Alert Streams in the LSST Era: Challenges and Opportunities

Eric Bellm
University of Washington
LSST DM Alert Production Science Lead

for M. Juric; on behalf of
the LSST Data Management Team
Scan the sky...
Scan the sky...
Find things that change.
Find things that change.
Follow them up!

- first P48 image
- first candidate loaded in database
- automated transient alert
- Keck spectrum
- SWIFT observation
- LCOGT observation
- CARMA triggered
- EVLA triggered
- LCOGT triggered
- SWIFT triggered
- second candidate loaded in database
- second P48 image

Hours since 03:30:00 September 3 [PST]
When making decisions, watch out for junk.

But there are lots of real events, too…

Masci+ 2017
Find exotic explosions...

Gamma-ray bursts

Superluminous Supernovae

Tidal Disruption Events
... binary neutron star mergers...

**Fermi**
Reported 16 seconds after detection

**LIGO-Virgo**
Reported 27 minutes after detection

NASA/LIGO/ESA

Villar+ 2017
... interstellar visitors & “killer” asteroids...
... and weird stars.

Kepler Light Curve for KIC 8462852

Boyajian+ 2016
Many surveys are already active.
Events are sorted and reported a wide variety of ways.

private databases & scripts

public webpages

email lists
Astronomer’s Telegram
GCN
IAU circulars

Transient Name Server

VOEvent Network
Events are sorted and reported in a wide variety of ways.

- Private databases & scripts
- Public webpages
- Email lists
  - Astronomer’s Telegram
  - GCN
  - IAU circulars
- Transient Name Server
- VOEvent Network

More manual, target & science specific

More automated, general purpose
The Large Synoptic Survey Telescope will produce an alert stream of greater scale and generality than any survey to date.

An automated 8.4 meter telescope that for 10 years will image half the sky every ~3 days, generate ~50 PB of (raw) imaging data, issue real-time alerts to any changes in the sky (~10 million/night), measure properties of ~40 billion objects in the sky (~1000 times each), and make the results available in a web-accessible database.

First Light: 2019
Operations: 2022
LSST is located in Cerro Pachon, Chile.
The summit, April 2015.
The summit, February 2018.
LSST is a database of the optical sky.

LSST data, including images and catalogs, will be available with no proprietary period to the astronomical community of the United States, Chile, and International Contributors. LSST’s alerts are immediately world-public.

LSST is a public facility: all science will be done by the community (not the Project!), using LSST’s data products.

The ultimate deliverable of LSST is not the telescope, nor the instruments; it is the fully reduced data. LSST is a facility that delivers data products and data access and analysis services.
We are building a multi-continent Data Management System.

**Satellite Processing Center**
(CC-IN2P3, Lyon, France)
- Data Release Production (50%)

**Archive Site**
- Archive Center
  - Alert Production
  - Data Release Production (50%)
  - EPO Infrastructure
  - Long-term Storage (copy 2)
- Data Access Center
  - Data Access and User Services

**Chilean Sites**
- Telescope and Camera
- Data Acquisition
- Crosstalk Correction
- Long-term storage (copy 1)
- Chilean DAC Entry-point

**HQ Site**
- Science Operations
- Observatory Management
- Education and Public Outreach
LSST has three data processing modes.

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.

For more details, see the “Data Products Definition Document”, [http://ls.st/dpdd](http://ls.st/dpdd)
Prompt: Time-Domain Alerts

We expect a high rate of alerts, **approaching 10 million per night**. We’ll also provide an *alert filtering service*, to select subsets of alerts, as well as serve the full stream to external *event brokers*.

Each alert will include the following:
- **Alert and database ID**: IDs uniquely identifying this alert.
- **The photometric, astrometric, and shape characterization of the detected source**
- **30x30 pixel (on average) cut-out of the difference image (FITS)**
- **30x30 pixel (on average) cut-out of the template image (FITS)**
- **The time series (up to a year) of all previous detections of this source**
- **Various summary statistics (“features”) computed of the time series**

The goal is to *quickly* transmit nearly everything LSST knows about any given event, enabling downstream classification and decision making.

Prompt processing also includes nightly identification of Solar System Objects.
Prompt Processing: System Architecture

Time

Camera

Visit 1
Snap 0

Shutter open/close (not to scale)
Prompt Processing: System Architecture

Time

Camera

Visit 1
Snap 0

Shutter open/close (not to scale)

T-36
NightMOPS Prediction

T-33
Template/Catalog Pre-Load
Prompt Processing: System Architecture

- **Visit 1 Snap 0**
- **Shutter open/close (not to scale)**
- **T-36 NightMOPS Prediction**
- **Template/Catalog Pre-Load**

**Time**

0 10 20 30 40 50 60 70 80 90 100 110 120 130 140
Prompt Processing: System Architecture

Time

Camera

Readout

Shutter open/close (not to scale)

Transmit

Visit 1
Snap 0

NightMOPS Prediction

Template/Catalog Pre-Load

T-36

T-33

T-15

Compress + WAN
Prompt Processing: System Architecture

![Diagram of the system architecture showing various components and time intervals.]
Prompt Processing: System Architecture
Prompt Processing: System Architecture

- Time
- Camera
- Readout
- Shutter open/close (not to scale)
- Transmit
- Visit 1 Snap 0
- Visit 1 Snap 1
- NightMOPS Prediction
- Template/Catalog Pre-Load
- T-36
- T-33
- T-15
- T-5
- T-10
- Compress + WAN
- Per-Snap

60 sec Alert Deadline
T=0
T+60
Prompt Processing: System Architecture

Time

Camera

Visit 1
Snap 0

Visit 1
Snap 1

Shutter
open/close
(not to scale)

 Transmit

Readout

T=0

T+60

60 sec Alert Deadline

NightMOPS Prediction

Template/Catalog Pre-Load

Compress + WAN

Per-Snap

Compress + WAN

Per-Snap
Prompt Processing: System Architecture

Time

Camera

Visit 1
Snap 0

Visit 1
Snap 1

Readout

Shutter open/close
(not to scale)

Transmit

T-36

NightMOPS Prediction

Template/Catalog Pre-Load

T-33

T-15

T-10

Compress + WAN

Per-Snap

T-5

T+3

Compress + WAN

Per-Snap

T+8

T+13

Image Char & Differencing

T+37

60 sec Alert Deadline

T=0

T+60

REAL-TIME DECISION MAKING • BERKELEY, CA • FEB. 26, 2018
Prompt Processing: System Architecture

Time

Camera

Visit 1
Snap 0

Visit 1
Snap 1

Shutter open/close
(not to scale)

Transmit

Readout

T-36

NightMOPS Prediction

Template/Catalog Pre-Load

T-15

T-10

Per-Snap

Compress + WAN

T-5

T+3

T+8

T+13

T+37

Image Char & Differencing

Src Assoc

T+39

60 sec Alert Deadline

T=0

T+60

Compress + WAN

Per-Snap

REAL-TIME DECISION MAKING • BERKELEY, CA • FEB. 26, 2018

21
Prompt Processing: System Architecture
Prompt Processing: System Architecture
Prompt Processing: System Architecture

Time

Camera

Readout
Shutter open/close (not to scale)
Transmit

Visit 1
Snap 0
Visit 1
Snap 1
Visit 2
Snap 0
Visit 2
Snap 1

T=0
60 sec Alert Deadline
T+60

T-36
NightMOPS Prediction

T-33
Template/Catalog Pre-Load

T-15
Compress + WAN
Per-Snap

T-10
Compress + WAN
Per-Snap

T-5
T-3

T+8
T+13
T+37

T+39
Src Assoc
Alert Gen
SDQA

T+55
T+47

String 1

String 2

NightMOPS Prediction
Template/Catalog Pre-Load

Image Char & Differencing
Src Assoc
Alert Gen
SDQA
Prompt Processing: System Architecture

- **Time:**
  - 0:00 to 140:60
  - T=0: Start of process
  - T+60: 60 sec Alert Deadline

- **Camera:**
  - Visit 1: Snap 0
  - Visit 1: Snap 1
  - Visit 2: Snap 0
  - Visit 2: Snap 1
  - Visit 3: Snap 0
  - Visit 3: Snap 1

- **String 1:**
  - T-33
  - T-36
  - T-10
  - T-15
  - T-13
  - T+3
  - T+5
  - T+8
  - T+13
  - T+37
  - T+47
  - T+55
  - T+62
  - T+67
  - T+75

- **String 2:**
  - T-33
  - T+3
  - T+8
  - T+13
  - T+37
  - T+47

- **Processes:**
  - NightMOPS Prediction
  - Template/Catalog Pre-Load
  - Image Char & Differencing
  - Src Assoc
  - Alert Gen
  - SDQA
LSST has three data processing modes.

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.

A catalog of ~37 billion objects (20B galaxies, 17B stars), ~7 trillion observations (“sources”), and ~30 trillion measurements (“forced sources”) accessible through online databases.

Reduced single-epoch, deep co-added images.

For more details, see the “Data Products Definition Document”, http://ls.st/dpdd
Data Releases provide the most thorough processing.

Made available in *Data Releases*
- Annually, except for Year 1
  - Two DRs for the first year of data

Well calibrated, consistently processed, catalogs and images
- Catalogs of objects, detections, detections in difference images, etc.

Complete reprocessing of all data, for each release
- Every DR will reprocess *all* data taken up to the beginning of that DR

Projected catalog sizes:
- *18 billion objects* (DR1) $\rightarrow$ *37 billion* (DR11)
- *750 billion observations* (DR1) $\rightarrow$ *30 trillion* (DR11)
Data Release Catalog Contents

Object characterization (models):
- Moving Point Source model
- Double Sérsic model (bulge+disk)
  - Maximum likelihood peak
  - Samples of the posterior (hundreds)

Object characterization (non-parametric):
- Centroid: \((\alpha, \delta)\), per band
- Adaptive moments and ellipticity measures (per band)
- Aperture fluxes and Petrosian and Kron fluxes and radii (per band)

Colors:
- Seeing-independent measure of object color

Variability statistics:
- Period, low-order light-curve moments, etc.
LSST has three data processing modes.

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.

A catalog of \(~37\) billion objects (20B galaxies, 17B stars), \(~7\) trillion observations (“sources”), and \(~30\) trillion measurements (“forced sources”) accessible through online databases.

Reduced single-epoch, deep co-added images.

Services and computing resources at the Data Access Centers enabling limited analysis, production, and federation of added value products.

Web APIs enabling the use of remote analysis tools.

Public LSST pipeline code for deeper insight into LSST data products.

For more details, see the “Data Products Definition Document”, [http://ls.st/dpdd](http://ls.st/dpdd)
LSST is planning a ten-year survey.

Survey in ugrizy bands, with ~825 visits per pointing

Wide-Fast-Deep: 2x/night every three nights over 18,000 square degrees

Special programs:
- Deep Drilling
- Galactic Plane
- North Ecliptic Spur
- South Celestial Pole

Ongoing cadence development & evaluation: https://github.com/LSSTScienceCollaborations/ObservingStrategy
A series of software pipelines produces the LSST alert stream.

**Single Frame Processing**

**Alert Generation**

**Alert Distribution**

**Forced Processing**

---

**Nightly Processing Pipeline**

- Crossstalk Corrected
- Calibration
- PSF and Background Characterization
- Image
- External sources
- DIA Objects
- Objects
- SS Objects
- Ephemeris Calculation
- Image Differences
- Difference Image
- DIA Sources
- Source Association
- DIA Image
- DIA Sources
- DIA Image
- DIA Objects
- Postage stamp generation
- Message Queue
- Simple Filter Service
- VOEvent Writer
- To LSST SALT
- Forced photometry for DIA objects
- DIA Objects
- DIA Forced Sources
- DIA Objects
- DIA Forced Sources
- DIA Image
- Precovery of new DIA objects
- DIA Objects

---

**LSST LDM-151:** Data Management Applications Design

ls.st/LDM-151
Single-Frame Processing provides calibrated exposures.

LSST LDM-151: Data Management Applications Design

ls.st/LDM-151
Alert Generation detects and associates transients.

Alert Generation

Coadd → Template generation → Template

Image

External sources → DIA Objects

Objects → SS Objects → Ephemeris Calculation

Image Differencing

DIA sources

Difference Image

DIA Sources

DIA Objects

ls.st/LDM-151

LSST LDM-151:
Data Management
Applications Design

ls.st/LDM-151
Alert Distribution packages and sends alerts.

LSST LDM-151:
Data Management
Applications Design

ls.st/LDM-151
LSST’s alert stream differs in scale and motivation from current astronomical databases.

Primary interface is an alert stream, not a batch query
Real-time, low-latency, naturally distributed & decentralized

All* subtraction candidates are streamed at low latency
“turn the database inside out”
(“alert” is somewhat of a misnomer…)

Events sent in (world-public!) rich alert packets enable standalone classification

Users find events of interest through classification & filtering systems
full stream to community brokers: ANTARES, ALeRCE, etc.
simple LSST “mini-broker” filtering service
key decision: is this an object I want to follow up?
LSST uses rich alert packets to minimize follow-up queries.

Each alert will at least include the following:

- **alertID:** An ID uniquely identifying this alert. It can also be used to execute a query against the Level 1 database as it existed when this alert was issued

- **Level 1 database ID**

- **Science Data:**
  - The DIASource record that triggered the alert
  - The entire DIAObject (or SSOBJECT) record
  - All previous DIASource records \(\rightarrow\) last 12 months
  - A matching DIAObject from the latest Data Release, if it exists, and its DIASource records

- Cut-out of the difference image centered on the DIASource (10 bytes/pixel, FITS MEF)

- Cut-out of the template image centered on the DIASource (10 bytes/pixel, FITS MEF)
DIASource and DIAObject records contain a wide range of measurements.

**DIASources:**
- Position
- aperture/PSF/dipole/trailed fluxes
- moments
- likelihoods, extendedness, spuriousness

**DIAObjects:**
- linkages to DIASources [-> light curve], Data Release Objects
- time series statistics

**SSObjects:**
- linkages to DIASources
- variety of solar system parameters
LSST alert distribution requires a new community ecosystem.

At ~20 full sized events per visit per user (or summarizing the lightcurve for all events in ~40 numbers) we can serve ~500 simultaneous users for the cost of a single full data stream.
LSST is testing new technologies for alert distribution.

1. Transport system: Apache Kafka
   - Scalability
   - Replication
   - Allows stream "rewind"

2. Data formatting: Apache Avro
   - Fast parsing with structured messages (typing)
   - Strictly enforced schemas, but schema evolution
   - Allows postage stamp cutout files

3. Filtering/ processing: Apache Spark
   - Direct connection to transport system
   - Stream interface similar to batch
   - Allows for Python or simple SQL-like queries

Maria Patterson, UW
Transport prototyping: Apache Kafka

- Distributed log system/messaging queue
- Reinvented as strongly ordered, pub/sub streaming platform
- Highly scalable, in production at LinkedIn, Netflix, Microsoft
- Great clients + connectors, including Python - good usability
Data formatting: Apache Avro

- Schemas defined with JSON
- Dynamic typing - strict adherence
- Flexible format - schema evolution
- Also used in production, science, recommended by Kafka
Filtering/Processing: Apache Spark

Applications

Environments

Data Sources

DataFrames / SQL / Datasets APIs

RDD API

Spark Core

Spark SQL  Spark Streaming  MLlib  GraphX
Community brokers will enhance the LSST alert stream.

- cross-match with other catalogs and alert streams
- classify events (the LSST Project can only characterize)
- redistribute alert packets
- filter alerts
- provide user interfaces
- enable community coordination
- trigger followup resources and manage that data
- provide storage and archiving
- provide annotation & citation
- manage “discovery”
- ...probably more?

A finite number of brokers will be selected by a proposal process to receive the full stream.
LSST will provide a “mini-broker” service

User-defined filters that act *only* on alert packet contents

Access to the filtered stream through LSST’s Science Platform

Cap of ~20 alerts per user per visit; some limits on computing capacity
Simple single-alert filters can enable a lot of science.
ZTF provides a near-term opportunity to prototype time-domain brokers on an LSST-like alert stream.

First light October 2017

Survey begins March 2018

Planning an LSST-like public alert stream Q2 2018
ZTF & LSST are quite different…
ZTF provides a natural stepping stone to LSST.

ZTF: 1M alerts/night
LSST: 10M alerts/night
The LSST alert stream presents both opportunities and challenges.

Opportunities

• a powerful new facility; huge discovery space
• rich data products to enable general-purpose inference: “batteries included”
• naturally distributed, BYOC

Challenges

• large data volumes and event rates
• sparse & irregular sampling due to LSST cadence
• faint targets; limited followup resources
• need to join with heterogeneous data sets, other alert streams
• LSST survey and tools must serve many science goals
• key scientific capabilities delegated to community brokers not directed by the LSST Project
• how is information shared in a distributed ecosystem?
Conclusions

LSST will deliver an alert stream of unprecedented scale and great scientific potential.

We are prototyping industry-proven technologies to deliver the alert stream.

Discovery and followup of time-sensitive events requires new community-developed decision-making infrastructure.

ZTF will use prototype versions of LSST tools to provide an LSST-like alert stream and filtering service this year.
Are we building a firehose?
or a community fountain everyone can play in?

credit: M. Patterson