



# LSST:UK DAC User Documentation

## *Work Package 2.5, Deliverable D2.5.2*

**Project Acronym** LUSC-B  
**Project Title** UK Involvement in the Legacy Survey of Space and Time  
**Document Number** LUSC-B-36

<b>Submission date</b>	24/MAR/2023
<b>Version</b>	1.0
<b>Status</b>	Final
<b>Author(s) inc. institutional affiliation</b>	Bob Mann and George Beckett (University of Edinburgh)
<b>Reviewer(s)</b>	Sebastian Hoenig (University of Southampton), Gavin Ramsay (Armagh Observatory)

<b>Dissemination level</b>	
Public	

## Version History

<b>Version</b>	<b>Date</b>	<b>Comments, Changes, Status</b>	<b>Authors, contributors, reviewers</b>
0.1	21/DEC/22	Initial draft	RGM
0.2	14/FEB/23	Complete draft	RGM
0.3	15/FEB/23	Minor changes	MGB
0.4	24/MAR/23	Minor changes in response to reviews	RGM
1.0	27/APR/23	Marked as Final following Executive Group approval	TMS

## Table of Contents

<b>VERSION HISTORY</b> .....	<b>2</b>
<b>1 EXECUTIVE SUMMARY</b> .....	<b>4</b>
<b>2 INTRODUCTION</b> .....	<b>5</b>
2.1 PURPOSE.....	5
2.2 GLOSSARY OF ACRONYMS.....	5
<b>3 REQUIREMENTS FOR UK DAC DOCUMENTATION</b> .....	<b>6</b>
3.1 RESOURCES REQUIRING DOCUMENTATION .....	6
3.2 EXISTING SOURCES OF DOCUMENTATION .....	6
3.3 THE NEED FOR DOCUMENTATION SPECIFIC TO THE UK DAC .....	7
3.3.1 <i>Lasair documentation</i> .....	7
3.3.2 <i>Documentation relating to data and software from UK in-kind contributions</i> .....	7
3.3.3 <i>IRIS-specific documentation</i> .....	8
<b>4 GUIDELINES FOR PRODUCING UK DAC DOCUMENTATION</b> .....	<b>8</b>
<b>5 REFERENCES</b> .....	<b>11</b>

## **1 Executive Summary**

In the light of the Covid-induced delay to the start of survey operations, the scope of this Deliverable has been changed, from a release of documentation through the UK's Data Access Centre (DAC) to a discussion of the requirements for such documentation in future. The areas in which UK DAC documentation is likely to be required are identified, and a set of principles – non-duplication, sufficiency, currency and use of similar format, jargon and style to Rubin documentation – for their development is outlined. These principles should inform the creation of a UK DAC documentation plan early in Phase C, which will provide a specification of each planned piece of documentation and which is intended, thereby, to aid their production.

## 2 Introduction

At the time that the LSST:UK Science Centre (LUSC) Phase B proposal was written, in 2018, it was expected that the LSST survey would begin in late 2022, and that the UK Data Access Centre (DAC) would be supporting users analysing data from what has become the Rubin Data Preview programme (plus, possibly, Rubin Commissioning) before the end of Phase B. The Phase B DAC Roadmap [1], therefore, included a phased development of user documentation and this Deliverable was foreseen as being the release note for a suite of documentation to be used by a broad community of UK users.

The delay in the Rubin Observatory schedule due to the Covid-19 pandemic and the decision that Commissioning data will be accessed through only the US DAC have had a knock-on effect on the schedule for the development of the LSST:UK Data Access Centre (DAC), while the adoption of the in-kind model for international contribution has affected its scope, since it will eventually be open to Rubin data rights holders outside the UK, under the terms of one of the UK's in-kind contributions (UKD-UKD-S3) . These factors led to a revision of the DAC Roadmap[2], which now envisages only a small set of users – primarily LUSC staff, plus a few external scientists involved in the validation of data products generated by the Phase B DEV programme – having access to the UK DAC before the end of Phase B.

In light of those changes, there is no need for delivery of a full suite of user documentation for the UK DAC prior to the end of Phase B – and, indeed, the system is still not sufficiently mature for producing that to be sensible – so the scope of this Deliverable has been changed, better to reflect what is feasible and useful at this point in the project.

### 2.1 Purpose

The purpose of this Deliverable is now to capture high-level requirements for future UK DAC documentation and to outline the DAC team's current thoughts on how that should be developed and deployed. A more detailed UK DAC documentation plan will be developed early in Phase C ahead of the first UK DAC release to be made accessible to a wider user community; Phase C Deliverable D2.6.1 comprises support material for Data Preview 1.

### 2.2 Glossary of Acronyms

DAC	Data Access Centre (or UK Data Access Centre work packages)
DEV	Development work packages
DPDD	Data Products Definition Document
LSST	Legacy Survey of Space & Time (was Large Synoptic Survey Telescope)
LUSC	LSST:UK Science Centre
RSP	Rubin Science Platform

## 3 Requirements for UK DAC documentation

### 3.1 Resources requiring documentation

As discussed in greater detail in [2], the UK DAC is expected to contain

- Images and catalogues from LSST Data Releases;
- A copy of the LSST Prompt Products Database (TBC);
- The Lasair<sup>1</sup> event broker; and
- Data products and software generated by UK in-kind contributions.

Lasair may retain its own web interface, but, otherwise, these resources will be accessed through the Rubin Science Platform (RSP[4]). The RSP is a set of online services, developed by the Rubin Observatory for its own DACs, and featuring three user-facing *aspects*:

- The **Portal** aspect: based on the Firefly[5] framework, this is a browser-based tool for interactive exploration of LSST data;
- The **Notebook** aspect: this is a tailored deployment of the JupyterLab[6] system, providing a Python scripting environment for analysing LSST data; and
- The **API** aspect: a set of services implementing International Virtual Observatory Alliance (IVOA[7]) protocols, to support bulk access to LSST data products.

All these resources will need to be fully documented if they are to be used effectively by the UK DAC user community.

### 3.2 Existing sources of documentation

The RSP is intended to be self-documenting. The current RSP deployment on the US Interim Data Facility - <https://data.lsst.cloud/> - includes an in-built set of documentation (<https://data.lsst.cloud/docs>), which is divided into three sections, for *Data documentation* (currently a description of the DP0.2[8] dataset), *Science platform documentation* (currently covering the Portal and Notebook aspects of the RSP), and *Software documentation* (currently linking to the documentation for the LSST Science Pipelines[9]).

In addition, the Rubin Observatory is developing a Rubin Science Platform Documentation site (<https://rsp.lsst.io/>), while a wealth of less formally structured documentation relating to the RSP is provided by the Rubin Community Engagement Team through the <https://community.lsst.org> web forum.

More general Rubin/LSST documentation is provided through the <https://www.lsst.io> site, which provides a single, searchable front-end to the large body of public Rubin/LSST documents. These are mainly collected into series, each of which provides a document code prefix – e.g., LSST Systems Engineering documents have codes with prefix LSE – and this technical documentation collection comprises a comprehensive body of reference information on the Rubin Observatory and the LSST, some of which will be invaluable to users. For example, the Data Products Definition Document (LSE-163) [12] provides the definitive reference for the Prompt and Data Release Data Products; it includes the definitions and formats of each attribute that will be included in the data

---

<sup>1</sup> See <https://lasair-ztf.lsst.ac.uk> for the current version of Lasair, processing alerts from the Zwicky Transient Facility[3]

release catalogues, which is information of crucial importance to all astronomers who use them. This technical documentation collection also functions as the collective memory of the Rubin construction project, so many of the documents are intended to be read primarily by Rubin staff, so the front page of the [www.lsst.io](http://www.lsst.io) site also includes a set of guides that provide users of LSST software and data with a sign-posted route into the detailed documentation.

Most Rubin documentation is produced using the Sphinx Python Documentation Generator[10] from source files in GitHub and built using *LSST the Docs* platform. That is described in SQR-006 [13], which provides a useful general discussion of best practices in documentation for Open Source software, based on the experience of the DM team during Rubin construction. Its key insight is that many benefits follow from adopting the same approach to the development of documentation as used for code development. For example, keeping the documentation in the same Git repository as the code to which it refers provides a natural way of keeping them in step and also enables the use of continuous integration tools to ensure that examples continue to work as the codebase is developed.

These key principles are implemented by the *Read the Docs* [11] system, and it was only details of the LSST DM build environment – notably the way that the pipeline software is built from code residing in multiple repositories – that necessitated development of the *LSST The Docs* platform, rather than simple adoption of *Read the Docs*. *LSST the Docs* generate URLs in the `lsst.io` domain from the relevant document codes – for example, the Data Products Definition Document (DPDD), being LSE-163, can be found at <https://lse-163.lsst.io/> - while short URLs for many important documents have been defined using the `ls.st` URL shortener; for example, the DPDD can also be found at <https://ls.st/dpdd>.

### 3.3 The need for documentation specific to the UK DAC

Comparison of Sections 3.1 and 3.2 reveals that most of the resources to be hosted in the UK DAC will already be documented by the Rubin Observatory, and, as discussed in Section 4, we plan to direct UK DAC users to existing Rubin documentation where it exists, rather than produce duplicate documentation of our own. There are, however, three categories of documentation specific to the UK DAC that will be required.

#### 3.3.1 Lasair documentation

While the Community Brokers[14] selected to receive the full LSST alert stream will have the same input data, they will process it in very different ways and produce differing sets of derived products and services, so user documentation specific to Lasair must be produced by the LUSC team and published as part of the UK DAC. The Lasair team are currently assessing the user communities that they will support – which may extend beyond the Rubin Data Rights Holders – and the outcome of that exercise will inform the style and content of the documentation to be provided for Lasair.

#### 3.3.2 Documentation relating to data and software from UK in-kind contributions

A number of the UK's in-kind contributions will be producing data to be hosted, and/or software to be run, within the UK DAC. For example, UKD-UKD-S5 (which corresponds to Phase B WP3.5 and Phase C WP 3.2) will be producing LSST-VISTA optical/near-infrared data products to be published through the UK DAC, while UKD-UKD-S9 (Phase B WP3.11/Phase C WP3.3) will produce software to perform cross-matches between catalogues from LSST Data Release and ancillary surveys. In addition, UKD-UKD-S10 (Phase B WP3.2/Phase C WP3.7) will return spectroscopic classifications of transients for publication through Lasair, which will also publish information on solar system objects developed as part of Phase C WP3.6 (Adler, which has yet to be formally accepted as a UK in-kind contribution). If so requested by their respective Recipient Groups, these data

products and software packages may be available elsewhere, but their primary home will be the UK DAC, which must, therefore, include documentation for their users. The primary target audience for this documentation will be Rubin Data Rights Holders, although those in-kind contributions that make data available through Lasair will have to consider its potentially wider audience, if that data is to be available beyond the community of Data Rights Holders.

### 3.3.3 IRIS-specific documentation

The UK DAC will be hosted on a range of resources provided through the IRIS<sup>2</sup> project, so there is likely to be some IRIS-specific documentation needed for the UK DAC. Its scope is, however, currently unclear, since it is hoped that many of the IRIS-specific aspects of the implementation of the UK DAC will be invisible to users. For example, it is hoped that no special authentication or authorisation steps will be required to access the IRIS resources running the UK DAC; they are likely to be using the IRIS Identity and Access Management (IAM[15]) service, but trust relationships with analogous services elsewhere should see users able to use their institutional credentials to access the UK DAC, as they will for LSST DACs. However, the UK DAC will be providing batch computing facilities to LSST data rights holders and, while that is one aspect of DAC provision that has not been developed significantly as yet by the Rubin Observatory, it is unlikely that can be done in a way that makes the resources being employed totally transparent to the user, so there is likely that the UK DAC team will have to provide specific documentation for that. It may also be necessary to make UK DAC users aware of IRIS usage and security policies through UK DAC documentation. The target audience for this documentation will be Rubin Data Rights Holders accessing IRIS resources as part of the UK DAC in-kind contribution, plus, possibly, LUSC team members using IRIS resources as part of their own in-kind contribution work, if that is performed through the UK DAC.

---

<sup>2</sup> <https://www.iris.ac.uk>



## 4 Principles for producing UK DAC documentation

It seems likely that only one LUSC WP (Phase B WP3.7/Phase C WP3.4, supporting low surface brightness science) will be sufficiently integrated with the LSST Stack to require documentation through *LSST The Docs* – and, indeed, its software may be documented as part of the DM pipeline documentation – but the introductory sections of SQR-006 remain a good reference to those planning documentation for the UK DAC, such as the DEV teams that will produce software to be run on data releases in the DAC.

An early<sup>3</sup> deliverable in Phase C will be a documentation plan, which will specify that documentation in detail, but it is clear already that there are a number of basic principles to which it must adhere:

- **Non-duplication.** LSST data products and software will be produced, and documented, by a very large Rubin team, while the UK DAC team will possess limited staff effort for documentation, so it must resist the temptation to produce tailored versions of Rubin documentation. The default position must be to point users to Rubin documentation and only produce original documentation where unique features need to be described.
- **Sufficiency.** The flipside of the non-duplication principle is the requirement that the documentation produced for the UK DAC must be sufficient for its intended goals. Those goals should be expressed in terms of a purpose for a target audience, which may, in turn, be characterised through assumed knowledge (e.g., of key Rubin documents, or academic journal papers). For example, a goal could be something like “*This document is a user guide for version 2.1 of software package X. Supplemented by the DPDD and the contents of Smith & Jones (2023), this document should be sufficient for data rights holders to obtain, and assess the validity of, cross-matches between LSST DR1 and data release Y of ancillary survey Z*”. The sufficiency of the documentation with respect to a particular goal should be testable, to the extent possible, given its nature. In some cases, this will require nothing more than a satisfactory report from agreed reviewers whereas, for some software, it may include the successful running of tests with specified inputs and target outputs. This is clearly related to the acceptance process for deliverables (e.g., from in-kind contributions).
- **Currency.** Keeping documentation current is clearly a key concern, while documentation for different data and software products will have different cadences on which releases are made and, hence, on which documentation may need to be updated. Most will be related to the LSST Data Release schedule, but those (e.g., Lasair and Adler) based on the alert stream will have to determine their own documentation update schedule.
- **Use of similar format, jargon and style.** When UK DAC documentation is required, it should use format, jargon and style similar to Rubin documentation. The second of these is the most important principle, as users may be confused if technical terminology found in UK DAC documentation differs from that in related Rubin documents, so the former must, typically, refer to the latter for the definition of terminology used. Use of similar format and style is preferred, to aid the ready comprehension of UK DAC documentation by users familiar with Rubin documents, but there may be instances adoption of a different style, to establish a UK branding, is appropriate.

---

<sup>3</sup> The first release of the documentation plan should be early enough that it can drive the preparation of the UK DAC documentation that will accompany Data Preview 1, which is currently scheduled for September or October 2024.

In view of these principles, the UK DAC documentation plan should:

- List the documentation to be produced for the UK DAC, and, for each item on that list, record:
  - its intended goals, including the reference documents that support it;
  - the criteria under which it will be assessed as sufficient;
  - the schedule on which it is intended that it will be reviewed and updated;  
and
  - its format and delivery mechanism.

The documentation plan will itself be subject to revision, in the light of user experience. It is not intended to be too formal a document, but, rather, a useful specification of each piece of documentation, which will aid its development.

## 5 References

- [1] *LSST:UK DAC Roadmap and Integration Plan for DEV Activities*, Project Deliverable D2.1.1
- [2] *Update to D2.1.1 based on in-kind agreement*, Project Deliverable D2.1.4
- [3] Bellm, E., et al, 2019, *The Zwicky Transient Facility: System Overview, Performance, and First Results*, PASP, 131, 018002
- [4] Rubin Science Platform Documentation, <https://rsp.lsst.io/>
- [5] Firefly GitHub repository, <https://github.com/Caltech-IPAC/firefly>
- [6] JupyterLab documentation, <https://jupyterlab.readthedocs.io/en/latest/>
- [7] International Virtual Observatory Alliance, <https://www.ivoa.net>
- [8] Vera C. Rubin Observatory Documentation for Data Preview 0.2, <https://dp0-2.lsst.io/>
- [9] LSST Science Pipelines, <https://pipelines.lsst.io/>
- [10] Sphinx Python Documentation Generator, <https://www.sphinx-doc.org>
- [11] Read the Docs, <https://readthedocs.org/>
- [12] Juric, M., et al, Data Products Definition Document, <https://lse-163.lsst.io/> (v3.7, 2021-10-18).
- [13] Sick, J., SQR-006: The LSST the Docs Platform for Continuous Documentation Delivery, <https://sqr-006.lsst.io/>
- [14] Information for Rubin Observatory Alerts and Community Brokers, <https://www.lsst.org/scientists/alert-brokers>
- [15] IRIS Identity and Access Management, <https://www.iris.ac.uk/portfolio/iam/>