Large Synoptic Survey Telescope (LSST)

Summary of Data Management Principles

Steven M. Kahn

LPM-151

Latest Revision: June 30, 2015

Change Record

Version	Date	Description	Owner name
1	6/30/2015	Initial publication	Steven M. Kahn

Table of Contents

The LSST Summary of Data Management Principles

Experiment Description

The Large Synoptic Survey Telescope (LSST) will be a wide-field, large-aperture, ground-based telescope designed to perform an imaging survey of the entire southern hemisphere of sky every few nights. The design incorporates an effective 6.7-m diameter primary mirror and a 9.6 square-degree, 3.2 Gigapixel camera, equipped with 6 optical filters covering the wavelength band 320 – 1050 nm. Over 10 years of operation, LSST will perform a minimum of 825 visits of every part of the southern sky, where a single visit consists of a pair of 15 s exposures separated by a 4 s gap. LSST is presently under construction on Cerro Pachon in central Chile, with first light planned for mid-2020, and the official start of survey operations in late 2022.

The telescope will acquire ~ 5 million images, leading to the detection of ~ 37 billion discrete astronomical objects, 20 billion galaxies and 17 billion stars. A large, complex data management system is under construction to retrieve, process, analyze, and archive what will be a massive data volume, approaching several hundred Petabytes. The primary data archive facility will be at the National Center for Supercomputing Applications (NCSA) at the University of Illinois at Urbana-Champaign. A parallel data processing facility will be developed at CC-IN2P3 in Lyon, France. Data access centers, which will enable scientific users to query the database and perform scientific analyses on the data, will be hosted at NCSA and in La Serena, Chile, as well as possible other sites around the world.

LSST will enable a wide range of complementary scientific investigations utilizing a common database. However, four main scientific themes have been used to motivate the flowdown of scientific requirements: (1) Taking a census of moving objects in the solar system; (2) Mapping the structure and evolution of the Milky Way; (3) Exploring the transient optical sky; and (4) Determining the nature of dark energy and dark matter. The techniques associated with these four themes stress the system design in complementary ways.

DOE's Roles in the Experiment:

Under the auspices of the Office of High Energy Physics, the Department of Energy is supporting the development and fabrication of the LSST Camera, and will contribute ~ 25% of the cost of operating the full facility. For DOE HEP, LSST is a Stage IV Dark Energy Experiment. DOE is also supporting the LSST Dark Energy Science Collaboration (DESC).

Partnerships:

LSST is a public/private interagency project with international participation. The lead federal agency is the National Science Foundation (NSF), which is funding the construction of the telescope and site facility, the data management system, and the education and public outreach components as a Major Research Facility and Experiment Construction (MREFC) Project. This is being undertaken by the Association of Universities for Research in Astronomy (AURA) under a Cooperative Agreement with NSF.

The total not-to-exceed cost for the NSF construction project is $473M.

DOE is the secondary federal agency, with responsibility for the delivery of the LSST camera, which is being fabricated as a Major Item of Equipment (MIE). SLAC National Accelerator Laboratory is the lead laboratory for the LSSTCam Project, which has a baselined MIE cost of $168M. A Memorandum of Understanding exists between NSF and DOE spelling out the commitment by each agency to the overall LSST program.

In addition, the Project has benefitted from the successful solicitation of ~ $40M of private funding by the Large Synoptic Survey Telescope Corporation (LSSTC). LSSTC is 501(c)(3) non-profit corporation, headquartered in Tucson, AZ, with ~38 institutional members comprised mostly of universities and other research organizations. The largest single donor was the Charles and Lisa Simonyi Foundation for the Arts and Sciences. The private money was used to enable the fabrication of the primary/tertiary mirror, the procurement of the secondary mirror blank, early leveling of the site, and the procurement of prototype sensors for the camera.

Internationally, the LSST Project is collaborating with the Chilean astronomical community, given the planned location of the telescope in Chile. In addition, significant technical contributions to the camera development are being made by a suite of laboratories in France associated with the Institut National de Physique Nucleaire et de Physique des Particules (IN2P3).

The organization appropriate to the operations phase of the program is still under discussion. At present, it is planned that NSF will contribute ~ 50% of the operations costs, DOE will contribute ~ 25%, and the remaining 25% will come from a variety of additional international partners under the auspices of LSSTC.

Organization – Agency/Lab Level:

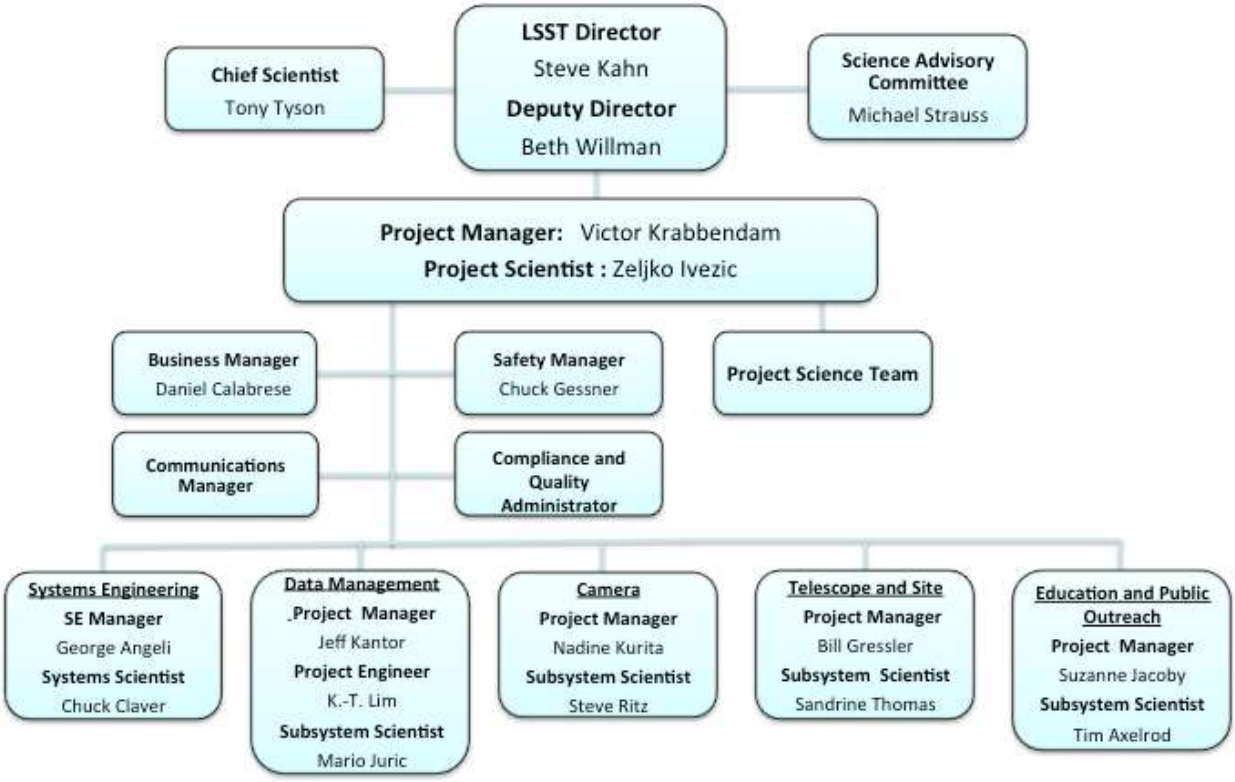
AURA leads the MREFC Project for NSF, and is likely to hold the cooperative agreement for the NSF contribution to operations as well.

SLAC leads the MIE Camera Fabrication for DOE, and will be the host laboratory for the DOE contribution to operations.

SLAC is the host laboratory for the LSST Dark Energy Science Collaboration.

Organization – Experiment Level:

The org chart below shows the functional organization for the overall LSST Project. The LSST Director reports to both the President of AURA and the Director of SLAC, and is responsible to both federal agencies. The LSST Camera Project Manager and Camera Subsystem Scientist support the LSST Director in managing the LSSTCam Project.



Collaboration:

Given that the science of LSST is so broad, it did not make sense to form a single collaboration representing all of the diverse scientific areas. Rather, several separate collaborations have formed, each with its own distinct scientific focus. The collaboration most relevant to the DOE interests is the LSST Dark Energy Science Collaboration (DESC). The DESC was first created in June 2012. Its membership currently consists of ~ 350 scientists, from a range of institutions, mostly within the U.S. However, as international partnerships are formed providing access to LSST data to institutions around the world, we expect the DESC membership to grow. The DESC has an elected Spokesperson who serves for a two-year nominal term. The Spokesperson appoints most of the other leadership positions within the collaboration.

Data Policy Management:

The LSST Project (in consultations with its sponsoring agencies) is responsible for setting all policies on access to both raw and processed data produced under its auspices.

Individual LSST collaborations (including the DESC) may produce secondary data products, which they are free to make available as they see fit.

Data Description & Processing:

There are three types of processed data that will be generated for LSST:

Level 1: These result from difference imaging performed real-time as the data are acquired. Level 1 data products include a stream of ~ 10 million time-domain events per night, detected and transmitted to event distribution networks within 60 s of observation. The Level 1 pipeline will also flag moving objects in the data stream, and determine their orbits. This is expected to result in a catalog of orbits for ~ 6 million small bodies in the solar system.

Level 2: These utilize both the raw images and accompanying housekeeping and calibration data to derive calibrated photometry, astrometry, shape information, and lightcurves for all detected sources. The Level 2 catalogs will eventually contain ~ 37 billion sources, 20 billion galaxies and 17 billion stars. In addition, forced photometry will be performed at the positions of transient sources, yielding ~ 30 trillion individual measurements. These will be accompanied by deep co-added images of every part of the southern sky. The Level 2 data products are released annually, except for the first two, which will be released 6 months apart.

Level 3: Individual collaborations and scientific users will construct additional Level 3 software to reduce the processed data for particular scientific analyses. The LSST Data Access Centers will offer services and computing resources to enable such user-specified custom processing and analysis to be performed, as well as software and APIs to enable development of the requisite codes within the LSST data management framework.

Data Products and Releases:

The LSST Project is responsible for processing the raw LSST images and producing the Level 1 and Level 2 data products. These are made available to the community through the Data Access Centers, which are located in Urbana, Illinois and La Serena, Chile. All raw and processed data are archived at both NCSA in Urbana and La Serena, as well as at CC-IN2P3 in Lyon.

All data acquired by LSST to date are reprocessed annually to produce the Level 2 data products. At any given time, both the current and the prior data release are stored on spinning disks, available for use by the community.

For two years after they are produced, LSST Level 2 data projects are held proprietary for use by the U.S. and Chilean communities, and those of selected other international partners who are contributing to operations. After this two year period, Level 2 data products will be made available to the world, although the Project reserves the right to charge requestors for the costs of distributing the data and/or making them accessible for analysis.

Level 1 data products (time-domain alerts) will be made available to a limited number of event distribution services, as they are generated. These event “brokers” will filter the events to identify particular classes of variable stars and transients, for subsequent release to the community at large. The time-domain event streams are not proprietary, however it will be up to the event distribution services to determine who gets access to their filtered lists.

Plan for Serving Data to the Collaboration and the Community:

The LSST Project will serve data to the U.S., Chilean, and international partner communities through its Data Access Centers. Given the very large data volume, in many cases it will make more sense for users to run their analyses on the computing facilities associated with the Data Access Centers, rather than to download the data to their home facilities. However, the larger collaborations may want to set up their own computing infrastructure to handle the large suite of Level 3 analyses they hope to perform. We expect this to be the case for the DESC; the detailed computing model for the DESC is under development.

If their authors choose to make them available and if the cost impacts are minimal, Level 3 data products can also be federated with the Level 2 data products and released to the broad community, or to a prescribed subset of users.

Plan for Archiving Data:

All raw and derived data for LSST will be archived at both NCSA in Urbana and La Serena, as well as at CC-IN2P3 in Lyon, at least for the duration of the operations phase of the facility, and most likely, for many years afterward.

Plan for Making Data Used in Publications Available:

Access to derived data presented in LSST publications is at the discretion of the authors of those publications. The LSST Project does not take responsibility for this aspect of data release policy. However, consistent with current SC policy, we expect that all data points shown in published graphs in LSST publications will be made available in machine readable form on a website listed in the publication.

Responsiveness to SC Statement on Digital Data Management:

This data management plan fully follows the SC Statement on Digital Data Management.