D2.1.3: Operational plan for UK Data Facility

D2.1: DAC Management

Project Acronym	LUSC-B
Project Title	UK Involvement in the Large Synoptic Survey Telescope
Document Number	LUSC-B-46

Submission date	27/OCT/23
Version	1.0
Status	Final
Author(s) inc. institutional affiliation	George Beckett, University of Edinburgh Bob Mann, University of Edinburgh
Reviewer(s)	T. Naylor (Exeter), J. Mullaney (Sheffield)

Dissemination level	
Public	

Version History

Version	Date	Comments, Changes, Status	Authors, contributors, reviewers
0.1	4/SEP/23	Document skeleton created	MGB
0.2	11/OCT/23	Further revisions and additional content	MGB
0.3	23/FEB/24	Updated in light of reviewer feedback	MGB
0.4	8/MAR/24	Further updates in light of reviewer feedback	MGB
1.0	03/APR/24	Final Version, meta-data updated following Executive Group approval	TMS

Table of Contents

Index of Figures

Index of Tables

1 Executive Summary

The UK Data Facility makes several computationally significant in-kind contributions to the Vera C. Rubin Observatory, which require careful planning and implementation. These contributions will typically exist for the lifetime of the survey: from the time of writing until at least 2036 and are supported by in-principal funding agreements from STFC and facilitated by the IRIS programme.

At the time of writing, the Observatory is preparing to begin operations in mid-2025, requiring LSST:UK to accelerate preparations over the coming 12 months towards contributing hardware and staff effort to:

- fulfil 25% of Data Release Processing
- operate a Full Independent Data Access Centre
- support the operation of the Lasair Community Broker
- support the production of derived datasets for identify UK science priorities

During LSST:UK Phase B, a team of experts has been assembled, who have developed experience and expertise with the key DRP and IDAC technologies, tailored to the specific environment of IRIS.

- A basic DRP framework has been deployed onto GridPP sites at Lancaster and RAL, with key technologies (Rucio, PanDA, and FTS) configured and tested, ready to commence scale-up processing tests.
- A technology preview of the Rubin Science Platform is operating on the IRIS/ LSST:UK cloud platform, Somerville, serving Rubin Data Previews and early outputs from DEV activities based on precursor surveys.
- A high-level operational model has been developed for three, key, downstream contributions: the Lasair community broker, the Near-IR data fusion pipeline, and the crowded-field crossmatch workflow, plus initial scoping has begun for several other activities that have begun in Phase C.

During the early part of LSST:UK Phase C, a management structure has been put in place to support local development in the UK and appropriate integration with Rubin operational teams and the peer data facilities in the USA and France.

Based on the anticipated workload, staff roles have been defined representing a total contribution of approximately 10 FTE during early operations, and work is underway to recruit and on-board staff into the identified roles.

Priorities have been identified for the next twelve months, to ensure the LSST:UK Data Facility is ready for the beginning of Rubin operations, focused on:

- Creating a model for involvement of multiple GridPP sites in the UK contribution to DRP.
- Producing a holistic Data Management Plan that explains the flow of datasets over the lifecycle of a data release.
- Building expertise and experience with handling very large relational databases (containing billions of rows with hundreds of columns).
- Defining an operational plan for timely creation of UK-contributed derived datasets and services.
- Identifying options for the UK Data Facility to enhance its contribution to the Rubin Observatory to create greater resilience and flexibility for Rubin Operations overall.

2 Introduction

2.1 Document Purpose

This document contains a high-level, operational plan for the UK Data Facility, which encapsulates significant computational and staff contributions to the Rubin Observatory – focusing on the UK contribution to preparing science-ready data products from telescope observations and of serving those products to the astronomy community via a UK-based data facility.

In this section, we provide a brief introduction to the Legacy Survey of Space and Time and a short history of the UK's contributions to date. These contributions are elaborated on in Sections 3, which explains where LSST:UK is at and where LSST:UK expects to be by the start of survey operations. Section 4 describes a high-level plan for how LSST:UK intends to prepare for operations. Finally, in Section 5 we look at topical priorities, plus risks and opportunities.

2.2 The Legacy Survey of Space and Time

The Legacy Survey of Space and Time (LSST) will be delivered to the science community as a collection of science-ready data products [5], on three timescales:

- Data Release (DR) products – typically produced annually and providing a comprehensive suite of science-ready data products including the accumulated set of processed visit images to date, along with deep coadd images and catalogues of measurements for detected objects and sources.
- Alert Products (AP) – issued in near-real-time during observing, as a data stream containing alerts for identified transient activities in the sky. The stream, which is expected to contain around ten-million alerts per night, will be transmitted to a small number (at the time of writing, seven) of designated Community Brokers (external to the Observatory) within 60 seconds of observing, for further filtering, classification, and refinement.
- Prompt Products (PP) – issued within 24 hours of an observing night and providing richer scientific measurements and imaging for identified transient events (as detected in AP) deposited in a Prompt Products database to aid Community Brokers and the wider community in understanding observed transient activity.

Within this document, the Data Release and Alert products are of most interest, as the LSST:UK will contribute to the creation and/ or handling of these products. In contrast, at the time of writing, Prompt Products are only expected to be available directly from the Project.

2.3 Data Rights

The Observatory will produce a mix of proprietary (that is, restricted access) and public datasets. At a high level, Alert Products are public, whereas Data Release and Prompt Products are restricted for a period of two years from initial publication. Except for a small subset that will be made available for education and public outreach (EPO), only approved Rubin Data Rights Holders (DRH) are permitted to use DR and PP for scientific discovery during the two-year restricted period. The data rights policy is described in [11].

All astronomers based in the USA and Chile automatically have access to data rights. Further, astronomers from the international community (for example, in the UK, as introduced below) can realise data rights in return for a Rubin-approved in-kind contribution.

In advance of Operations, the Observatory is generating various datasets (both real and simulated), as part of efforts to commission and scientifically validate the telescope and data-processing systems. Some of these datasets have value to the astronomy community – for example, to help them prepare for early operations – and so, are published as a series of Data Previews (DP). As for the majority of operational datasets, the DPs are restricted to DRH.

2.4 The LSST:UK Consortium

The LSST:UK Consortium (the Consortium) has proposed a substantial package of in-kind contributions, consisting of seventeen distinct elements [6], aiming to secure data rights for the whole UK astronomy community. These in-kind contributions are to be undertaken with funding from STFC in a programme called the LSST:UK Science Centre (LUSC).

At the time of writing, formal agreement on the value of the in-kind contribution is still to be realised, though it is assumed the Consortium will be granted data rights for at least 300 astronomers (Principal Investigators) along with 1,200 associated Junior Associates (as defined in [11]).

Some of the elements of the in-kind contribution depend on the availability of significant computing infrastructure (and LUSC staff) for part or all the operational phase of the telescope. These are listed in Table 1, with more details to be found in [6]. The computing infrastructure for these contributions is collectively referred to as the UK Data Facility (DF).

This report represents a high-level plan for LSST:UK intended to ensure that these in-kind contributions can be delivered to the Rubin community in line with what has been proposed to realise the required data rights for the UK astronomy community.

Table 1: In-kind Contributions that are incorporated into this LSST:UK Operational Plan

Contribution ID	Title
UKD-UKD-01	LSST:UK Contribution to Annual Data Release Processing
UKD-UKD-03	LSST:UK's operation of an international DAC
UKD-UKD-04	Science Software development: Lasair Transient Broker
UKD-UKD-05	Science Software development: Near infrared data fusion
UKD-UKD-09	Science Software development: Cross-matching
UKD-UKD-10	Science Software development: spectroscopic classification of transients and 4MOST spectra
UKD-UKD-16	Adler – Solar System Transient Classification

2.5 LSST:UK Funding

LUSC is funded by the STFC Astronomy Programme as a series of typically 4-year awards. These are referred to in the Consortium as *phases* of funding. At the time of writing, it is anticipated that the complete programme will involve five phases, named Phase A, Phase B, …, Phase E.

The Phase A award, which ran during July 2015–March 2019, had a vision to secure leadership for LSST:UK in the Rubin Observatory community. The funding enabled UK astronomers to start contributing to Rubin International Science Collaborations and supported the successful prototyping of a UK LSST Data Access Centre, plus initial development of the analysis software that the UK community would require to exploit the huge scientific potential of the unprecedented LSST survey.

Phase B, which ran from April 2019 to March 2023, allowed LSST:UK to participate in Rubin Commissioning and to build on the successes of Phase A by turning the proto-DAC into a pre-production facility and deploying within it the analysis software needed for the start of survey operations.

At the time of writing, the programme is mid-way through the first year of Phase C, which runs for four years from April 2023 until March 2027. It will span commissioning and the first two years of survey operations, including the production of the first two data releases; Phase C is, therefore, the core of the 20-year LUSC programme, delivering computing infrastructure and science analysis software for the full ten-year survey.

2.6 Glossary of Acronyms

BEIS	[Department of] Business, Energy, & Industrial Strategy
DAC	Data Access Centre
DF	Data Facility
DMO	Data Management Operations
DP	Data Preview
DR	Data Release
DRH	Data Rights Holder
DRP	Data Release Processing
EPO	Education and Public Outreach
FTE	Full-time Equivalent
IDAC	Independent Data Access Centre
IPC	(Rubin) In-kind Program Committee
LSST	Legacy Survey of Space and Time
LUSC	LSST:UK Science Centre
PVI	Processed Visit Image
RSP	Rubin Science Platform
UGP	User-generated Products
ZTF	Zwicky Transient Facility

3 Context

3.1 Rubin Timeline

The Rubin Observatory has published and periodically updates a timeline for early operations [2], which we use to guide operational planning in LSST:UK. This includes a schedule for producing Data Preview (DP) and Data Release (DR) products. The published schedule, at the time of writing, is reproduced in Figure 1.

Rubin Early Data Release Scenario	Jun 2021	Jun 2022	Jun 2023 - Sep 2023	Dec 2024 - Apr 2025	Aug 2025 - Mar 2026	Feb 2026 - Nov 2026	Feb 2027 - Nov 2027	Feb 2028 - Sep 2028	Feb 2029 - Sep 2029
	DP0.1	DP0.2	DP0.3	DP1	DP2	DR1	DR2	DR3	DR4
Data Product	DC2 Simulated Sky Survey	Reprocessed DC2 Survey	Solar System PPDB Simulation	First Light LSSTCam Data	LSSTCam Science Validation Data	LSST First 6 Months Data	LSST Year 1 Data	LSST Year 2 Data	LSST Year 3 Data
Raw images	☑	☑	☐	☑	☑	☑	☑	☑	☑
DRP Processed Visit Images and Visit Catalogs	☑	☑	☐	☑	☑	☑	☑	☑	☑
DRP Coadded Images	☑	☑	☐	☐	☑	☑	☑	☑	☑
DRP Object and ForcedSource Catalogs	☑	☑	☐	☐	☑	☑	☑	☑	☑
DRP Difference Images and DIASources	☐	☑	☐	☐	☑	☑	☑	☑	☑
DRP ForcedSource Catalogs including DIA outputs	☐	☑	☐	☐	☑	☑	☑	☑	☑
PP Processed Visit Images	☐	☐	☐	☐	☐	☑	☑	☑	☑
PP Difference Images	☐	☐	☐	☐	☐	☑	☑	☑	☑
PP Catalogs (DIASources, DIAObjects, DIAForcedSources)	☐	☐	☐	☐	☑	☑	☑	☑	☑
PP SSP Catalogs	☐	☐	☑	☐	☑	☑	☑	☑	☑
DRP SSP Catalogs	☐	☐	☐	☐	☐	☑	☑	☑	☑

Figure 1: Summary of data products expected in each data preview and early survey data releases (as of June 2023).

There is still some uncertainty over the release date (and content) of future DPs and DRs. Because of this, the timeline is reviewed on a regular basis by the LSST:UK Executive Group and the impact of any changes to the schedule to LSST:UK's plans is considered.

3.2 Rubin In-kind Contribution

As noted in Section 2, LSST:UK has proposed an ambitious programme of in-kind contributions to the Observatory.

Each in-kind contribution has one or several designated Recipient Groups, who are expected to realise significant benefit from the in-kind contribution and who will testify to the value of the in-kind contribution. Usually, the recipient group is either the Observatory or one of the eight Rubin international science collaborations[1].

In-kind contributions are overseen by a body called the In-kind Program Coordination (IPC), which has established a regular review mechanism along with processes for accommodating any changes or deviations in a contribution that may be required (see Section 4.2).

For each of the contributions that are the subject of this plan (listed in Table 1), we note below key details that are significant for operational planning.

3.2.1 LSST:UK Contribution to Annual Data Release Processing

The Observatory proposes to run a 10-year survey and to publish this survey as a series of eleven Data Releases – a first release (DR1) capturing the first six months of observations, followed by ten further released (DR2, …., DR11) representing cumulative

[1] See https://www.lsstcorporation.org/science-collaborations (last accessed 6th September 2023)

annual updates to observations. For example, DR2 captures the first year of observations, DR3 captures the first two years, and so on.

The preparation of each DR, a process called Data Release Processing (DRP), is a substantial piece of work involving significant computational and storage infrastructure. It is a multi-stage process, in which images progress through various processes, and for a dataset of the scale of a DR, it is expected to take many months to complete.

LSST:UK has proposed to process 25% of each Data Release as its contribution to DRP. This will be done alongside the Observatory, which will complete 35% of the processing, and the Institut national de physique nucléaire et de physique des particules (IN2P3) in France, which will complete 40% of the processing.

The LSST:UK contribution to DRP includes both computing infrastructure and staff effort[2], sized based on information presented in [10]. Specifically, LSST:UK has proposed to provide staff effort (initially, 4.25 FTE from April 2023, rising to 5.25 FTE from April 2024, for the duration of Operations) and one quarter of the storage and compute infrastructure identified as being required for DRP in [10]. A detailed breakdown of this contribution is provided in [2].

3.2.2 LSST:UK's operation of an international DAC

Once validated, each DR is made available to the science community via a small number of Data Access Centres. The Project is planning to operate two DACs: one hosted on the Google Science Cloud with some elements at the USDF, for US-based and international Data Rights Holders; and one in Chile, focused on the Chilean astronomy community.

Furthermore, as part of the in-kind programme, the Project invited international parties to propose further DAC-like services (IDACs), serving the needs of the general or a specific subset of the Rubin science community. The requirements for an in-kind IDAC contribution are documented in [9].

LSST:UK has committed to running a *Full IDAC*, as defined in [9], serving the two-most recent DRs in full, similar to the Project DACs. Furthermore, LSST:UK has committed to provide sufficient capacity to serve 20% of the anticipated Rubin community.

This contribution earns data rights based on storage space and compute time, though not staff effort. The LSST:UK resourcing plan is based on the model described in [9] and [10], and is documented in [2].

The Rubin Observatory is developing a software suite, called the Rubin Science Platform (RSP), which encapsulates the software required for a Rubin DAC. The RSP software is available to IDAC operators and a requirement of a full IDAC. Therefore, LSST:UK is planning to run the RSP as the primary interface to DRs and related products.

In addition to providing access to DRs, the UK IDAC will host in-kind contributions from UKD-UKD-S04 (the Lasair service), UKD-UKD-S05 (VISTA-LSST fused products), and UKD-UKD-S09 (crossmatch catalogues from LSST and selected other surveys), at least. The implications of this hosting are documented in [1] and [14], and estimated resource requirements are included in [2].

[2] It also includes some funding for power costs and local, system-administration effort at the contributing site.

The UK IDAC is also likely to support UKD-S10 (4MOST follow-up) and UKD-S16 (Adler), though the details of this support are still to be developed. Two DAC-DEV integration exercises are required to address this omission.

3.2.3 Science Software development: Lasair Transient Broker

The Lasair Community Broker[3] is a platform for astronomers to work effectively with the LSST transient alert stream. While Lasair itself is not considered, by the Rubin IPC, as an in-kind contribution attracting data rights, it is critical to at least one other UK in-kind contribution (UKD-UKD-S10 (4MOST follow-up)) and likely to be a significant advantage for UKD-UKD-S16 (Adler solar-system classifier). It is also of significant value to those in the Rubin community who are interested in time-domain astronomy.

The LSST alerts are expected to begin to stream towards the end of the first year of telescope operations and then to continue throughout the life of the survey. In advance of the LSST stream, Lasair is working with an alert stream from the Zwicky Transient Facility, which is a recognised precursor telescope for LSST. Further, the Observatory is expected to make available, in the run-up to Operations, a test stream that mimic what a broker will receive during Operations.

Lasair will be (and, in its current guise for ZTF, is) hosted alongside the UK RSP on the UK IDAC. It is expected to interface with some of the services offered by the RSP, though the details are still to be finalised.

Part of the staff effort to maintain and develop Lasair is drawn from the UK DAC budget – specifically focused on Lasair infrastructure provision – though, it is recognised that the work is likely to be distinct from work towards general UK IDAC development and operations.

Further, the infrastructure requirements for Lasair are part of the overall infrastructure sizing model [2]; and will be refined periodically, in the run-up to the beginning of telescope operations.

3.2.4 Science Software development: Near-infrared data fusion

In-kind contribution UKD-UKD-S05 is developing a pipeline workflow for a User-generated Product (UGP) that will fuse image data from each LSST DR with corresponding data from the VISTA surveys (specifically, at the time of writing, VIKING, VIDEO, and VHS).

As part of the delivery of this UGP to the science community, the UK IDAC team will run the pipeline workflow for each DR (at least, DR1, DR2, and DR3) and publish the resultant products via the UK DAC. This commitment is detailed in [1].

In addition, in support of the validation and promotion of the LSST-VISTA fused dataset, the UK IDAC team will work with the DEV team to test and refine the processing and publication of the UGP during Phase C, based on DP1 and DP2.

As with the LSST DRs, the two most recent version of the LSST-VISTA fused dataset will be made available to DRH via the UK IDAC. The estimated infrastructure requirements for doing this (including the processing) are captured in [2]. These requirements will be refined based on findings from the joint processing of DP1 and DP2.

[3] https://lasair-ztf.lsst.ac.uk/ (accessed 7th September 2023)

3.2.5 Science Software development: Cross-matching

In-kind contribution UKD-UKD-S09 is developing an algorithm (and a software implementation called Macauff) that can reliably identify counterparts to LSST objects in relevant ancillary surveys (as well as objects for which no likely counterpart is found). The algorithm is designed to handle the relatively crowded object fields that are expected in LSST, due to the sensitivity of the telescope.

As part of the in-kind contribution, the UK IDAC team will run Macauff to crossmatch a selection of around five ancillary surveys (selected in conjunction with the recipient groups) against each DR (at least, DR1, DR2, and DR3) and publish the resulting catalogues via the UK RSP. This commitment is detailed in [1].

In addition, in support of the validation and promotion of the crossmatch catalogues, the UK IDAC team will work with the DEV team to test and refine the processing and publication of the crossmatch catalogues during Phase C, based on DP1 and DP2.

The crossmatch catalogues for the two most recent DRs will be made available to DRH via the UK IDAC. The estimated infrastructure requirements for doing this (including the processing) are captured in [2]. These requirements will be refined based on findings from the joint processing of DP1 and DP2.

3.2.6 Science Software development: spectroscopic classification of transients and 4MOST spectra

In-kind contribution UKD-UKD-S10 is developing a service to instigate spectroscopic follow-up of interesting transient events, detected in the LSST alert stream, using observing time (for the TiDES project) on the 4MOST telescope. The team plan to write suitable alert queries for the Lasair Alert Broker to identify interesting transients that are within the topical 4MOST observing field and then automatically submit observation requests to the 4MOST scheduler.

The service that identifies events and communicates with 4MOST may be hosted alongside Lasair on the UK IDAC though has modest infrastructure requirements, comparable to those of a workstation computer. Further, significant effort from the UK IDAC team is not required for operation of the services.

Timely follow-up of events is, however, important, so the service will require infrastructure with a high level of availability.

3.2.7 Adler – Solar System Transient Classification

In-kind contribution UKD-UKD-S16 is developing an Open Source LSST Solar System transient classifier, closely aligned and potentially integrated into Lasair.

Adler will utilise the LSST alert stream through Lasair and the Prompt Products to provide users with the information needed to identify truly changing Solar System bodies. Adler will flag potential dynamically new comets and interstellar objects first entering our Solar System, identify possible planetesimal collisions, break-up events, or the onset of cometary activity and cometary outbursts. Adler will fit rotational light curves and phase curves and identify outlying photometric points; Adler will also automatically flag suitable Comet Interceptor fly-by targets, and identify cometary coma/tails, through multi-aperture photometry of the alert images and source extendedness.

3.3 Organisation of Rubin Data Production

Within the Rubin Observatory, the production, verification, and publication of the survey, along with the software and services required to support this, is the responsibility of Data Management Operations (DMO). As described in [8], DMO has four functions:

- Data Production – responsible for the production software (the Pipeline) and execution of each processing campaign.
- Data Services – responsible for the Rubin Science Platform software and associated services.
- Data Abstraction – responsible for data engineering along with campaign middleware.
- Data Facilities – responsible for the underpinning infrastructure and supporting services required by the other three functions.

UK in-kind contributions to DRP will involve collaborative working within DMO, with effort focused on Data Facilities, though with a smaller contribution to Data Production.

3.3.1 Production Processes

DMO will typically run two processing pipelines (see Figure 2):

- the Data Release Pipeline, for periodically generating Data Releases, represented by the left branch in the figure
- the Difference Imaging Pipeline that produces the nightly alert stream (which will be forwarded to Community Brokers, such as Lasair) and prompt products, for identifying transient behaviour in the sky, represented by the right branch

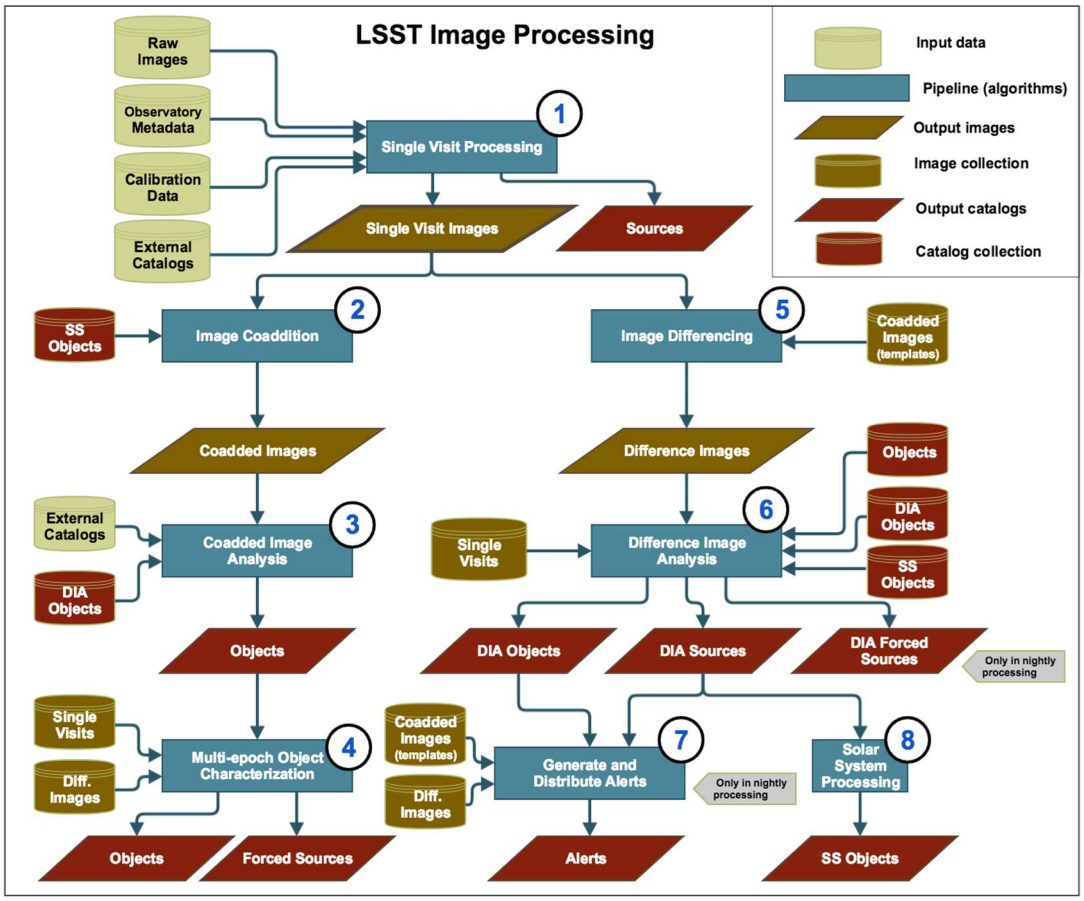

Figure 2: Stages of the Rubin Imaging Processing Pipeline

The UK DRP in-kind contribution only involves the DR Pipeline: as noted earlier, difference-imaging and prompt-products production is undertaken entirely on the USDF.

A typical DR Pipeline run, called a Campaign, is expected to take 12 months to complete[4] and the Campaign is overseen by the Campaign Management team within Data Production.

The likely timeline of a Campaign is presented in Figure 3, with the UK, US and French DFs contributing resources, in line with their commitment, throughout the campaign.

First, a small test campaign is run – suggested to be around 2% of the sky coverage – which should complete in around one week. This test campaign is used to check that the pipeline configuration is working correctly and that supporting services (such as monitoring and validation) are functioning.

The core of the campaign takes around 8 months to complete, involving a sequence of (at the time of writing, seven) processing *Steps* that can mostly be completed independently by each DF (that is, without the need to share intermediate products between DFs). Each Step involves a large number (potentially millions) of discrete *Tasks* that progressively refine images and generate scientific measurements.

At three points in the campaign, synchronization of data products between the three DFs is required, indicated by the cross-cutting, red, vertical bands in Figure 3[5]. A DF cannot continue production past a synchronisation point until it has received the data products it needs from the other two facilities.

Throughout the Campaign, the topical outputs and intermediate products will be checked by a science verification team (also part of Data Production). They will test the outputs aiming to identify, early on, any potential issues with the Campaign in terms of scientific validity of the results.

The final two months are reserved for preparing and serving the data to the intended IDAC sites, ready to be published to the community at the end of Month 12.

[4] DR1, which captures the first six months of observations, is a special case expected to take 6 months.

[5] The division of work between the three Data Facilities has changed since this figure was produce. At the time of writing, The French DF will undertake 40% of DRP; the US DF will undertake 35% of DRP; and the UK DF will undertaken 25%.

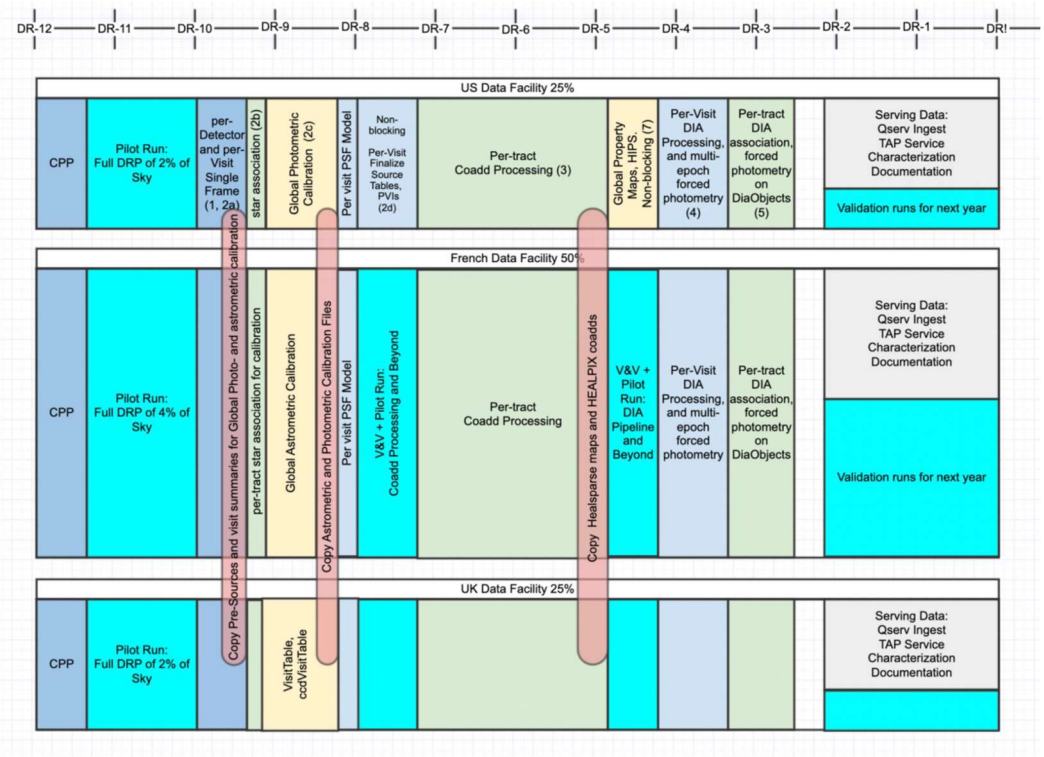

Figure 3: Timeline for a typical Data Release Processing Campaign[6]. The scale represents the months in advance of the Data Release (e.g., DR-6 means six months prior to the Data Release).

DMO has, up until now, had little direct involvement in the plans for setting up an operating Independent Data Access Centres. DMO does encompass the software and services (the RSP) that will be used on the Project DACs and Full IDACs (possibly, others). Further, DMO Data Facilities has responsibility for the setup and operation of the infrastructure for two Project DACs (a hybrid instance across the Google Science Cloud and SLAC, and a small on-premises instance in Chile, respectively), but the additional considerations required for the successful operation of a network if IDACs is delegated to the IDAC contributors themselves. To this end, the IPC has convened an IDAC Coordination Group to facilitate the collaborative development of good practice, experience, tools, and techniques for successful IDAC operation.

3.4 IRIS

The LSST:UK in-kind contributions highlighted in Section 3.2 require substantial computing infrastructure. LSST:UK will realise this infrastructure via the IRIS programme[7] — a collaboration of significant STFC-funded projects and facilities that have a common interest in substantial computing infrastructure to support their science cases. The IRIS programme has some direct, capital funding, which is used to deliver a baseline level of computing infrastructure, along with generic technology solutions addressing topics such as data movement, authentication and authorisation, and science-platform tooling. However, projects with requirements that go beyond this baseline level of infrastructure are required to contribute funding into the IRIS programme, based on an expectation of more effectively and efficiently being able to realise their infrastructure requirements utilising the mature practices and processes (for example, procurement and system administration) in place at IRIS partners sites.

[6] https://confluence.lsstcorp.org/display/DM/Multi-Site+DRP+Diagram (accessed 6th September 2023)

[7] https://www.iris.ac.uk/

In 2022, STFC prepared a business case on behalf of and with support from LSST:UK, which they submitted to the (at the time) UK Government Department of Business, Energy & Industrial Strategy (BEIS). The business case summarised LSST:UK's operational plans, as detailed in the in-kind contribution proposal [6], along with cost estimates of operations costs (including staff effort) and capital costs (effectively, computing infrastructure). The business case was approved by BEIS at the end of 2022 and, as a result, the estimated costs (derived in [2]) have been included in the forward estimates of the STFC Astronomy Programme.

The STFC Astronomy Programme will provide capital funding for LSST:UK's use on an annual basis (during April—March), in line with the cost projections in [2]. To expend the funding, LSST:UK will submit resource requests to be reviewed by IRIS, as per IRIS review processes and a report containing the outcomes of this review will be provided by IRIS to STFC. STFC will then prepare an award (or awards) to the nominated institutions that will host the infrastructure. LSST:UK will be required to report on progress with expending funds to STFC IRIS and STFC's LSST:UK Oversight Committee (see Section 4.1).

3.5 Phase B Contributions

During LUSC Phase B, significant progress was made with the design and implementation of many of the contributions defined here. That progress is summarised below, to confirm the starting point for Phase C work.

3.5.1 LSST:UK Contribution to Annual Data Release Processing

LSST:UK's role in DRP emerged relatively late on in the programme, during early negotiations of in-kind contributions. Despite that, significant progress has been made to onboard the LUSC team and to begin work on the contribution, including:

- Appointment of a UK Data Facility liaison (George Beckett, U. of Edinburgh) at 0.25 FTE.
- Creation of a DRP team, initially with 2.0 FTE of effort, rising to 4.0 FTE from April 2023
- Allocation of GridPP infrastructure (via IRIS) to participate in early experimentation with candidate technologies and techniques from high-energy physics (for example, configuration of storage (Rucio Storage Elements) and compute (grid compute resources accessible via PanDA)
- Set up of management processes within LSST:UK and interacting with peers in the US and French DFs, including: weekly, joint-data facility meetings involving all-three DFs, fortnightly technical working group meetings focused on storage, data transfer and job submission; fortnightly UK-team meetings.
- Significant progress on a high-level architecture for the DRP solution, which is captured in a series of Joint Data Facility workshops (starting in June 2021 with a roughly six-month cadence)[8].

From April 2024, the UK DRP team that will progress into Operations will be formalised and a call to the LSST:UK community to determine that team will run in late 2023.

[8] Workshop proceedings available to Project members via https://confluence.lsstcorp.org/x/Ro9cC.

3.5.2 IDAC Architecture

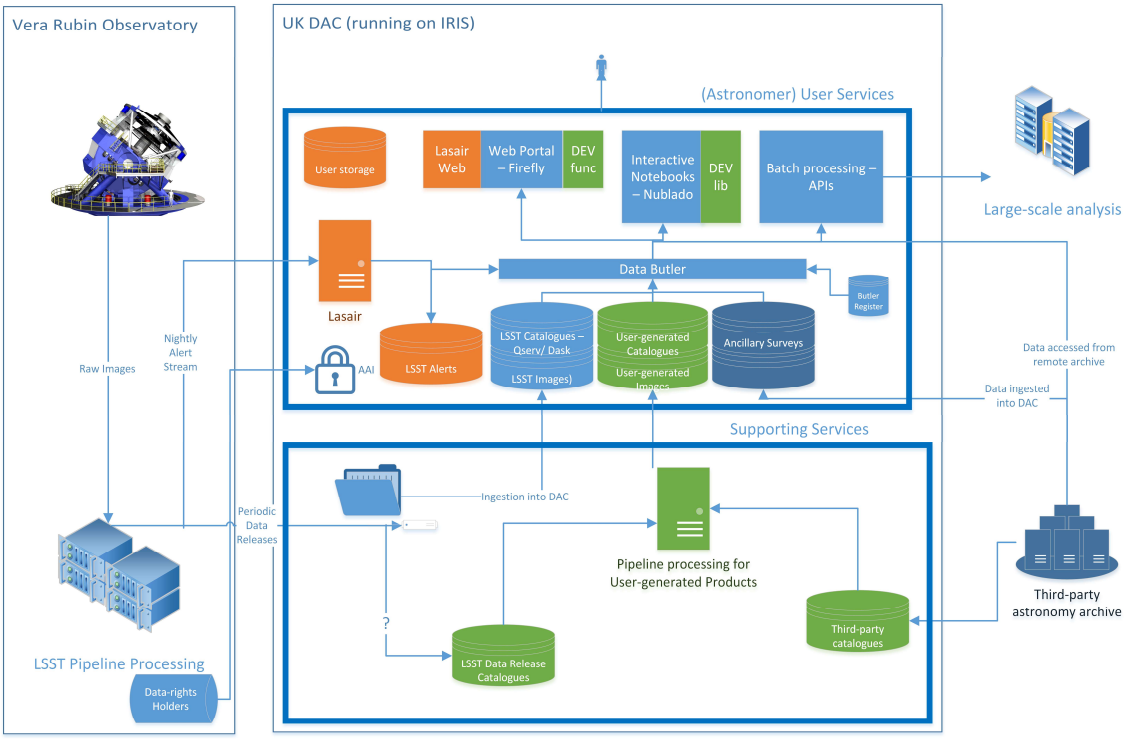

Figure 4: Proposed IDAC Architecture

Significant progress has been made, during Phase B, to develop a UK IDAC solution. The proposed architecture is presented in Figure 4 and a preview implementation of this architecture has been deployed onto the Somerville cloud platform in Edinburgh.

A DAC team has been assembled, at the time of writing with 4.0 FTE of effort, though expected to rise to 5.0 FTE early in Phase C.

The UK DAC team has developed experience with key IDAC technologies, including:

- Rubin Science Platform – the DAC team is collaborating with the Project SQuaRE team who develop the RSP and is running a current instance of the RSP on the UK IDAC, accessible to around 20 early users.
- Qserv – the DAC team is collaborating with the Project Qserv developers and is running a production instance on the UK IDAC, hosting various scientific datasets. The UK IDAC team also runs various test and development instances, intended to grow experience with the technology and to refine the deployment architecture.
- Data Butler – the DAC team has deployed several instances of the Data Butler as part of the UK IDAC offering of scientific images to early users.
- Data transfer and ingestion – the DAC team is experimenting with transferring and ingesting the substantial datasets that will be involved in telescope operations, using data from UK DEV activities to inform these experiments.
- Various survey datasets have been ingested into the UK IDAC, including LSST DP0.1, UKD-S5 fused HSC-VISTA data, and UKD-S9 Gaia-catWISE crossmatch data. Building on these, early work has begun to build a user community.

A candidate infrastructure layout has also been defined on which to run the IDAC service, based on (Kubernetes-architected) containerised services running on top of OpenStack virtualised hardware with a mix of (Ceph) network-attached storage for curated datasets and fast, local storage for caching priority databases.

More information about the UK IDAC experiments during Phase B can be found in [13].

Interactions between the various IDAC contributions that have been proposed has also begun. In March 2023, the Rubin In-kind Program Coordination team organised a two-day workshop introducing representatives from IDAC with representatives from International Science Communities (and the wider Rubin astronomy community). Beckett was part of the organising committee for the meeting, which identified a range of further work to be undertaken. A final report from the workshop is still pending.

3.5.3 Lasair Transient Broker

During Phase B, a candidate Lasair platform has been designed, implemented and tested using precursor data from the Zwicky Transient Facility. This service has been made available to the science community and has seen strong uptake with, at the time of writing, around 300 registered users and has been cited in over 70 scientific papers.

The Lasair service is running on the Somerville cloud platform, alongside the prototype UK IDAC, with similar underlying supporting technologies. Integration of Lasair and the UK IDAC (both services and data holdings) is currently limited, though work to develop this has commenced at the time of writing.

Lasair has piloted the use of the Rubin Community Forum to provide support to early users. The Rubin Community Forum is the intended route by which data-rights holders in the UK and beyond, will obtain science support for all Rubin and associated products.

More information about Lasair can be found in [14].

3.5.4 Science Software development: Near-infrared data fusion

The DAC team has worked with the UKD-S5 team during Phase B to:

- develop an HPC implementation of the processing workflow for running the data-fusion pipeline (based on the LSST stack)
- develop an ingestion and publication workflow for publishing the output datasets to the Rubin community via the UK IDAC
- to identify technical requirements of both workflows
- to agree on a high-level plan for undertaking these activities during operations.

This work is described in [3] and the referenced reports.

3.5.5 Science Software development: Cross-matching

The DAC team has worked with the UKD-S9 team during Phase B to:

- develop an HPC implementation of the processing workflow for cross-matching pairs of survey catalogues
- develop an ingestion and publication workflow for publishing the output database to the Rubin community via the UK IDAC
- to identify technical requirements of both workflows
- to agree on a high-level plan for undertaking these activities during operations.

This work is described in [3] and the referenced reports.

3.5.6 Science Software development: spectroscopic classification of transients and 4MOST spectra

The UKD-S10 team has been active during Phase B, though has not explicitly worked with the DAC team. Their infrastructure requirements have been modest (a small amount

of storage and compute on the Somerville cloud platform) and their interaction with other DAC and related services has been contained to using the Lasair public interfaces.

4 Operational Plan

4.1 Roadmap to Operations

The execution of in-kind contributions for DRP, IDAC and selected DEV activities is dependent on the Rubin Observatory timeline (introduced in Section 3.1) and, derived from this, LSST:UK maintain a roadmap showing data-rights holders when various services and datasets are expected to be available to the community via the UK Data Facilities. A topical snapshot of this roadmap is included in Annex A.

The roadmap forms the backbone for delivery of most of the LSST:UK in-kind contributions and the operational plan described here is structured around that roadmap. For example, targets for the availability of DR and DP products dictates when DRP campaigns need to be completed, when data products will be ready to publish in the IDAC and when generation of (DEV) user-generated products can begin.

Any changes to the Rubin Observatory timeline trigger changes to the LSST:UK roadmap, which then initiates scheduling changes for many of the DRP, IDAC, and DEV contributions, which will be handled through the change management process (see 4.1).

4.2 Project Organisation

UK involvement in Rubin proceeds through two organisations:

- The *LSST:UK Consortium* comprises all UK astronomy groups with a scientific interest in LSST (at the time of writing, every astronomy group in the UK).
- The *LSST:UK Science Centre* (LUSC) is a distributed organisation that conducts funded work on behalf of the LSST:UK Consortium.

Figure 5 shows the principal bodies set up by the Consortium.

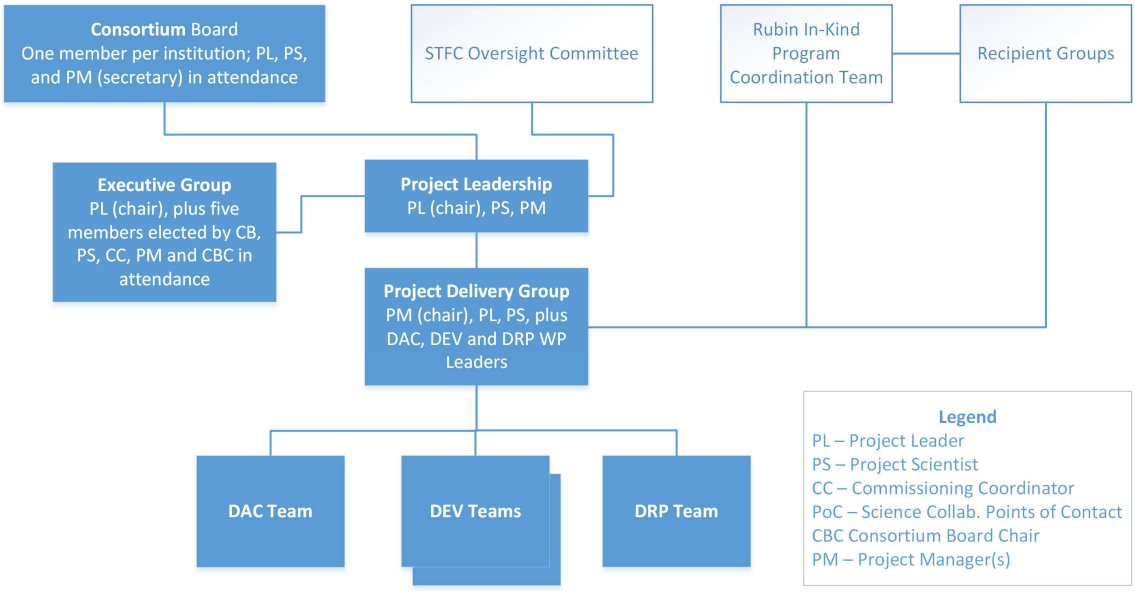

Figure 5: Governance Structure of LSST:UK

The Consortium Board is the sovereign entity in LSST:UK. It comprises a representative from every UK astronomy group and acts as a proxy for the UK astronomical community, who are the principal beneficiaries of the LUSC programme, since that is securing data rights for them and providing (software, data, computing) resources to support their scientific exploitation of that data.

Foremost amongst the groups to which the Board delegates responsibility is the Executive Group, which comprises five members elected by the Board. The Executive Group advises the Project Leadership team.

The Project Delivery Group comprises the Project Managers and the Work Package Leaders. The Group has an information-sharing function and helps the Project Managers to fulfil their reporting responsibilities to the Executive Group and the Board.

The programme of work is divided into three work packages, which can be categorised into three contributing areas: DAC, DEV, and DRP.

Each Work Package has a nominated Work Package Leader, who is responsible for ensuring effective progress in achieving the work-package objectives within the constraints of the available resources and timeline for the project.

The Project Leadership reports on project progress; significant exceptions; requests to draw on working allowance, etc., through two mechanisms:

- A Project Assurance Report is submitted to the STFC Office every six months.
- An STFC Oversight Committee convenes twice each year to receive reports from the Project Leadership, regarding project progress.

Full detail of the Project Governance is provided in the Project Management Plan[9].

The Project Delivery Group reports on each in-kind contribution to the Rubin Observatory every three months. The report is assessed by the relevant recipient group(s) through a process that is coordinated by the Rubin IPC and once finalised, this realises the number of Data Rights Holder positions agreed in the in-kind agreement. Full details of the in-kind reporting mechanism is provided in [7].

4.3 Responsibilities

The operational plan is based on a set of identified, high-level responsibilities that need to be fulfilled by the Science Centre team. These are derived from relevant UK science requirements [12] – which, in turn, are derived from key Rubin Observatory documents including [9], [10] and [11]. These responsibilities can be categorised naturally into five functional areas as described below (with the related UK science requirements noted in square brackets[10]).

4.3.1 UK Data Facility Coordination

The high-level responsibilities within UK Data Facility Coordination are, as follows:

- To ensure the UK in-kind contribution to DRP is resourced and managed appropriately.
- To ensure the UK Data Facility is resourced and managed appropriately, liaising with IRIS to effectively spend (and, potentially, supplement) capital funding from STFC.
- To work with IRIS to ensure appropriate computing infrastructure is provisioned for the UK DF. [R0.2.1.5, R0.2.2.6]

[9] https://lsst-uk.atlassian.net/l/cp/0puHo9pi (last accessed on 11th October 2023). https://lsst-uk.atlassian.net/l/cp/0puHo9pi (last accessed on 11th October 2023).

[10] At the time of writing, the Science Requirements do not address DRP, which became part of the UK responsibility after the most recent revision. This omission will be addressed in the next update.

- To maintain the forecast infrastructure requirements (c.f., [2]), adjusting for changes to the operational requirements of the Observatory and the opportunities and threats that are likely to arise – e.g., technological, economic, and political. [R0.2.1.4]
- To work with IRIS infrastructure providers and relevant third parties to ensure their infrastructure roadmaps and Digital Asset work is aligned to UK DF requirements.
- To identify synergies and pursue opportunities for collaboration with peer experiments (e.g., SKA) in the IRIS programme.
- Maintain a data-management plan that: catalogues UKDF-related data assets; attributes value, longevity, and risk to data assets; and ensures sufficient backup and recovery processes are in place to protect Rubin and UK DRP data assets.
- To support and promote collaborative working across the three DFs, to create a resilient and flexible platform for DRP.
- To engage with other IDAC providers to develop a shared plan for supporting the Rubin astronomy community and to highlight and devise solutions for additional computational challenges that will face IDAC providers (for example, related to movement and ingestion of data products).

4.3.2 Data Release Processing

The purpose of the Data Release Processing in-kind contribution is to complete the assigned 25% of processing work in a timely manner, during telescope operations, including the design, implementation and testing of an appropriate technology solution on IRIS infrastructure.

In the run-up to telescope operations, a series of campaigns of increasing scale and complexity is envisaged, with the objective to prove the infrastructure and software configuration at and between each DF.

To this end, high-level DRP team responsibilities are, as follows:

- To ensure suitable storage is provisioned, within the UK DF, to receive input images (raw and intermediate) from the US and French Data Facilities as required by Campaign Management.
- To ensure sufficient and appropriate computing capacity and working storage is available, within the UK DF, to perform processing tasks in a timely manner, as required by Campaign Management.
- To ensure output data is transmitted to the US and French DFs, from the UK DF, as required by Campaign Management.
- To contribute to the science validation of each in-progress Campaign, in collaboration with the V&V team.
- To provide reasonable assistance to the US and French DFs, to resolve localised or inter-facility operational issues that may impact on the timely completion of a Campaign.
- To install, configure, and run ancillary and supporting services and middleware (for example, management databases, monitoring/ logging systems, and data access mechanisms) as required by Campaign Management and by Verification and Validation, to allow them to track progress of the subset of the Campaign running at the UK DF and to identify and diagnose any issues with a particular Campaign.
- To liaise with IRIS and UK DF infrastructure providers to mitigate any maintenance or other planned infrastructure outages that may be required from time-to-time (for example, for replacement of equipment at end of life).
- To represent the UK DF at Data Production planning and progress meetings.
- To liaise with key technology providers (e.g., Rucio, PanDA, and FTS) to ensure DRP requirements are considered, and to formulate a software-hardware configuration that can fulfil the UK in-kind contribution to DRP.

- To work with IRIS infrastructure providers to mitigate incidents (e.g., unexpected service outages) and problems that are identified, putting the UK DF contribution to DRP in jeopardy.

4.3.3 Independent Data Access Centre

The purpose of the UK IDAC is to make available to the Rubin science community topical LSST science data products (Data Releases and Date Previews) along with computing capabilities to support the anticipated science requirements of 20% of the Rubin science community.

To this end, high-level IDAC team responsibilities are, as follows:

- To host a *full* copy of the two most recent LSST Data Releases including Processed Visit Images, Deep Coadd Images, and science catalogues [R0.2.4.5, R0.2.4,6].
- To ingest DP and DR datasets in a timely manner and publish in a form that can then be exposed to end-users via the agreed user interfaces [R0.2.2.1, R0.2.2.3, R0.2.2.4].
- To align the service offering of the UK IDAC, where possible, with the Project DAC at the US DF (to provide a consistent experience for users and promote opportunities for closer collaboration on the development and operation of DAC services) [R0.2.1.2, R0.2.4.12].
- Maintain an instance of the Rubin Science Platform (RSP) with sufficient capacity and functionality to accommodate 20% of the anticipated international Rubin community [R0.2.4.2, R0.2.4.8, R0.2.4.9, R0.2.4.10].
- Support a 24×7 mode of use with a level of reliability consistent with peer research infrastructures [R0.2.1.3].
- To limit access to the Rubin datasets (and derived datasets) – as described in [11] – held on the UK IDAC to Rubin Data Rights Holders [R0.2.1.6, R0.2.4.7, R0.2.4.11].
- To enable periodic scale-up of the UK IDAC supporting infrastructure to meet expected capacity growth during early operations, without impacting the user experience [R0.2.4.4].
- To contribute to Rubin science support activities, ensuring users of the UK IDAC have a comparable support experience to users of the Project DAC [R0.2.5.1—R0.2.5.8].
- To liaise with IRIS and UK DF infrastructure providers to mitigate any maintenance or other planned infrastructure outages that may be required from time-to-time (for example, for replacement of equipment at end of life).
- To work with IRIS infrastructure providers to mitigate incidents (e.g., unexpected service outages) and problems that are identified, putting the UK IDAC in jeopardy.

4.3.4 Other In-kind Contributions

The UK IDAC will host both additional services as well as derived data products (UGPs), developed by LUSC DEV activities and encapsulated in in-kind contributions, to enhance the opportunities for scientific discovery in priority UK science areas.

To this end, the UK IDAC team are likely to have additional responsibilities, as follows:

- Provide sufficient capacity and capabilities to enable funded DEV work packages to progress their programme of work [R0.2.1.1, R0.2.4.13—R0.2.4.15].
- Provide access to the required input datasets or Rubin services to support the production of those user-generated products that form an in-kind contribution, as documented in the relevant DAC-DEV interface requirements document [R0.2.2.5, R0.2.5.9, R0.2.4.13].
- Provide sufficient capacity and capability to fulfil specified in-kind contributions (that is, LSST:UK DEV activities) for generation and hosting User-generated Products or

ancillary services that enhance the Rubin Science Platform, as documented in the relevant DAC-DEV interface requirements document [R0.2.4.3, R0.2.4.14]. In some instances, this may require engagement with the IRIS Resource Scrutiny and Application Panel (RSAP) to secure appropriate and sufficient infrastructure for UK in-kind contribution (commonly known as DEV activities) that deliver derived science products or services [R0.2.1.5].

- Ensure sufficient capacity in UK IDAC to host the previous 12 months of LSST Nightly Alerts, along with derived data, images, and output files, in support of the Lasair community broker.
- Provide effort to support/ undertake the preparation of user-generated products for agreed in-kind contributions (that is, DEV activities), as documented in the relevant DAC-DEV interface requirements document.

4.3.5 Networking

The UK Data Facility as outlined above will handle and move significant amounts (tens of Petabytes) of science data, both temporary/ working data, and persistent data with longer-term value, in various forms – including file-/ object-based, database, and streaming, which will require performant and resilient network connections between IRIS site and connecting IRIS into the wider Internet.

Based on discussions with representatives from IRIS, it is understood that data movement at the scale required by DRP can be realised on existing network capacity on the public Internet and relevant academic networks, including the Janet network in the UK[11], the GÉANT network across Europe[12], and Rubin-provisioned networking in the USA (including ESNET[13]).

The Observatory organises regular Rubin Observatory Network meetings (typically, once every six months) at which the topical network requirements are considered in conjunction with network-provider infrastructure plans and conflicting requirements from other major science experiments.

LSST:UK, and the two other Data Facilities, are represented at these meeting, which provide an opportunity to monitor the networking environment and raise an alert in the event that Rubin networking requirements are likely to deviate from the capabilities and capacities that are available.

4.4 Staffing

Staff resources to fulfil the programme of work for LUSC are defined at the work-package level, and organised into, typically, four-year funding phases. At the time of writing, we are in the first year of LUSC Phase C (which runs from April 2023 until March 2027). Within Phase C, funding has been secured for four areas of activity: project management, DAC (that is, finalising and running the UK IDAC), DRP (that is developing, implementing and operation the UK contribution to Data Release Processing); and DEV covering the various research-and-development oriented activities. Most important for operational planning are the DAC and DRP effort levels, which are:

- 5.0 FTE for the UK IDAC (DAC) work packages
- 5.25 FTE for the DRP work packages

[11] https://beta.jisc.ac.uk/janet

[12] https://network.geant.org/

[13] https://www.es.net/

—organised as described below.

4.4.1 DRP Staffing

To fulfil the UK's in-kind commitment to Data Release Processing, both during and in preparation for Rubin Operations, a UK-based team with experts in compute-oriented astronomers, as well as experts in distributed compute and data management. From April 2024, it is intended that the team will realise 5.25 FTE of effort, being assembled and expanded during the first year of the LSST:UK Phase C (that is, April 2023—March 2024).

The team roles have been developed in collaboration with DRP experts from the Rubin Observatory as well as US and French Data Facility Liaisons and captured in a Rubin project-internal Work Breakdown Structure (WBS).

- **Production Scientist** (0.5 FTE) – to contribute to the science validation of DRP and help ensure that Data Release Products are fit for purpose. Specifically, the Data Release Production Scientist will:
  - Oversee UK processing activities, to identify processing problems, off-spec data products, and underperforming processing stages.
  - Undertake a real-time (that is, during production) assessment of the astronomical validity of the data release processing undertaken at the UK facility.
  - Coordinate with other Production Scientists (that is, at US and French DFs) regarding data issues that are not specific to UK DRP.
  - Understand the scientific intent of the processing pipeline and be able to troubleshoot issues with elements of the pipeline configuration.

  The Production Scientist is responsible for Section 4.3.2, Item 4, and has a role to play in Section 4.3.2, Item 8.

- **Processing Scientist** (0.5 FTE) – to be responsible for the completion of DRP in the UK and ensure the timely delivery of data products to maintain processing momentum. The Processing Scientist will be responsible for:
  - Coordinating the curation of datasets via the Data Butler as well as other required database platforms for DRP.
  - Monitoring day-to-day progress with the UK contribution to DRP and measure real-time DRP performance.
  - Undertaking basic validation of UK processing activities, to identify and coordinate resolution of day-to-day processing problems, off-spec data products, or under-performing processing stages.
  - To prepare resource estimates and forecasts for IRIS provisioning
  - To support the QA team by providing suitable sample products and data to enable them to validate UK processing contributions.
  - To liaise with the Data Wrangler, Workflow Manager and Operations Support to prepare for each new Data Release campaign, confirming software and hardware configurations to be used and to collate feedback from UK pre-Data Release testing and validation activities.

  The Processing Scientist is responsible for Section 4.3.2, Item 8 and has a role to play in the following activities: Section 4.3.2, Items 1, 2, 4, and 6.

- **Data Wrangler (Data Curation – Rucio)** (1.0 FTE) – to configure, operate, and maintain the data-distribution, staging and curation, required for DRP—e.g., using Rucio. The Data Wrangler will be responsible for:
  - Contributing to the overall operation of Rucio across the campaign.

- Oversee any bulk-download/ transfer operations and audit location and availability of data assets, in line with campaign requirements.
- Configuring UK endpoints and interfaces to Rucio, to allow UK infrastructure to effectively contribute to DRP.
- Problem-solving for Rucio workflows and configuration elements.
- Liaising with UK (storage and transport) infrastructure providers to enable effective use of Rucio.

The Data Wrangler is responsible for Section 4.3.2, Items 1, 2 (data management) and 3, has a role to play in the following activities: Section 4.3.2, Items 2, 5, and 9.

- **Workflow Manager (Workload Management – PanDA)** (1.0 FTE) – to configure, operate, and maintain the (compute) processing workflow, required for DRP. It is expected that the processing workflow will be implemented using PanDA. The Data Wrangler (WM) will be responsible for:
  - Contributing to the overall operation of PanDA across the campaign.
  - Configuring UK endpoints and interfaces to Panda, to allow UK infrastructure to effectively contribute to DRP.
  - Install and maintain Pipeline software and dependencies on UK infrastructure.
  - Setting up and monitoring UK elements of the Data Butler service.
  - Problem-solving PanDA workflows and configuration elements.
  - Liaising with UK (computing) infrastructure providers to monitor and improve effectiveness and efficiency of PanDA configuration in the UK.

  The Workflow Manager is responsible for Section 4.3.2, Items 2, 6 (workflow) and has a role to play in the following activities: Section 4.3.2, Items 5, 7, 9, and 10.

- **DRP Operations Support** (1.0 FTE) – a Research Software Engineer who will have a thorough understanding of the whole UK contribution to DRP and be able to trouble-shoot and contribute to processing tasks as required.
  - Addressing issues with the configuration and operation of the DRP environment
  - Liaise with network providers (transit and endpoints) regarding efficient transport of data products to processing sites, etc.
  - Liaising with the Rubin Infrastructure Team regarding implementation, configuration and optimisation of the DRP environment.

  The DRP Operations Support role is responsible for items in Section 4.3.2 (specifically, Items 5 and 9) and has a role to play in Section 4.3.2, Item 10.

- **Fabric Administration** (1.0 FTE) – effort assigned to IRIS provider sites to address fabric management, as required for the LSST:UK contribution:
  - VO management, security, accounting, operations,
  - Networking, Security and incident response, Accounting.

  Fabric Administration has a role in the following items: Section 4.3.2, Items 5 and 6; and Section 4.3.3, Item 4.

- **Data Facility Advisor** (0.25 FTE) – Advisor to the Rubin Observatory regarding Data Production across the three Data Facilities. Move into DF level (over DAC and DRP)
  - Coordinating with peer roles at other Data Facilities to ensure progress of the overall campaign.
  - Overseeing the integration of UK DRP and IDAC services.

  The Data Facility Advisor is responsible for the items noted in Section 4.3.1, Section 4.3.5, as well as Section 4.3.2, Item 10, and Section 4.3.3, Item 10.

4.4.2 IDAC Staffing

To fulfil the UK in-kind commitment to prepare for and to run a full IDAC for the duration of the survey requires a mixed team of computational scientists, DevOps, and computational astronomers. Based on guidance in [9] and, building on experience from DAC activities in LUSC Phases A and B, we plan to provide 5.0 FTE of effort from April 2023, as follows:

- **Catalogue Wrangler** (1.0 FTE) – effort to oversee the ingestion and publication of catalogues into the science platform database (Qserv)
  - Managing needed Data Butler instances, as well as other required database platforms for DRP.
  - Transfer, ingest and registration (with TAP) of catalogues from DRs (and DPs)
  - Transfer, ingest and registration (with TAP) of catalogues from UGPs
  - Maintain Qserv database, Butler registry database (likely, PostgreSQL), Prompt Products database (if required)
  - Testing, profiling and optimisation of database configurations

  The Catalogue Wrangler is responsible for Section 4.3.3, Items 1 (databasing), 2 (databasing) and has a role to play in Section 4.3.2, Items 3 and 6.

- **Image Wrangler** (1.0 FTE) – effort to oversee the ingestion of image (and similar) products from processing runs into the science platform storage
  - Transfer and registration (with a Butler) of data products from DRs (and DPs)
  - Transfer and registration (with a Butler) of data products from UGPs
  - Testing, profiling and optimisation of the Butler registry and repository configurations

  The Image Wrangler is responsible for Section 4.3.3, Items 1 (images) and 2 (images), and has role to play in Section 4.3.2, Items 3 and 6.

- **RSP Administrator** (0.5 FTE) – maintain the RSP service deployment on the UK DAC
  - Keep the UK RSP instance in sync with the Project RSP
  - Tailor the UK RSP instance to the UK IDAC hosting environment (for example, selecting and configuring suitable OpenStack implementations of supporting services)
  - Maintain RSP supporting services, such as the Vault service and AAAI module.
  - Tailor the UK RSP instance to the locally offered services (for example, documentation of UK-specific UGPs and software)

  The RSP Administrator is responsible for Section 4.3.3, Items 3, 4, and 6.

- **DAC Service Developer** (0.5 FTE)
  - Provision and maintain supporting services for the DAC (including, the RSP, image and database holdings, and ingestion and transfer services)
  - Advising and supporting astronomers in the use of the RSP APIs to support large-scale, batch processing campaigns on third-party infrastructure (for example, secured through separate IRIS awards or HPC-time awards).

  The DAC Service Developer is responsible for Section 4.3.3, Items 5 and 7.

- **DEV Support** (2×0.5 FTE)
  - Provision and maintain supporting services for the Lasair Community Broker.
  - Profile and optimise the core Lasair dataflows.

- Maintain and refine the DAC-DEV interface definition and implementation of resulting workflows.
- (Potentially, in collaboration with relevant DEV teams), process and publish agreed UGPs from UK in-kind contributions.

DEV Support has responsibility for Section 4.3.4, Items 2, 4, and 5.

- **Science Support** (1.0 FTE)
  - Develop training material (including sample notebooks, presentations, and interactive demonstrators) to support use of the UK IDAC.
  - Be familiar with and able to employ Rubin core training material.
  - Contribute to handling of support queries on the Rubin Community Forum.
  - Lobby on behalf of the user community for improvements to the UK IDAC (or Project DAC) offerings.

  Science Support has responsibility for Section 4.3.3, Item 8.

4.5 Infrastructure

Infrastructure requirements, for those in-kind contributions that require significant computing resources are tracked in a sizing model [2].

At the time of writing, the infrastructure requirements of several contributions are still to be defined – notably, UKD-UKD-10 and UKD-UKD-16. The requirements for these contributions are still to mature to a level that can be factored into the sizing model, though this is expected to happen during 2023—2024, leading to a new version of the sizing model being produced.

5 Priorities

At the time of writing, we are at the beginning of LUSC Phase C, during which time LSST: during which time the Rubin Observatory will enter operations and LUSC contributions will change from development and testing to operations and science support. Based on progress during Phase B (described in Section 3), the following topical priorities have been identified:

1. Data Release Processing is expected to utilise tooling and workflows from the high-energy physics community (in particular, the Large Hadron Collider). Given this, the intention is to contribute UK infrastructure to DRP from the UK grid infrastructure (GridPP[14]), secured through IRIS. At the time of writing, experiments are proceeding using storage and compute resources at two GridPP sites: RAL and Lancaster. This is slightly different to how the French DF and US DF are operating. They effectively contribute resources from a single site (instead of two). Utilising multiple GridPP sites has potential advantages, in terms of resilience and flexibility, though **how multiple GridPP sites are interfaced by DRP campaign management** is unclear. For example, should each site be regarded as a separate endpoint or should the UK be regarded as a single site, with assignment of work to different GridPP sites being handled internally. It is important to resolve this question in the coming months, since it is likely to have a significant impact on future DRP planning (in UK and for the wider Rubin project).
2. Each Data Release will include processed versions of all observations to date. So, for example DR3[15] will include both processed observations from Year 2 and reprocess observations from Year 1 (that were previously processed and published in DR2). It seems reasonable, therefore, **to keep the raw images from each DRP campaign in storage in the UK** (to reduce the amount of data to be transferred for each DR to just the new observations). One option would be to curate raw images on Echo storage at RAL. However, the implications of and mechanism for doing this has not yet been investigated.
3. The UK Data Facility is expected to span multiple IRIS sites. For example, current planning involves DAC services running at Edinburgh and RAL, DRP processing at Lancaster and RAL, and DEV post-processing at Cambridge. For this setup to be effective, it needs to be easy and quick to **move significant data between Data Facility sites**. For example:

   - Moving processed data from Cambridge into the DAC
   - Moving raw images from curated storage in RAL into GridPP for processing (see previous point)
   - Backing up valuable data from the UK DAC onto a reliable, offsite backup (for example, the RAL tape archive).

   At the time of writing, Rucio is the candidate distributed-date management technology for doing this, though the implications of and mechanism for doing this has not yet been investigated.

4. The intention is to use Qserv (a distributed MySQL database developed by the Rubin construction project) to host and serve the astronomy catalogues within a DR as well as User-generated Products from UK in-kind contributions. During Phase B, the DAC team has gained significant experience with Qserv (which is

[14] https://www.gridpp.ac.uk/

[15] Recall DR1 is a special case, covering just the first six months, so the first full-year release is numbered DR2.

a complicated software suite) to the point where we can confidently deploy, upgrade, size and ingest data into the service. However, experience so far suggests that we do not understand how to size the hosting environment appropriately to provide a balanced and performant user experience. For example, some whole-catalogue queries take significantly longer on the UK IDAC than on the Project instance of the RSP, as is the ingestion of catalogue data into Qserv. To this end, effort needs to be invested to **profile and optimise the deployment configuration of Qserv**.

5. At the end of DRP, the Data Release needs to be ingested into the Project DACs, as well as being distributed to and ingested into in-kind IDACs, to be made available to the Rubin astronomy community in a timely way. This is envisaged to be a significant and challenging task that has, to date, not received significant attention from either Rubin or in-kind contributors. There is an urgent need to begin experiments, at scale, for **DR distribution and ingest**, to ensure DR publication is not significantly delayed. This item is also covered in the Phase C Risk Register[16] (see Risk R63).
6. At the time of writing, the UK IDAC team has a good understanding of the infrastructure and handling requirements for UKD-S5 and UKD-S9. However, two other in-kind contributions that potentially have significant and distinct requirements on **UK DF infrastructure are UKD-S10 (TiDES) and UKD-S16 (Adler Alerts)**. DAC-DEV investigations, analogous to those undertaken with UKD-S5 and UKD-S9, should be undertaken in the coming months.
7. In operational discussions with the Rubin Data Production team, a suggestion has been made to **replicate the Prompt Products data from the USDF to the UKDF** to provide greater resilience for the USDF and also greater capacity to support user demand. The Prompt Products service has substantial infrastructure requirements in its own right, plus would place additional demand on UK IDAC staff resources to operate. A feasibility study should be undertaken to determine whether or not it is realistic for the UK IDAC team to take on this responsibility.

[16] https://lsst-uk.atlassian.net/l/cp/qSV2cjzx (last accessed on 10th February 2024).

6 References

[1] G. Beckett, *D2.5.3: Collection of Final Reports for Mini-projects with DEV Activities*, Project Technical Report LUSC-B-38 (June 2023)

[2] G. Beckett and R.G. Mann, *D2.4.3 Infrastructure-requirements forecast for Data Releases during Phases C and D*, Project Technical Report LUSC-B-42 (July 2023)

[3] G. Beckett and M. Read, *D2.5.3 Collection of final reports for mini-projects w/ DEV activities*, Project Technical Report LUSC-B-38 (June 2023)

[4] L.P. Guy et al., *Rubin Observatory Plans for an Early Science Program*, Rubin Technical Report RTN-011, available at https://rtn-011.lsst.io/ (updated 21st June 2023)

[5] M. Jurić, et al., *Data Products Definition Document*, Rubin Technical Report LES-163 (updated 10th July 2023)

[6] LSST:UK Consortium, *LSST:UK In-kind Contributions to the Vera C. Rubin Observatory Legacy Survey of Space and Time*, internal project report (updated 10th July 2023)

[7] P. Marshall et al., *Manual for In-kind Contributors and Recipients*, Rubin Technical Report RDO-41, available from https://docushare.lsst.org/docushare/dsweb/Get/RDO-041 (last updated 28th July 2021).

[8] W. O'Mullane, *Management and Execution plan for Data Management Operations*, Rubin Technical Report RTN-046, available from https://rtn-046.lsst.io/ (last updated 10th July 2023)

[9] W. O'Mullane et al., *Guidelines for Rubin Independent Data Access Centers*, Rubin Technical Report RTN-003, available at https://rtn-003.lsst.io/ (updated 7th August 2021)

[10] W. O'Mullane et al., *DM sizing model and cost plan for construction and operations*, Rubin Technical Report DMTN-135, available at https://dmtn-135.lsst.io/ (updated April 2023)

[11] Rubin Observatory, *Vera C. Rubin Observatory Data Policy*, Rubin Project Report RDO-013, available at https://docushare.lsst.org/docushare/dsweb/Get/RDO-013 (updated 22nd July 2021)

[12] S. Smartt et al., LSST:UK Science Requirements Document, LSST:UK technical report available at https://lsst-uk.atlassian.net/l/cp/7gjTq1B5 (updated June 2022).

[13] S. Voutsinas et al., *D2.4.2: Lessons Learned on Setup and Operation of a Proto-DAC for Commissioning and Data Preview*, Project Technical Report LUSC-B-26 (November 2022)

[14] R. Williams et al., *D2.3.2: Design of Infrastructure for LASAIR*, Project Technical Report LUSC-B-31 (August 2022)

Annex A: LSST:UK IDAC Roadmap

	UKDAC3	UKDAC4	UKDAC5	UKDAC6	UKDAC7
Delivery Date	2023-03-31	2024-03-31	2025-06-30	2026-03-31	2027-03-31
Platform	IRIS	IRIS	IRIS	IRIS	IRIS
Functionality	•RSP components: all	RSP components: all - Inc. Nublado Ver. 3	RSP components: all	RSP components: all	RSP components: all
Data	UKDAC2 holdings, plus WP3.11 (Gaia-CatWISE catalogues)	UKDAC3 holdings plus: • DP0.2 (cats and images) • DP0.3 (cats) • VISTA-HSC (im+cats)	UKDAC4 holdings plus: •DP1	UKDAC5 holdings plus: •DP2	UKDAC6 holdings plus: •DR1
Access	•Authentication: GitHub credentials •Accounts: admin, invited users involved in DEV WPs (e.g. science teams related to WP3.5) and other activities (e.g. Lasair usage).	Auth: GitHub credentials (tbc) Accounts: Small number (~20) interested users	Auth: OpenID Connect (as per project) Accounts: ~100, depending on demand	Auth: OpenID Connect (as per project) Accounts: ~400 (UK/ Intl)), depending on demand	Auth: OpenID Connect (as per project) Accounts: ~1,000 (UK plus Intl), depending on demand
Documentation	Project-provided, supplemented by UK-specific online documentation to support authorised users	Project-provided, supplemented by UK-specific online documentation to support authorised users	Project-provided, supplemented by UK-specific online documentation to support authorised users	Project-provided, supplemented by UK-specific online documentation to support authorised users	Project-provided, supplemented by UK-specific online documentation to support authorised users
Helpdesk	Community forum	Community forum	Community forum	Community forum w/ active engagement from UK DAC staff	Community forum w/ active engagement from UK DAC staff