Southern Horizons in Time-Domain Astronomy
Proceedings IAU Symposium No. 339, 2019
E. Griffin, ed.

doi:10.1017/S1743921318002685

Towards Science with LSST: Data Products and Communications

Workshop 7

M. L. Graham

Department of Astronomy, University of Washington, Seattle, WA, USA
email: mlg3k@uw.edu

Abstract. The main purpose for holding a Workshop about the Large Synoptic Survey Telescope (LSST) was to move all participants further towards answering the question, "How will I do my science with LSST data?" Presentations included (i) the planned pipelines and products of the data management team, and (ii) the existing channels for communication within the science community and between the community and the LSST Data Management team. In between the formal presentations, small groups discussed matters such as how to select the data products or communications resources that were best suited to individual science goals. The latter discussions were designed both to facilitate engagement with the material and to foster collaboration. Participants should thus have become better equipped to continue on their respective individual paths towards science with LSST.

Keywords. Surveys, catalogues, methods: data analysis

1. Introduction

LSST is a major new-generation instrument that incorporates many novel features, concepts and technologies. The main motivation for this Workshop was to instruct intending users as to its capabilities, potential, and operating systems, and to explain how the data management pipelines and products are validated in order to ensure that they meet the needs of the scientific community. The slides that were used are available online at [doi.org/10.5281/zenodo.1067258]. Participants attending the Workshop grouped themselves according to their main science interests – Galactic variables, extragalactic variables, Galactic transients, extragalactic transients.

A presentation giving an overview of the LSST project included hardware specifications such as primary mirror size, field of view, filters, and survey qualities (e.g., single-visit image depth and saturation, number of visits per field, total area); those are also available in Ivezić *et al.* (2008) and at http://lsst.org.

That was followed by a brief introduction to the LSST Data Management (DM) team, with a focus on the responsibility of the Subsystem Science Team to meet the desired high scientific standards. It was emphasised that additional questions and feedback about DM data pipelines and products are welcome at any time, because those communications both help the future user community to prepare for science with LSST, and also help the DM team ensure that its plans will meet the community's needs. The Workshop then allowed space for the the self-selected groups of participants to discuss their specific LSST science goals.

2. LSST Data Pipelines and Products

The Data Management presentation focused mainly on the current plans for data processing; those are described fully in Jurić *et al.* (2015) and Jurić *et al.* (2017a). This

Table 1. List of online LSST resources relevant to the content of this report

Ref.	Description	URL
(A)	Database Schema Browser	https://lsst-web.ncsa.illinois.edu/schema/index.php?sVer=baseline
(B)	'For Scientists' Webpage	https://www.lsst.org/scientists
(C)	Data Management	http://dm.lsst.org
(D)	'Ways to Participate' Webpage	https://www.lsst.org/participate
(E)	Transients & Variable Stars	https://tvs.science.lsst.org
(F)	`Community` Forum	https://community.lsst.org

report summarises the contents of the presentation, but Jurić *et al.* (2015) and Jurić *et al.* (2017a) should be consulted for the most up-to-date version of the data processing plans. Another resource for understanding the contents of a given catalogue is the Database Schema Browser (see resource A in Table 1).

Difference Imaging and Analysis (DIA) and Alert Production (AP) pipelines will run within 60 seconds of each shutter closure. This prompt processing includes:

- single-visit image processing (e.g., instrument signature removal)
- creating difference images (subtracting the new image from a template)
- detecting sources with $|\text{signal/noise}| > 5$ in the difference image
- characterising detected sources (flux, FWHM, dipole, trailed)
- associating detected sources with known objects by coordinate
- updating object variability parameters to include latest information
- issuing an Alert for each detected source
- updating DIA source and object catalogues in the Data Access Center

Alerts will be generated for each source detected in the difference image, whether or not it is a known variable object. Alerts will contain a full record of the source (e.g., its coordinates, flux, shape parameters, etc.), the 1-year history of DIA sources, and any objects associated with it from the catalogues produced by the DIA or DRP (see below), plus image stamps: a 2×3 array of difference and template with flux, variance, and mask, of at least $6'' \times 6''$ in size (but extended to the source's footprint), including metadata such as WCS, zero-point, and PSF. Additionally, within the following 24 hours, forced photometry is performed on the past 30 days of images for all new sources with no previous detections, and the processed single-visit images become available in the Data Access Center. The Moving Object Pipeline Software (MOPS) will then run on the source catalogue to identify and characterise moving objects.

The Data Release Pipeline (DRP) will run on a yearly basis on all of the existing data, up to the cut-off date for that release. The DRP will provide deeply stacked co-added images and catalogues of objects detected in both single-visit and co-added images. Those catalogues will also include forced photometry for all objects in all single-visit images, characterisation parameters for the shape and variability of all objects, and value-added quantities such as photometric redshift estimates. The DIA pipeline will also be re-run on a yearly basis on all of the existing data up to the cut-off date for that release, using the most up-to-date versions of all codes and algorithms. The database tables for all catalogues will contain cross-associations and will be capable of table joins.

For science goals that are not fully met by DM's planned data products, user-driven processing might be necessary, and LSST will provide support in that area. Specifically, LSST will not write science-case specific algorithms for processing data, *but* the software used for LSST data processing will be made available (under the GPLv3 license) for the community to use as a basis for custom processing pipelines, *and* LSST will commit an added ∼10% of its computing resources towards enabling user-driven analysis and data product creation. Unless it requires a computational need that exceeds a significant

https://doi.org/10.1017/S1743921318002685 Published online by Cambridge University Press

fraction of that 10%, such user-driven processing will be possible through the Science Platform, which is described in full in Jurić *et al.* (2017b). The Science Platform is designed to make the LSST data products and processing resources available to community members with data access rights by serving as a toolkit, a workspace, and a portal to the data and computational resources. The Science Platform includes proprietary processing in its core design.

The Workshop participants, still in their small groups, then identified the pipelines and products that will be appropriate for their own science goals, and discussed what kind of challenges they might encounter – and what questions still remained.

3. LSST Communications Resources

The Workshop's second formal presentation highlighted the current on-line information available about LSST and Data Management. It covered (a) the main webpage 'for scientists' (see resource B in Table 1), which contains information about properties such as filter throughputs, cadence optimisation, planned archive capabilities, and image simulations. It also covered (b) the main DM webpage (see resource C in Table 1), which contains information about the software stack, with links to `GitHub` repositories and documentation, and (c) the 'Participate in LSST' webpage (see resource D in Table 1), which has links to a wide variety of options for getting involved, such as joining Science Collaborations (applicable to astronomers in America or Chile, or in countries with a signed MOU). The current activities of the Transients and Variable Stars Science Collaboration (see resource E in Table 1) – the most relevant collaboration for the largest fraction of participants – were presented as examples of science engagement.

Two new avenues for project–community interaction were also highlighted. In an effort to open up communications further between LSST and the science community, the DM Subsystem Science Team had recently appointed liaisons to each of the Science Collaborations, and had started a new category in the `Community` online forum (see resource F in Table 1) called 'Data Q&A'. This new category is specifically intended as a site for people to ask science questions related to DM's data products and pipelines, and is monitored by DM team members who are committed to providing answers in a timely fashion.

After this summary of the main LSST communications resource, the small groups of Workshop participants discussed which of the presented resources would help them meet specific challenges and answer questions both in the short term and over the longer term.

4. Summary

The formal presentations in this Workshop covered the main LSST processing pipelines and data products that are currently being build by the Data Management team, and also the current communications resources used by the science community and the project. In between the formal presentations, small group discussions were encouraged in order to facilitate engagement with the material and with community networking. Frequently-heard questions and points of discussion among the participants included what the next steps regarding the observing strategy would be and how they would affect the science goals, the algorithms used for difference imaging, the triggering criteria and content of Alerts, the options for accessing the public Alert Stream, general data access policies, and science enabled by the user processing pipelines. The two main themes of feedback from Workshop participants were that the information was useful and relevant to their current concerns about doing science with LSST, and that they appreciated the opportunity to interact with others who are also interested in similar science domains.

https://doi.org/10.1017/S1743921318002685 Published online by Cambridge University Press

5. Acknowledgements

This material is based upon work supported in part by the National Science Foundation through Cooperative Agreement 1258333 managed by the Association of Universities for Research in Astronomy (AURA), and the Department of Energy under Contract No. DE-AC02-76SF00515 with the SLAC National Accelerator Laboratory. Additional LSST funding comes from private donations, grants to universities, and in-kind support from LSSTC Institutional Members.

References

Ivezić, Ž., Tyson, J.A., Abel, B., *et al.* 2008, arXiv:0805.2366

Ivezić, Ž., *The LSST Science Collaboration: LSST Science Requirements Document*, (LPM-17), https://ls.st/LPM-17

Jurić, M., *et al.* 2015, arXiv:1512.07914

Jurić, M., *et al.* 2017a, *LSST Data Products Definition Document* (LSE-163), https://ls.st/LSE-163

Jurić, M., *et al.* 2017b, *LSST Science Platform Vision Document* LSE-319), https://ls.st/LSE-319