Large Synoptic Survey Telescope

www.lsst.org

LSST Data Products and User Interfaces

Richard A. Shaw[1], T. Axelrod[2], A. C. Becker[3], S. Bickerton[4], M. Juric[5], J. Kantor[6], K. S. Krughoff[3], R. H. Lupton[4], S. Van Dyk[7], and the LSST Data Management and Simulation Teams

[1]National Optical Astronomy Observatory, [2]Univ. Arizona, [3]Univ. Washington, [4]Princeton Univ., [5]Harvard Univ., [6]LSST Corp., [7]IPAC/California Institute of Technology

The LSST will produce the richest sets of astronomical data ever created, which will open up an unparalleled temporal discovery space. The data products will include deep imaging of half of the sky in 6 passbands; catalogs of all detected sources including stars, galaxies, solar system objects; lightcurves of variable objects; and alerts of transient sources that will be generated within a minute of their detection. The LSST Project is currently prototyping a scalable Data Management System capable of processing, archiving, and serving these data to the astronomical community. We anticipate that individual investigators and research teams will, during the course of their analysis, generate scientific datasets using data products from LSST (possibly combined with data from other resources) that will be of great value to the LSST community. The LSST Project plans to support these community-based science activities by providing direct computing and storage resources and the use of portions of the LSST software stack, and the development of user and programmatic interfaces that enable the discovery, exploration, and analysis of LSST data products. In addition to science data products, a number of data products will be generated to assess science quality. Although science data quality assessment will be highly automated, even the limited human interaction required to assess and diagnose problems drives the need to prototype user interfaces that enable efficient data exploration and analysis. Much of this capability is also needed for generating and evaluating calibration products, documenting survey progress, and supporting science analysis for users.

1. Pipeline Productions

The LSST Data Management System (DMS) organizes pipeline processing into a series of *productions* (Fig. 1) that generate data products on different timescales of relevance. The **Alert Production** runs nightly to generate time-critical products including alerts, single-visit calibrated images, catalogs of moving and transient/variable sources, and data quality assessments. The **Data Release Production** runs annually on the entire data set, to generate master source and object catalogs, image co-adds, and other products. The **Calibration Products Production** runs periodically to generate master calibration frames for tracking changes in the instrumental signature over time.

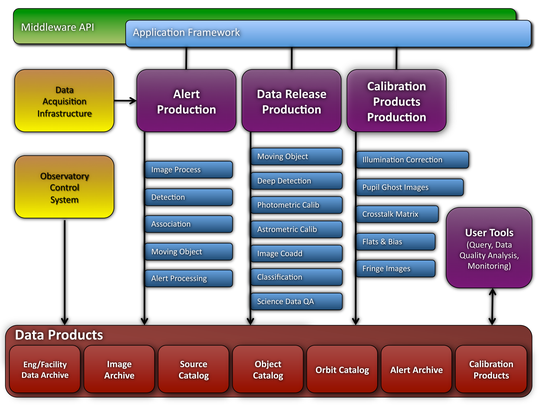


Figure 1: The LSST DMS organizes the generation of data products into a series of *productions* over different timescales of scientific interest.

The data products that will be generated by the DMS system include the following:

Images

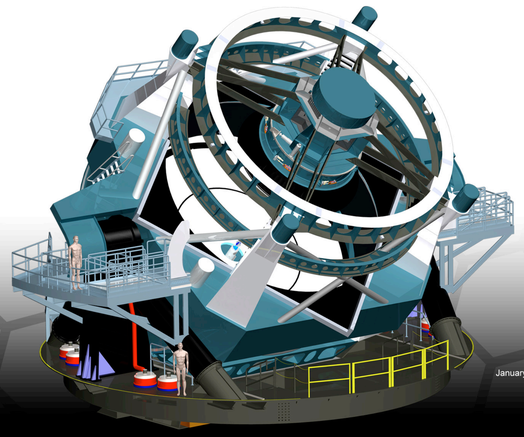
The archive of image data will include raw images, deep detection co-added images, per-band image templates (with moving objects and transients removed), calibration reference images, and the ability to generate single-visit calibrated images and difference images on demand. Multi-band (color) images will also be generated to support visualization for the science and outreach communities.

Catalogs

The science database will include catalogs of astrophysical *objects*, detections of *sources* with individual measurements of objects, and an *orbit* catalog of solar system objects. Each catalog will include an extensive set of measured properties, including position, various photometric measurements in all 6 bands, photometric redshift, and shape.

Alerts

An alert is a notification to the community that a transient event has been observed with LSST. The DMS will generate alerts of changes in the static sky within 60s of the shutter closing, based on detections in difference images relative to the per-band image templates. Triggers will include transient and variable objects, as well as uncataloged moving objects. The alert feed will use community standards to convey source information including position, the available photometric measurements in all passbands, object shape, and postage-stamp images to aid follow-up observations. Users who subscribe to the alert feed will be able to apply filters to select event signatures that are of greatest scientific interest to them.



2. Data Archive and Portal Services

The LSST data access centers (DAC) will provide science data products through a variety of web-based tools and services. User interfaces (UI) are planned for data discovery and retrieval, catalog queries, visualization, and analysis; examples of some successful UIs in use today are shown in Fig. 2. Users will be able to generate a variety of diagrams such as multi-band light curves, color-magnitude diagrams, reduced proper motion diagrams, etc., in a virtual work environment hosted by the LSST DAC, without the necessity to move large quantities of data to their local compute environment.

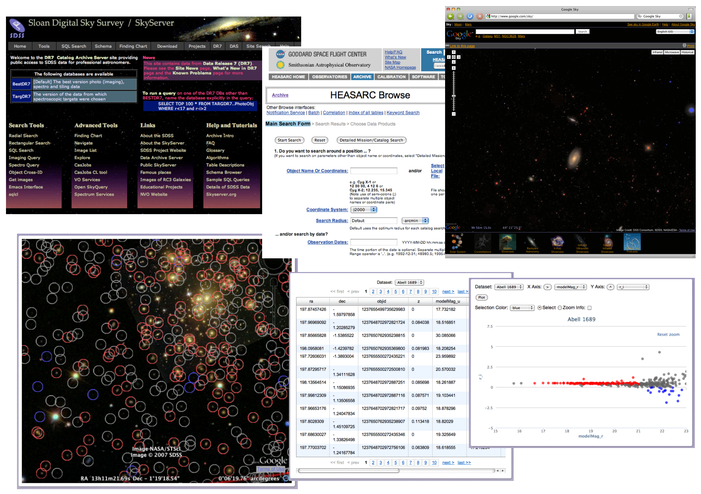


Figure 2: User Interfaces for the Science Archive will draw upon best-in-breed designs and web-based tools to offer access to data products, advanced visualization, and analysis over the internet to support science and community outreach.

In recent years many researchers and collaborations have turned to programmatic interfaces to large datasets. The LSST DACs will support access to data products with scripts and third-party or home-grown software, to perform highly specialized queries or direct processing on images and catalogs to achieve their science goals.

3. Community Science Data Products

While a great deal of science will be possible with the DMS-generated data products, we anticipate that individual investigators, research teams, and major international science collaborations will, during the course of their investigations, generate scientific datasets of great value to the LSST community. Examples might include object classifications, identification of lensing arcs, morphology of resolved galaxies, eclipsing binary parameters, temporal analysis of spatially resolved variable sources (SN light echoes, symbiotic nebulae, pulsar wind nebulae), etc. Generating such data products are beyond the scope of the LSST Project, but are of obvious scientific value.

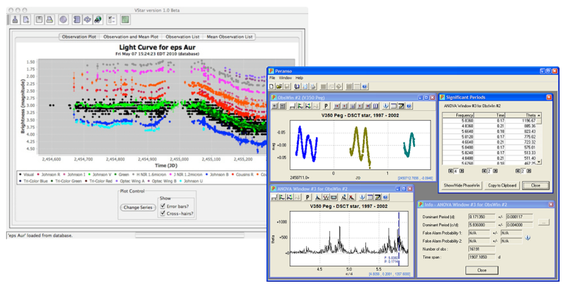


Figure 3: Object classifications and phase-folded light curves are examples of potential community-generated data products that are beyond the scope of the planned DMS, but would have broad science appeal when federated with the LSST Archive.

The LSST DMS has resources in its baseline plan to assist the scientific community with the generation of high value datasets, and with enabling community access to them, including:

- Providing 10% of the available compute resources at the DACs to support processing and storage of community science data products
- Making the full DMS software stack open source, with APIs designed for scientist developers to leverage the software through APIs
- Programmatic access to all DMS-generated data products
- The ability to execute third-party and user-written code on DAC resources
- Federating community-generated datasets with the LSST archive

Particularly valuable products may be selected to become part of the standard DMS processing. This would require finding additional resources to support the code involved, as well as the storage and other resources necessary to persist and serve the data products for each data release. Federating or incorporating community science products will leverage the science potential of the LSST survey, enabling new science for the entire community.

4. Science Data Quality Visualization

The requirements for science quality for LSST data are driven by the ambitious science goals for the survey, including the ability to make robust discoveries of exceedingly rare phenomena in the time domain. With such large data rates and accumulated data volumes in the archive, data quality assessment must be continuous, highly automated, and comprehensive over all elements of the DMS and the data products. The need for human verification of data quality, whether to monitor survey progress, diagnose problems, or to assure community investigators of the scientific integrity of their results, requires efficient user interfaces, classical and machine-learning quality evaluation, and advanced visualization techniques. Example diagnostic diagrams that have been prototyped for DMS are shown in Fig. 4.

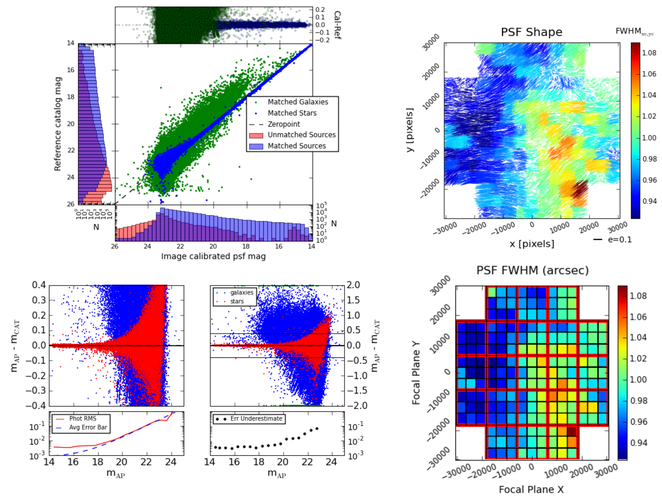


Figure 4: Visualization of science data quality over a variety of measured quantities in a single focal plane. Analysis of Image Simulations in the prototype DMS system shows the completeness of object recovery (UL), photometric accuracy (LL, stars in red), shape and orientation of point sources across the focal plane (UR), and variation in PSF width (LR).

January 2012

LSST is a public-private partnership. Design and development activity is supported in part by the National Science Foundation. Additional funding comes from private gifts, grants to universities, and in-kind support at Department of Energy laboratories and other LSSTC Institutional Members.