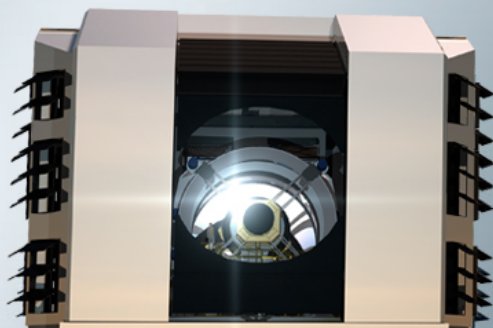




LSST Science Operations (and Beyond!)

Beth Willman
LSST Deputy Director

SAC Telecon
April 4, 2016





What activities and resources must LSST include in its proposed Science Operations directorate, to satisfy our operational requirements? How might we approach estimating the required level of effort within this scope?

What additional support, services and capabilities should be provided [for example: by NOAO, an LSSTC Science Institute, DESC], to enable community research with LSST data? [Research = Publications]

What am I asking the SAC to do?



1. Given LSST's operational requirements: Evaluate and provide advice on strengthening the scope and level of effort we're proposing for our Science Operations directorate.
2. Provide a prioritized set of suggestions for services, capabilities, and support - outside of this boundary of science operations – that the community will need to do research with LSST data.*

*There is a parallel, complimentary study of the ground-based OIR capabilities needed to maximize LSST science.



What activities and resources must LSST include in its proposed Science Operations directorate, to satisfy our operational requirements? How might we approach estimating the required level of effort within this scope?

What additional support, services and capabilities should be provided [for example: by NOAO, an LSSTC Science Institute, DESC], to enable community research with LSST data? [Research = Publications]



The Boundaries and Normalization of Science Operations is driven by one requirement:

We must deliver a science-ready dataset to the community.



The Boundaries and Normalization of Science Operations is driven by one requirement:

We must a dataset to the community.

These words are straightforward to define.

“community” – The US, Chilean and international contributor community

“dataset” – See next slides

Scale of the Dataset



Final Image Collection – All DRs	515 PB	All Data Releases Includes Virtual Data (475 PB)
Final Image Collection – DR11	114 PB	Data Release 11 (Year 10) Includes DR11 Virtual Data (88 PB) and all raw images from all years
Final Database	16 PB	Data Release 11 (Year 10) Includes Data, Indexes, and DB Swap
Final Disk Storage	375 PB	Archive and Base Sites
Final Tape Storage	121 PB	Single Site, Single Copy Only
Peak Number of Nodes	1750	Archive and Base Sites Compute and Database Nodes
Number of Alerts Generated	28 billion	Life of survey

- Virtual Data is data that is dynamically recreated on-demand from provenance information



There are virtually infinite options on what quantities one can measure in images and store in catalogs.

1. “Maximize science enabled by the catalogs”

- Working with images takes time and resources; a large fraction of LSST science cases should be enabled by just the catalog.

2. “Provide simple but useful, commonly used, external or derived, quantities”

- Example: $E(B-V)$ values for each object.
- Example: Photo- z using well known, published, algorithms.

3. Minimize information loss

- Provide (as much as possible) estimates of likelihood surfaces, not just single point estimators

4. Provide and document the transformation (the software)

- Measurements are becoming increasingly complex and systematics limited; need to be maximally transparent about how they’re done



There are virtually infinite options on what quantities one can measure in images and store in catalogs.

1. “Maximize science enabled by the catalogs”

- Working with images takes time and resources; a large fraction of LSST science cases should be enabled by just the catalog.

2. “Provide simple but useful, commonly used, external or derived, quantities”

- Example: $E(B-V)$ values for each object.
- Example: Photo- z using well known, published, algorithms.

3. Minimize information loss

- Provide (as much as possible) estimates of likelihood surfaces, not just single point estimators

4. Provide and document the transformation (the software)

- Measurements are becoming increasingly complex and systematics limited; need to be maximally transparent about how they’re done

Products to be Delivered by the LSST Project



- A stream of ~10 million time-domain events per night, detected and transmitted to event distribution networks within 60 seconds of observation.
- A catalog of orbits for ~6 million bodies in the Solar System.

Nightly
(Level 1)

- A catalog of ~37 billion objects (20B galaxies, 17B stars), ~7 trillion single-epoch detections (“sources”), and ~30 trillion forced sources, produced annually, accessible through online databases.
- Deep co-added images.

Annual DRS
(Level 2)

- Services and computing resources at the Data Access Centers to enable user-specified custom processing and analysis.
- Software and APIs enabling development of analysis codes.

Added Value
(Level 3)

Products to be Delivered by the LSST Project



- A stream of ~10 million time-domain events per night, detected and transmitted to event distribution networks within 60 seconds of observation.
- A catalog of orbits for ~6 million bodies in the Solar System.

Nightly
(Level 1)

See the Data Products Definition

- A catalog of single-epoch detections (“sources”), and ~30 trillion forced sources, produced annually, accessible through online databases.
- Deep co-added images.

Document, <http://ls.st/dpdd> for full details

Annual DRS
(Level 2)

- Services and computing resources at the Data Access Centers to enable user-specified custom processing and analysis.
- Software and APIs enabling development of analysis codes.

Added Value
(Level 3)

What about “deliver” and “science-ready”?



The Boundaries and Normalization of Science Operations is driven by one requirement:

We must **deliver** a **science-ready** dataset to the community.

- In addition to providing high quality, calibrated data products, we must validate them against quantitative Science and System requirements;
- We must provide well-documented data products and software;
- We will need a science assurance system that continually monitors, supports, and responds to science pipelines, science products, and survey strategy.
- In addition to providing services to explore and extract data, we need to explicitly ensure that the community can use the products to accomplish their scientific goals,

Currently envisioned scope of Science Operations



Survey Performance and Monitoring group

- Simulations

- Survey Evaluation and Optimization

- Survey Monitoring Tools

Currently envisioned scope of Science Operations



Survey Performance and Monitoring group

- Simulations

- Survey Evaluation and Optimization

- Survey Monitoring Tools

Community Support group

- User Help Desk

- Level 3 Resource Allocation

- SUI/T and Level 3 User Support

- Documentation

- User's Committee

- Communication with Scientific Community

- Science User Interface and Toolkit

Currently envisioned scope of Science Operations



Survey Performance and Monitoring group

- Simulations

- Survey Evaluation and Optimization

- Survey Monitoring Tools

Community Support group

- User Help Desk

- Level 3 Resource Allocation

- SUI/T and Level 3 User Support

- Documentation

- User's Committee

- Communication with Scientific Community

- Science User Interface and Toolkit

Science Products group

- Nightly (Level 1) Data Verification and Software

- Annual (Level 2) Release Scientific Verification and Software

- Science Data Quality Assurance Tools

Trade offs for required Level of Effort



Tradeoffs of FTEs vs time:

- To evaluate products vs. requirements and reacting to results
- For operations team members to identify bugs
- To fix bugs
- To respond to help requests
- To evolve the software
- To deliver data releases

Tradeoffs of FTEs vs ability to mitigate risks:

- Science can be done with LSST-delivered resources
- Observing strategy performing relative to science goals
- System understood well enough to meet science goals
- Community support for facility (relates to usability)



What activities and resources must LSST include in its proposed Science Operations directorate, to satisfy our operational requirements? How might we approach estimating the required level of effort within this scope?

**What additional support, services and capabilities should be provided [for example: by NOAO, an LSSTC Science Institute, DESC], to enable community research with LSST data?
[Research = Publications]**



What activities and resources must LSST include in its proposed Science Operations directorate, to satisfy our operational requirements? How might we approach estimating the required level of effort within this scope?

**What additional support, services and capabilities should be provided [for example: by NOAO, an LSSTC Science Institute, DESC], to enable community research with LSST data?
[Research = Publications]**

I'm interested to focus this discussion on NOAO and its possible contributions



- LSST has been instructed by the NSF to deliver a description of the “full scope” of “enhanced community services” needed to enable the US community to conduct research with LSST data products.
- Building on the foundation of the Elmegreen report’s recommendations, the NSF has decided that NOAO will be central to their investment in LSST beyond core research infrastructure (on the non-computing side of things).
- As NOAO is starting to develop its concept for an LSST Community Science Center, it is working with the Elmegreen report’s recommendations, and is looking to what will come from the upcoming Kavli-support study of the ground-based OIR capabilities needed to maximize LSST science.
- Thoughts from the SAC on priorities for outside the operations boundary would be helpful for these developing plans.



What services, support and capabilities will the U.S. science community will to enable research with LSST data, during:

The early years (commissioning/early operations – first two years of operations)

Years 3 – 7 of Operations

Years 8 – 10+2 of Operations

Looking ahead – What about a SAC-led white paper for the 2020 Decadal Review?

What am I asking the SAC to do?



1. Given LSST's operational requirements: Evaluate and provide advice on strengthening the scope and level of effort we're proposing for our Science Operations directorate.
2. Provide a prioritized set of suggestions for services, capabilities, and support - outside of this boundary of science operations – that the community will need to do research with LSST data.*

*There is a parallel, complimentary study of the ground-based OIR capabilities needed to maximize LSST science.

Extra Slides



LSST's Required Management of this Dataset



- **Archive Raw Data:** Receive the incoming stream of images that the Camera system generates to archive the raw images.
- **Process to Data Products:** Detect and alert on transient events within one minute of visit acquisition. Approximately once per year create and archive a Data Release, a static self-consistent collection of data products generated from all survey data taken from the date of survey initiation to the cutoff date for the Data Release.
- **Publish:** Make all LSST data available through an interface that uses community-accepted standards, and facilitate user data analysis and production of user-defined data products at Data Access Centers (DACs) and external sites.