



D2.5.5: Interface Requirements for DAC-DEV interactions

WP2.5: Science Support

Project Acronym LUSC-B
Project Title UK Involvement in the Large Synoptic Survey Telescope
Document Number LUSC-B-24

Submission date	25/AUG/22
Version	1.2
Status	Final
Author(s) inc. institutional affiliation	George Beckett, Mike Read
Reviewer(s)	Chris Frohmaier (Soton); Ken Smith (QUB)

Dissemination level	
Public	

Version History

Version	Date	Comments, Changes, Status	Authors, contributors, reviewers
0.1	06/JAN/22	MGB prepared cover document to introduce interface requirements so far developed for WP3.5 and WP3.11.	MGB
1.0	17/JUN/22	MGB updated in light of recent experiments in WP3.5 and WP3.11	MGB
1.1	1/AUG/22	Updated in light of comments from reviewers	MAR and MGB
1.2	25/AUG/22	Minor correction to section descriptions in introduction	MAR

Table of Contents

VERSION HISTORY	2
1 INTRODUCTION	4
1.1 PURPOSE.....	4
1.2 GLOSSARY OF ACRONYMS.....	5
2 INTERFACE REQUIREMENTS FOR WP3.11	6
2.1 DESCRIPTION OF INTERFACES.....	6
2.2 STATUS	7
2.3 FUTURE PLANS	7
3 INTERFACE REQUIREMENTS FOR WP3.5	9
3.1 STATUS	9
3.2 FUTURE PLANS	10
4 REFERENCES	11

Index of Figures

Figure 1: Workflow for generation of crossmatch catalogues.	7
Figure 2: High-level workflow for generation fused data products.	10

1 Introduction

1.1 Purpose

A unique element of the LSST:UK programme, and a particular advantage for UK astronomy, is a portfolio of research and development activities (commonly referred to as DEV activities), which are funded as part of the Phase B programme and are working to enhance and extend the astronomy potential of the baseline Rubin Observatory software, services, and data products, for priority astronomy topics within the UK.

Some of these DEV activities need to *interface* with the LSST:UK Data Access Centre (DAC), in order that their outputs can be made available to science users during operations. For example, a DEV activity may produce a piece of software that enables astronomers to analyse LSST data in a different way or may produce a User-generated Data Product (in Rubin Observatory nomenclature) to be co-hosted alongside official LSST Data Release products.

A goal of the WP2.5 team has been to engage early with the affected DEV teams to develop, test and document how the DEV products will interact with the DAC and be supported by the DAC team. An important element of this activity is the production of a set of DAC-DEV Interface Definitions, which will define the interfaces between the DAC and the DEV products in sufficient detail to ensure seamless operation, and to document where the responsibility for different aspects of the interface will lie (that is, with the DAC team or the DEV team).

Based on the anticipated Rubin Observatory timeline, when the Phase B project was being planned, it was expected that sufficient information would be available at the end of 2021 to allow these interface definitions to be finalised. However, the significant delay to construction, plus the change of remit for some DEV activities because of the Rubin in-kind package negotiations, has meant the interfaces were, as at the end of 2021, still being finalised.

Two DEV work packages have been identified, with products (software and/ or datasets) to be integrated into the LSST:UK DAC. These are:

- WP3.11: Crossmatching and Astrometry at LSST Depths
- WP3.5: LSST and Near-Infrared Data Fusion

The DAC team is working with these DEV activities to understand, agree and develop the required interfaces. The topical state of this work is captured in two documents:

- WP3.11 DAC-DEV Interface Requirements [1]
- WP3.5 DAC-DEV Interface Requirements [2]

and summarised in this report.

Both documents are works in progress, reflecting that research and development is ongoing within the respective DEV teams and recognising that some of the underlying DAC services and technologies (which are being produced by the Rubin Observatory) are still in active development.

The above-noted versions of the DAC-DEV interface requirements are intended to document the topical maturity levels of the interfaces, ensure common agreement on roles and responsibilities across the two teams, as well as highlight areas of weakness and focus points for further work.

D2.5.5: INTERFACE REQUIREMENTS FOR DAC-DEV INTERACTIONS

The development of the interface definitions for the two activities are following a similar path:

1. Agree a high-level architecture (or workflow/implementation, as is appropriate) of the interface (engaging with the Rubin Observatory staff or relevant Science Collaborations, to resolve tactical-level uncertainties).
2. Design and run a first data challenge to test the high-level architecture – potentially engaging with early science users, if appropriate.
3. Refine the architecture, considering experience from the first data challenge, engaging with Rubin/ Science Collaborations to address issues with underpinning software/ requirements (for example, Rubin-provided services).
4. Set up a DEV rehearsal to demonstrate that the DEV outputs can be integrated into the DAC.
5. Make any final adjustments and finalise the interface definition.

The two DEV activities are at different stages of maturity. For WP3.11, work is underway on the second iteration (that is, Step 3) of the DAC-DEV Interface Requirements, with efforts focused on: (a) optimising the crossmatch implementation; (b) working with the Recipient Group to precisely define the work-package outputs. For WP3.5, a first data challenge (Step 2) is in progress, though significant changes to the LSST Stack software have delayed this activity.

Sustainability is an important consideration for both DEV activities, as the (DEV) teams who are developing the products are unlikely to continue to develop the products during Rubin Operations. Some of the responsibility for ensuring sustainability will rest with Recipient Groups (through the In-kind Agreement) and this is briefly discussed in the In-kind Manual [3]. However, some of the responsibility will lie with LSST:UK (in particular, the DAC team) and, to help us address this, we have engaged with the Software Sustainability Institute, who piloted a new course (Intermediate Research Software Development) with LSST:UK in December 2021 [4], and with whom we are sharing the in-progress DAC-DEV Interface Requirements documents for advice and guidance.

In Section 2 and Section 3, we introduce the two Work Package-specific documents and outline the expected timeline to finalising the respective interface definition.

1.2 Glossary of Acronyms

DAC	Data Access Centre
DEV	Research and Development Activities
SSI	Software Sustainability Institute
RSP	Rubin Science Platform
TVS	Transient and Variable Stars Science Collaboration

2 Interface Requirements for WP3.11

LSST catalogues will be so crowded (even far from the Galactic Plane) that standard algorithms for cross-matching objects observed by Rubin with those from other surveys will fail. Hence, the team will provide (through the DAC) a service which uses state-of-the-art cross-matching algorithms that include the effects of crowding and partially mitigates them. The team has proposed to provide matches for six surveys – namely, VISTA, Gaia, WISE, SkyMapper, Schmidt, and 2MASS – with the ability to extend to EUCLID, giving the UK a significant advantage in exploiting LSST data. In addition, the team will provide algorithms which calculate the effects of crowding on proper motions and parallaxes.

The objectives of WP3.11 are:

- O1: Understanding the effects of crowding and proper motions on LSST astrometry
- O2: Using that understanding to create a model of the astrometric uncertainties
- O3: Combining that model with photometric information to provide crossmatches between LSST and other catalogues

This work has been incorporated into the LSST:UK in-kind contribution (UKD-UKD-S9) to the Rubin Observatory as non-directable software-development effort, to deliver the software (call Macauff), plus documentation, ready to be used by the UK DAC (and Recipient Group). The principal recipient group is the Transient and Variable Stars Science Collaboration (TVS).

The Work Package has been extended through until the end of Phase B (that is, March 2023).

2.1 Description of Interfaces

Work between the DAC team and the WP3.11 team has identified and documented seven interfaces (with a possible eight). These address a workflow (illustrated in Figure 1) and described in full detail in [1]. The workflow starts with the selection of an ancillary survey against which a Rubin Data Release will be cross-matched and completes once a new crossmatch table is ingested into and published via the UK DAC.

Both the ancillary survey and the LSST Object Catalogue need to be pre-processed (reduced) into a lean CSV form, containing only the required information. Given this pre-processing needs to be done for each input catalogue, it needs to be simplified and automated as much as possible.

The core of the crossmatch process is to be run in parallel on IRIS HPC resources (at the time of writing, benchmark runs are being conducted on the Cambridge CSD3 system). The aim is to complete the crossmatch process in two weeks or less, to ensure timely publication into the UK DAC.

Once processed and validated, the crossmatch tables will be ingested into a Qserv database running on the UK DAC and published to end users via the Rubin Science Platform.

Each crossmatch catalogue (at least, those involving an LSST Data Release) are subject to LSST data-access restrictions, so may only be published to Rubin Data Rights Holders.

2.2 Status

As alluded to in Section 2.1, a high-level understanding of the interface requirements has been documented (see Figure 1), though uncertainty about the precise specification of the in-kind contribution that WP3.11 will make to the TVS has delayed finalisation of interface definitions.

A data challenge was completed during 2021 to test the high-level requirements. This highlighted two issues:

- The current implementation of the crossmatch algorithm cannot process LSST-scale datasets in a sufficiently timely manner. Work is underway, at the time of writing, to address this, though that work is not expected to significantly affect the interfaces.
- Because there is no standard for publishing astronomy survey data, each ancillary survey to be crossmatched against LSST during Rubin operations will require dedicated effort to transform the survey data into a canonical form (referred to in the interface requirements as skinny-table format). For legacy surveys, it is possible for the WP3.11 team to do this preparatory work, though the inclusion of any ancillary surveys that are contemporary to LSST will require effort from the DAC team and/ or representatives from the astronomy community.

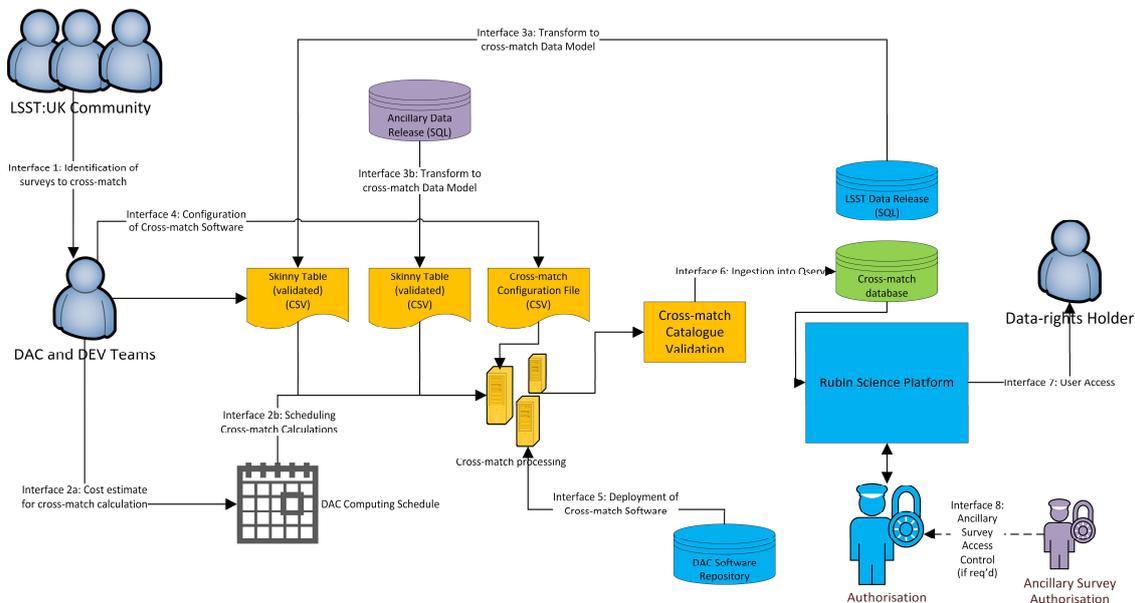


Figure 1: Workflow for generation of crossmatch catalogues.

2.3 Future Plans

From an interface-definition perspective, the following tasks remain:

- Formalise details of in-kind contribution, potentially as a set of acceptance tests to be passed by the team, and sustainability roles for DAC and RG. (August 2022)
- Define a generic process for preparing ancillary surveys to be crossmatched against LSST. (September 2022)
- Profile and optimise the Macauff code – aiming to achieve timely processing. (July 2022)
- Complete a full-scale data challenge using a release candidate of the crossmatch software, to confirm the operational model for the outputs. (October 2022)

D2.5.5: INTERFACE REQUIREMENTS FOR DAC-DEV INTERACTIONS

- Add IRIS infrastructure requirements, for completing this in-kind contribution, to the UK DAC infrastructure sizing model, to ensure requirements are fed into IRIS procurement and provisioning forward estimates.
- Confirm process for ingesting output, crossmatch catalogue into UK DAC, including publishing through the RSP. (November 2022)

3 Interface Requirements for WP3.5

The Work Package 3.5 team is developing an extension to the LSST Stack software (sometimes referred to as the LSST Science Pipeline) to support joint processing of LSST pixels together with pixel data from the ESO VISTA Public Surveys, to produce merged optical and near-infrared catalogues. These catalogues are called (in Rubin Observatory nomenclature) User-generated Products. They will significantly extend the capabilities of the baseline LSST data releases and have already been accepted as part of the LSST:UK in-kind contribution to the Rubin Observatory (UKD-UKD-S5).

VISTA is an ideal candidate because VISTA surveys coincide well with the LSST survey, in terms of sky coverage. Further, the pixel sizes of the two surveys are similar, simplifying the joint-processing task.

The in-kind contribution focuses on high-latitude, extragalactic surveys (from VISTA). Notably VHS, VIKING, VIDEO and VEILS, which all overlap with the LSST Wide, Fast Deep Survey and the LSST Deep-Drilling Fields. These VISTA surveys are all due to be completed in advance of the beginning of Rubin Observatory operations.

WP3.5 will produce a collection of VISTA deep coadd images, which coincide with LSST images and extend the range of available observing bands, supplemented by object catalogues in the LSST schema plus new measurement catalogues that exploit information from the combined VISTA-LSST coadd image sets. These products will be made available to the Rubin Community via the UK DAC.

The primary recipient group for the contribution is the NOIRLab Community Science and Data Centre (the Galaxies and AGN Science Collaborations are also recipients).

3.1 Status

At the time of writing, a high-level workflow for producing merged catalogues and processed VISTA images has been defined (see Figure 2). Further, the first part of the workflow (ingesting and processing VISTA images) has been tested with a topical version of the LSST Stack.

Progress has been delayed a little due to changes in the underlying LSST Stack software. Specifically, the Rubin Observatory Data Management team is making significant changes to the LSST Stack middleware (moving from Generation 2 to Generation 3) and these changes have required that existing workflow steps are revalidated. The change has also affected the way in which VISTA and LSST catalogues are merged and published into the UK DAC.

D2.5.5: INTERFACE REQUIREMENTS FOR DAC-DEV INTERACTIONS

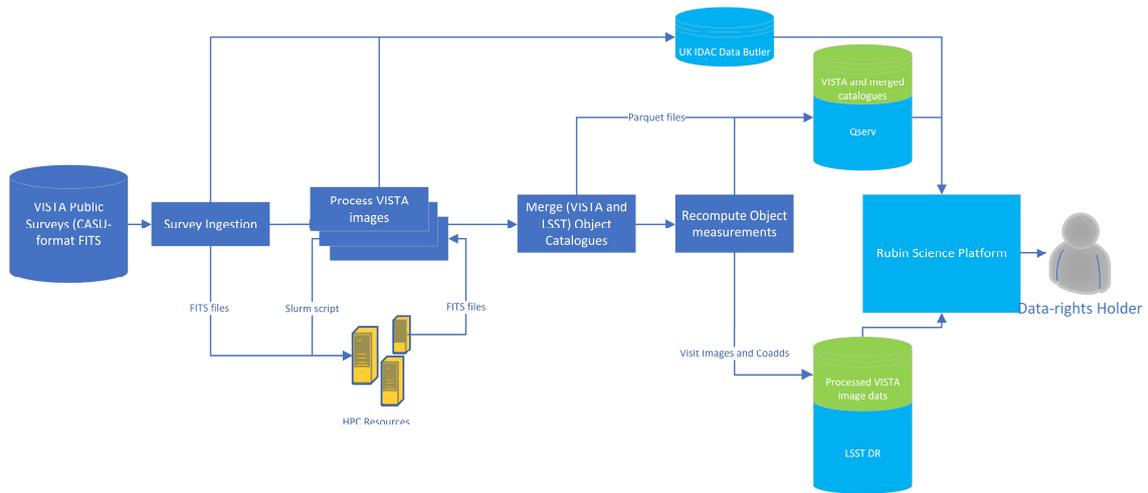


Figure 2: High-level workflow for generation fused data products.

3.2 Future Plans

From an interface-definition perspective, the following tasks remain:

- Complete a full-scale processing run, merging appropriate VISTA surveys with processed data from the Hyper Suprime-Cam (HSC) telescope, which is a precursor for LSST. (June 2022)
- Finalise approach to merging VISTA and LSST object catalogues, and ingesting pipeline outputs into the DAC database service (called Qserv). (August 2022)
- Ingest images (Butler repository) for HSC and processed VISTA survey images into UK DAC (September 2022)
- Confirm process for ingesting output, crossmatch catalogue into UK DAC, including publishing through the RSP. (November 2022)

4 References

- [1] G. Beckett et al., “WP3.11 DAC-DEV Interface Requirements”, LUSC-B-20 (draft, last updated November 2021)
- [2] G. Beckett et al., “WP3.5 DAC-DEV Interface Requirements”, LUSC-B-22 (draft, last updated November 2021)
- [3] Vera C. Rubin Observatory, “Manual for In-kind Contributors and Recipients”, RDO-41 (July 2021)
- [4] The Software Sustainability Institute, “Intermediate Research Software Development in Python”, web-based resource <https://carpentries-incubator.github.io/python-intermediate-development/> (accessed 7th January 2022)