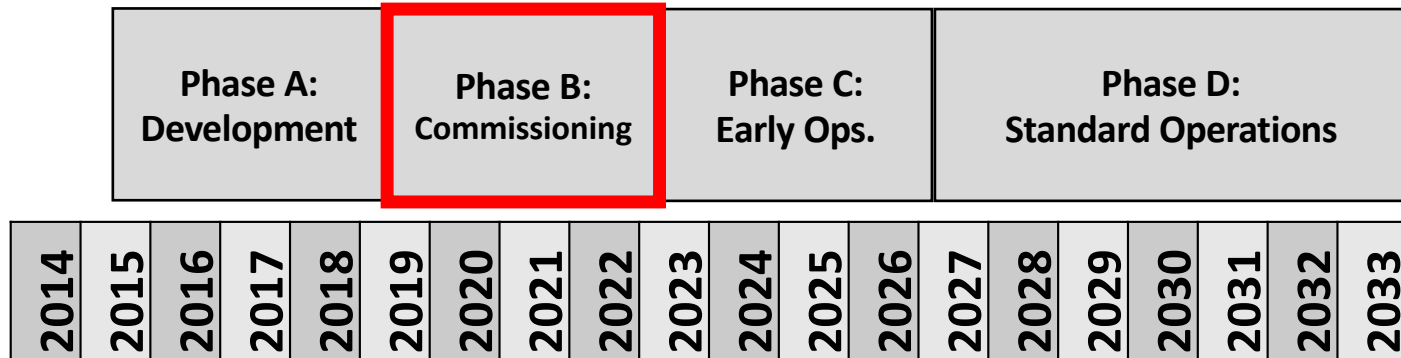
- Brief introduction to the UK Consortium
- Funding and science priorities
- UK DAC resources
- Topical activities
- Future objectives
- Some questions/ ideas for discussion

LSST:UK Consortium

- Representation from every (35) UK astronomy groups
- Coordination of LSST Data Rights allocations (100 PIs)
- Cooperating on funding generation for joint LSST:UK DAC/ DEV Programme
 - DAC—capabilities and infrastructure for a UK-based DAC
 - DEV—capabilities and technologies for priority User-generated Data Products



LSST:UK Science Centre



- Four-phase programme
 - Phase A (Jul'15—Mar'19) – development
 - Phase B (Apr'19—Mar'23) – commissioning
 - Phase C (Apr'23—Mar'27) – early operations
 - Phase D (Apr'27—Mar'33) – standard operations
- Infrastructure funded separately
 - Allocated from shared PPAN infrastructure, called IRIS

UK DAC Team

- Make-up and size
 - Drawn from Edinburgh Wide-field Astronomy Unit
 - Already operate UKIDSS and VISTA archives
 - Involved in DM for Gaia and Euclid
 - Equiv. 3—4 FTE in LSST:UK funded effort (Phase B)
- Activities/ technologies overlap with, for example:
 - Science platform for Gaia
 - Large-scale compute for Euclid
 - European VO activities (ASTERICS, ESCAPE)
 - HPC and data-intensive research through EPCC (UK supercomputing centre)—ARCHER UK National Service systems and new World-class Data Infrastructure

UK Science Priorities

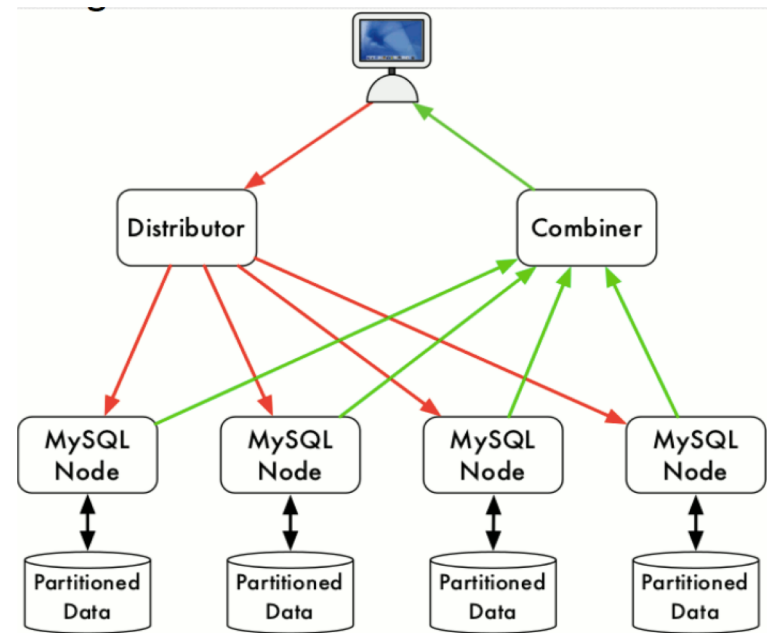
- Interests in all areas of LSST science
 - DESC, Transients, and Galaxies are most active in UK (greatest SC membership)
- Aspects with most direct impact on UK DAC
 - Multi-wavelength analyses (integration w/ other surveys)
 - Transients (broker for alerts)
 - Triggering/ preparing follow-up observations – e.g. transients and multi-object spectroscopy

UK DAC Team Objectives

- During Phase A
 - Understand and prototype technologies for operating UK DAC
 - Develop implementation plan for UK DAC during Phase B
 - Build experience, in community, of LSST tools and services
 - Leverage IVOA influence to ensure standards align to LSST requirements
- Complemented by Phase A development activities
 - Towards UK-relevant User-generated Products
- Aim to run LSST Project software and services as much as possible
 - Tailoring as required to key use cases

Evaluation of Qserv

- Maintain deployment of Qserv on UK testbed
 - Track roadmap and functional developments
 - Engaging w/ Community site re. progress and issues
- Ingest non-LSST data (UKIDSS and SDSS)
 - Grow understanding of ingest process (ready for DR)
 - Optimise for UK environment (e.g. IRIS)
 - Enable experiments w/ multi-wavelength analysis
- Develop benchmark suite of science-realistic queries
 - Assess Qserv functionality vs. UK science requirements
 - Develop strategy for cross-matching other catalogues



Cloud deployment of Qserv


- Containerised deployment, using cloud-like storage
 - Set up of Qserv databases on Ceph cluster
 - Head node and worker node on OpenStack/ Kubernetes
- Automated deployment and catalogue partitioning
 - Rancher/ Magnum for deployment
 - Scripted chunking and ingest
- Ceph experiments
 - CephFS vs Block Storage
 - Ceph hardware optimisations: SSD-hosted journals; RAID vs. replication; interconnect
- Balance performance vs. ease of administration

Lasair—Pilot Broker for LSST Alert Stream

- Operating broker for ZTF
 - Treat as pre-cursor for LSST Prompt Products
- Ingest nightly alerts into relational database using Kafka
 - Filter, annotate, and follow-up
- Publish via web interface
 - Includes standard queries, plus cross-match to other surveys
 - Watch lists and individual alerts
- Pilot analysis platform using Jupyter Notebooks
- Plus, early experiments with stream-based and NoSQL alternatives for DB

lasair.roe.ac.uk

Lasair



[Home](#)

[Query Objects](#)

[Cone Search](#)

[Coverage](#)

[Watchlists](#)

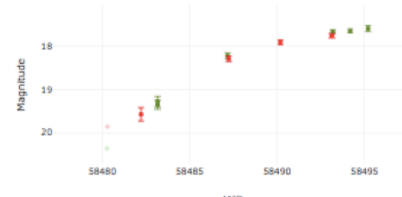
[Jupyter](#)

[Ingestion Status](#)

[Release Notes](#)

[Team and Contact](#)

Object ZTF18adbntwo



- Object has 10 candidates, at mean position:
 - (RA, Dec) = (53.575547, 1.082069)
 - (RA, Dec) = (03:34:18.131, 01:04:55.450)
 - (l, b) = (183.875397, -41.746524)
- Classified as SN at distance 8.67 arcsec.
- The transient is possibly associated with 1237666302167613543/2MASXJ0334; a B=16.04 mag galaxy found in the SDSS/NED/GLADE/2MASS catalogues. It's located 7.72 S, 4.86 E (8.1 Kpc) from the galaxy centre. A host $z=0.048$ implies a transient $M = -17.28$.
- Information on this webpage also [available as JSON](#).
- Conesearch Links (at 5 arcsec): | [Simbad](#) | [NED](#) | [Transient Name Server](#)

Comments


Lasair Bot	Jan. 17, 2019, 3:21 p.m.	In TNS as SN2018les at 0.1 arcsec, discovered 2018-12-30 05:17:04 (MJD 58482.00) by ZTF, ATLAS
------------	--------------------------	--

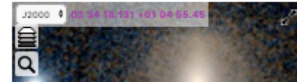
You must be signed in to post comments.

Crossmatches

rank	ID	Catalog	Type	Separation
1	1237666302167613543/2MASXJ0334	SDSS/NED/GLADE/2MASS	galaxy	8.67

AladinLite

Image layer is PanSTARRS DR1; use the layers icon to change it . You can also overlay PanSTARRS and/or Gaia DR2 catalog.



Stream-based approach

- Prototype infrastructure for low-latency event stream processing
 - Inc. support for end-user analysis components and workflows
- Distributed architecture designed to scale to data rate and volume
- Trialling “big data” stream processing technologies (SMACK)
 - Spark
 - Mesos
 - Akka
 - Cassandra
 - Kafka

Jupyter Notebook Platform

- Suite of custom Jupyter notebooks
 - set up with astronomy/ LSST software tools and database access
- Deployed on cloud platform (OpenStack/ Kubernetes)
- Question about AuthN/ AuthZ

```
figtitle += ' = '+objectdict[i]['iau_name']
if 'class' in objectdict[i]:
    figtitle += ' ('+objectdict[i]['class']+)'
plt.title(figtitle)

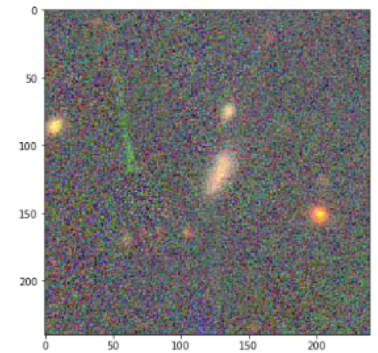
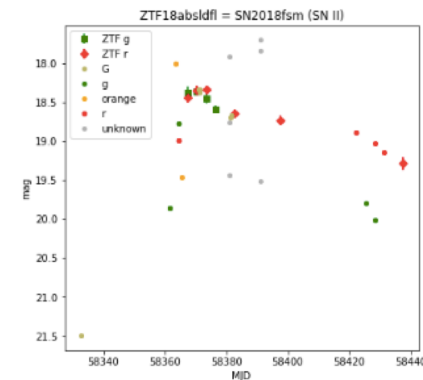
# show cutout
if 'cutout' in objectdict[i]:
    if os.path.exists(objectdict[i]['cutout']):
        plt.subplot(122)
        img = mpimg.imread(objectdict[i]['cutout'])
        plt.imshow(img)

print('https://lasair.roe.ac.uk/object/%s/' % i)
plt.show()

n += 1

print('Done.')
```

<https://lasair.roe.ac.uk/object/ZTF18absldfl/>



DESC Data Wrangling

- Contribution to Data Challenges
 - Develop/ automate simulation workflow—e.g. in DiRAC/ Parsl
 - Optimisation of image simulation software (ImSim)
 - Port DCs to UK resources (WLCG/ IRIS)
- Plan to undertake data-transfer experiments w/ IN2P3
 - Ingest DC2/ DC3 outputs (catalogues) into UK DAC
 - Understand process for receiving Data Release Products

Plans for Phase B

- Deploy DAC (technology preview) for early experiments
 - Possible for Commissioning (assuming UK secures role)
- Achieve LSST Community Broker status
 - Develop Lasair as production service for ZTF
- Finalise infrastructure provision on IRIS
 - Shared-access cloud infrastructure
- Mature plans for community-wide User-generated Products
- Contribute to International DAC Network

Questions

- How best for UK DAC team and LSST DM team to interact?
 - Are UK activities useful/ interesting for DM team?
 - In process of publishing set of Phase A Technical Reports
 - Can UK DAC team help?—e.g. testing
- How can we best track LSST DM Progress?
 - E.g. monthly status reports/ Community discussions
- Could we deploy proto-DAC in UK (for internal use)?
 - How best do access latest DM software outputs
- How do we prepare to ingest a Data Release?
- How should we proceed with AuthN/ AuthZ?
- Is there a sizing model (for DR) that we can see?