Large Synoptic Survey Telescope (LSST)

LSST Data Policy

LDO-013

Latest Revision: November 1, 2019

Executive Summary

This document defines policies controlling LSST data rights and access for U.S. and Chilean scientists, international members, the worldwide scientific community, and the public. An individual's rights to access, analyze, publish, and/or share the full range of LSST data products and services are described. The difference between a *proprietary* data product and a *derived* data product is described and clarified, with examples.

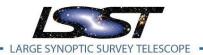The potential consequences of policy violations are stipulated, and the relevant committees that will be needed during operations to arbitrate policy matters and allocate potentially limited resources are outlined.

Two operations phase committees to help manage data rights and data resources are described. These are the Data Policy Committee (DPC) and the data Resource Allocation Committee (RAC).

Revision History

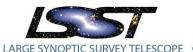| Version	Date	Description	Authors
0	2017-02-13	Initial version	W. O'Mullane
1.0	2018-06-28	Released to community for input	Willman, B., Graham, M., O'Mullane, W., Petravick, D., Johnson, M.
1.1	2018-09-03	Community feedback incorporated	Willman, B., Graham, M., O'Mullane, W., Petravick, D., Johnson, M.
2.0	2018-11-28	Significantly revised to make less restrictive to publications and derived data products	Bianco, F., Blum, R., Bolton, A. Hawley, S., Marshall, P., Strauss, M.
3.0	2018-11-30	Management Board version	Bianco, F., Blum, R., Bolton, A. Hawley, S., Marshall, P., Strauss, M.
4.0-4.3	2018-12-28	Revised following MB input on D3.	Blum
5.0	2019-01-01	Minor wording changes, further MB input, input from NCOA Dir	Blum
5.1	2019-01-23	DPOL 510 added caveat on bulk down loads (consistent with ODF too). Minor wording addition in 10.1 para 1, item (c).	Blum, Macfarlane
6.0	2019-05-24	In-kind data rights model	Blum
6.2	2019-06-10	Significant derived data products removed	Blum

6.3	2019-07-08	Minor edits, accepted prior suggestions	Blum
6.4	2019-07-27	SAC, SCs, AURA input	Blum, Bianco, Mandelbaum, Strauss et al., Lehr
7.0	2019-08-16	Add PPDB DPOL.	Blum, Slater , Dubois-Feldman
7.1	2019-08-21	Modify DPOl-404	Blum, Lehr
7.2	2019-10-20	Final input by NSF and DOE prior to public draft release.	Blum
1.0	2019-10-22	LDO-013 v1	Blum
1.0.1	2019-10-25	Applied DOE-proposed patches, clarifying this as an LSST document not needing explicit agency approval	Marshall
1.0.2	2019-10-29	Applied NSF-proposed patch, clarifying approach to publishing telescope metadata	Marshall
1.0.3	2019-11-01	Patches from DM team including products provided to Minor Planet Center	Marshall

Table of Contents

1. Introduction

The purpose of this document is to establish the policies during commissioning and operations that will govern LSST data access for US and Chilean scientists, international members, the worldwide scientific community and the public. These policies will inform the technical implementation of user access mechanisms during operations and will serve as guidelines for the user community in terms of data access, data product sharing, and publishing rights. The policies contained herein are subject to approval by the LSST Operations Management Board[1]. Overall approval for the full operations plan, including this Data Policy, will be provided by both NSF and DOE as part of their review and approval process.
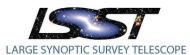
The Large Synoptic Survey Telescope (LSST) is an astronomical project that will generate a data set of unprecedented volume and complexity. Most LSST data products are subject to a proprietary period, with immediate access granted to all scientists in the US, a diverse set of scientists in Chile, and named individuals or communities in foreign countries.

The LSST Project will deliver a deep survey that covers $\sim 18,000$ square degrees in the southern sky and will detect ~ 40 billion stars and galaxies [1]. A total of ~ 825 visits to each part of the sky within this area will be made in six filters, *ugrizy*, over 10 years. About 10% of the observing time will be devoted to special programs that extend the areal coverage, depth, and/or sampling cadence (e.g., Galactic Plane, "deep drilling" fields). LSST will acquire ~ 20 terabytes of raw data each night and process it in real time, distributing alerts within 60[2] seconds on objects that vary significantly (5-sigma) in brightness or position, delivering processed images and updated object catalogs in 24 hours[2], and releasing a yearly reprocessed data set that will also include deep image stacks. To enable science by end-users, the LSST Data Management (DM) team is creating the LSST Science Platform (LSP): a user access and analysis environment with a web-based service for data access (via Data Access Centers described below), analysis, and processing that includes software tools and computational resources.

After a proprietary period of two years, all the LSST data become public. Alert packets of transient and variable data derived from prompt nightly processing will be public immediately (the term public is defined in DPOL-301, below). The public nature of these alert packets was agreed upon in the original MREFC proposal [2] and in an early policy document (Document-13380, [3]), which this document supersedes. Although the data will

---

[1] The Management Board is comprised of representatives of the operations partners representatives: AURA, and SLAC, with an at-large consensus representative proposed by AURA.

[2] This is a lower limit based on technical considerations. NSF and DOE may specify a longer latency.

be public after two years, access to the data will not be made available through LSST Data Access Centers (DACs) in the US and Chile for non-data rights holders.

Access to the data products from the LSST survey is critical to doing science. The current operations plan specifies that two Data Access Centers for LSST Users (see section 4 below) will be built. One will be located in La Serena (the Chilean DAC), and the other is planned for NCSA (the US DAC). Each will provide access and tools to analyze data, provide user support, and provide means for additional high-level compute resources through a merit-based allocation process; see section 10.

Other DACs might be constructed, for example by international scientists[3]. To provide access to LSST proprietary data, such DACs will need to comply with this present policy. For purposes of this policy, data access will be considered to be through LSST DACs which include the US and Chilean DACs and any potential, compliant, international DACs.
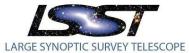
The high-level principles from which the proposed LSST data rights policies in this document have been derived include:

- To maximize the science output of LSST.
- To provide open access to data products for all LSST Data rights holders
    - To support and ensure access to junior scientists.
    - To support scientists from small institutions and historically underrepresented institutions.
- To incentivize international partners to join as contributors to operations.

2. Policy statements and document structure

In this document, *data policy* paragraphs are assigned a **DPOL**-number to allow for easy reference, and all other text is for clarification or to give examples of interpreting the policy. In Section 3, a set of high-level policies are introduced to define the terms used in this document. In Section 4, a detailed explanation of exactly who has LSST data rights is provided. A full description of which LSST data products and services are public or proprietary is provided in Section 5. In Section 6 the difference between LSST data products and *derived* data products is defined. Section 7 discusses the policies related to co-authorship of publications based on LSST data. Section 8 contains a set of science-driven examples that demonstrate the data sharing and co-authorship policies. Section 9 briefly discusses the potential consequences of violations of policy, and Section 10 describes the

---

[3] LSST DACs will need to meet minimum technical guidelines set by the LSST Operations team as well as to demonstrate the capability to comply with the present data rights policy.

two policy committees, one for data access and one for computational resource allocation, that will be needed during LSST Operations.

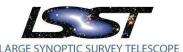## 3. High-Level Policies and Definitions of Terms

DPOL-301 The term *public* means that a data product can be shared with anyone, anywhere, worldwide. The term public does *not* mean that a data product is served publicly through any specific LSST interface at any particular level of service. The term *proprietary* means that a data product cannot be shared with anyone that does not have data rights. A full discussion of which LSST data products are public, and which are proprietary, is given in Section 5. In Section 6 we define the concept of *derived* data products (DDP), which can be made public.

The alert stream is public immediately. Prompt data products available within 24hr of the time taken at the telescope are proprietary for 2 years after the data release in which they are contained, and then become public. Data releases are a consistently processed version of raw images obtained before the release date and include catalogs and meta-data associated with the image set. Data releases are proprietary for two years following their release through the DACs then become public (see section 5).

DPOL-302 The term *data rights* is defined as the permission to work with and publish results based on **proprietary** LSST data, and the term *data access* as the ability to access LSST data and services through the LSST DACs. Anyone with rights also has access. Access to public LSST data and proprietary data may be through different mechanisms. Proprietary data rights holders are also called LSST Users in this document; non-users refers to everybody else. Access to public LSST data by non-users and proprietary data by LSST Users may be through different mechanisms.

Beyond scientific use of the LSST data as it is obtained, the benefits of data rights include:

- The right to apply for membership in LSST-sponsored Scientific Collaborations (membership in Scientific Collaborations is entirely up to the Collaboration);
- Access to the LSST help desk and engagement with community science support groups;
- Consideration for membership in LSST community based committees (e.g., Science Advisory Committee, Users sub-committee, Cadence sub-committee)
- Participation of LSST staff at User meetings, similar to participation during Construction (on a best-effort basis).

- Access to the LSST Science Platform on an authorized LSST DAC (at present these include US and Chilean DACs);

● User account management with compute and database allocations.

The term **science derivatives** includes analyses, interpretations, and discussions about scientific phenomena using LSST data, as well as derived data products (see DPOL-701) – in other words, the typical contents of the analysis and discussion of a scientific journal article. LSST will not require that LSST Users obtain approval from an LSST publication board prior to submitting journal articles based on proprietary LSST data (but note that some LSST Science Collaborations do have publication policies and boards for papers they produce). Science publications based on unreleased LSST data are not allowed by anyone at any time, including during the commissioning phase (Section 5.4).

DPOL-303 No Reserved Science: All types of scientific endeavors are equally open to all individuals with data rights. No area of research with LSST data is protected; the Science Collaborations or sub-groups thereof **do not have rights to any particular area of research.**
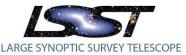
This applies to LSST Users regardless of their membership in, or level of contributions to, a Science Collaboration, and also applies to scientists working with public LSST data. However, specific software tools or derived data products created by a group of scientists may be kept proprietary as they decide (see DPOL-601) unless otherwise provided in a federal funding agreement.

DPOL-304 LSST Users (see section 4) will be entitled to an account with a Data Access Center (DAC) that gives access to the proprietary data and processing and analysis tools, and other DAC services such as web Application Programming Interfaces (APIs), help desks, and computational resources. A Data Policy Committee (DPC) will be charged by the LSST Operations Director to help manage this policy and assist Users in its interpretation. A Resource Allocation Committee (RAC) will allocate access to resources beyond this basic resource set as discussed in Section 10.2

DPOL-305 Acceptable Use: All users of LSST data and services, whether they have data rights or not, are required to abide by the LSST General Acceptable Use Policy [4] or its equivalent document during LSST Operations.

The above requirement applies to users' scientific use of LSST data and services, and to their behavior when accessing LSST-supplied services like online forums.

DPOL-306 Acknowledgement: All papers that use LSST proprietary data should include an acknowledgment that it was prepared in a manner consistent with the Data Rights and Access policies laid out in this document.

The intent of the above requirement is to raise awareness of the LSST data rights and access policies, and to avoid accidental violations. An example of an acceptable statement is *"This publication is based in part on proprietary LSST data, and was prepared in accordance with the LSST data rights and access policies. All authors of this publication meet the requirements for co-authorship of proprietary LSST data."* A minimum acceptable acknowledgement for papers including proprietary or public LSST data will include acknowledgement of NSF and DOE funding of LSST science as well.

DPOL-307 Use Agreement: Individuals with data rights will be asked to sign a use agreement that restricts the redistribution and use of LSST proprietary data products.

The intent of the above requirement is to promote awareness of the LSST data policies and to enable users to take responsibility for adhering to them. The LSST Project will not monitor the actions of its Users for potential violations to data rights policies. Every User is personally accountable for understanding and abiding by the LSST data rights and data access policies. Suspected violations should be reported to the DPC (see Section 10.1).

DPOL-308 Acceptable Use Policy: Any violation of the Acceptable Use Policy [4] or the Data Rights and Access Policies in this document can result in revocation of LSST data rights and data access. The potential consequences for policy violations are discussed further in Section 9.

4. Defining Who Has Data Rights

All US, Chilean, and approved international members as defined in Sections 4.1, 4.2, and 4.3 are LSST Users, meaning that they have data rights and access to proprietary LSST data. Any ambiguities or clarifications on who is a User will be addressed by the DPC as needed or appropriate.

4.1 US Scientists

DPOL-401 US Scientists: All US scientists as defined below are LSST Users. LSST Users will access the proprietary data through the LSST DACs which also provide limited dedicated computational resources for analysis.

This document defines **US scientist** as any professional scientist who is employed (paid the majority of their salary) or holds a research grant/fellowship at or is enrolled at a US educational or research institution that wishes to access the LSST data through the science platform in order to (i) do non-profit scientific research at the level appropriate for inclusion in a peer-review journal and/or (ii) develop new citizen science projects for LSST Education and Public Outreach (EPO). Amateur astronomers who wish to do non-profit

scientific research may also submit requests for verification of eligibility to the DPC, which may require a justification of why the EPO-provided resources are insufficient (Section 5.3).

4.2 Chilean Scientists

DPOL-402 Chilean scientists are LSST Users [5] and will access the proprietary data through the LSST DACs including the Chilean DAC [6].

This document defines **Chilean scientists** as those professionals, their research associates, and their students who have access to Chilean telescope time (as defined by SOCHIAS). Other Chilean scientists will be included as appropriate through general agreement with LSST, or by petitioning the DPC in separate cases.
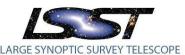
4.3 International Members

International members include countries, institutions, or groups of scientists with agreements with NSF and/or DOE (or as delegated to SLAC and/or AURA by NSF and DOE) for use of LSST. International member scientists (principal investigators and their research associates and students identified by class, name, and/or number) are identified through approved institutional Memoranda of Agreements (MOAs) with NSF, DOE, SLAC, or AURA, which define in-kind contributions to LSST in exchange for a certain number of named Principal Investigators (PIs) with data rights. Typically, each PI may have up to four junior associates (postdoctoral fellows and/or graduate students) who also have data rights. It is left to the international members to name the individuals with data rights. All international members are ultimately approved by NSF and DOE, which are the sole entities that may confer data rights.

4.4 Alternative Routes to Data Rights

Although it is common practice for astronomers with access to a proprietary data set to use this as a resource to form collaborations with others based on in-kind contributions, this is not the case with LSST data. LSST Users are not allowed to distribute proprietary LSST data products to non-users (although they are allowed to share derived data products; see Section 6). This applies to all individual Users and to the LSST Science Collaborations. Only the LSST project can assign data rights through in-kind or other contributions as delegated by NSF and DOE and their awardees AURA and SLAC.

DPOL-403 Builders: Individuals who have attained LSST Builder status (as conferred by the LSST construction project management) shall be LSST Users, regardless of their location or institution. LSST Builder status may be conferred after 2 full-time equivalent years of contributions to the LSST Construction Project and does not expire. However, LSST builders are not immune to revocation of data rights if they should violate LSST guidelines

The contents of this document are subject to configuration control and may not be changed, altered, or their provisions waived without prior approval of the LSST Change Control Board.

11

(DPOL-308). The current list of LSST Builders may be found at
https://project.lsst.org/contact/builders-list.

4.5 Grace Period

For the majority of scientists who are not at US or Chilean institutions, being identified as an "LSST User" is a non-permanent state. As described above, the LSST data rights for most US and Chilean scientists will be conferred by association with an institution, whereas the LSST data rights for international scientists will be conferred to individually named PIs under MoAs that are specific to each international participant. Since LSST data rights can be conferred, they can also be lost.

DPOL-404 Change of Residency: Junior LSST Users (grad students, postdocs, research scientists and junior faculty) who lose their data rights due to a change in their residency, employment, or institutional affiliation in the US or Chile, will retain their data rights for a grace period of two years that starts from the time of the change. During the grace period, the individual maintains data rights only to those data products available at the time they had left their institution.

Petitions for extensions to the default period can be made to the DPC (Section 10.1).

Prior to each regular data release, the lists of LSST Users with data access in the Data Access Center will be cleaned up by removing all accounts which cannot be verified to belong to valid LSST Users (including those whose grace period has expired). An account will only be removed if it is expired or repeated attempts to contact the account owner have failed to elicit a response.

5. LSST Data and Services

In order to enable efficient use of the ~20 terabytes of data acquired each night by the LSST camera, the LSST Data Management (DM) team will deliver software for a wide variety of data pipelines, products, and services that are defined by the Data Products Definitions Document [7]. All released LSST data will be verified by quality assessment (QA) metrics, and no data will be made available, proprietary or not, without first passing QA (the level of QA varies from alerts to data release data products). Full records of data provenance will be included with each data release. Data access to scientific data products for LSST Users will be through the LSST DACs via authorized and authenticated accounts; only LSST Users will be able to access data through the LSST DACs.

The purpose of the following sections is to cover all of the LSST data products and services, denoting for each whether they are public or proprietary, and describing the access

The contents of this document are subject to configuration control and may not be changed, altered, or their provisions waived without prior approval of the LSST Change Control Board.

12

mechanisms for LSST Users. In Section 5.1 the general categories of data and services are described. In Section 5.2 the computational resources of the DACs are discussed. In Sections 5.3, 5.4, and 5.5 the special cases of data for Education and Public Outreach (EPO), commissioning, and special programs are presented, respectively.

5.1 General Data and Services

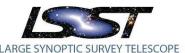The following descriptions apply to LSST data products stored and served from LSST DACs [11].

DPOL-501 Telescope and Site Metadata: Much of the metadata from the mountain in the Engineering Facility Database (EFD) is proprietary and will be accessible through the LSST DACs. A subset of this information will be public as obtained, e.g., weather logs, all-sky camera and seeing, and observing patterns. The public metadata from the telescope and site might be published on, e.g., the LSST website, whenever practical. Metadata appropriate to a data release becoming public will also become public.

DPOL-502 Raw and Calibration Data: The entire set of raw images and calibration files used to produce processed images are proprietary and will be accessible to LSST Users through the LSST DACs.

DPOL-503 Alert Stream: The real-time alert stream is public. The contents of the alerts database that records and stores all issued alerts is public; however, access to the LSST alert database at the LSST DACs will be restricted to LSST Users.

The complete alert stream will not be distributed directly to individual investigators, but rather will be delivered to a selected set of community brokers. While brokers will not be required to make the alert stream, in whole or in part, openly accessible, the capability and commitment to do so will be an important criterion in the broker selection process (the selection process will guarantee that at least one broker will make alerts public to comply with DPOL-503). Individual investigators will be able to access a subset of the alert stream through the Alert Filtering Service, as well as using the community brokers. Alert distribution and policies for alert brokers are discussed in LDM-612 [8].

DPOL-504 Prompt Processing Catalog Products: The image differencing source (DIASource), forced photometry (DIAForcedSource), and object (DIAObject and SSObject) records produced in Prompt Processing are public. Access to these catalogs held at the LSST DACs will be restricted to LSST Users, but the information in the Prompt Products Database (PPDB) tables listed here in DPOL-504 can be shared by broker teams or others with access rights to the LSST DACs. Observations of previously known and newly identified Solar

System objects (calibrated astrometry and photometry) will be reported to the Minor Planet Center (MPC) within 24 hours. Data reported to the MPC are public.

DPOL-505 Prompt Processing Image Products: The raw images, processed single visit images, difference images, and template images are proprietary, except for the alert postage stamps distributed in the alert stream and recorded in the alert database.

As described in LSE-163[7], the products of Prompt Processing correspond to a subset of the products of Data Release processing, which occurs annually. For this reason, the proprietary period for Prompt Products does not expire until two years after the same data are reprocessed and included in a Data Release. In other words, the products of Prompt Processing do not become public two years after the date of their initial release, but rather two years after their inclusion in the associated Data Release.

DPOL-506 Data Release Products: The single-visit images, difference images, stacked images, and associated source and object catalogs that are reprocessed and released on an annual or semi-annual (e.g. commissioning data) basis are proprietary, and will be accessible through the LSST DACs.

DPOL-507 LSST Science Platform: This portal to the data will provide a software framework for astronomical research including query, visualization, and analysis tools, plus the associated computational, storage, and communications infrastructure needed to enable science. The LSST Science Platform is accessible only to LSST Users through the LSST DACs.

The LSST Science Platform is described in LSE-319[9] and LDM-554 [10]. The computational resources provided at the DAC are discussed further in Section 5.2.

DPOL-508 Software Stack: The LSST Data Management team's Software Stack (Python and C++ software to process images and manipulate other types of data, as well as the source code for the Science Platform) is open source (public) and available online[4].

5.2 DAC Computational Resources

DPOL-509 DAC Accounts: All data access through the LSST DACs will be with authorized and authenticated accounts; there will be no non-authenticated or anonymous access. LSST User accounts on LSST DACs will have a default amount of storage and CPU quota.

---

[4] https://pipelines.lsst.io/index.html

The initial allocation of computational resources to each LSST User account will be enough to perform basic scientific queries and analysis on catalogs, as well as some image processing, and to store the results.

DPOL-510 User Resource Allocation: LSST Users may need more storage or CPU than their initial allocation. A mechanism will be put in place to apply for additional resources. A Resource Allocation Committee (Section 10.2) will make science merit-based allocations of extraordinary resources available to Users at LSST DACs.

DPOL-511 Bulk Downloads: LSST expects to support bulk data download services for significant fractions of the proprietary or public data through a process to be determined in the future. LSST reserves the right to charge for the full cost of providing this service or to restrict access to this capability for policy or programmatic reasons.

5.3 Education and Public Outreach (EPO)

EPO data products will be accessed through a system separate from the LSST DACs; accessing EPO products will not allow the user to interact with the LSST DACs.

DPOL-512 EPO Data Products: A predetermined fraction of data will be made available to be used as EPO data products as specified in LSE-131 [12], and all EPO data and services are public and will be openly accessible through an LSST EPO-specific interface.

DPOL-513 Metadata: Only metadata reviewed and approved by LSST will be distributed via the EPO system. The LSST-funded tools for some citizen science programs may require data access which is available to LSST Users only. Leaders of Citizen Science projects may petition to the DPC for data rights as needed and appropriate. Further details regarding EPO-related activities involving LSST data can be found in LSE-131[12].

5.4 Commissioning Data

In this section, the term *commissioning* refers to the final stages of the LSST Construction Project, starting from first light and continuing through Science Validation, up to the start of full Operations. Commissioning data include data from the single-raft commissioning camera (ComCam) and the full LSST camera. Production of data products from commissioning is the responsibility of the commissioning team and will have QA levels that could differ from those in operations. Provision of commissioning data products to LSST Users is the responsibility of the LSST (pre)-Operations team and will be at a level that resources permit, with a goal to provide access to Users via an early version of the US DAC and LSP.

DPOL-514 Commissioning Team: The commissioning team may select and include non-staff members[5] that will perform analysis of commissioning data that is necessary for successful LSST commissioning.

DPOL-515 Commissioning Data: Release of commissioning data to LSST Users will be as resources permit and at the discretion of the LSST Operations team. Commissioning data will be proprietary and subject to the same policies as LSST data during operations.

DPOL-516 Science Data from Commissioning: Scientific analysis of the commissioning data will be an integral and necessary part of the science verification process. All commissioning data used for science will be released to all LSST Users prior to any publication by anyone. Members of the commissioning team may not submit science papers to a journal and/or the arXiv based on commissioning data prior to the release of those data to LSST Users, but they may undergo the LSST Publication Board process (this board is part of the construction project, not operations) in advance of the release of those data.
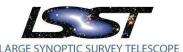
5.5 Special Programs Data

In the context of LSST, Special Programs refers to data obtained through initiatives such as mini-surveys, target of opportunity observations, or deep-drilling fields. These data will be processed by the DM system as described in LSE-163 [7].

DPOL-517 Special Program Data: Data from these Special Programs will be subject to the same policies as data from the Wide-Fast-Deep main survey unless specifically noted. LSST, for purposes of providing science opportunities for its community, may enter into mutual agreements with other projects or surveys to gain access to scientifically useful resources such as other surveys at other wavelengths in jointly observed fields (see also DPOL-602).

DPOL-518 Special Program Resource Allocation: There may be Special Programs that require a processing pipeline and/or computational resources above and beyond what LSST Data Management has designed and/or allocated. Special processing requests will be submitted to the RAC for merit-based allocation of resources (section 10.2). Data products resulting from such requests may be proprietary or public. The classification as proprietary or public will be made at the time of the RAC allocation.

---

[5] Non-staff commissioning team members are selected by the commissioning team based on needed and available skills from LSST Users in the LSST science community.

6. Derived Data Products and Target Lists

Only LSST Users will be allowed to access and work with proprietary LSST data. However, LSST Users may produce *derived data products* from the proprietary data that can be shared with non-users and published in papers as described next.

DPOL-601 Data are considered an **LSST *derived data product* (DDP) if the data product is derived from LSST proprietary data but cannot be used to recreate[6] any proprietary LSST data product(s).** Examples of DDPs include, but are not limited to, color-color and color-magnitude diagrams, and mass maps generated from shear measurements. These derived products do not include the catalog entries for the sources in the DDP with the exception of a list of sources up to 1000 in number[7]; see DPOL-603 below. Any DDP may be shared with anyone.

LSST Users may collaborate on projects and share DDPs between collaborators (Users and non-users alike). Such data may remain proprietary to themselves, shared with the LSST User community, or made public at the discretion of the collaborators.

DPOL-602 In-kind Data Sets and Derived Data Products DPOL-601 applies to most cases of LSST Users working with groups of collaborators including non-users. DPOL-601 is not intended to apply to the case of major collaborations between LSST and other large surveys which aim to share data and data access generally, or produce significant data products as part of a joint science program. Large collaborative programs should propose joint sharing of survey images and catalogs, co-processing of images, and other significant collaborative aspects to the DPC. Recommendations from the DPC to enter into mutually beneficial arrangements will be reviewed and decided upon by the LSST Operations Director and approved by the Management Board.

Final approval of collaborative agreements between LSST and other large surveys lies with NSF and DOE. Further, NSF and DOE reserve the right to enter into collaborative agreements with other surveys through their awardees AURA and SLAC. Such agreements may include conferring data rights upon new international members with the express purpose of gaining access to new data sets or deriving new data products. Such data sets or derived data products would be made available to all LSST data rights holders and would be subject to the data rights policy described in this document.

---

[6] In general, recreate means either re-derive or reverse engineer as well as "replace" in an equivalent way. E.g. producing a catalog of positions and magnitudes that are not exactly the same as the LSST catalog entries but are equivalent would not be allowed.

[7] The numerical value of this threshold will be kept under review by the DPC.

DPOL-603 Target Lists In addition to sharing DDP, Users are allowed to share target lists of small samples of objects with non-users to encourage and facilitate collaboration with non-users for purposes of following up LSST observations with observations from other telescopes and in the course of doing their normal research with collaborators (for example, see section 8.7). Such lists are limited to 1000 objects and may include coordinates (RA, Dec) and magnitudes (*ugrizy*) from the proprietary catalog. Larger target lists, for example for multi-object spectroscopy (MOS) with multi-thousand simultaneous spectra can be requested via the DPC. Users who have access to MOS capabilities need not petition the DPC as long as they do not share the proprietary data (e.g., target lists) with non-users.

7. Publications

The following refer to any form of publishing or making public science results from LSST proprietary data.
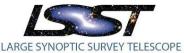
DPOL-701 Publication: No restrictions will be placed on LSST Users' ability to publish science results and DDP from proprietary LSST data consistent with use described in section 6. DPOL-516 restricts members of the LSST commissioning team from submitting science papers based on commissioning data to a journal prior to the release of those data to the LSST Users.

DPOL-702 Non-users as Co-Authors: Non-users may co-author papers which use results from proprietary LSST data, but this does not give non-users general access to the proprietary data themselves. DDP (DPOL-601) are public and can be shared with all co-authors.

DPOL-703 Reproducibility: For the purposes of publication, reproducibility of scientific results should be accomplished by including, describing or citing the queries used to generate the LSST data set on which the analysis was made, and not by posting or publishing copies of the proprietary LSST data. LSST plans to explore the technical feasibility of maintaining data accessibility for reproducibility purposes beyond the two-year proprietary period, likely at reduced service levels.

8. Data Policy Use-Case Examples

This section provides examples to demonstrate the practical application of the LSST data policies presented in this document. First, two generic examples are provided to illustrate how to collaborate with non-users (DPOL-603 and DPOL-702). Below, Sections 8.2 to 8.7

provides a suite of specific scientific scenarios[8], each with an interpretation based on the applicable data rights policies.

8.1 Collaborating with non-users

Scenario 1: General Follow-up Observations – LSST Users are free to collaborate with any User or non-user. Users cannot share access to the LSST DACs or proprietary data with non-users but are free to share derived data products generated from proprietary data products. This includes coordinates of up to 1000 objects for follow-up observations (DPOL-603). Such lists should include only targets appropriate to the observing period and time request in the proposal. Proposal here means a single project with a well defined science objective. The 1000 object sample or object list is meant to cover multiple observing runs for the project to the same telescope or different telescopes.

Scenario 2: Data or Code that Enable a Scientific Publication – A non-user contributes computer code, simulations, or scientific ideas to a paper using LSST data, but never touches the LSST data. This is unambiguous; they are allowed to be co-authors.

8.2 A Comet Discovery in Proprietary Data

A new comet is found by a LSST User in stacked deep drilling images, but is not found by the LSST Solar System Processing System and not included in the public Prompt data products (e.g., the alert stream or through the Minor Planets Center). This LSST User has a colleague who could obtain immediate follow-up for scientific analysis (e.g., color evolution, gas outflow measurements), but this colleague is a non-user. The comet's ephemerides are a DDP which may be shared with the colleague.

8.3 The Discovery of a New Local Group Galaxy in Proprietary Data

(a) An LSST User creates stacked images out of the single-visit images produced by the Prompt processing pipeline, prior to the annual data release. These stacked images are considered proprietary LSST data because they are equivalent to the stacked images that will be produced by the Data Release pipeline, which are proprietary. These stacked images are not a DDP.

(b) The LSST User analyzes their stacked images, discovers a new Local Group galaxy, and identifies stars belonging to this galaxy. The catalog of star coordinates and magnitudes are considered proprietary LSST data (regardless of whether the catalog numbers < 1000 objects or not, but see below) because they are equivalent (see DPOL-601) to the Object catalogs that will be produced by the Data Release pipeline, which are proprietary. This

---

[8] With gratitude to individuals in the scientific community who suggested use case scenarios.

catalog is not a DDP. Plots of objects in the catalog used for analysis and publication are DDPs and can be shared with collaborators and published. Candidate objects (up to 1000) identified specifically for spectroscopic follow-up taken from the galaxy catalog are considered a target list (RA, Dec, magnitudes) and can be shared (DPOL-603).

(c) The LSST User analyzes their catalog of stars to measure physical properties of the new galaxy. These physical properties are a DDP because they cannot be used to recreate the star catalog.

(d) The LSST User publishes a paper on this discovery within the two-year proprietary period. This paper may contain the image stacking algorithm and queries used to construct the stacked images and the star catalog (for reproducibility), but not those proprietary data products themselves. Images suitable for enhancing publications can be obtained from the EPO data available publicly.

8.4 Broker Classifications Based on Alerts and Proprietary Data

A community broker run by a LSST User accesses proprietary data products in the DAC, e.g., the forced-photometry light curves, and uses its own algorithm to classify objects (i.e make a determination of the nature of the object beyond its characteristics conveyed in the alert packet. E.g., "high redshift galaxy"). The classifications are a DDP and can be shared publicly. The forced-photometry is public, but the access to the prompt products database is still restricted to the LSST Users (see DPOL-504).
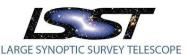
8.5 Brokers That Save and Serve All Alerts

A community broker saves every alert from the public stream and builds an alerts database and/or a reconstruction of the DIAObject, DIASource, and DIAForcedSource catalogs.

While the LSST Alert Database and Prompt Processing Catalog Products are public, access to them through the LSST DACs is still only available to LSST Users. A broker's copy of these datasets is also public and may be made available to anyone at the broker's discretion.

8.6 Photometric Redshifts Based on Proprietary Data

The LSST Data Release pipeline will include photometric redshift estimates in its object catalogs, and these photo-z's are proprietary. Any use of the LSST proprietary data to produce an equivalent set of photo-z's is constrained by DPOL-602.

The contents of this document are subject to configuration control and may not be changed, altered, or their provisions waived without prior approval of the LSST Change Control Board.

20

8.7 Contributing Proprietary LSST Data to Enable a Publication

An individual without LSST data rights discovers an interesting transient in their own low-*z* all-sky survey and is preparing a paper on which they are first author. The first author discusses their discovery, follow-up, and preliminary physical interpretation with a colleague who is an LSST User. The User searches the LSST proprietary data and finds additional photometric data points that provide evidence for precursor activity that was not detectable by the first author's survey. The colleague's science derivative – that the progenitor exhibited outbursts prior to explosion – would not be publishable on its own, but substantially contributes to the non-user first author's paper. The User may share their LSST derived data product (i.e the pre-outburst behavior and photometry) with the first author and be included as a co-author.

9. Concerning Violations of Policy

During Operations, LSST will not actively monitor the actions of LSST Users and will not need to review derived data products prior to their publication (Users may reach out to the DPC if they have doubts). It is expected that all LSST Users will abide by the Acceptable Use Policy ([LPM-123][4]) and the policies laid out in this document. Suspected violations should be reported to the Data Policy Committee (Section 10.1).

DPOL-901 LSST reserves the right to sanction Users who violate the policies in this document or facilitate non-users in accessing proprietary data up to and including removal of data rights and/or access. More typically, sanctions might include removing data access privileges from the individual for a period of at least one data release. If data rights are revoked, reinstatement may be petitioned to the Data Policy Committee (Section 10.1).

In order to actively prevent violations, LSST will: (1) make both LPM-123[4] and this document easily found by web searches and publicly accessible; (2) charge the Data Policy Committee (Section 10.1) with providing prompt answers to users' questions about data policy. Users will have to acknowledge this policy when instantiating their LSST DAC accounts. LSST in consultation with the DPC will also continue to add to the use cases described here during operations.

10. Policy Committees during Operations

10.1 Data Policy Committee

To ensure that the LSST User community can responsibly use LSST data products, and to allow for exceptions or new LSST reciprocal agreements with other large surveys, a Data Policy Committee (DPC) will be needed. Examples of DPC activities are to (a) provide timely

and adequate answers to questions that arise; (b) arbitrate cases of disputed violations to data policy; (c) evolve the data policies as needed for approval by the LSST Operations Management Board; and (d) consider collaborative opportunities not explicit in the present policy, based on their potential for high scientific impact and return for LSST science. The DPC will be formed before the start of Operations. The DPC will be advisory to the LSST Operations Director who will appoint the committee members to regular terms of participation and charge them with their oversight responsibilities.

It is expected that the committee membership will generally be drawn from Users representing the managing partners AURA and SLAC (and others as deemed appropriate by AURA, SLAC and LSST). These committee members will be well versed in basic operations issues and data policy considerations. When petitioned to decide on large collaborative proposals (DPOL-602) brought to the DPC, the committee chair will be empowered to convene (as appropriate and needed) a committee of LSST expert scientists drawn from the community to advise and make recommendations on collaborative survey proposals and reciprocal agreements.

10.2 Resource Allocation Committee

It is anticipated that some individuals or groups of LSST Users will require storage and/or computational processing resources in excess of the basic quota allocated to all DAC accounts, and therefore, a Resource Allocation Committee (RAC) is needed to develop and implement a process for handling such requests. The RAC will be analogous to a Time Allocation Committee (TAC) for a telescope, except that here the limited resource would be compute cycles, disk space for storage, and potentially significantly large bulk data downloads. As with a TAC, the RAC would consider the scientific justification for increases to basic quota for LSST Users and be advisory to the LSST Operations Director. For this reason, its membership should include representatives from the LSST science community. RAC members will be appointed for regular terms of participation by the Operations Director who will also charge them with allocation of resources to support the community during survey operations. Operations staff members will assist the RAC by assessing the technical feasibility of requests as appropriate.

11.   References

[1] Ivezic, Z., et al., 2008, ArXiv e-prints (`arXiv:0805.2366`)

[2] **[Document-10549]**, Wolff, S.C., Kahn, S.M., Krabbendam, V.L., Sweeney, D.W., Tyson,

J.A., 2011, *Proposal to the National Science Foundation*, Document-10549, URL

`https://ls.st/Document-10549`

[3] **[Document-13380]**, Wolff, S., Kahn, S., 2013, *Data Rights and Data Management Policy*,

Document-13380, URL `https://ls.st/Document-13380`

[4] **[LPM-123]**, Petravick, D., 2017, *General Acceptable Use Policy*, LPM-123, URL

`https://ls.st/LPM-123`

[5] **[Document-11622]**, Smith, W., Vera, V.P., 2011, *Supplementary and Clarifying Agreement*

between the Universidad de Chile and AURA covering the use of the LSST on Cerro Pachon,

Document-11622, URL `https://ls.st/Document-11622`

[6] **[LDM-572]**, O'Mullane, W., Petravick, D., 2017, *Chilean Data Access Center*, LDM-572, URL

`https://ls.st/LDM-572`

[7] **[LSE-163]**, Jurić, M., et al., 2017, *LSST Data Products Definition Document*, LSE-163, URL

`https://ls.st/LSE-163`

[8] **[LDM-612]**, Bellm, E., coauthors, 2018, *Plans and Policies for LSST Alert Distribution*,

LDM-612, URL `https://ls.st/LDM-612`

[9] **[LSE-319]**, Jurić, M., Ciardi, D., Dubois-Felsmann, G., 2017, *LSST Science Platform Vision*

Document, LSE-319, URL `https://ls.st/LSE-319`

[10] **[LDM-554]**, Ciardi, D., Dubois-Felsmann, G., 2017, *Science Platform Requirements*,
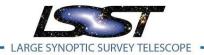
LDM-554, URL `https://ls.st/LDM-554`

[11] **[LPM-251]**, William O'Mullane, M.G., Beth Willman, 2018, *Policy for Independent Data Access Centers*, LPM-251, URL `https://ls.st/LPM-251`

[12] **[LSE-131]**, Jacoby, S., Emmons, B., Selvy, B., 2017, *Interface between Data Management and Education and Public Outreach*, LSE-131, URL `https://ls.st/LSE-131`

12. Acronyms

Acronym	Description
API	Application Programing Interface
AURA	Association of Universities for Research in Astronomy
CPU	Central Processing Unit
DAC	Data Access Center
DESC	Dark Energy Science Collaboration
DPC	Data Policy Committee
DDP	Derived Data Product
DM	Data Management
EPO	Education and Public Outreach
EFD	Engineering Facility Database
LDM	Light Data Management
LPM	LSST Project Management (Document Handle)
LSE	LSST Systems Engineering (Document Handle)

LSST	Large Synoptic Survey Telescope
MOS	Multi Object Spectroscopy
MPC	Minor Planet Centre
MREFC	NSF's Major Research Equipment and Facilities Construction
NCSA	National Center for Supercomputing Applications
NSF	National Science Foundation
PI	Principal Investigator
PPDB	Prompt Products Database
QA	Quality Assurance
RAC	Resource Allocation Committee
US	United States