UK Ambitions for Astronomy in the LSST Era

George Beckett, University of Edinburgh

7th March 2019

Disclaimer:- I am not an astronomer!

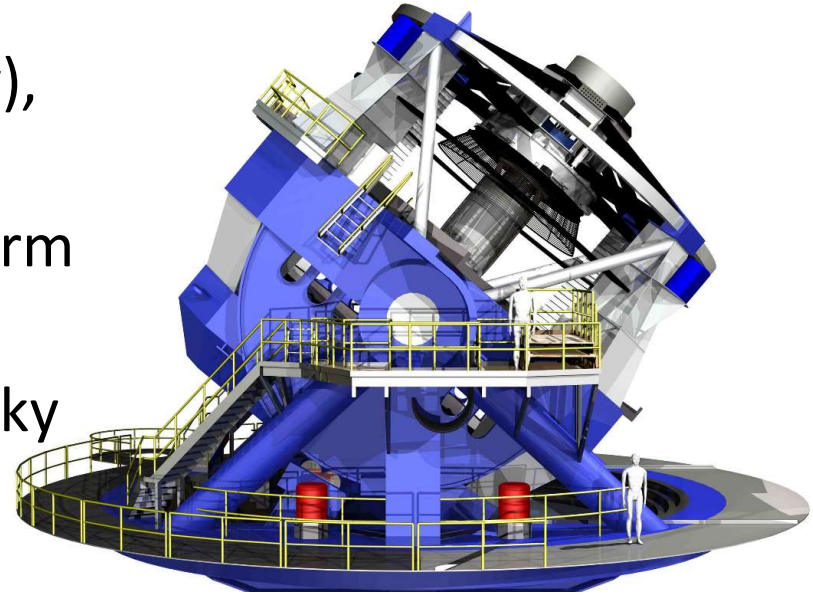
Background in HPC and scientific computing.

Focused on technical challenge of *connecting* astronomers with LSST, maximising their opportunities for new science.

LSST Overview

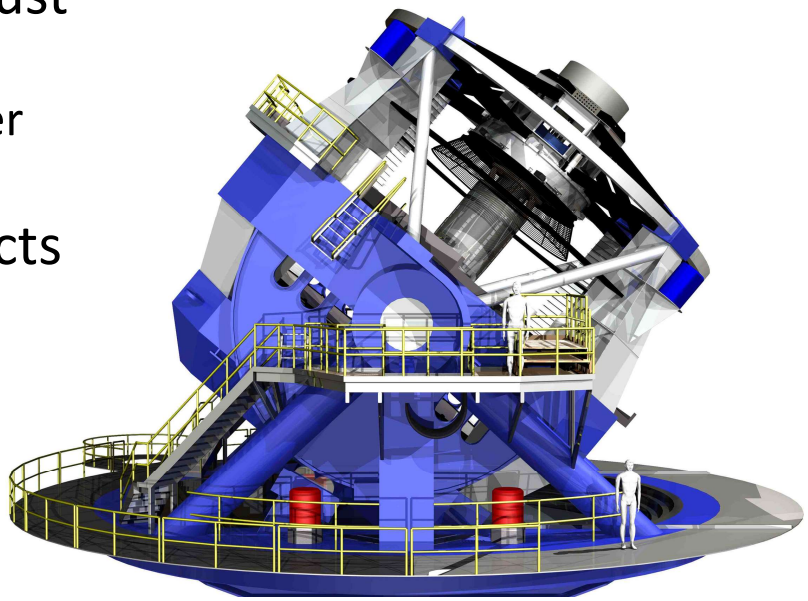
Large Synoptic Survey Telescope (LSST)

- New optical telescope being constructed in northern Chile (Cerro Pachón)
- *Large* – significant width, depth (sensitivity), and speed at which it can image sky
- *Synoptic* — imaging in 6 colour-bands to form overall picture of sky
- *Survey* – creating an atlas of entire visible sky



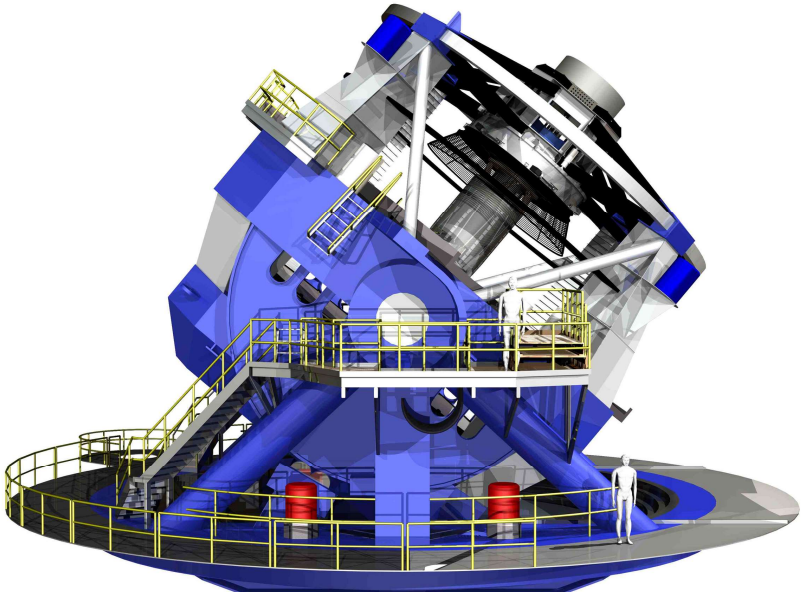
Large Synoptic Survey Telescope (LSST)

- 10-year survey, starting in 2022
- Telescope will image 30,000 sq. deg. of sky in just three nights
  - It will observe each patch more than 800 times over 10 years of survey
- Stacked images help to identify very faint objects
  - Expect 20Bn galaxies and 17Bn stars
- Regular, frequent visits allow creation of time-series for dynamic objects
  - Astronomy of transients
  - Detection of near-earth objects



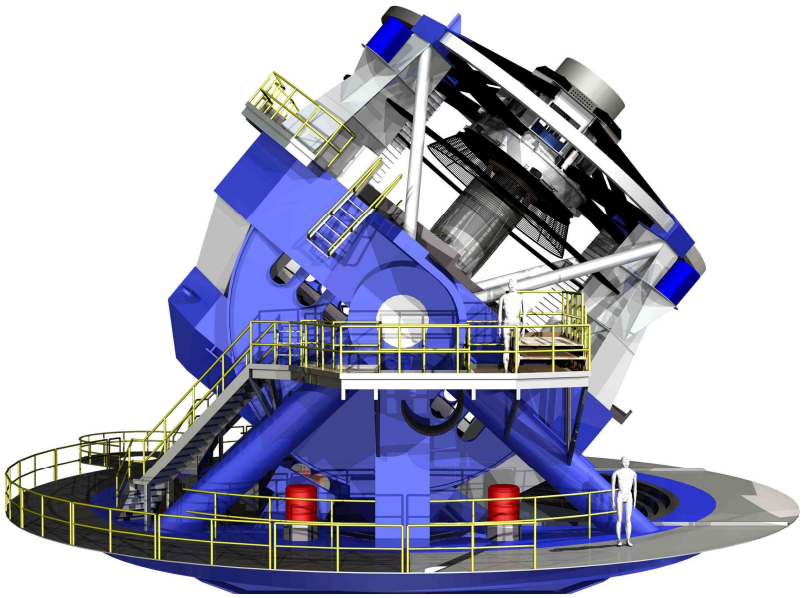
Large Synoptic Survey Telescope (LSST)

- Largest digital camera ever made
  - 3.2 Gigapixels
- Field of view of 9.6 sq. degrees
- Schedule
  - 90% of time for survey
  - 10% for very deep and fast survey, looking at 8 specific fields



Large Synoptic Survey Telescope (LSST)

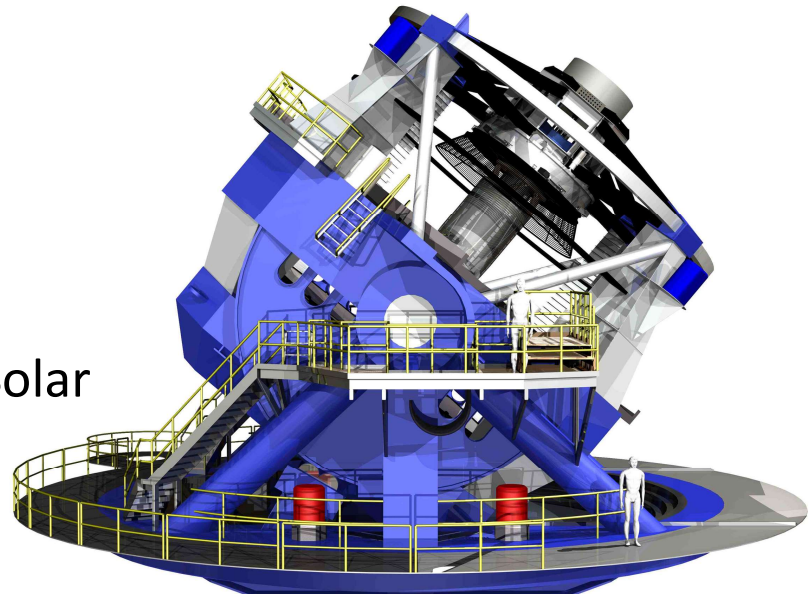
- Costs
  - Construction budget of $640M
  - Operational budget of $370M
- Funding
  - US National Science Foundation
  - US Department of Energy (camera technology)
  - International investment (for $100M of operational costs)



Primary Science Drivers (for Project)

- Highlights
  - Understanding Dark Energy and Dark Matter
  - Creating inventory of solar system
  - Cataloguing transient sky
  - Mapping Milky Way
- International Science Collaborations
  - Galaxies; Stars, Milky Way and Local Volume; Solar System; Dark Energy; Active Galactic Nuclei; Transient and Variable Stars; Strong Lensing; Informatics and Statistics

    https://www.lsstcorporation.org/science-collaborations



Data

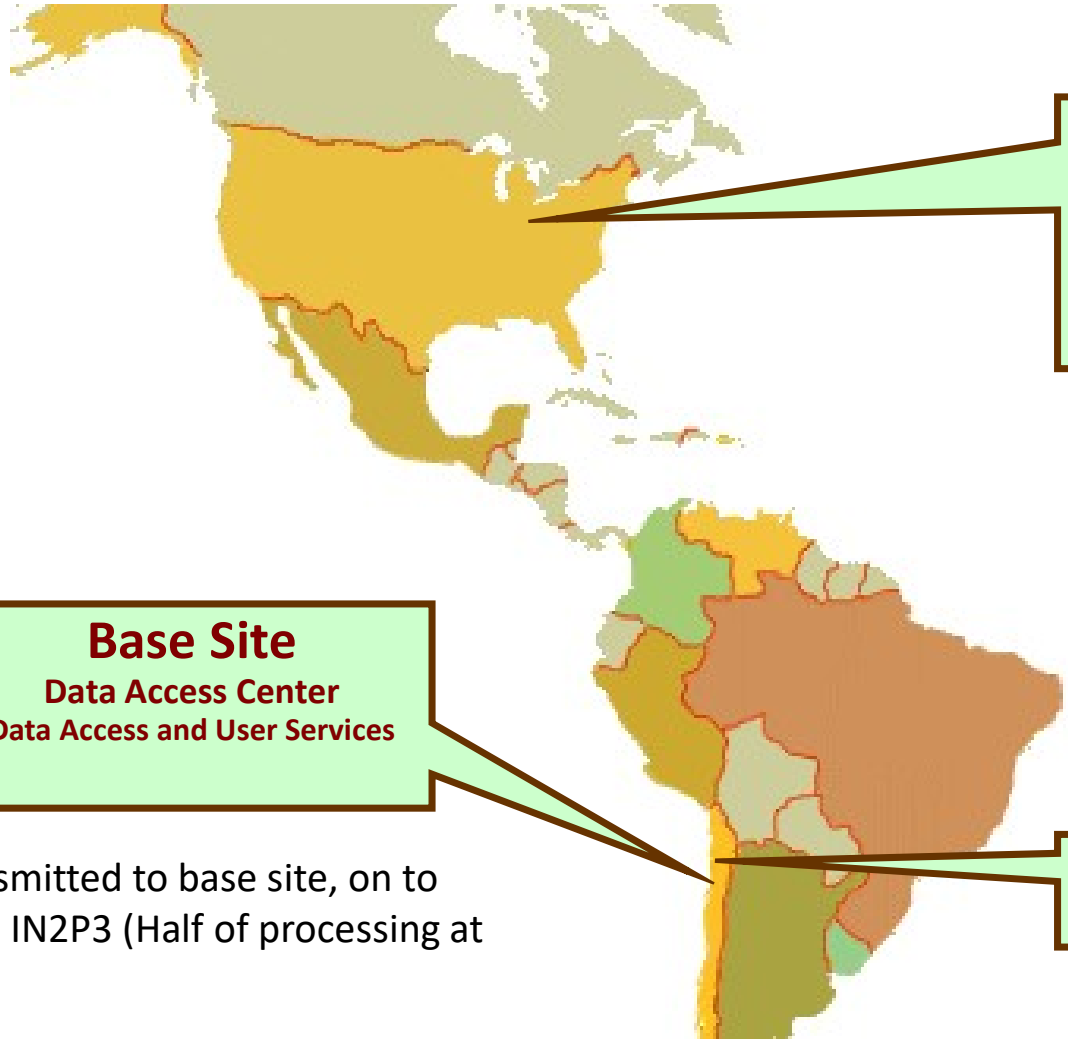
Data Management

- Data access restricted to LSST Data Rights holders, for 2 years
  - US and Chile-based astronomers, plus subscribers from international community
- Data will be hosted in Data Access Centres:
  - in Chile – hosting and local computing
  - in USA (NSCA) – processing, hosting, and local computing
  - in France (IN2P3) – processing and hosting
- UK plan to operate Data Access Centre
  - Hosting and local computing
  - possibly supporting more than UK astronomy

Data Classification

- Three classes of data
  - Prompt Products – difference imaging data w.r.t. reference
    - Released on nightly basis, detail objects that have changed brightness or position
    - ~10^7 events published each night (metadata and postage-stamp image)
  - Data Release Products – image products (reduced and calibrated) and catalogues
    - Released annually as incremental Data Releases
    - D/R grows from 10PB in Yr 1 to 70PB in Yr 10 (uncompressed)
  - User-generated Products – datasets derived from Prompt and D/R Products
    - Not part of survey output: created by community
    - Estimate 10% of computing and storage requirements for User-generated products

Key Locations



Archive Site

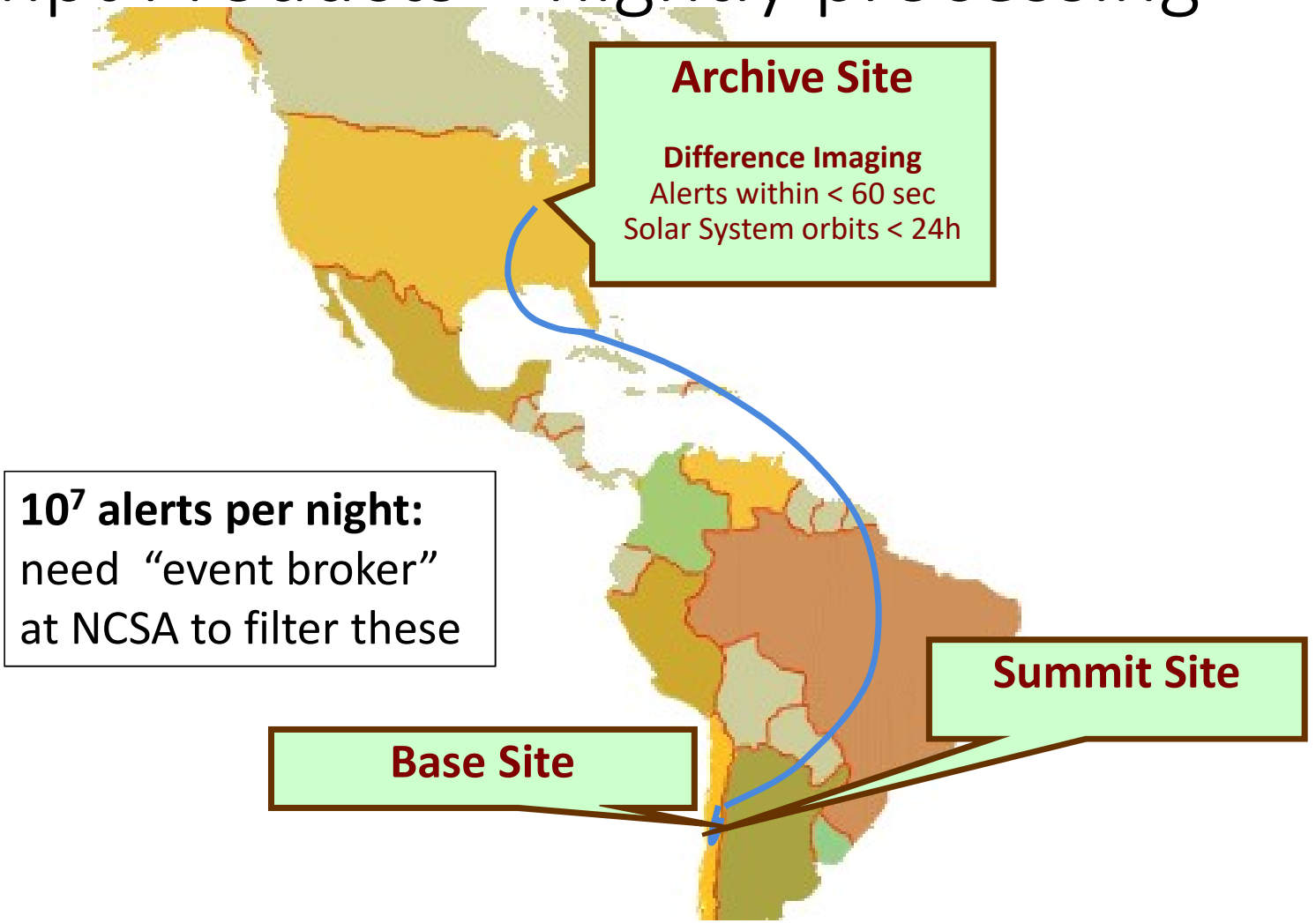
Difference Imaging
Alerts within < 60 sec
Solar System orbits < 24h

Base Site
Data Access Center
Data Access and User Services

Summit Site

Observations transmitted to base site, on to NCSA, and then to IN2P3 (Half of processing at IN2P3 in France)

Prompt Products – nightly processing

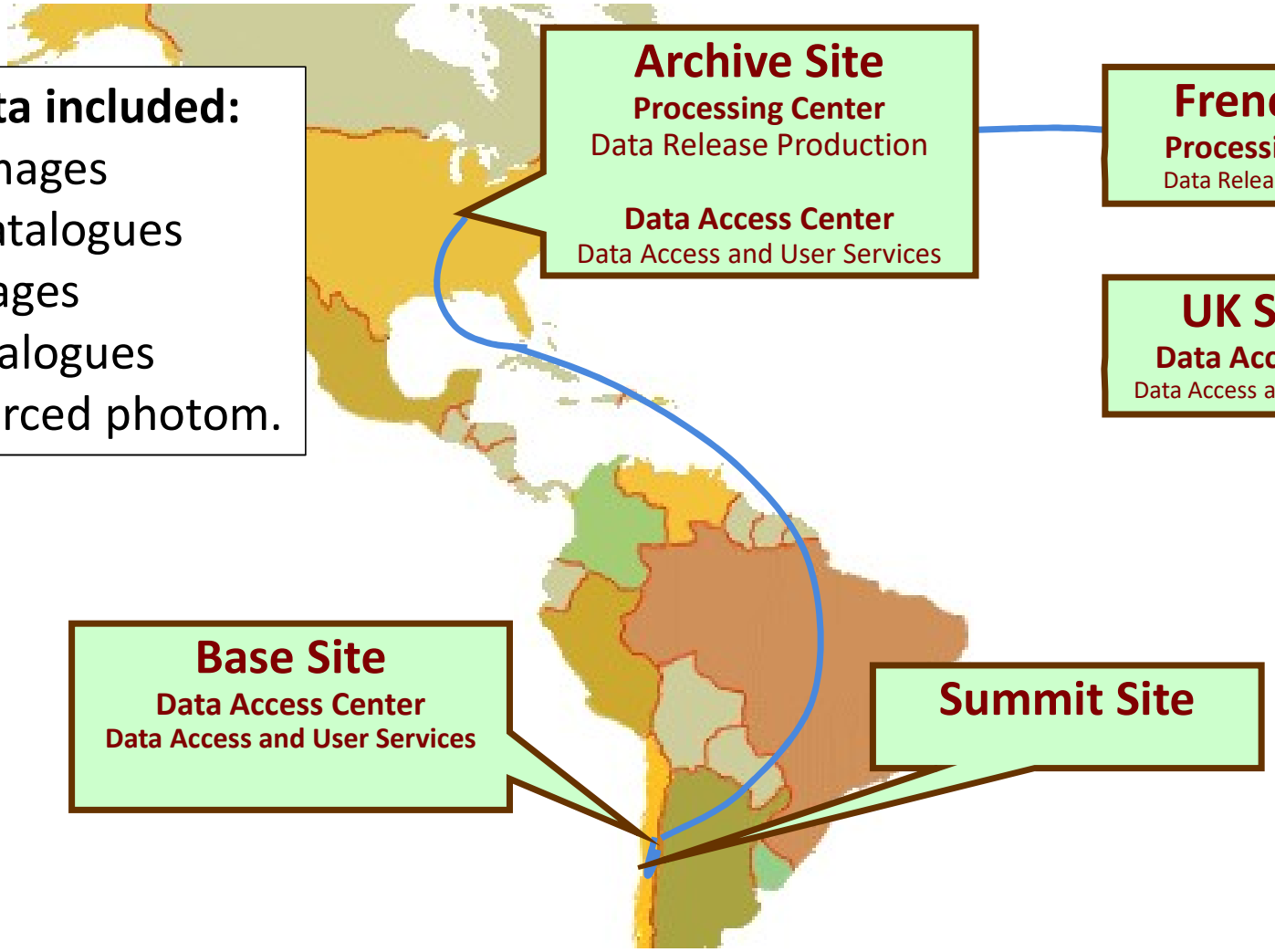


10^7 alerts per night:
need "event broker"
at NCSA to filter these

Data Release Products – annual data release

All extant data included:

- per-visit images
- per-visit catalogues
- co-add images
- co-add catalogues
- Per-visit forced photom.



Archive Site
Processing Center
Data Release Production

Data Access Center
Data Access and User Services

French Site
Processing Center
Data Release Production

UK Site (?)
Data Access Center
Data Access and User Services

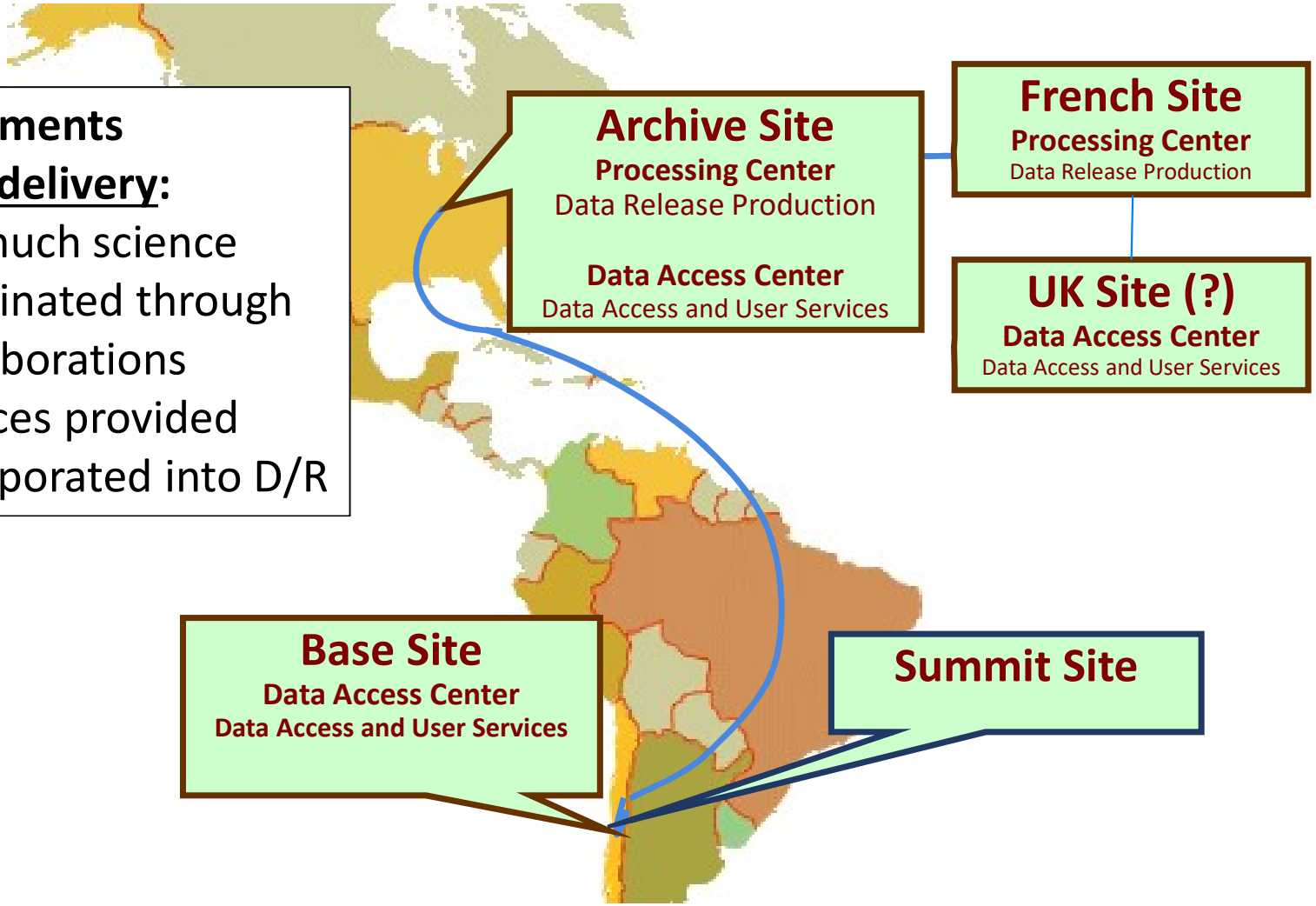
Base Site
Data Access Center
Data Access and User Services

Summit Site

User-generated Products

Beyond requirements of LSST <u>project delivery</u>:

- needed for much science
- mainly coordinated through Science Collaborations
- some resources provided
- may be incorporated into D/R



LSST Data Product Categories

PROJECT

Prompt
Formerly "Level 1" data products

Real Time Difference Image Analysis (DIA)

- A stream of ~10 million time-domain events per night (Alerts), detected, characterized, and transmitted to event distribution networks with 60 seconds of shutter close.
- A catalog of orbits for ~6 million bodies in the Solar System

Data Release
Formerly "Level 2" data products

Reduced single-epoch & deep co-added images, reprocessed DIA products

- A catalog of ~37 billion objects (20bn galaxies, 17bn stars), ~7 trillion observations ("sources"), and ~30 trillion measurements ("forced sources")
- Produced annually and accessible through online databases.

COMMUNITY

User Generated
Formerly "Level 3" data products

User-produced added-value data products

- Deep KBO/NEO, variable star classifications, shear maps, etc
- Enabled by services and computing resources at the LSST Data Access Centers (DACs) and via the LSST Science Platform

LSST Science Platform

LSST:UK

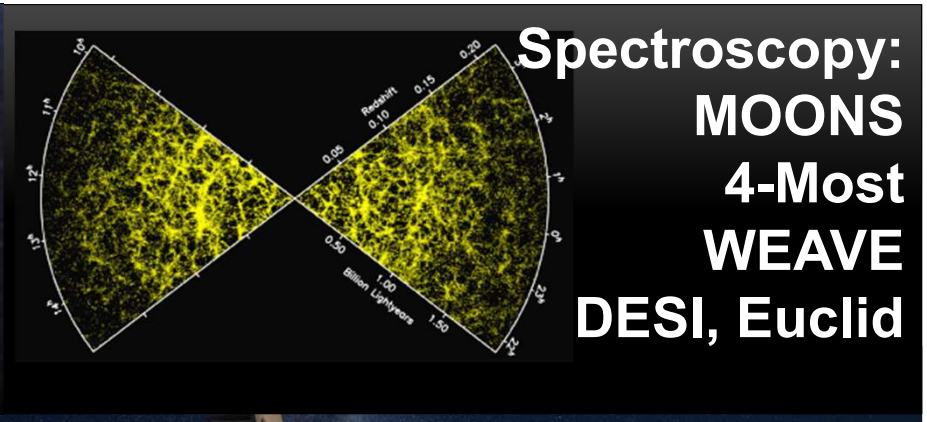
LSST:UK Consortium

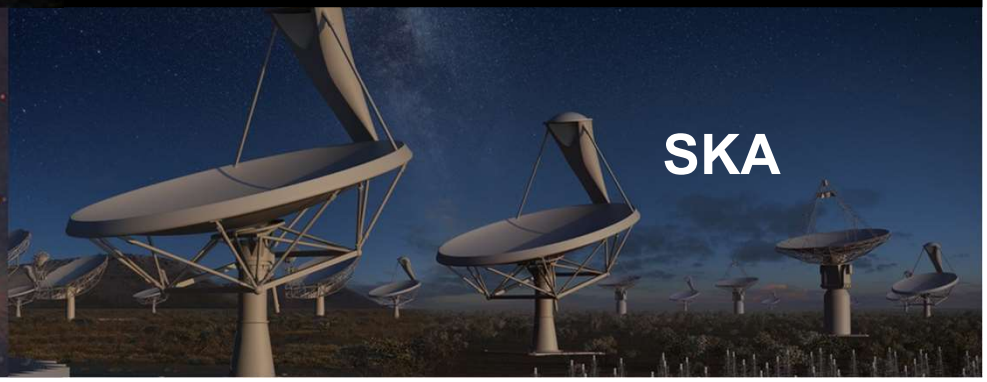
- Representation from every (35) UK astronomy groups

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 LSST Synergies



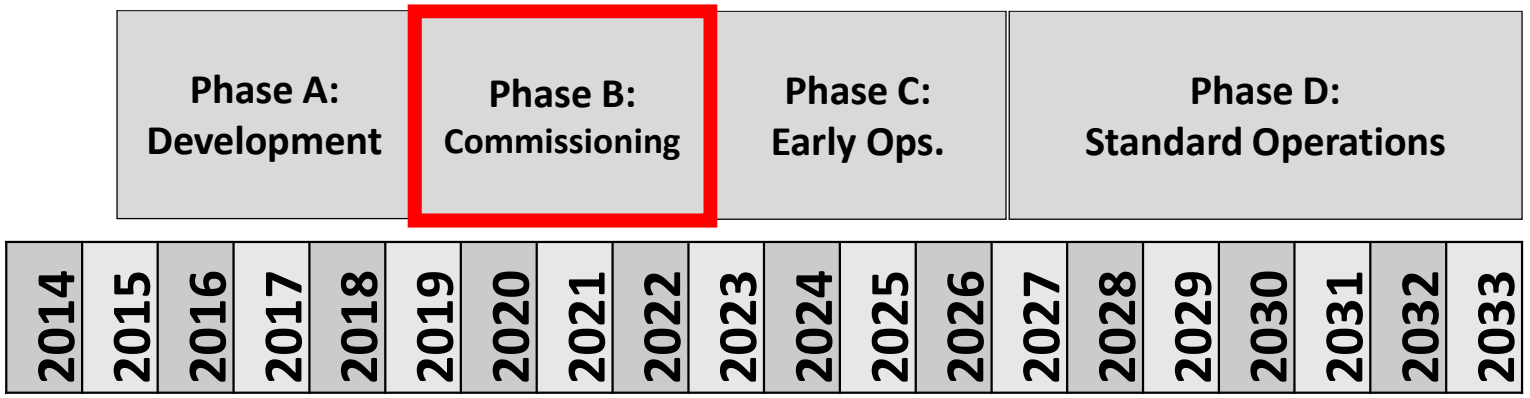


UK has unprecedented opportunity to lead the exploitation of this combined suite of facilities

LSST:UK Consortium

- Coordination of LSST Data Rights allocations
  - Funding secured for 100 Affiliate PIs (plus 400 Junior Associates)
- Cooperating on 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
- Forecast budget of £32M (not including capital for Data Access Centre)
- Infrastructure funded separately
  - Allocated from shared UK Gov't infrastructure, called IRIS

LSST:UK Phase A

- £2.7M for Phase A programme (July 2015—Mar 2019)
- Edinburgh is coordinator:
  - Bob Mann is Project Leader; MGB is project manager and technical lead
- LUSC-DAC: six staff-years (Edinburgh)
  - DAC testbed, Data Challenges, supporting LUSC-DEV
- LUSC-DEV: sixteen staff-years (Man, Cam, QUB, Soton, UCL, Oxf)
  - Weak lensing – analysis of galaxy intrinsic alignment, shape classification
  - Milky Way – star/galaxy separation, tidal stream detection
  - Transients – alert handling, classification
  - Solar System – postage stamps, light curves
  - Sensor characterization – image analysis systematics

UK DAC Team

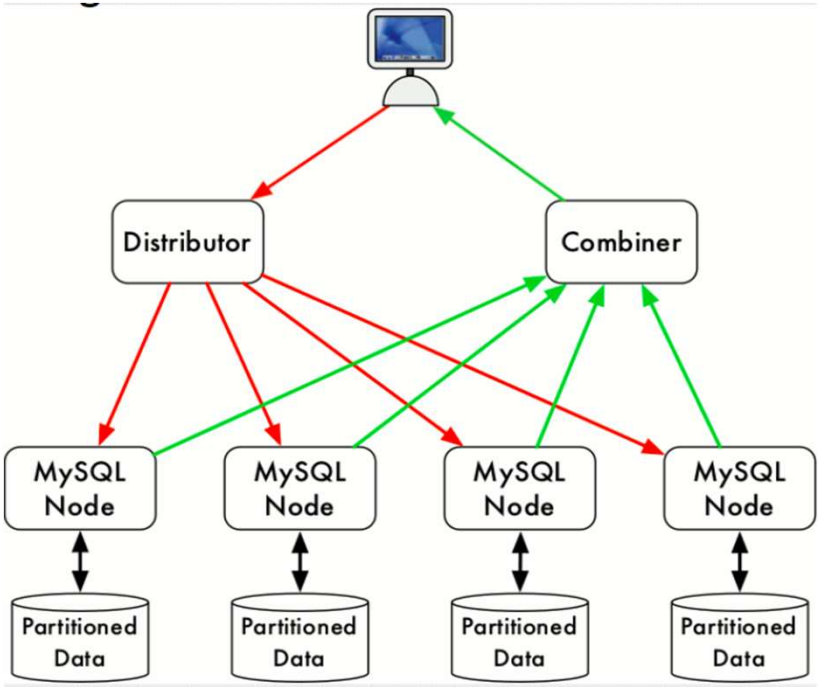
- Make-up and size
  - Drawn from Edinburgh Wide-field Astronomy Unit
    - Already operate UKIDSS and VISTA archives
    - Involved in DM for Gaia and Euclid
- Activities/ technologies overlap with, for example:
  - Science platform for Gaia
  - Large-scale compute for Euclid
  - European VO activities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

UK 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

- Candidate database for LSST catalogues
- 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



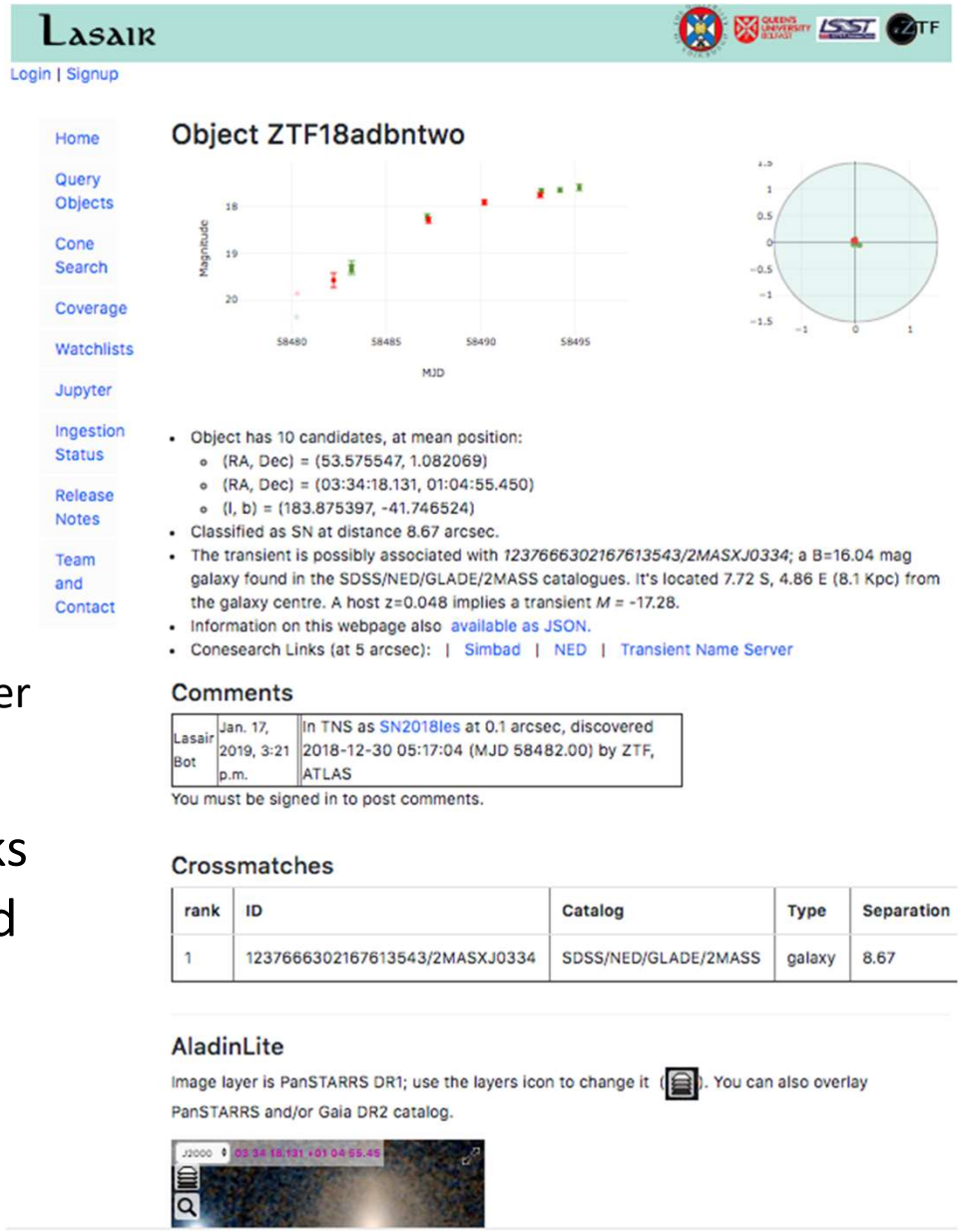
Cloud deployment of Qserv

- 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 Zwicky Transient Facility
  - Treat as pre-cursor for LSST Alert Stream
- 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



Stream-based approach

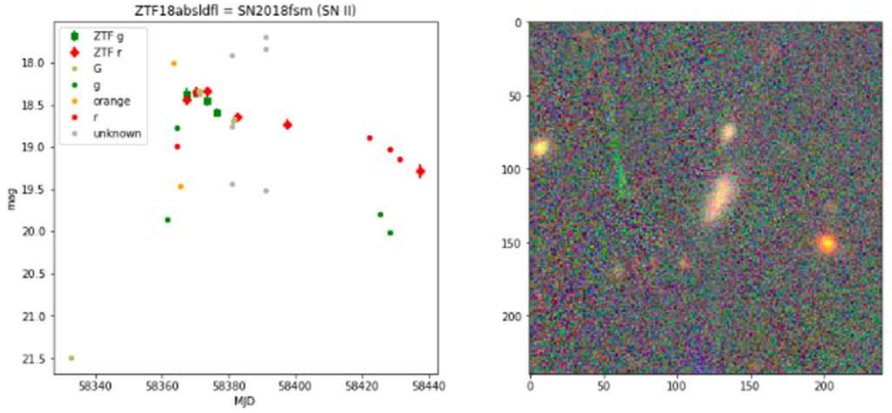
- Prototype infrastructure for low-latency event stream processing
  - Inc. support for end-user analysis components and workflows
- Distributed architecture designed to scale to data rate and volume
- Trialling “big data” stream processing technologies (SMACK)
  - Spark
  - Mesos
  - Akka
  - Cassandra
  - 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)
- Workstation-class resources
  - W/ access to catalogues and image repositories
  - W/ scale out to HPC via batch system

```
        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/
```



Dark Energy Science Collaboration

- Optimisation of LSST simulator
  - Multicore parallelism/ distributed-memory
- Contribution to Data Challenges
  - Develop/ automate simulation workflow—using HPC and Grid (LHC/ WLCG)
- Plan to undertake data-transfer experiments w/ IN2P3
  - Ingest outputs (catalogues) into UK DAC
  - Understand process for receiving Data Release Products

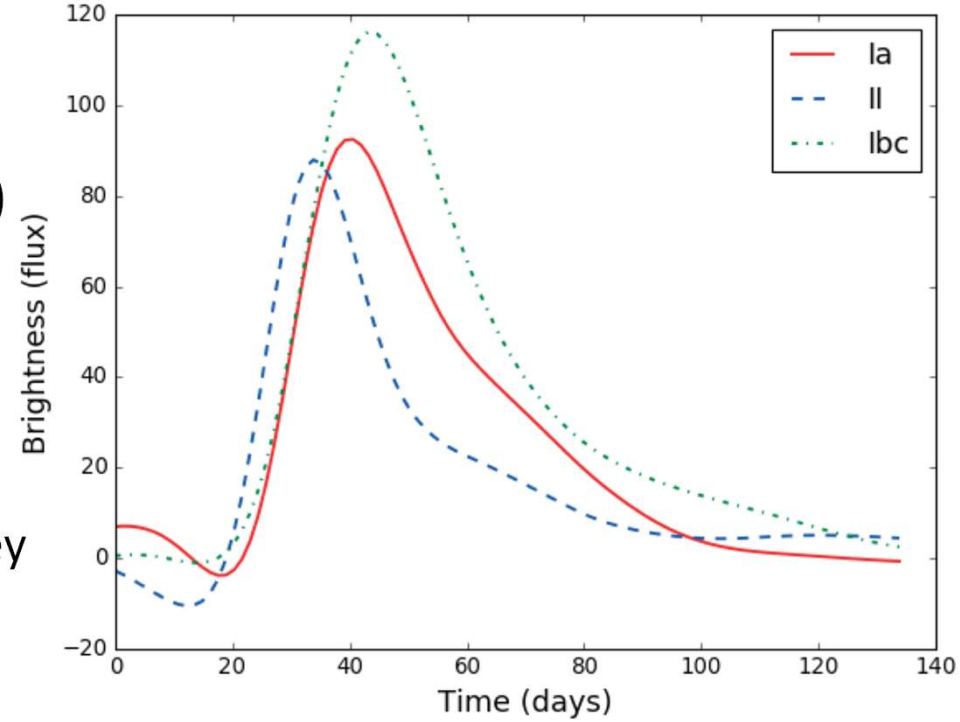
Characterisation of dark energy

- Weak-lensing
  - Statistical technique that determines distribution of dark energy based on observed distortion to distant galaxies
- Galaxy classification pilot
  - Shape determination for 30,000 galaxies, over 3 wave bands, from Dark-energy Survey (30 Terabytes)
  - Embarrassingly parallel – currently running on GridPP, in approx. 80 node-hours
  - LSST implies 1,000×data to analyse

Credit: Joe Zuntz/ Manchester

Classification of Supernovae

- Interest in Type 1a S/N
  - Standard Candles w/ known luminosity
  - Measure expansion of Universe
- Separating Type Ia from others (Ib, II, …) not straightforward
- For LSST, classification based on light-curve alone
- Team ran survey simulation
  - testing how schedule and cadence of survey affects ability to classify S/N
  - Use machine learning to enable large-scale S/N classification



Credit: Natasha Karpenka, Mark Sullivan/ Soton, Michelle Lochner, Hiranya Peiris, Jason McEwen, Ofer Lahav/ UCL

Sensor technologies

- LSST needs very special camera to achieve fast, high-resolution, wide view
- e2v CCD250 is custom CCD designed in UK with non-flatness of <5 µm
  - Designed for maximum resolution
  - High quantum effects

Credit: Ian Shipsey/ University of Oxford

Plans for Phase B

- Deploy DAC (technology preview) for early experiments
  - … and Commissioning
- Achieve LSST Community Broker status
  - Develop Lasair as production service for ZTF
- Finalise infrastructure provision
  - Funded separately from LSST:UK activities
- Mature plans for community-wide User-generated Products
- Contribute to International DAC Network

International DAC Network

- Project set up working group to look at options for network of DACs
  - New opportunities for science with LSST data
  - More computing infrastructure for user-generated analysis
  - Better scope to integrate user-generated products into D/R
  - Greater availability and resilience of DAC services
- Introduces technical challenges
  - Distribution of Data Releases
  - Increased work and complexity
  - Risk to enforcement of data-access policy
- W/G due to make recommendation to LSST in March
  - Watch this space!

Summary

- LSST is most ambitious optical sky survey ever conceived
- Applications from across astronomy, including:
  - Unprecedented opportunity to study transient behaviours
  - Catalogue solar-system objects
  - Explore galaxy expansion and dark universe
- UK astronomy significantly involved
  - Leading astronomy research activities
  - Sensor technology development
- Significant computational-science project
  - Esp. real-time monitoring of event stream and serving up multi-PB catalogues
- UK well-positioned with involvement in complementary telescopes like SKA, Euclid and GAIA

Thank you