LSST:UK
Requirements for Cloud

George Beckett

UK Data Facility Liaison/ LSST:UK Project Manager

University of Edinburgh

Vera C. Rubin Observatory

- New observatory in Chile (Cerro Pachón)
- Hosts 8.4m Simonyi optical telescope
  - 3-mirror configuration, extremely wide field of view
- **Legacy Survey of Space and Time** (LSST)
  - 10-year survey starting early 2025
  - Covering 20,000 sq.-deg. in u,g,r,i,z,y bands
  - Observes each field 800+ times during survey
- (Annual) Data Releases
  - Catalogues, Deep and Single-visit Images
  - Served to community via Independent *Data Access Centres*
- Nightly alert stream
  - Expect ~10M alerts per night
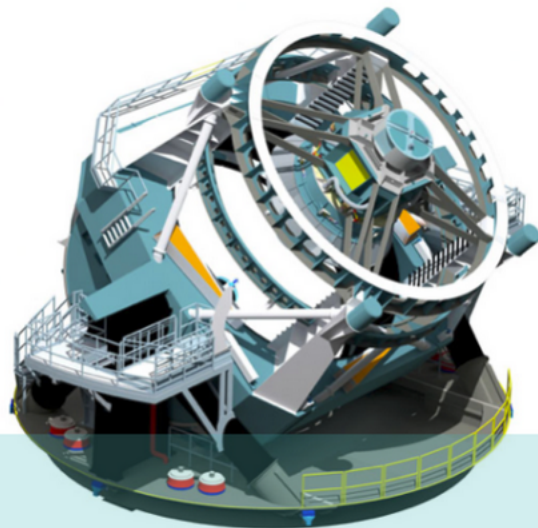 - Streamed to small number of *Community Brokers*

Acknowledgement: Rubin Obs/NSF/AURA

Overview of Data Management
System Vision

Raw Data: 20TB/night

Sequential 30s images covering the entire visible sky every few days

LSST

SCIENCE

PIPELINES

Prompt Data Products

- Alerts incl. science, template and difference image cutouts

- Catalogs of detections incl. difference images, transient, variable & solar system sources

- Raw & processed visit images (PVIs), difference images

Data Release Data Products

Final 10yr Data Release:
- Images: 5.5 million x 3.2 Gpixels
- Catalog: 15PB, 37 billion objects
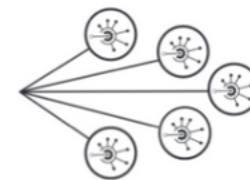
60s via Alert Streams

24h via Prompt Products

80h via Image Services

via Data Releases

Community Brokers

Rubin Data Access Centres (DACs)

USA (USDF)
Chile (CLDF)
France (FRDF)
United Kingdom (UKDF)

Independent Data Access Centers (IDACs)

Access to proprietary data and the Science Platform require Rubin data rights
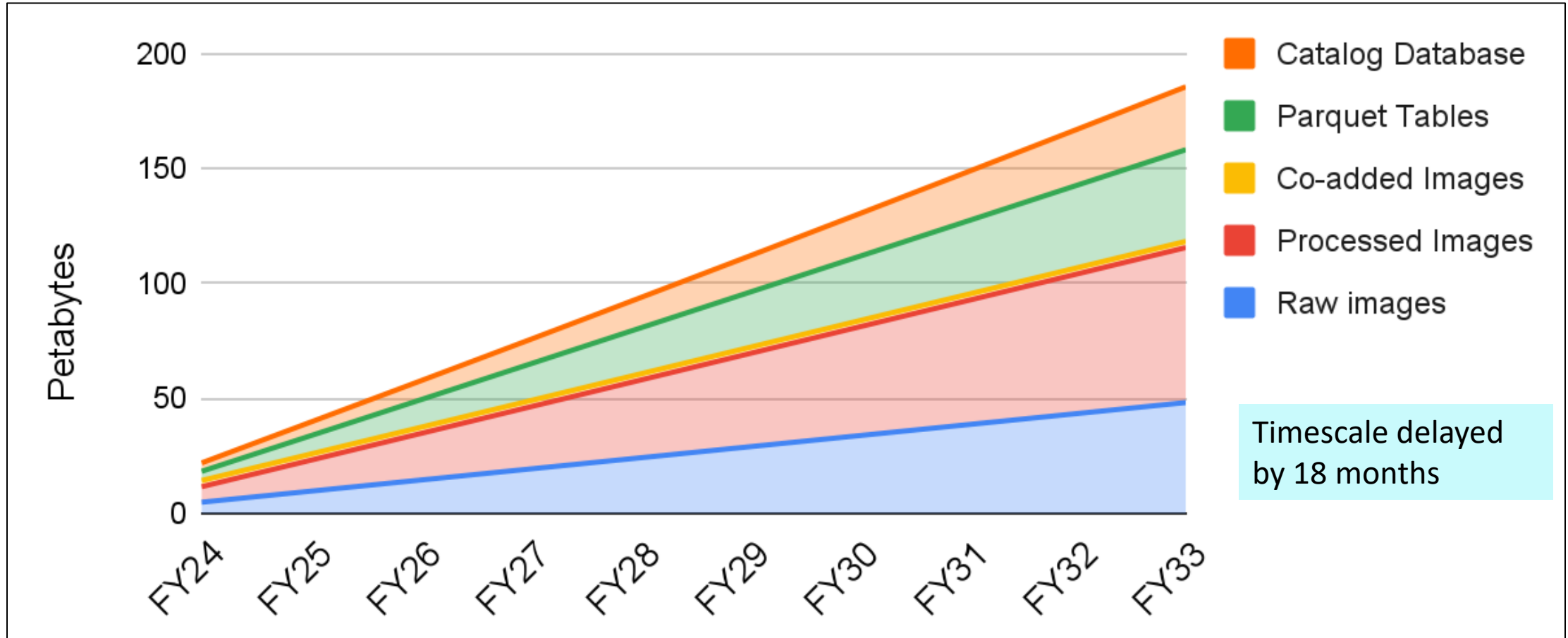
Rubin Science Platform

Provides access to LSST Data Products and services for all science users and project staff.

RUBIN SCIENCE PLATFORM

PORTAL   NOTEBOOKS   WEB APIS

DATA RELEASES   ALERT FILTERING SERVICE   USER DATABASES   USER FILES   USER COMPUTING   SOFTWARE TOOLS

Credit: Leanne Guy

Data Release Products

LSST:UK Consortium and Contribution

- Consortium of 36 UK institutions
  - Includes representation from all UK astronomy departments
- In-kind contributor to Vera C. Rubin operations
  - Data Release Processing (**DRP**) to process 25% of survey data (alongside CC-IN2P3, France and SLAC, USA)
  - Independent Data Access Centre (**DAC**) serving 20% of international Rubin community (est. 1,500 data-rights holders)
  - Community Alert Broker receiving/processing nightly alerts (**Lasair**)
  - Derived data products focused on UK science priorities (**DEV**)
- Various other contributions that are less computational
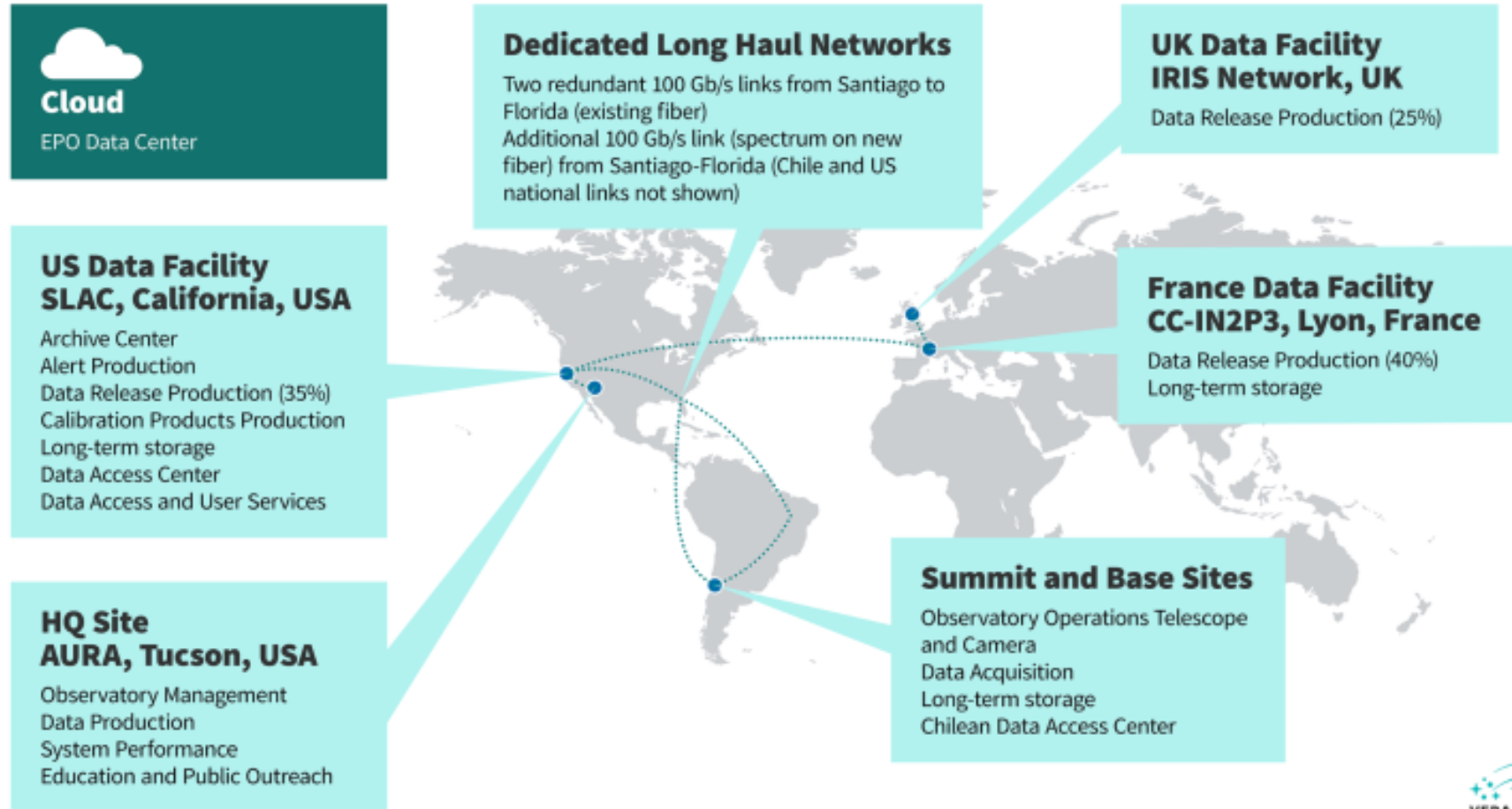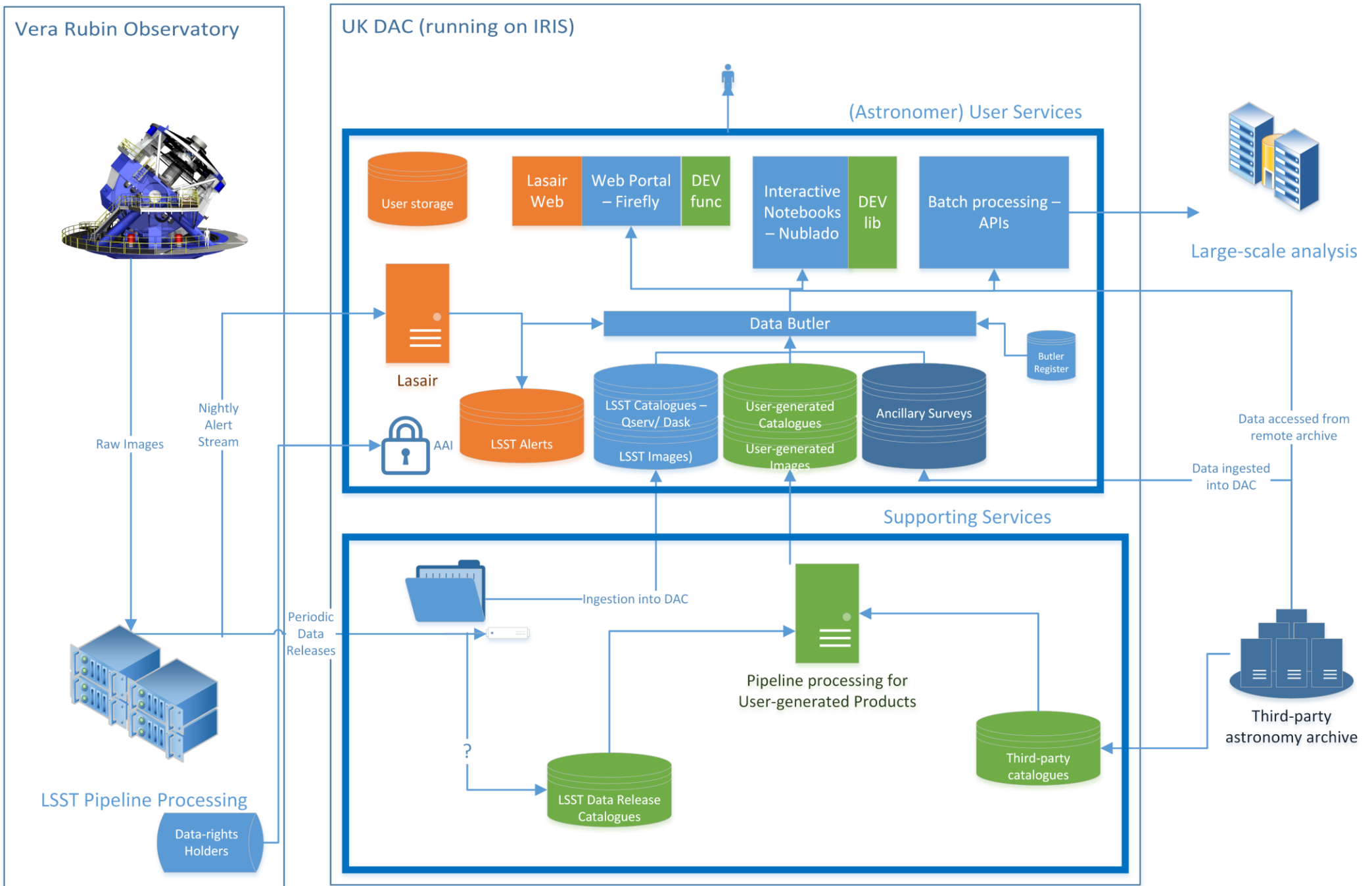
IRIS and LSST:UK

- LSST:UK In-kind contribution matched to IRIS capabilities
  - **DRP** – Grid and persistent storage
  - **DAC** – Cloud and persistent storage
  - **Lasair** – Cloud
  - **DEV** – HPC and persistent storage
- IRIS supporting services (IAM, accounting, security advisory)
- Networking critical for all aspects of operation

Data transfer

Cloud
EPO Data Center

Dedicated Long Haul Networks
Two redundant 100 Gb/s links from Santiago to Florida (existing fiber)
Additional 100 Gb/s link (spectrum on new fiber) from Santiago-Florida (Chile and US national links not shown)

UK Data Facility
IRIS Network, UK
Data Release Production (25%)

US Data Facility
SLAC, California, USA
Archive Center
Alert Production
Data Release Production (35%)
Calibration Products Production
Long-term storage
Data Access Center
Data Access and User Services

France Data Facility
CC-IN2P3, Lyon, France
Data Release Production (40%)
Long-term storage

HQ Site
AURA, Tucson, USA
Observatory Management
Data Production
System Performance
Education and Public Outreach

Summit and Base Sites
Observatory Operations Telescope and Camera
Data Acquisition
Long-term storage
Chilean Data Access Center

Sizing Model

- LSST:UK maintaining long-term sizing model (2023—2036)
  - Input to IRIS RSAP and Long-term Plan

DRP	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)	6	11	21	30	40	50	63	73	83	93	103	100	50
Normal/ Object (PB)	8.0	9.0	16.0	23.0	30.0	37.0	44.0	51.0	58.0	65.0	72.0	61.4	0

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

Sizing Model (cont.)

Lasair	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	4	4	4	4	4	4	4	4	4	4	4	4
Normal/ Object (PB)	0.0	0.3	0.5	0.8	1.1	1.3	1.6	1.8	2.1	2.4	2.6	2.6	2.6
Fast/ NVMe (TB)	115	136	156	177	197	218	238	259	279	300	300	300	136

DEV	Preops	Survey Operations (based on commencement of survey in FY25)										Post-ops	
Capacity	FY24	LOY1	LOY2	LOY3	LOY4	LOY5	LOY6	LOY7	LOY8	LOY9	LOY10	FY35	FY36
CPU (M core hrs)	2	5.3	6.5	6.5	6.5	6.5	6.5						
Normal/ Object (PB)	0.5	1.2	2.11	2.11	2.11	2.11	2.11						

LSST:UK Data Management Plan

- Data Release
    - Astronomy Catalogues – high-demand; held at IDAC; working copy
    - Deep Coadd Images – high-demand; held at IDAC; working copy
    - Visit Images – infrequent/ batch demand; held on Echo?; working copy

LSST:UK Data Management Plan

Dataset	Size (PB)	Usage pattern	Usage Scale	Gold Copy?	Hosting
Astro Catalogues	O(1—10)	Frequent	Interactive queries	No	Somerville
Deep Coadd Images	O(1—10)	Frequent	Interactive/ batch	No	Somerville
Visit Images	O(10—100)	Infrequent	Batch/ interactive	No	Echo
Parquets	O(1—10)	TBD	Batch/ interactive	No	???
Nightly Alerts	O(0.01—0.1)	Frequent	Interactive	No	Somerville
Prompt Products	*TBD*	*Frequent*	*Lasair*	*No*	*US Data Facility*
Image Service (PP)	*TBD*	*Frequent*	*Lasair*	*No*	*US Data Facility*
UK DEV (Cat)	O(0.01)	Frequent	Interactive	Yes	Somerville (Echo)
UK DEV (Images)	O(0.1—1)	Frequent	Interactive/ batch	Yes	Somerville (Echo)

- Interactive access via Rubin Science Platform (Somerville)
- Lasair access via Somerville
- Batch processing via IRIS HPC

Federation, Potential Use Cases

- Improved resilience and up-time for Lasair and RSP
  - Work around scheduled maintenance or unexpected outages
  - HA or on-demand spin up of services on alternative host

- Burst-out capacity to cope with spikes in demand (for RSP)
  - For new Data Release, for example
  - Works around limitations of smaller cloud instance with few projects

- Replication or remote exposure of large datasets
  - Backup for gold copy datasets
  - Mechanism to support less common/ time-limited analysis patterns

Summary

- LSST:UK is making significant contribution to Rubin operations
  - To earn data rights for UK astronomers

- Four strands of in-kind contribution require significant computing
  - DRP – taking 25% share of annual data release processing workload
  - DAC – operation of Data Access Centre
  - Lasair – Community Broker to receive and process nightly alert stream
  - DEV – derived data products and analysis software

- Reliant on IRIS to provide significant computational requirements
  - Engaging actively with IRIS to ensure success
  - Looking to exploit distributed nature of IRIS resources for capability/ resilience
  - Significant commonality with other major astronomy projects/ facilities

Thank you

Motivation for Cloud

- On-demand, elastic and scalable

- Delineation of responsibilities

- Abstraction from underlying technologies

- Ability to benefit from emerging technologies and techniques