



D2.6.6 Documentation for near-IR and crossmatch data products on UK RSP

WP2: LUSC-DAC

Project Acronym LUSC-C
Project Title UK Involvement in the Large Synoptic Survey Telescope
Document Number LUSC-C-03

Submission date	21/DEC/2023
Version	1.0
Status	Final
Author(s) inc. institutional affiliation	George Beckett, Amanda Ibsen, Mike Read, Stelios Voutsinas – University of Edinburgh
Reviewer(s)	Elham Saremi (Southampton), Philip Holloway (Oxford)

Dissemination level	
Public	

Version History

Version	Date	Comments, Changes, Status	Authors, contributors, reviewers
0.1	21/DEC/23	Initial write-up	MGB
0.2	5/JAN/24	Minor revisions	MGB, MR
0.3	5/FEB/24	Updated in light of reviewers' feedback	MGB
1.0	11/MAR/24	Approved by Executive Group. Meta-data updated prior to publication	TMS

Table of Contents

VERSION HISTORY	2
1 BACKGROUND	4
1.1 GLOSSARY OF ACRONYMS.....	4
2 DOCUMENTATION OVERVIEW	5
2.1 OVERVIEW DOCUMENTATION ON SQUARE ONE.....	5
2.2 SCHEMA BROWSER	6
2.3 TECHNICAL DOCUMENTATION VIA NUBLADO NOTEBOOKS	7
3 CONCLUSIONS AND FURTHER WORK	9
4 REFERENCES	10

Index of Figures

Figure 1: Rubin Science Platform landing page	5
Figure 2: Top-level documentation in RSP Square One.....	6
Figure 3: Firefly's built-in schema browser being used to interrogate schema for the Gaia-CatWISE crossmatch catalogues.	7

1 Background

One of the advantages of LSST:UK operating an Independent Data Access Centre (IDAC) is the opportunity to host additional datasets and tools focused on UK science priorities [4]. In addition to hosting these datasets, the UK IDAC team need to actively publicise and describe them, ensuring that astronomers can grasp the potential benefits of these datasets and confidently integrate them into their research. These datasets are expected to provide a valuable complement to Rubin data products published during the Early Science Program [2].

The deliverable outlines efforts to document two specific UK-provided datasets. The first dataset involves a crossmatch catalogue for the Gaia EDR3 and CatWISE surveys (an early output of the Rubin in-kind contribution, UKD-S9) [1]. The second dataset comprises a near-infrared fused dataset that integrates HSC imaging with imaging from several overlapping VISTA surveys (VIKING, VIDEO and VHS) [1].

The described work not only enhances the accessibility and usability of the specific datasets but also establishes a methodology for documenting other ancillary datasets and supplementary services.

1.1 Glossary of Acronyms

ADQL	Astronomical Data Query Language
API	Application Programming Interface
DAC	Data Access Centre
DF	Data Facility
DP	Data Preview
DR	Data Release
DRH	Data Rights Holder
DRP	Data Release Processing
IDAC	Independent Data Access Centre
IVOA	International Virtual Observatory Alliance
LSST	Legacy Survey of Space and Time
LUSC	LSST:UK Science Centre
RSP	Rubin Science Platform
TAP	Table Access Protocol

2 Documentation Overview

Astronomers are expected to engage with the LSST survey and related datasets using the Rubin Science Platform (RSP). This promotes the philosophy of bringing the compute to the data and provides three different interface classes:

- A browser-based, interactive, graphical interface called Firefly.
- A script-based notebook platform, known as Nublado.
- A suite of Application Programming Interfaces (APIs) based on International Virtual Observatory Alliance (IVOA) and equivalent standards.

Complementing the three access mechanisms, there are three obvious documentation opportunities within the science platform:

1. High-level overview and background documentation provided through the RSP landing page, called Square One.
2. Technical documentation of the catalogue structures accessible via the Firefly schema browser.
3. Technical documentation for programmers available through example (Nublado) notebooks.

For the two ancillary datasets under consideration, we have used all three of these documentation platforms.

2.1 Overview Documentation on Square One

The landing page for the Rubin Science Platform (Figure 1) provides direct access to several useful resources for end users, including documentation on both the science-platform's functionality and the datasets that are hosted on the platform.

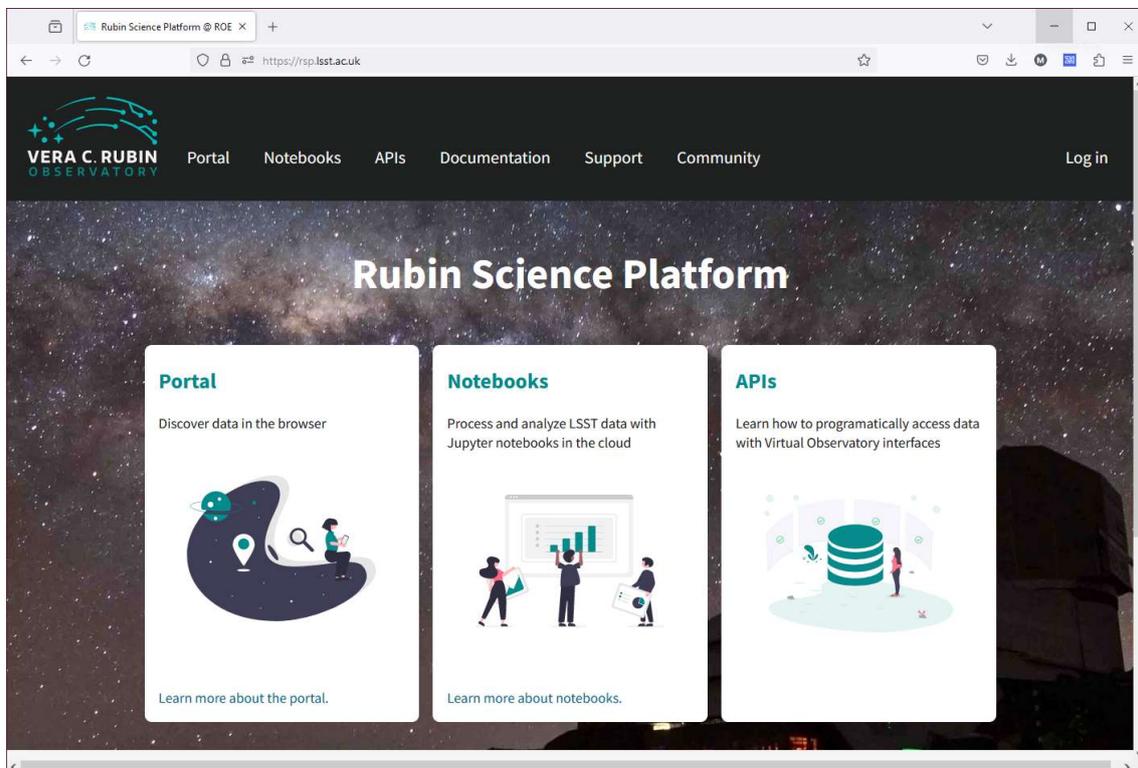


Figure 1: Rubin Science Platform landing page

Much of the documentation is inherited from the Rubin Data Management team and is accessible without significant intervention. However, there is an opportunity to tailor the documentation to site-specific requirements and configurations.

At the time of writing, we have made initial experiments with localisation on the UK RSP instance (available at <https://rsp.lsst.ac.uk/>), including:

- Removing references to datasets not currently hosted in the UK IDAC.
- Adding descriptions of the two ancillary datasets – that is, Gaia-CatWISE crossmatch data and VISTA-HSC fused data.

An example of the output is presented in Figure 2.

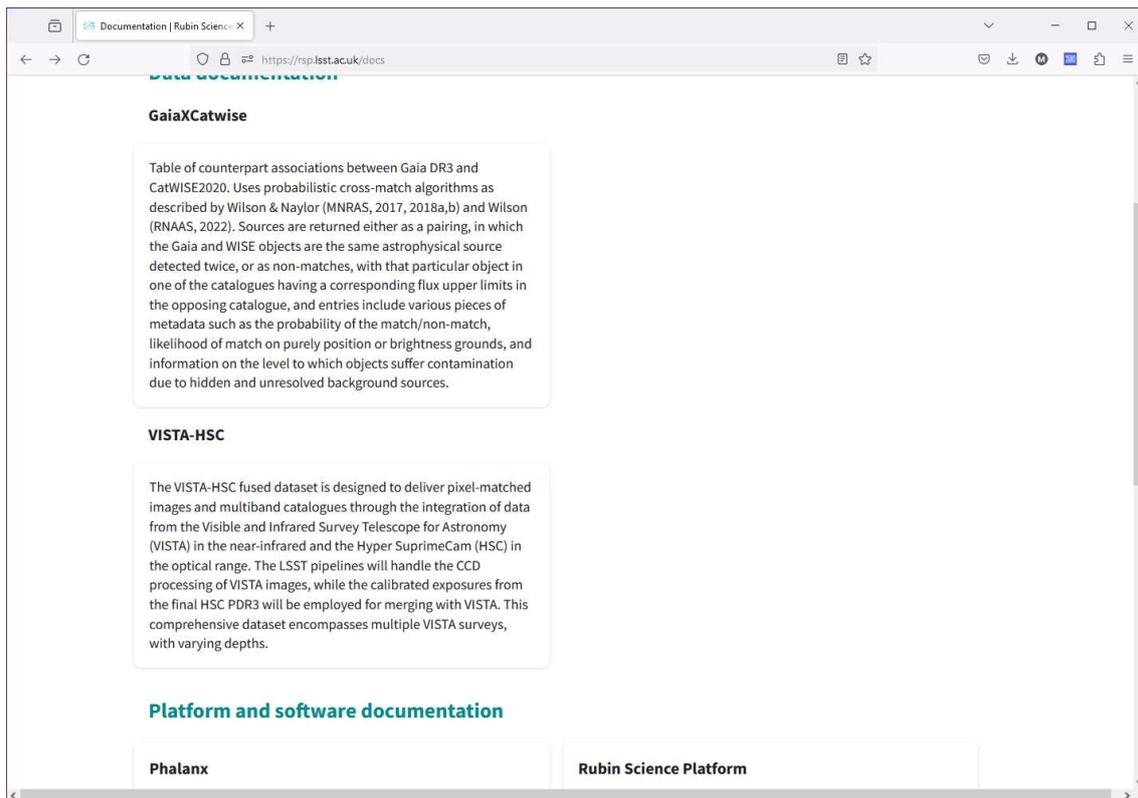


Figure 2: Top-level documentation in RSP Square One.

2.2 Schema Browser

The browser-based Firefly interface provides interactive access to astronomy catalogues and, to a lesser extent, images. In the RSP, catalogues are commonly accessed in a relational database, (usually, Qserv) based on Astronomical Data Query Language (ADQL) queries. Firefly includes a tailored query interface that allows astronomers to perform basic catalogue searches without writing actual ADQL. However, on many occasions, the astronomer may prefer or need to input ADQL directly to access the specific data of interest.

To help with building ADQL queries and, more generally, understanding the organisation of data in catalogues, Firefly includes a simple schema browser (Figure 3). The presentation of the schema can be enhanced by providing human-readable summaries of the contents of each field. While summaries are provided for Data Release Products, they need to be generated manually for derived data products such as Gaia-CatWISE crossmatch data and VISTA-HSC fused data.

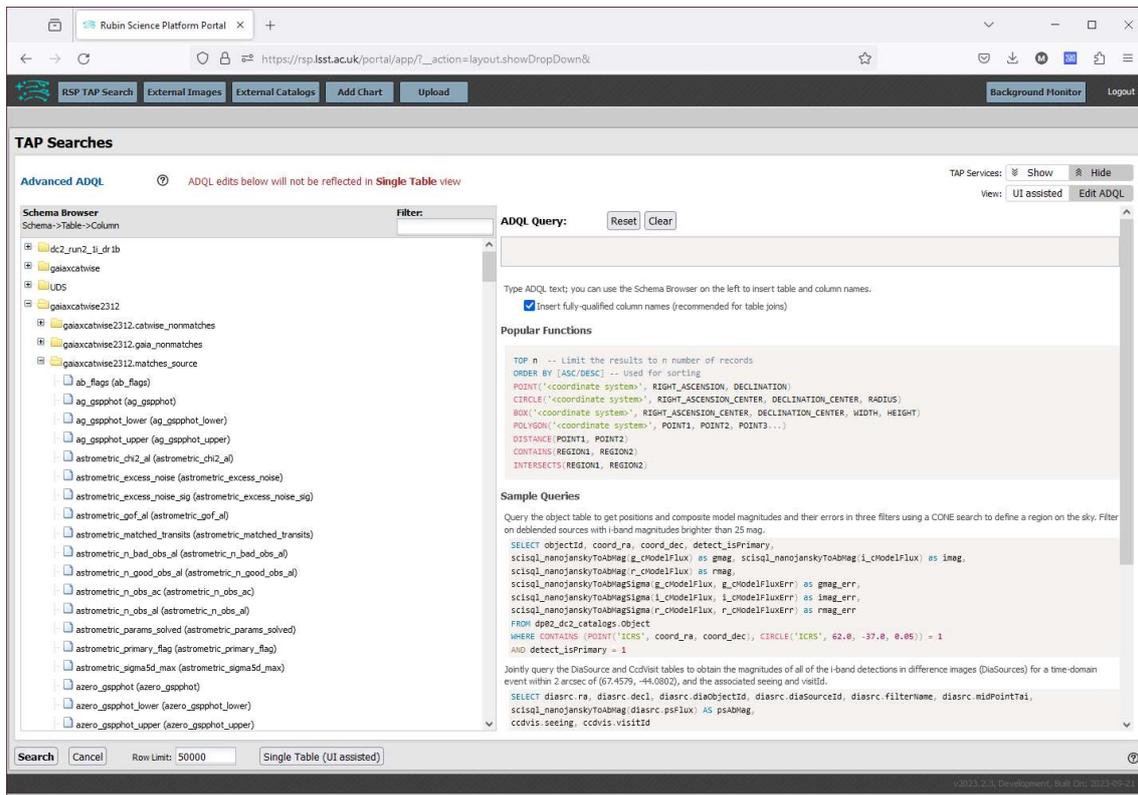


Figure 3: Firefly's built-in schema browser being used to interrogate schema for the Gaia-CatWISE crossmatch catalogues.

2.3 Technical Documentation via Nublado Notebooks

It is anticipated that most research studies, using LSST data, will be done via the Nublado notebook interface. To illustrate how an astronomer can access datasets, the UK IDAC team has developed a small collection of example notebooks. These notebooks are hosted on the LSST:UK GitHub repository and are available to check out into a user's home storage on the science platform.

At the time of writing, we have provided the following notebooks using either VISTA-HSC fused data or Gaia-CatWISE crossmatch data as an example:

- **introduction.ipynb** – a simple notebook that illustrates how to query a Data Butler.
- **image_display.ipynb** – how to identify and then access an image from the Butler registry.
- **cutout_images.ipynb** – creating a cutout image around an object of interest.
- **visualization_using_firefly.ipynb** – accessing the image visualization capabilities of Firefly from within a notebook session.
- **simpleTAPDemo.ipynb** – accessing a catalogue via the Table Access Protocol (TAP) interface.
- **Crossmatchusecase.ipynb** – accessing crossmatch information for a selection of objects of interest.

Information on how to check out the notebooks is provided on the LSST:UK Confluence wiki space, where it is accessible to early adopters.

In the next stage of the UK DAC rollout, we expect to make the repository hosting the example notebooks public. We also plan to move the documentation on how to clone the example notebooks into user space within the Square One environment.

Documentation for near-IR and crossmatch data products on UK RSP

The Rubin Data Management team also provide many sample notebooks. At the time of writing, many of these are tailored specifically to the Rubin instance of the RSP, so do not run correctly on the UK IDAC. In addition, the data holdings in the UK IDAC are quite different to those in the Rubin DAC, which adds further incompatibility issues.

Moving forward, we expect improving the alignment of data holdings in the UK IDAC with those in the Rubin DAC (for example, importing Rubin Data Previews). Subsequently, we plan to collaborate with the Rubin DM team to generalise the most interesting notebooks, ensuring compatibility with any DAC/ IDAC (provided the relevant data is present).

3 Conclusions and Further Work

The UK DAC team has built on existing functionalities in the Rubin Science Platform software stack to provide user documentation on several topical derived data products, targeting UK science priorities.

This has required collaboration with the RSP developers – for example, to enable local customisation of Square One documentation – and has exposed some limitations in the current RSP implementation – for example, limited visibility of table schema via the RSP schema browser.

The documentation that has been produced is sufficient early adopters, who are responsible for scientifically validating precursor datasets from in-kind contributions UKD-S5 and UKD-S9 [1]. However, more detailed information is likely to be needed to support a larger user base, with people less familiar with the generation and application of these datasets. For example, something comparable to [3]. The UK IDAC team will work with the in-kind contributors to provide this extra information.

Following on from this, the UK IDAC team intends to ingest current Rubin data previews (that is, DP0.2 and DP0.3) to bring the data holdings of the UK IDAC into closer alignment with those of the Rubin DAC. Following this alignment, the team plans to grant access to Rubin example notebooks for users of the UK IDAC, thereby expanding the available documentation resources.

Further, in collaboration with the developers of the crossmatch catalogues (from UKD-S9), we will look at options to improve the description of the crossmatch-catalogue schema. This improvement may be implemented within the Firefly schema browser if it is capable, or through alternative means if needed.

4 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] 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)
- [3] M. Jurić, et al., *Data Products Definition Document*, Rubin Technical Report LES-163 (updated 10th July 2023)
- [4] 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)