



Legacy Survey of Space and Time (LSST)
**Summary of Data Management
Principles**

Steven M. Kahn

LPM-151 (rel2.0)

Latest Revision: January 10, 2020

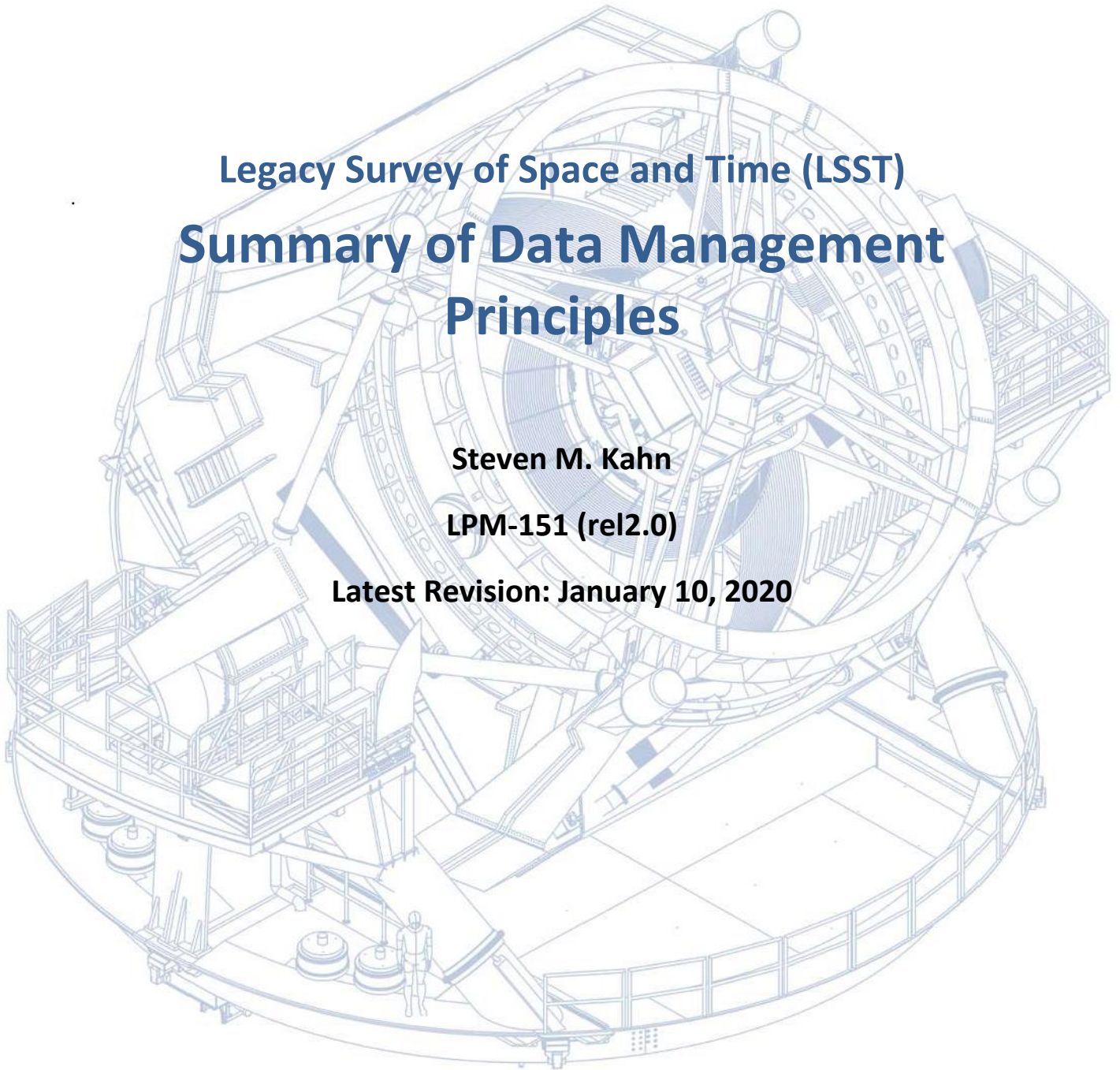




Table of Contents

- Change Record i
- 1 Experiment Description 1
- 2 DOE’s Roles in the Experiment 1
- 3 Partnerships 1
- 4 Organization – Agency/Lab Level..... 2
- 5 Organization – Experiment Level 2
- 6 Collaboration..... 3
- 7 Data Policy Management..... 3
- 8 Data Description & Processing..... 3
- 9 Data Products and Releases..... 4
- 10 Plan for Serving Data to the Collaboration and the Community 4
- 11 Plan for Archiving Data 4
- 12 Plan for Making Data Used in Publications Available 5
- 13 Responsiveness to SC Statement on Digital Data Management 5



The LSST Summary of Data Management Principles

1 Experiment Description

The Legacy Survey of Space and Time (LSST) will be a synoptic sky survey carried out by the wide-field, large-aperture, ground-based Simonyi Survey Telescope at the Vera C. Rubin Observatory that has been designed to perform an imaging survey of the entire southern hemisphere of sky every few nights. The telescope 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 accumulate some 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 Pachón in central Chile, with the official start of survey operations planned for 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 image data will be processed at LSST data facilities in the US and France, while data access centers, which will enable scientific users to query the database and perform scientific analyses on the data, will be hosted at the US data facility 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

2 DOE's Roles in the Experiment

Under the auspices of the Office of High Energy Physics, the Department of Energy (DoE) is supporting the development and fabrication of the LSST Camera, and will contribute ~ 50% of the cost of operating the full LSST survey. For DOE HEP, LSST provides the dataset that enables Stage IV Dark Energy investigations. As such, DOE is also supporting the LSST Dark Energy Science Collaboration (DESC), which will prepare for and carry out the cosmology analysis of the LSST data, including providing feedback about key aspects of the system affecting that cosmological analysis (e.g., observing strategy, image processing algorithms) to the LSST operations teams.

3 Partnerships

LSST operations will be managed jointly by the Association of Universities for Research in Astronomy (AURA) and SLAC, with the National Science Foundation (NSF) and DOE each contributing ~50% of the

operations costs.

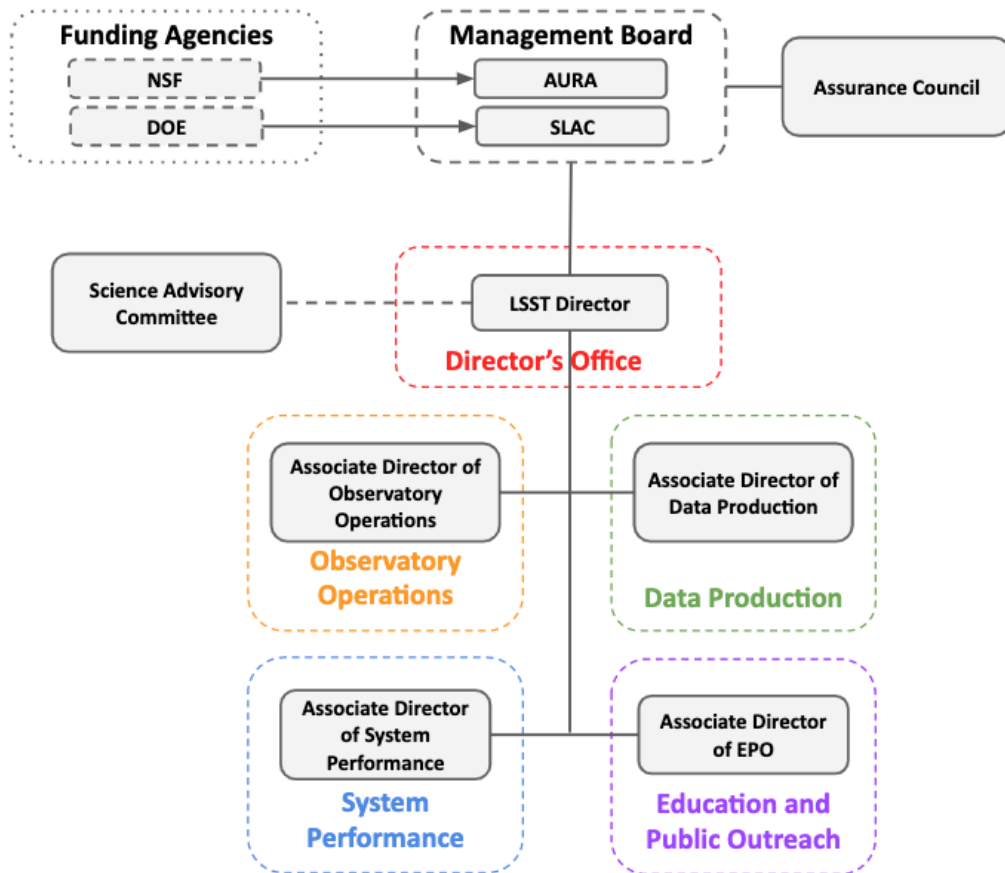
4 Organization – Agency/Lab Level

AURA leads the Vera C. Rubin Observatory (VRO) Major Research Equipment and Facilities Construction (MREFC) Project for NSF, and will hold the cooperative agreement for the NSF contribution to operations as well.

SLAC leads the Major Item of Equipment (MIE) LSST Camera Fabrication for DOE, and is the host laboratory for the DOE contribution to operations. SLAC is the host laboratory for DESC.

5 Organization – Experiment Level

The LSST Operations Director reports to the LSST Operations Management Board, consisting of the President of AURA and the Director of SLAC (or their designates). The LSST Operations Director is supported by a Deputy Director at each managing organization and an Executive Council of operations department heads.



6 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), which has its own Data Management Plan (available from <http://lsstdesc.org>). Seven further science collaborations, each focused on a different field of astronomy beyond cosmology, are recognized by LSST as being key components of the broader LSST science community; all of the LSST science collaborations provide vital input to the optimization of the survey design and valuable feedback on the data management system science pipelines, platform and products.

7 Data Policy Management

The LSST Operations team (in consultation with its sponsoring agencies) is responsible for setting all policies on access to both raw and processed data produced under its auspices. The current version of the LSST Data Policy can be obtained from <http://ls.st/LDO-013>.

Individual LSST users, as well as groups like the LSST science collaborations (including the DESC) may produce secondary “derived data products” (DDPs), which they are free to make available as they see fit. DDPs are defined and regulated in Section 6 of the LSST Data Policy.

8 Data Description & Processing

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

Prompt Products: These result from difference imaging performed real-time as the data are acquired. Prompt data products include a stream of ~ 10 million time-domain events (“alerts”) per night, detected and transmitted to event distribution filtering networks (“alert brokers”) within 60 s of observation. The prompt product 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.

Data Release Products: 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 data release 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 data release data products are released annually, except for the first two, which will be released approximately 6 months apart.

User-generated Products: Individual collaborations and scientific users will construct additional software to analyze the processed data for particular scientific investigations. 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. User-generated data products that meet the relevant

criteria may be designated as Derived Data Products, and distributed by their creators beyond the LSST data rights community, as specified in the LSST Data Policy.

9 Data Products and Releases

LSST Operations is responsible for processing the raw LSST images and producing the prompt and data release data products. These are made available to the community through the Data Access Centers. All processed data are archived at each of the US, Chilean and French data facilities.

All data acquired by LSST to date are reprocessed annually to produce the data release 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 data release products are held proprietary for use by the LSST data rights holders (all members of the U.S. and Chilean communities, plus those from other countries who hold LSST data rights agreements). After this two year period, the data release 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.

Prompt 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.

10 Plan for Serving Data to the Collaboration and the Community

LSST Operations will serve data to the U.S., Chilean, and international data rights holder 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 user-generated analyses they hope to perform. We expect this to be the case for the DESC, which maintains its own detailed computing model.

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

11 Plan for Archiving Data

Data products from LSST will be archived at all of the US, French and Chilean data facilities, at least for the duration of the operations phase of the facility, and most likely, for many years afterward.



12 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 (see Section 7 of the LSST Data 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.

13 Responsiveness to SC Statement on Digital Data Management

This data management plan fully follows the SC Statement on Digital Data Management¹.

¹ <https://science.osti.gov/Funding-Opportunities/Digital-Data-Management>