Designs and Requirements for LSST Time Domain

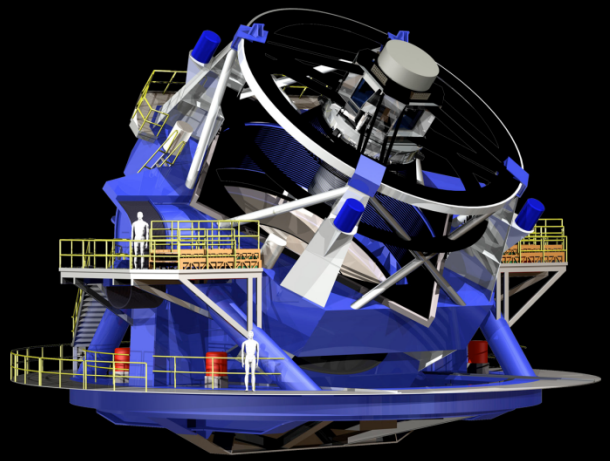
Mario Juric
LSST Data Management Project Scientist

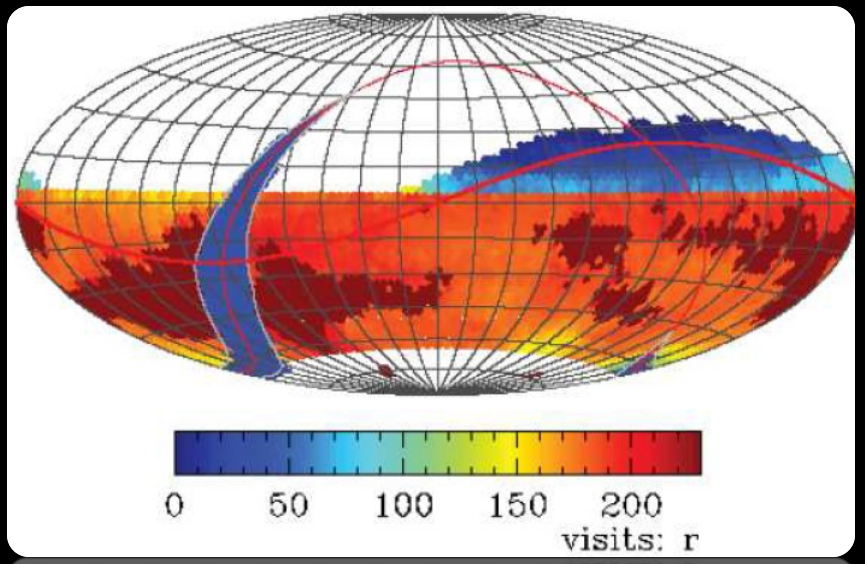
IVOA InterOp Meeting, Heidelberg
May 12-17th, 2013

LSST: A Deep, Wide, Fast, Optical Sky Survey

8.4m telescope optical (ugrizy) 0.5-1% photometry (sys)

3.2Gpix camera 2 x 15sec exp / 2sec read



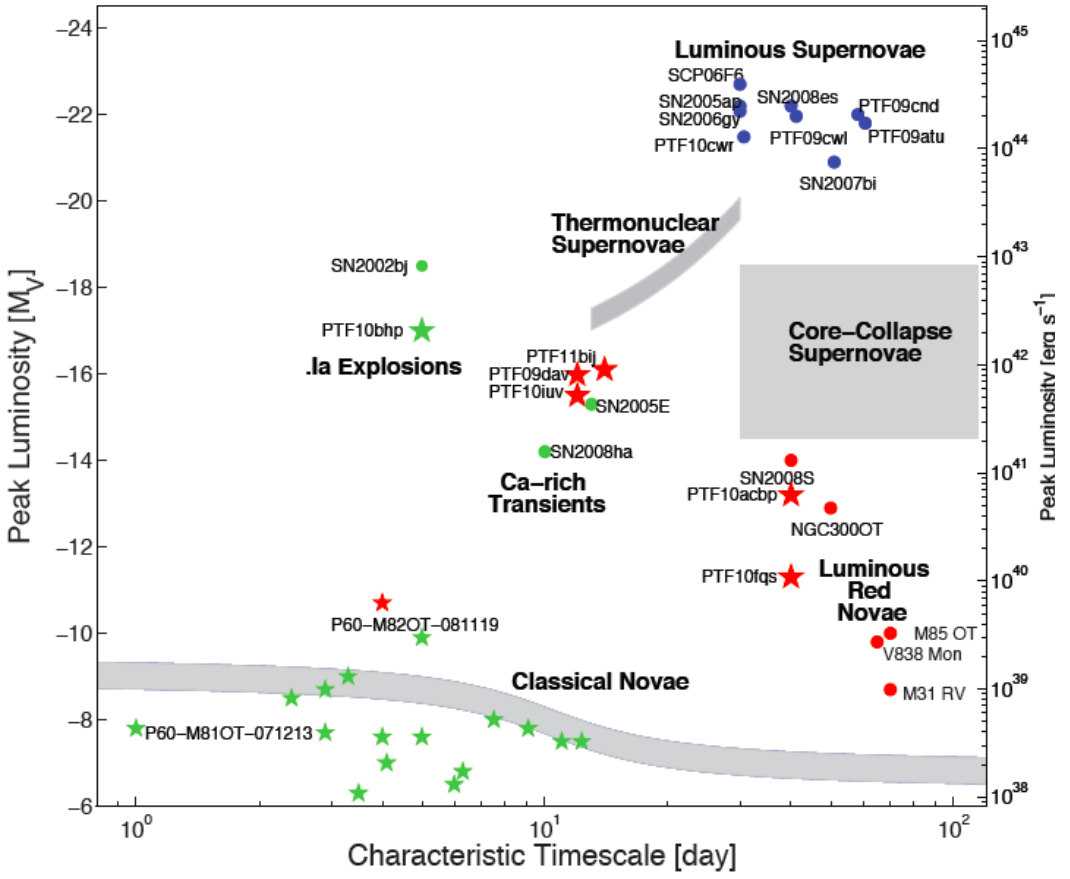


Location: Cerro Pachon, Chile Construction Start: July 2014 (planned)

First Light: October 2019 Operations: October 2021

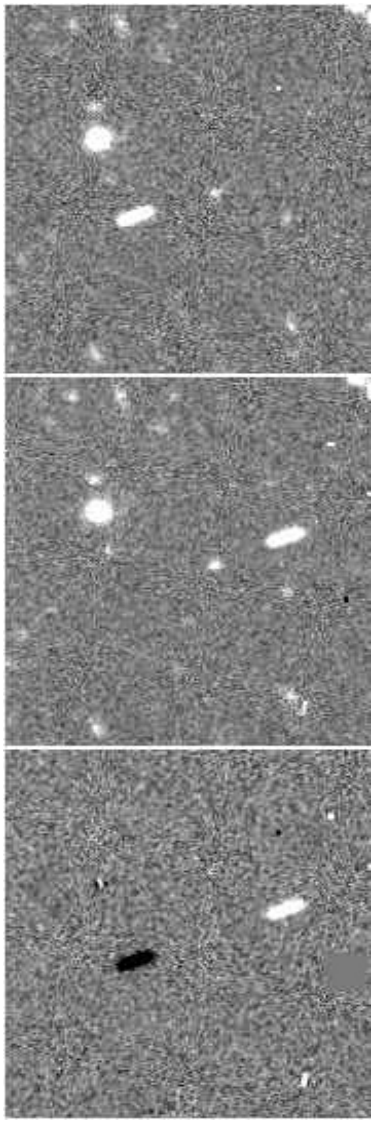
Motivating the LSST (1/4)

- Time domain science
  - Nova, supernova, GRBs
  - Source characterization
  - Instantaneous discovery
- Census of the Solar System
  - NEOs, MBAs, Comets
  - KBOs, Oort Cloud
- Mapping the Milky Way
  - Tidal streams
  - Galactic structure
- Dark energy and dark matter
  - Strong Lensing
  - Weak Lensing
  - Constraining the nature of dark energy



Motivating the LSST (2/4)

- Time domain science
  - Nova, supernova, GRBs
  - Source characterization
  - Instantaneous discovery
- Census of the Solar System
  - NEOs, MBAs, Comets
  - KBOs, Oort Cloud
- Mapping the Milky Way
  - Tidal streams
  - Galactic structure
- Dark energy and dark matter
  - Strong Lensing
  - Weak Lensing
  - Constraining the nature of dark energy



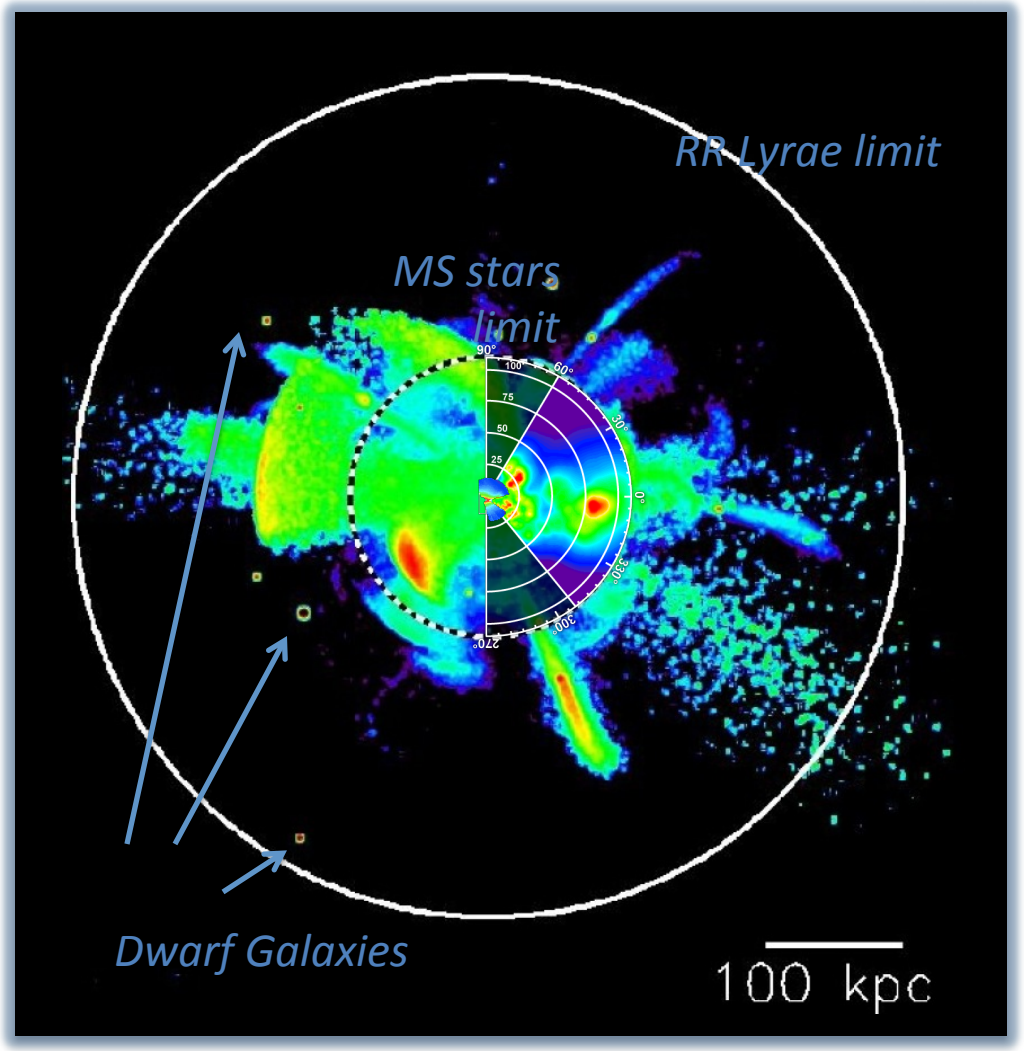
Exposure 1

Exposure 2

Exposure 1
-
Exposure 2

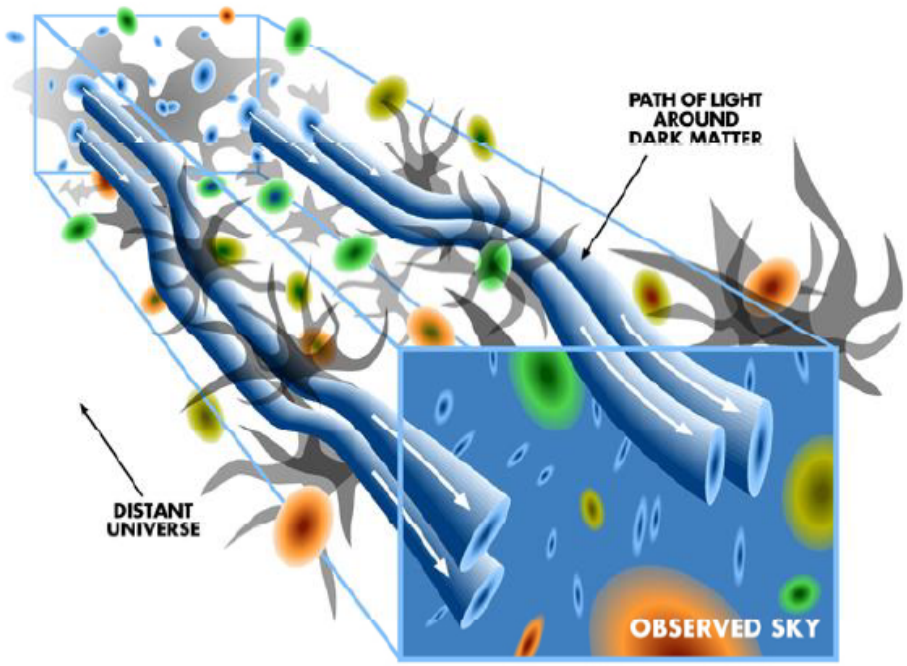
Motivating the LSST (3/4)

- Time domain science
  - Nova, supernova, GRBs
  - Source characterization
  - Instantaneous discovery
- Census of the Solar System
  - NEOs, MBAs, Comets
  - KBOs, Oort Cloud
- Mapping the Milky Way
  - Tidal streams
  - Galactic structure
- Dark energy and dark matter
  - Strong Lensing
  - Weak Lensing
  - Constraining the nature of dark energy



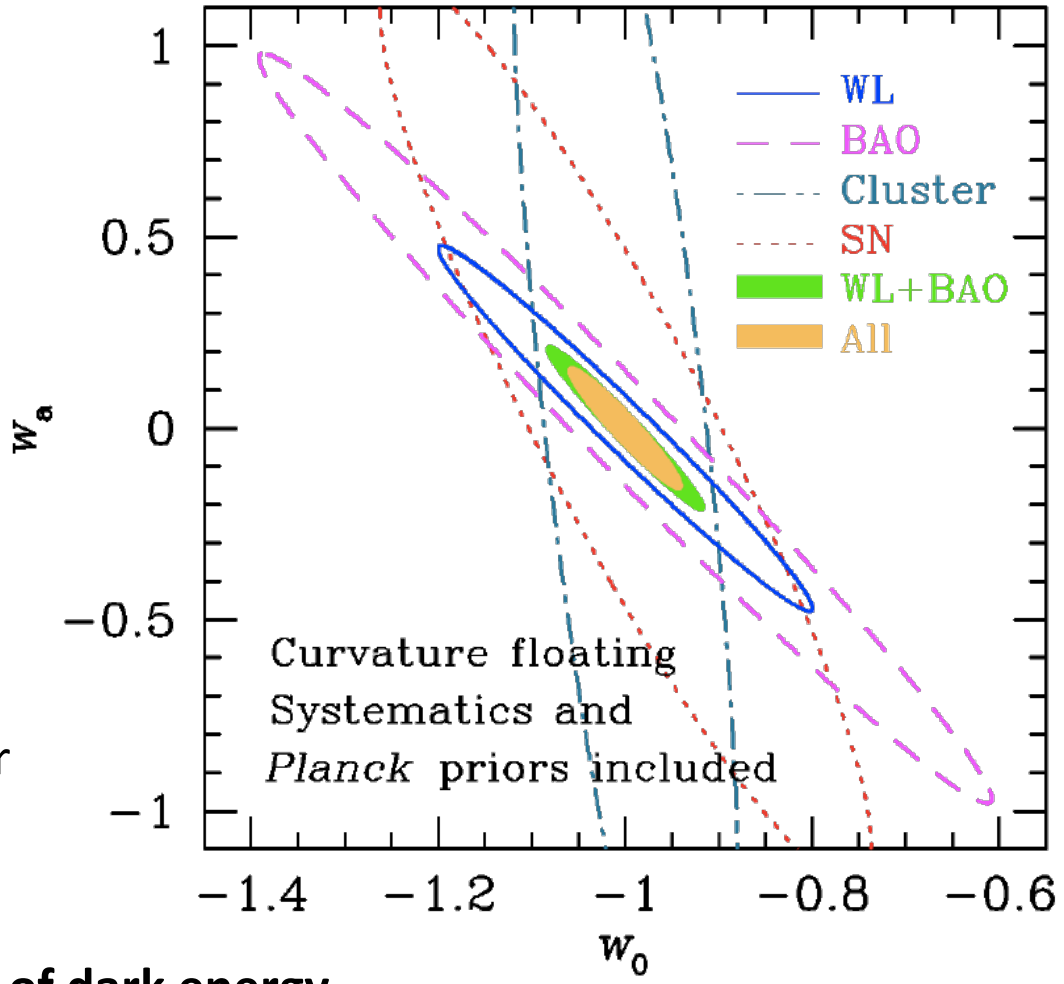
Motivating the LSST (4/4)

- Time domain science
  - Nova, supernova, GRBs
  - Source characterization
  - Instantaneous discovery
- Census of the Solar System
  - NEOs, MBAs, Comets
  - KBOs, Oort Cloud
- Mapping the Milky Way
  - Tidal streams
  - Galactic structure
- Dark energy and dark matter
  - Strong lensing
  - **Weak Lensing**
  - Constraining the nature of dark energy



Motivating the LSST (4/4)

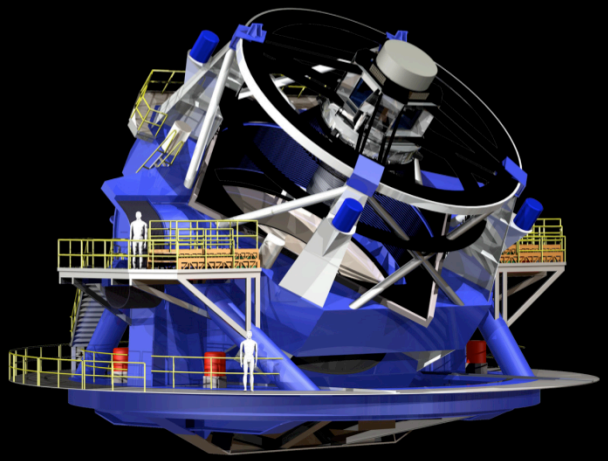
- Time domain science
  - Nova, supernova, GRBs
  - Source characterization
  - Instantaneous discovery
- Census of the Solar System
  - NEOs, MBAs, Comets
  - KBOs, Oort Cloud
- Mapping the Milky Way
  - Tidal streams
  - Galactic structure
- Dark energy and dark matter
  - Strong lensing
  - Weak lensing
  - **Constraining the nature of dark energy**



LSST: A Dedicated Survey

r<24.5 (<27.5@10yr) 18000+ deg^2 10mas astrom.

Imaging the visible sky, once every 3 days, for 10 years (825 revisits)



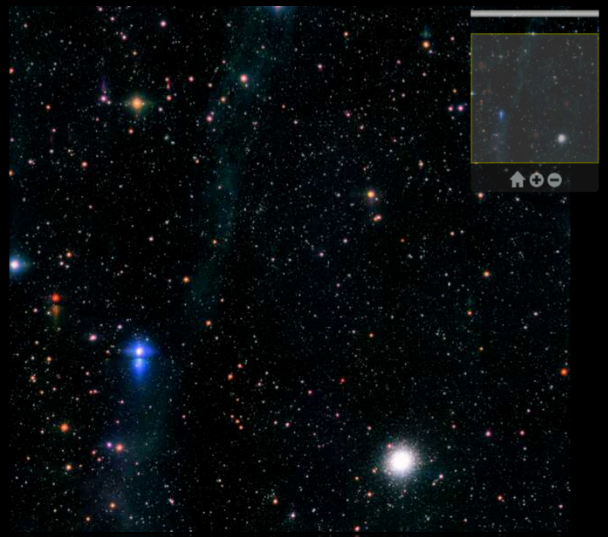
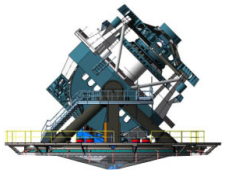


Table 4: Level 2 Catalog Obj

Name	Type	Unit
psRadecTai	double	time
psPm	float[2]	mas/yr
psParallax	float	mas
psFlux	float[ugrizy]	nmgy
psCov	float[66]	various
psLnL	float	
bdRadec	double[2]	degrees

Telescope → Images → Catalogs

"LSST" is the database. The Google Index of the Optical Sky



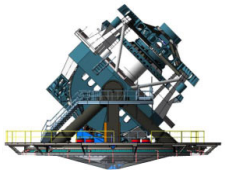
LSST: Data Volume

- One 6.4-gigabyte image every ~17 seconds
- ~1000 visits (two back-to-back images), per night
- 15 terabytes of raw scientific image data / night

- 8.4 terapixel image (movie) of the sky to ~27.5 mag in 6 bands

- A catalog of ~38 billion observed objects (24B galaxies, 14B stars)
- A catalog of ~32 trillion photometric measurements

- ~2000 events per observation (includes variables+asteroids)
- ~2 million events per night, for 10 years
- Requirement: Process & transmit alerts within 60 seconds



LSST DM: Languages and Tools

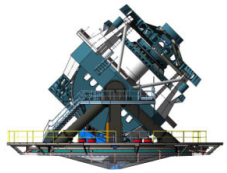
- LSST DM stack is written in Python 2.7, unless computational demands require the use of C++

- Languages
  - **C++**:
    - Computationally intensive code
    - Made available to Python via SWIG
  - **Python**:
    - All high-level code
    - Prefer Python to C++ unless performance demands otherwise

- ~60 packages (git repositories, ~corresponding to Python packages)

- Build system: **scons**
- Version control: **git**

- Free software
  - GPLv3 licensed, **http://dev.lsstcorp.org/cgit**



Software Architecture

A general purpose pipeline toolkit for OIR astronomy

Command-line driver scripts

Cluster execution middleware

Tasks (ISR, Detection, Co-adding, …)

Measurement Algorithms (meas_*)

Camera Abstraction Layer (obs_* packages)

…

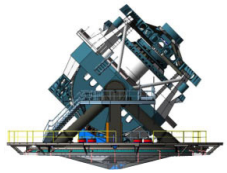
Application Framework (comp. intensive C++, SWIG-wrapped into Python)

Middleware (I/O, configuration, …)

External C/C++ Libraries (Boost, FFTW, Eigen, CUDA ..)

External Python Modules (numpy, pyfits, matplotlib, …)

Red: Mostly C++ (but Python wrapped); Blue: Mostly Python; Black: External Libraries



Levels of Data Products

http://ls.st/srd

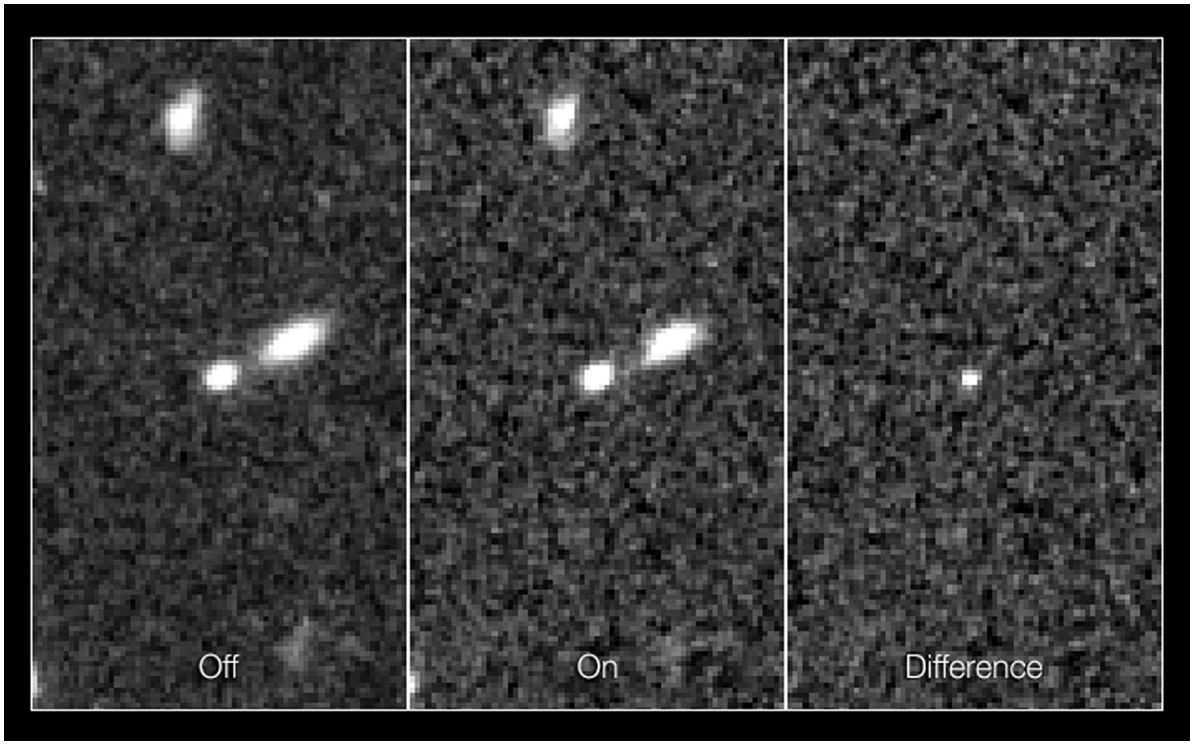
3.5 Data Processing and Management Requirements

Detailed requirements on data processing and management will be described in the LSST System Requirements Document (for example, specifications for catalog completeness and reliability). Here, only a rough guidance is provided. There will be three main categories of data products:

- **Level 1** data products are generated continuously every observing night, including alerts to objects that have changed brightness or position. → Nightly
- **Level 2** data products will be made available as annual Data Releases and will include images and measurements of positions, fluxes, and shapes, as well as variability information such as orbital parameters for moving objects and an appropriate compact description of light curves. → Annual
- **Level 3** data products will be created by the community, including project teams, using suitable Applications Programming Interfaces (APIs) that will be provided by the LSST Data Management System. The Data Management System will also provide at least 10% of its total capability for user-dedicated processing and user-dedicated storage. The key aspect of these capabilities is that they will reside "next to" the LSST data, avoiding the latency associated with downloads. They will also allow the science teams to use the database infrastructure to store their results. → User-created

LSST Level 1 Products: Time Domain

- Detection performed on image differenced against a deep template
- Measurement performed on the difference image and direct image
- Associated with pre-existing observations and stored in a RDBMS
- Transmitted as VOEvents to enable rapid follow-up

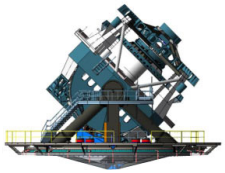


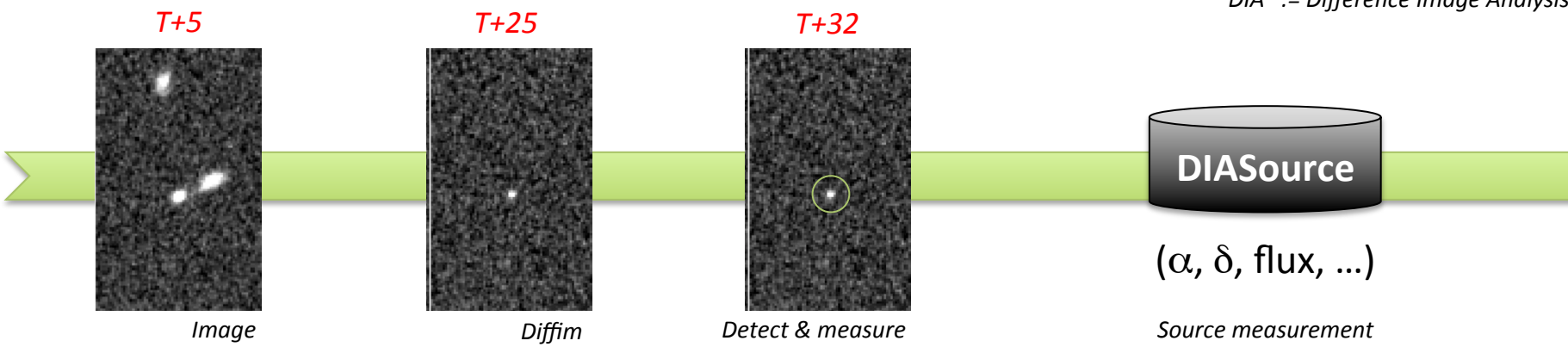
CANDELS (http://www.spacetelescope.org/images/heic1306d/)

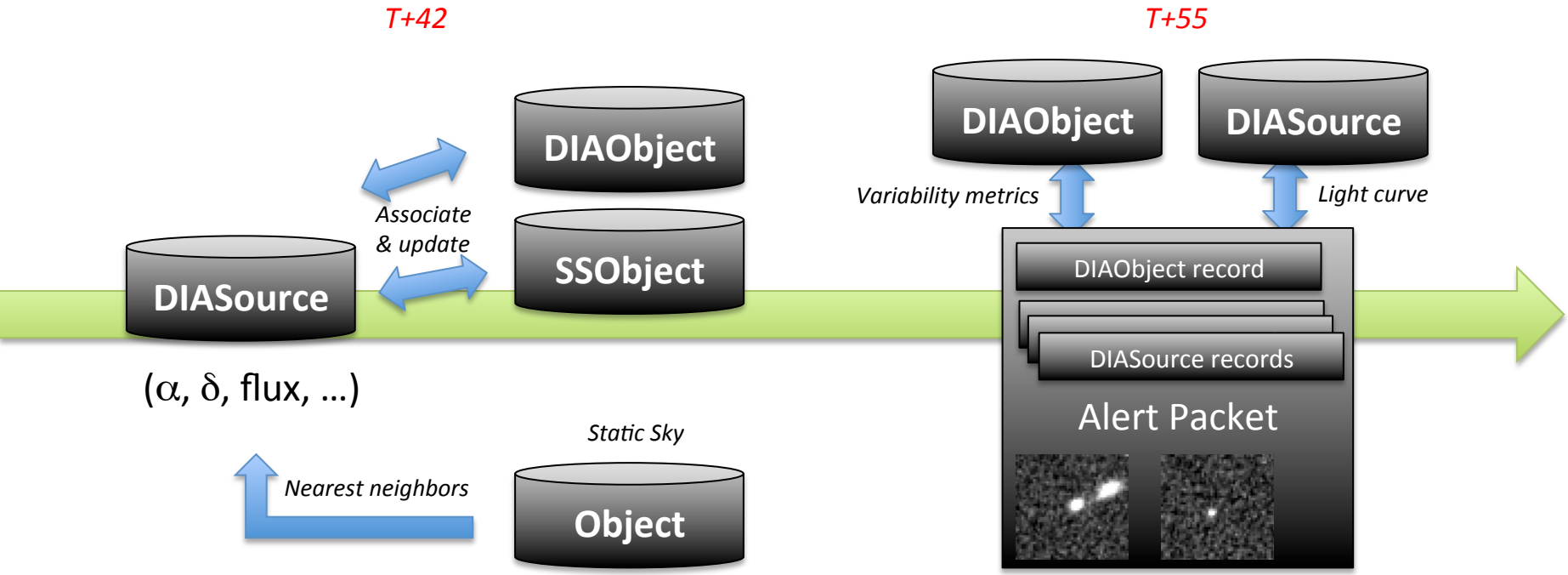
- *Also: time-domain from direct imaging (not covered in this talk)*

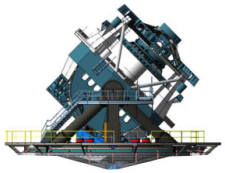
LSST Alert Processing Pipeline

"DIA" := Difference Image Analysis









Measurements

Measured from a difference image

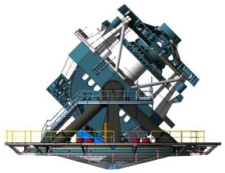
- **DIASource** records:
  - Position
  - Shape (adaptive Gaussian moments; Bernstein & Jarvis 2002)
  - Model fits:
    - Point source model
    - Trailed source model
    - Dipole model

Expecting ~2,000,000 / night (average)

Computed from the set of associated DIASources

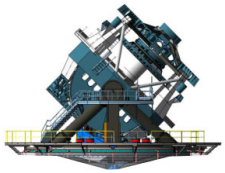
- **DIAObject** records:
  - Position, parallax, and proper motion fit
  - Point source flux
  - Variability characterization (eg., Richards et. al 2012 parameters)
  - Nearby static sky object IDs

Expecting ~few tens of million (entire survey)



Use Case #1: Enabling Rapid Follow-up

- Publish event information as a VOEvent. Include:
  - The DIASource record that triggered the alert
  - The associated DIAObject record
    - or SSObject record, if this was an asteroid
  - Associated DIASource records
    - this is really the “light” curve
  - Image cut-outs
    - 30 x 30 cut-out of the difference image (FITS MEF)
    - 30 x 30 cut-out of the template image (FITS MEF)
  - Metadata

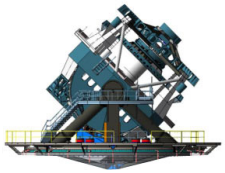


Use Case #2: Enabling Statistical Studies

- Relational database
  - Current baseline: mysql
  - Exploring alternatives: HBase, Cassandra, etc.

- Expected sizes:
  - **sizeof(DIASource) = 200 bytes**
  - **sizeof(DIAObject) = 600 bytes**
  - raw db update size per alert = 800 bytes/alert
  - db update size per night = 1.6 GB
  - db update size per year (300 nights) = 480 GB
  - **Time domain data size ~ 0.5 TB * (t/yr)**

- Level 1 dataset will be downloadable (small)
  - Enabling offline data-mining
  - Online: qserv/mysql



LSST and the Virtual Observatory

- **Access protocols**
  - Internal: native RDBMS protocol, HTTP(S)
  - Public: TAP, SIAP, VOSpace, OAuth/OpenID, …
- **Catalogs**
  - Internal: native RDBMS storage
  - Query: native SQL, ADQL
  - Bulk: compressed FITS binary tables (or HDF5)
- **Images:**
  - Internal: internal format (may be FITS)
  - Public: compressed FITS (MEF)
- **Time Domain Events**
  - Internal: native RBMS storage
  - External: VOEvent + VOEvent Transport Protocol

LSST DM Philosophy:

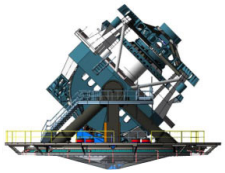
Prefer formats, protocols, and standards **widely supported and used** in the astronomy and physics communities, or industry (were appropriate).

Where no such standards exist, or are inadequate, propose a solution and work with the community to standardize it.

"Rough consensus and running code"

More philosophy:

Make everything **machine readable and machine accessible.**



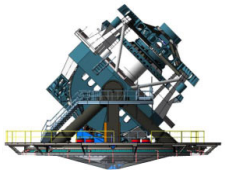
Emerging Issues: Libraries

- LSST has chosen Python 2.7 as its primary language
  - Rapid application development and scripting
  - Significant penetration in the astronomical community
  - Appears well on its way to replace IDL/IRAF-cl/… as a new de-facto standard
- VO community is largely Java oriented

- **Issue: No commonly agreed-upon, supported, documented, high-performance, Python(-wrapped) implementations of many VO standards**
  - Eg. TAP, ADQL, VOTable, SIAP
  - **Serious obstacle to adoption**

- Significant recent contributions by open source volunteers and VO programs
  - atpy, astropy, etc.
  - Various national VO efforts (eg. DaCHS)
  - New work underway in the US-VAO

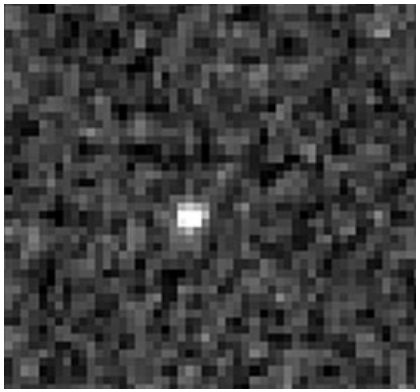
- **We would prefer an open, community-accepted and supported, solution.**

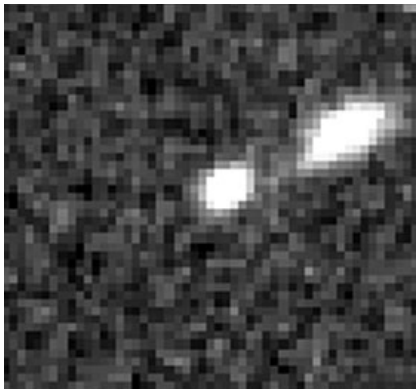


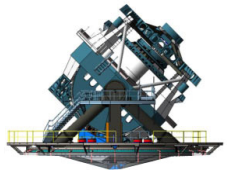
Emerging Issues: VOEvent Features

- For performance reasons, need to minimize alert-triggered callbacks to LSST database
  - Send all information in-band, included in the event packet
  - Avoid <Reference>s (that would trigger actual callbacks)

- **Currently cannot pack FITS images in VOEvent**
  - Could be solved by simple extensions







Emerging Issues: VOEvent Overhead

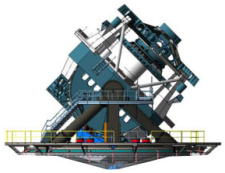
- Existing VOEvent format imposes significant overheads
- Example:
  - From http://wiki.ivoa.net/internal/IVOA/IvoaVOEvent/example1-v1.0.xml
  - Actual transmitted information:

    ~40 bytes
  - Bytes on the wire:

    787 (!!!)

```
▼<VOEvent xmlns="http://www.ivoa.net/xml/VOEvent/v1.0"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 id="ivo://raptor.lanl/VOEvent#23564" role="observation"
 version="1.0"
 xsi:schemaLocation="http://www.ivoa.net/xml/VOEvent/v1.0
 http://www.ivoa.net/internal/IVOA/IvoaVOEvent/VOEvent-v1.0.xsd">
  ▼<Who>
     <PublisherID>ivo://raptor.lanl/organization</PublisherID>
     <Date>2005-04-15T14:34:16</Date>
   </Who>
  ▼<What>
     <Param name="RA" ucd="pos.eq.ra" unit="deg" value="185.0"/>
     <Param name="Dec" ucd="pos.eq.dec" unit="deg" value="13.2"/>
     <Param name="magnitude" ucd="phot.mag;em.opt.R" unit="mag"
     value="18.2"/>
   </What>
  ▼<Why>
     <Concept>Fast Orphan Optical Transient</Concept>
   </Why>
 </VOEvent>
```

- **Issue: Overheads of over ~10-20% probably unsupportable**
- **More compact serialization, provisions for bulk delivery, needed**



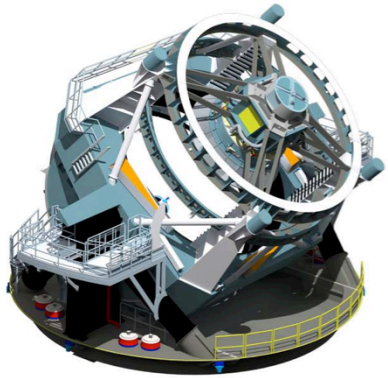
Summary

- LSST is committed to providing access to its data using formats, protocols, and conventions widely adopted by its user community.

- VO standards play a role, especially in time domain/alert distribution.

- We've identified potential scaling issues with VOEvent. We're looking forward to working with the transient community to find solutions to these.

Thank You!



LSST@Europe: September 9-12 in Cambridge, UK!
Conference web page: **http://ls.st/c13**

@LSST @mjuric http://lsst.org http://dev.lsstcorp.org