

## 2 Work Package Scope and Implementation

Work Package 3.11 will produce three types of output:

- **Cross-match Algorithm** – the team will develop and publish a novel algorithm for more accurate identification of cross-matches between catalogues, especially targeted at the challenging case of crowded fields. Specifically, the algorithms will allow objects in the LSST object catalogue to be matched confidently to corresponding objects in other, relevant, astronomy catalogues.
- **Cross-match Software** – the team will produce a software implementation of the cross-match algorithm that is capable of handling LSST-scale catalogues in production, as part of the UK Data Access Centre.
- **User-generated Products** – the cross-match software will be applied to a number of different third-party catalogues (to be determined) to create new datasets that capture the correspondence between LSST objects and their counterparts in the specified third-party catalogues in a form that can easily be incorporated into end-user analysis tasks. To maximise their relevance, these new datasets will be produced in a timely manner, once all of the relevant input data is available. Because LSST data products will not be available until after the end of WP3.11, the team will only demonstrate the ability to produce these User-generated products, based on representative precursor surveys and/ or LSST data previews.

For each third-party catalogue, a four-step workflow is envisaged:

1. A survey catalogue, to be cross-matched, will be identified.
2. Both the LSST data and the third-party catalogue will be pre-processed into a canonical format suitable for ingestion into the cross-match software. In part, this will involve extracting (reducing) the essential information needed to perform the cross-match.
3. The Cross-match Software will ingest and analyse the survey inputs, creating a new dataset that describes identified correspondences between the two surveys, as well as identifying those sources for which no identified counterpart exists.
4. The cross-match dataset will be added to the UK DAC, in a way that is accessible and useful to DAC users for their science.

Each of these four steps involves one or more interfaces, which are described in more detail in the next section.

The length of the LSST survey (expected to be over 10 years, during 2023—2033) and because, at the time of writing, the WP3.11 team only have funding to develop the cross-match algorithm and software during Phase B (effectively, until 2023), consideration needs to be given to the sustainability of the software.

Over the lifetime of the survey, various changes that will affect the cross-match software (and, possibly, algorithm) should be anticipated:

- The platform on which the software is run will change. Not only will hardware be deprecated and replaced, but also the form of the hosting platform and supporting technologies will change.
- It is possible that bugs will be identified in the software (or algorithm) beyond the end of WP3.11, which need to be fixed or worked around.
- It is possible that new functionality will be required, beyond the end of WP3.11, to address unanticipated opportunities and applications.

Software sustainability, without a funded development team, is very challenging. It is not possible to address the issue completely in this document, but there are steps that can and should be taken to reduce the potential threats, which include:

- It becomes impossible to use the software, because it becomes incompatible with available hosting platforms in the UK DAC.
- It becomes impossible to use the software, because bugs that make the outputs meaningless, cannot be resolved because the understanding of the software and algorithm has been lost.
- It becomes impossible to reproduce/ verify science outputs, because the conditions under which User-generated Products have been produced is not determinable.

To address this, the UK Data Access Centre is taking a number of steps, summarised as follows:

- Software Preservation is at the core of LSST:UK planning and provisioning.
- Software sustainability and accessibility is one of the criterion by which LSST:UK software deliverables are assessed.
- Where possible, software will be developed using an Open Source philosophy, to promote community engagement in the use and development of the software, and to avoid crucial information being inaccessible.
- Where practical to do so, open standards and/ or widely supported interfaces and technologies will be adopted in preference to proprietary or custom interfaces.
- Understanding of the software, and underpinning algorithms, is shared between multiple people, to reduce the risk due to staff movement and consequent knowledge loss.
- Algorithms and software will be supported by design and implementation documentation that is sufficient to allow the work to be reproduced (assuming effort is available to do so).

Key steps that should be taken for Work Package 11, to support these tactics is described in Section 3.

## WP3.11 DAC-DEV INTERFACE REQUIREMENTS

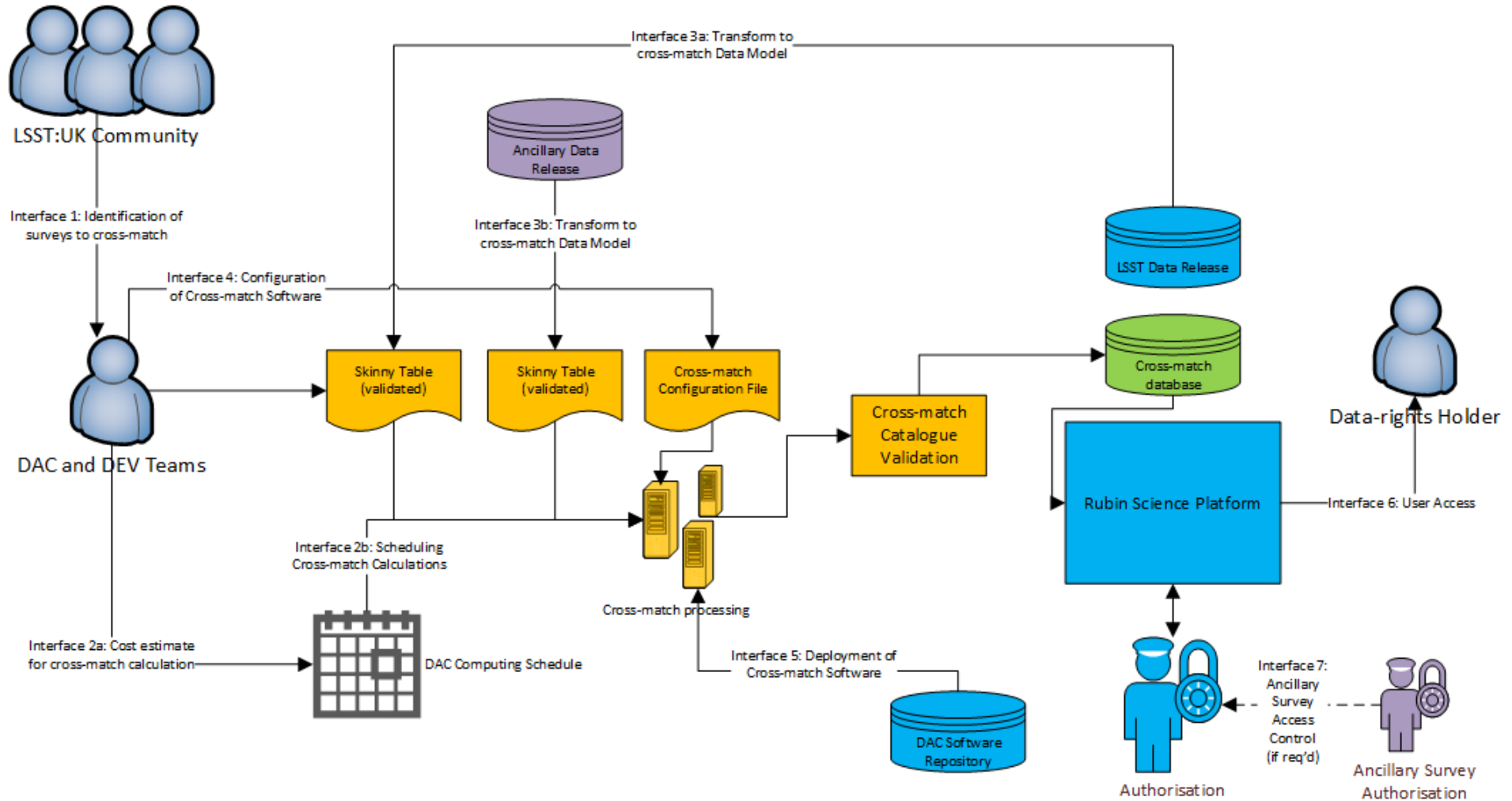


Figure 1: Workflow for generating a Cross-match Catalogue.