

Data Management Plan for the Large Synoptic Survey Telescope

Supplement to the LSST MREFC Proposal

The LSST is a proposed optical survey telescope that will collect 15 Terabytes of astronomical image data every night. Data will be processed at a dedicated, world-class computing facility and made available to the scientific community and to the general public. The LSST proposal for MREFC funding conforms fully to NSF policy as described in the [Award and Administration Guide, NSF 11-1 January 2011 Chapter VI - Other Post Award Requirements and Considerations, Section 4. Dissemination and Sharing of Research Results.](#)

As described in section 4.5, the LSST MREFC Proposal, dated January 31, 2011, provides sufficient funding to make all data collected by LSST, and all software that processes those data, available without proprietary period to scientists in the US and in Chile, which will host the telescope. World-wide access to information about transient phenomena is also supported by the current budget. Fully 30% of direct costs in the NSF MREFC request are devoted to building the data management system hardware and software.

1. Types of data and materials to be produced

The LSST data products are described in three groups:

- Level 1 data products include science images, difference source and object catalogs, transient alerts, and data quality metrics and statistics. They are generated by pipeline processing of the stream of data from the camera subsystem during normal observing. Level 1 data products are therefore continuously generated and/or updated every observing night.
- Level 2 data products include stacked science images, template and calibration images, deep source and object catalogs, alert statistics, and data quality metrics and statistics. They are generated as part of Data Releases, which are required to be performed at least yearly.
- Level 3 data products are derived at the initiative of scientific users from Level 1 and/or Level 2 data products to support particular science goals. The LSST is required to facilitate the creation of Level 3 data products, by providing suitable software interfaces and computing infrastructure, but is not itself required to create any Level 3 data product. Instead these data products are created externally to the data management system, using software written by researchers, e.g., science collaborations.

Finally, the LSST MREFC Proposal provides for funds to develop Education and Public Outreach programs that make use of LSST data.

2. Standards to be used for Data and Metadata Format and Content

LSST will adhere to the most widely used astronomical data and metadata format standards and will support distribution of the data products according to the following standards:

- [Flexible Image Transport System \(FITS\)](#)
- [Virtual Observatory \(SIAP, VOTable, VOEvent, etc.\)](#)

3. Policies for Access and Sharing

Level 1 data products will be made available within 24 hours of processing, Level 2 data products will be made available annually. The access policies for Level 3 data products will be product- and source-specific, and Level 3 products developed by others may be proprietary. If researchers choose to cache Level 3 products on LSST-managed computing and storage resources, the proprietary status will be enforced via automated access controls.

LSST key science deliverables will be enabled by providing computing resources co-located with the raw data and catalog storage. Researchers will be able to process up to petabyte-sized sections of the Level 1 and Level 2 data on LSST-provided dedicated supercomputer clusters or on local computers, institutional clusters, or across a scientific grid. The workload placed on the LSST-provided clusters by these users will be actively managed to ensure equitable access to all segments of the user community.

4. Policies and Provisions for Reuse, Re-distribution, and the Production of Derivatives

No restrictions are placed on reuse, re-distribution, and production of derivatives from LSST data. Further, to promote production of derivatives, all LSST data management software is today open and available, and will remain so throughout the construction, commissioning, and operations phases of the project. The software is available in source form from an online repository accessible via the public internet and research and education networks. The software is licensed under the Gnu Public License (GPL) 3.0 and the Limited Gnu Public License (LPGL).

5. Plans for Archiving Data, Samples, and other Research Products, and for Preservation of Access to Them

All data products will be archived for the entire operational life of the LSST Observatory. During this time they will be accessible via direct query and download or for fusion with other astronomical surveys. The scientific value of providing access beyond the lifetime of the 10-year survey is likely to be high but will be subject to additional review.