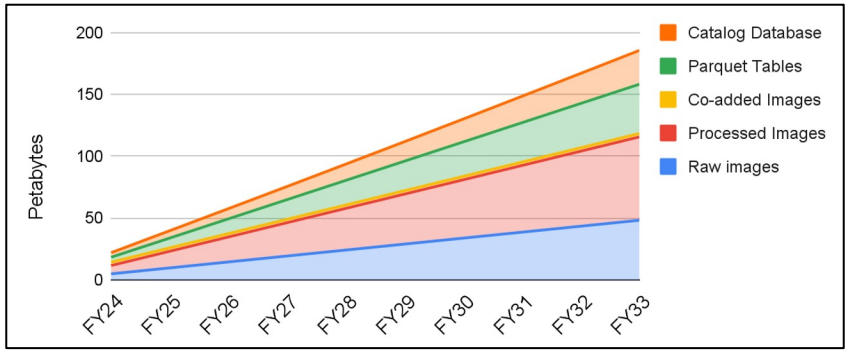
Delivering a UK Independent Data Access Centre for LSST

George Beckett, University of Edinburgh

LSST Data Access Centres

- LSST 10-year survey will produce ~200 PB of data products
  - Annual Data Releases – catalogues, deep and single-visit images
  - Nightly Alert Stream – ~10M alerts per night of transient activity
- Not suitable for traditional approaches to scientific analysis
  - Compute should be co-located with data
  - Specialised tools for handling and analysis
- Astronomer access via Data Access Centres
  - Rubin will operate two centres (in USA and Chile)
  - Sized for core requirements of local communities
  - International community to supplement with complementary resources



Rubin Data Products

Prompt Data Products

Real Time Difference Image Analysis (DIA)

- Stream of ~10 million time-domain events per night (Alerts), transmitted to event distribution networks within 60s of camera readout.
- Images, Object and Source catalogs derived from DIA, and an orbit catalog for ~6 million Solar System bodies within 24h.
- Enables discovery and rapid follow-up of time domain events

Data Release Data Products

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

- Catalogs of ~37 billion objects (20 billion galaxies, 17 billion stars), ~7 trillion sources and ~30 trillion forced source measurements.
- 11 Data Releases, produced ~annually over 10 years of operation
- Accessible via the LSST Science Platform & LSST Data Access Centers.

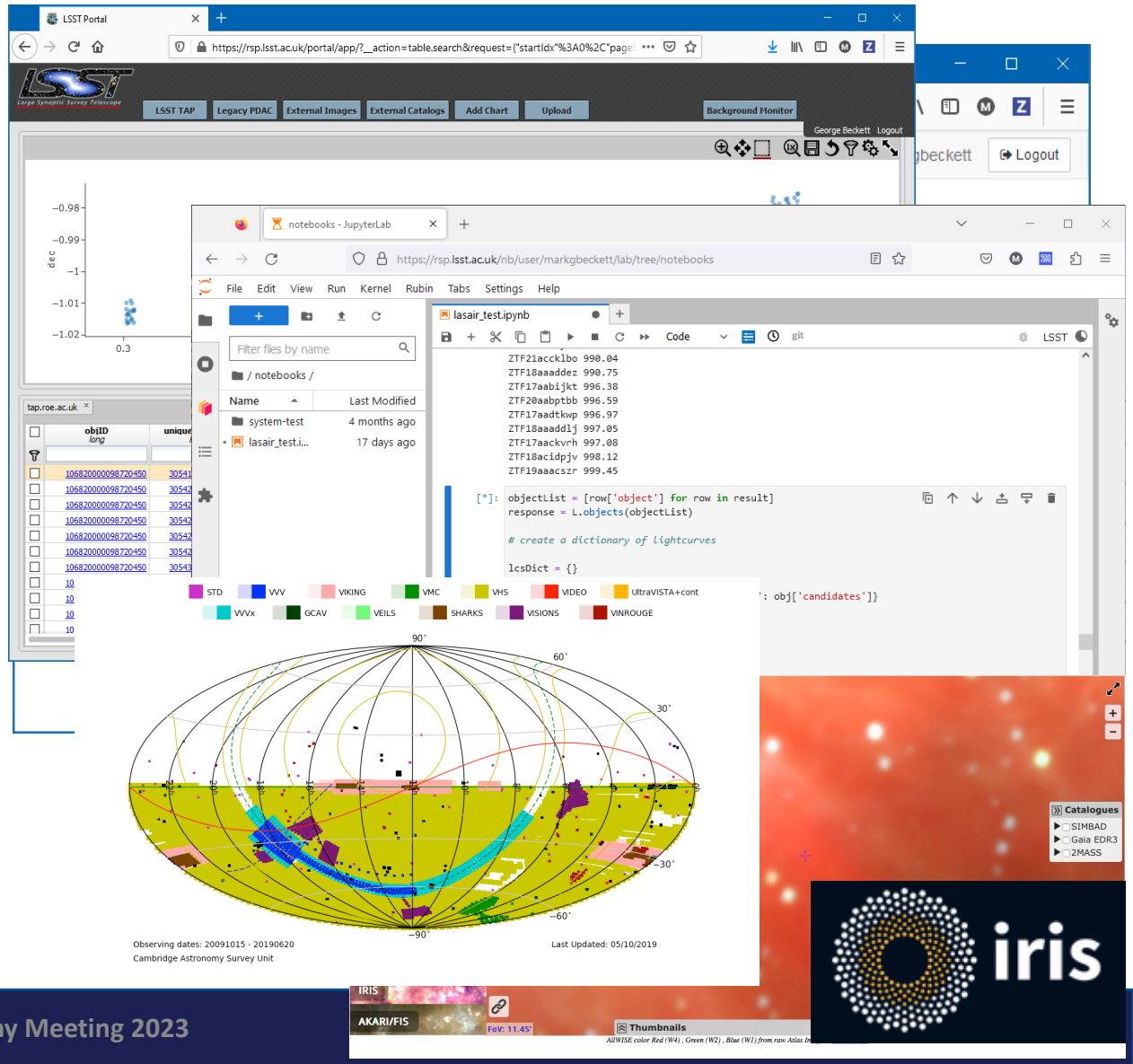
User Generated Data Products

User-produced derived, added-value data products

- Deep KBO/NEO, variable star classifications, shear maps, etc …
- Enabled by services & computing resources at the LSST DACs and via the LSST Science Platform (LSP).
- 10% of LSST computing resources will be allocated for User Generated data product storage & processing.

UK IDAC Plans

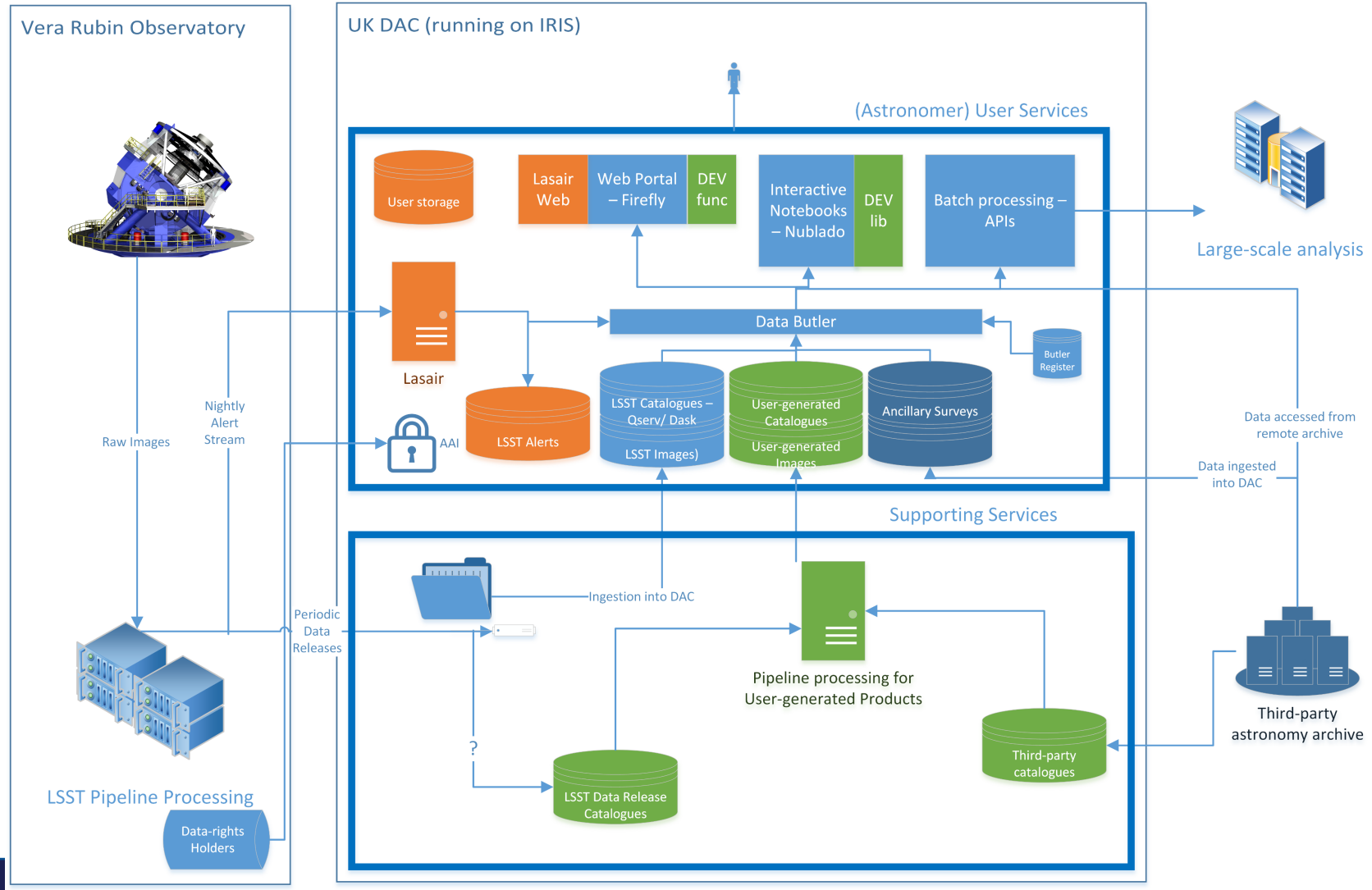
- Operating as a “Full IDAC”
  - Hosts complete LSST Data Releases (current and most-recent previous)
  - Running Rubin Science Platform
  - Co-located w/ UK Community Broker
  - To support full range of science
- Baseline resources for 1,000—1,500 users
  - Option to secure additional capacity via IRIS RSAP



IDAC Network

- In-kind programme includes 12 IDACs and 1 Science Processing Centre
  - Argentina, Australia, Brazil, Canada, Denmark, Slovenia, Spain, Japan, Korea, Mexico, Poland
  - Croatia
- Most expected to offer IDAC Lite services
  - Hosting subset of Data Release (catalogues-only, possibly reduced columns)
  - Potentially specialised – domain-specific expertise, co-located w/ other data
- Science Processing Centres primarily offer compute resources
  - limited storage, so potentially ‘close’ to other IDAC
- Network and user communities interacting through IDAC workshops
  - Organised via Rubin Community Forum (https://community.lsst.org/c/sci/idacs/44)

UK IDAC Architecture



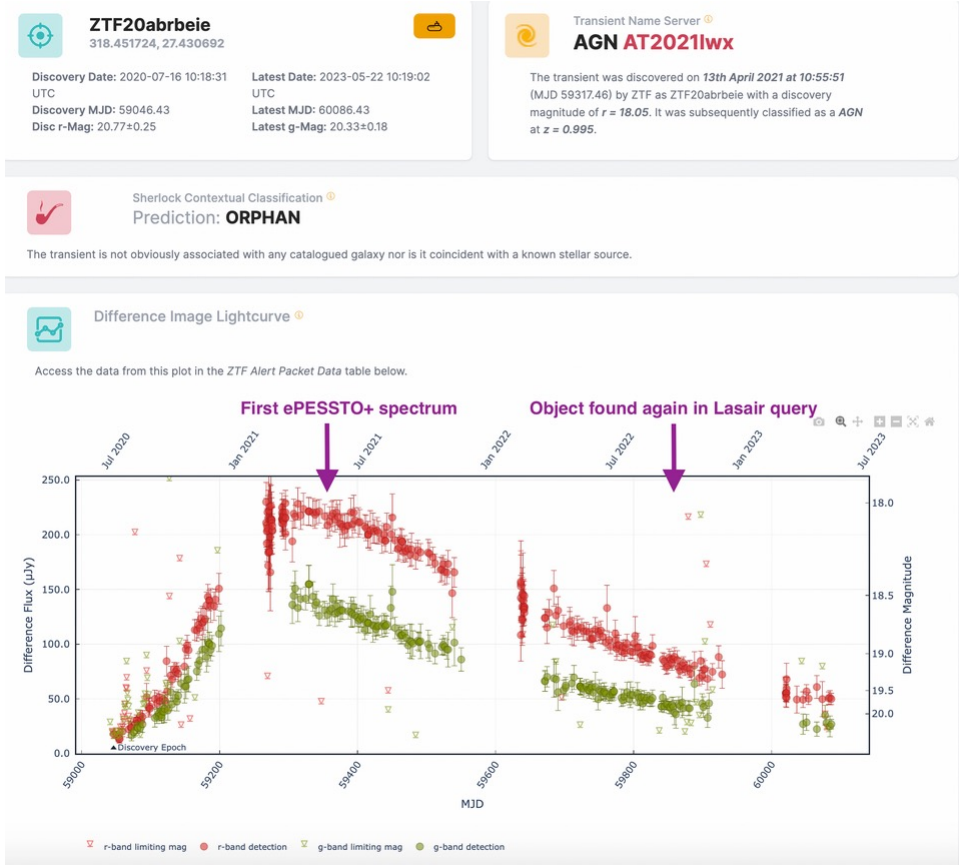
UK IDAC Resourcing

- UK IDAC hosted on IRIS
  - Core services are cloud-based (Scientific OpenStack)
  - Plus access to IRIS HPC – for example, running LSST Pipeline at scale
  - Costs factored into STFC long-term planning
- DAC team – 5.0 FTE by start of Operations
- Support provided primarily via Community Forum (as for Rubin)
  - Pilot based on Lasair

IDAC	Preops	Survey Operations (based on commencement of survey in FY25)										Post-ops	
Capability	FY24	LOY1	LOY2	LOY3	LOY4	LOY5	LOY6	LOY7	LOY8	LOY9	LOY10	FY35	FY36
CPU (M core hrs)	0.53	0.88	2.10	2.45	2.63	3.94	5.26	5.26	6.57	7.88	7.88	7.88	7.9
Normal/ Object (PB)	2.2	21.9	50.0	72.1	94.9	117.6	140.2	162.9	185.6	208.3	231.0	231.0	231.0

Lasair – LSST:UK Community Broker

- Community Broker for astronomical transients
  - One of seven confirmed for LSST
  - Prototype using alert stream from Zwicky Transient Facility – https://lasair-ztf.lsst.ac.uk/
  - Successfully supporting science (71+ citations)
- Status
  - Unified SQL for user-initiated and real-time streaming queries
  - Can incorporate user-provided "Watchlists"
  - “Sherlock” – contextual classification using ~40 catalogues
  - Web, API, Notebook interfaces
- Users can stream results in near-real-time
  - ePESSTO+ – uses Lasair to select targets for spectroscopic observation

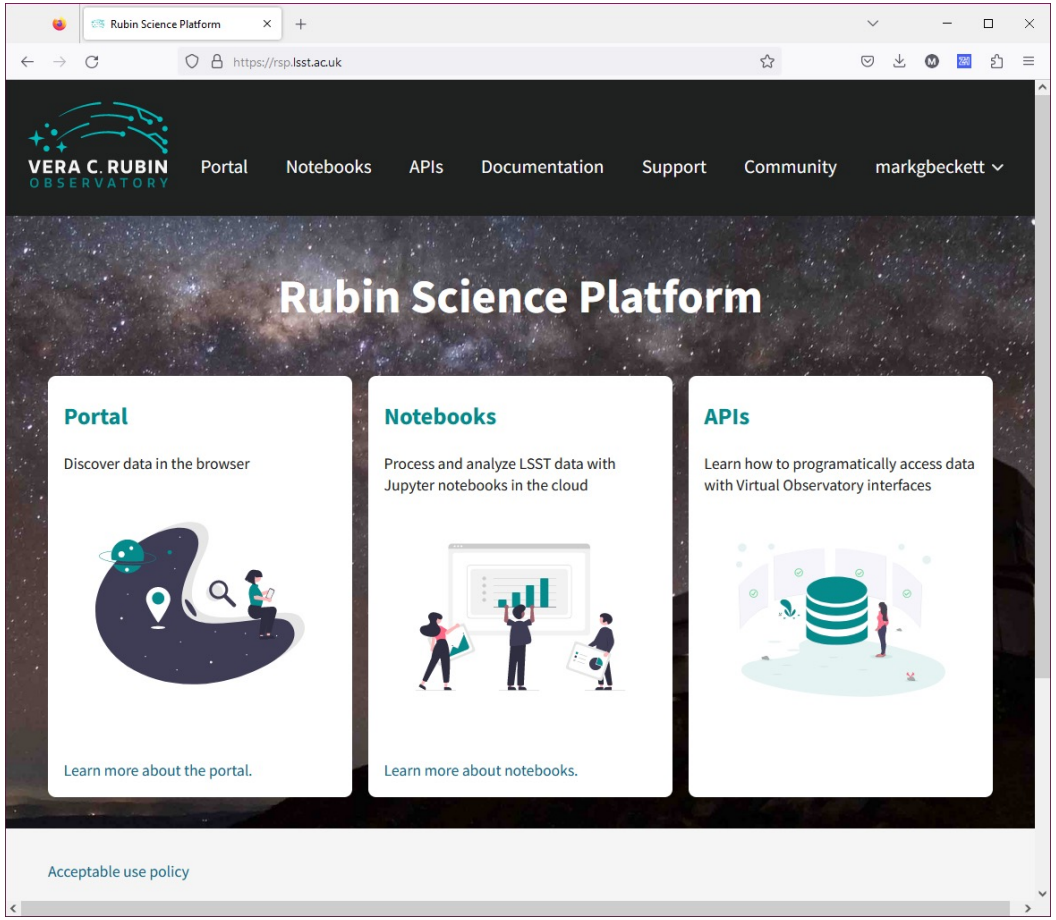


Running on IRIS Cloud in RAL and Edinburgh

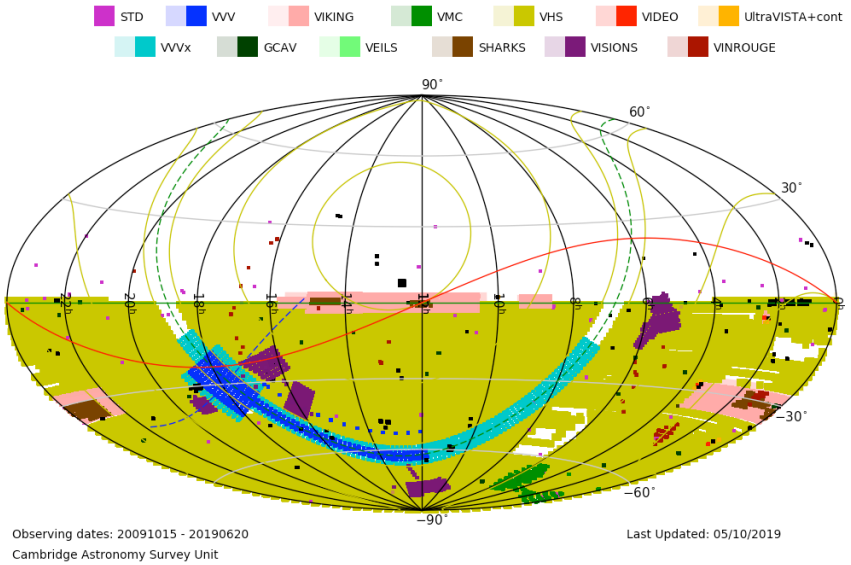
University of Edinburgh, Queen’s University Belfast, and University of Oxford

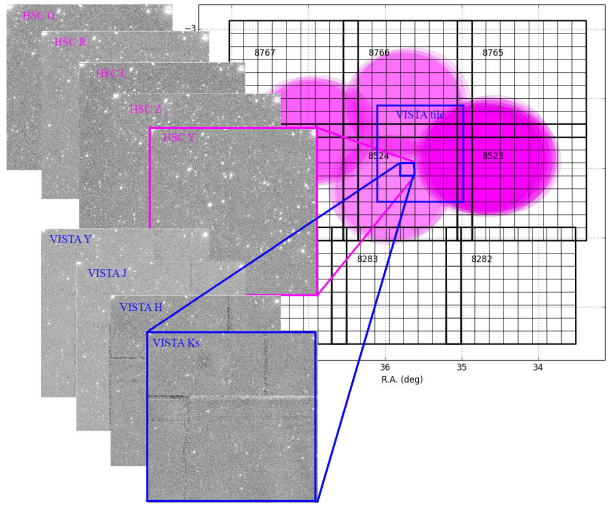
Status, Plans, Ways to Get Involved

- UK IDAC running as tech-preview
  - Available to early-access users
  - Hosting User-generated Products from Phase B (see following slides)
  - Phase C includes Data Previews
  - (DRH eligible to access Rubin DAC)
- Potential to be early-access user
  - Esp. if keen to access UGPs
- Working closely with Rubin SQuaRE and Qserv teams
- IDAC Network starting up
  - discussing common concerns and option for collaboration



Optical/near-infrared data fusion





- 4-m class
- 0.339 arcsec pixels
- 1.65 deg^2 field of view
- Z,Y,J,H,Ks near-IR bands

Joint processing of HSC and VISTA pixel data with LSST pipeline
Precursor to optical/NIR data products from LSST+VISTA

University of Southampton
University of Cambridge

Running on IRIS HPC resources at Cambridge

Catalogue cross-matching at LSST depths

- Probabilistic cross-matching at LSST depths is non-trivial
  - Especially in very crowded fields at low Galactic latitude
- Accurate algorithms are computationally expensive when run on billion-object catalogues

- Testing based on CatWISE × Gaia (EDR3)

Running on IRIS HPC resources at Cambridge

University of Exeter

Summary

- LSST:UK will provide full Independent Data Access Centre
  - Serving up two most recent Data Releases (in full)
  - Co-located with Lasair Community Broker
- UGPs and ancillary surveys for UK science priorities
  - See Friday sessions for more info on these
- Have developed strong relationship with Rubin DM teams
- Part of IDAC Network

Thank You