

Source: LSST

LSST-France data processing activities

fabio hernandez

Contents

- LSST data management overview
- Ongoing and foreseen activities related to data processing
- Summary

LSST data management

LSST subsystems

- Telescope and site
- Camera
- Data management

- IN2P3 involved in both camera construction and data management
- In this presentation we focus on data management

LSST data management

- Archival

*to **record, transport** and permanently **store raw data** issued by camera*

- Processing

*to **detect transients** and **emit alerts** within 60 seconds after observation*

*once per year, to **release a self-consistent, immutable dataset**, composed of processed data since the beginning of the survey (“Data Release Processing”)*

*to **develop the software** necessary for processing the data: image processing algorithms (calibration, point spread function, co-addition of images, characterization of objects, processing pipelines, ...), catalogue database, middleware (workload management, orchestration, ...), data transfer, etc.*

- Publication

*to **deliver the reduced data** (images + catalogs)*

to facilitate custom data reduction and individual data analysis

Data products

Nightly

Stream of 10M time-domain **events** per night, detected and transmitted to event distribution networks within 60 seconds of observation

Catalog of **orbits** for 6M bodies in the Solar System

Annual

Catalog of 37B objects (20B galaxies, 17B stars), 7T observations, 30T measurements, produced annually, accessible through databases

Deep co-added **images**

On demand

Services and computing **resources** to enable user-specified custom processing and analysis

Software and APIs enabling development of analysis code

Source: LSST

Data volume

- Raw data

7.2 GB per image

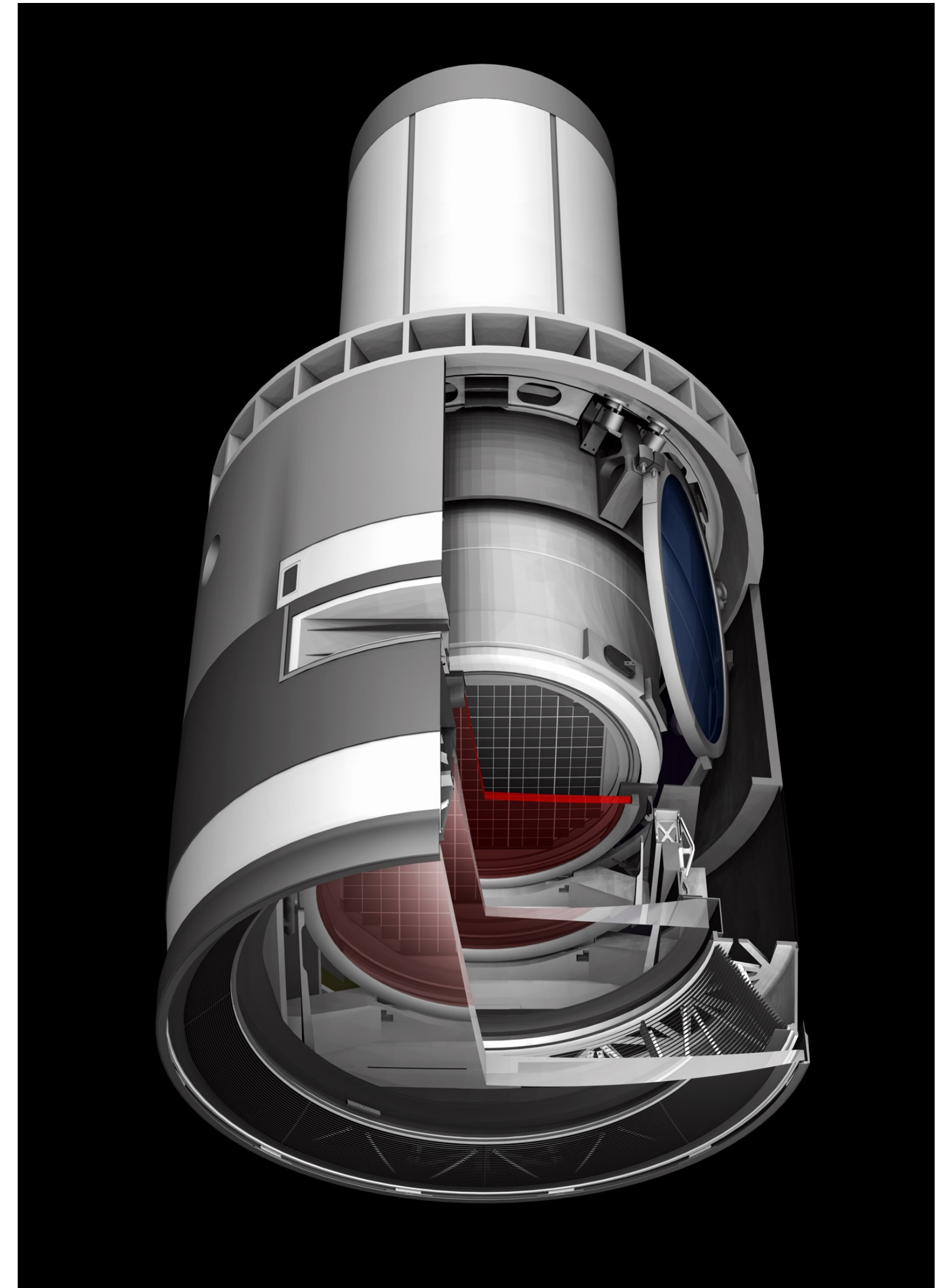
2000 science images + 450 calibration images per night, 300 nights per year

15 TB per night, 4.5 PB per year

- Aggregated data over 10 years of operations, including derived data

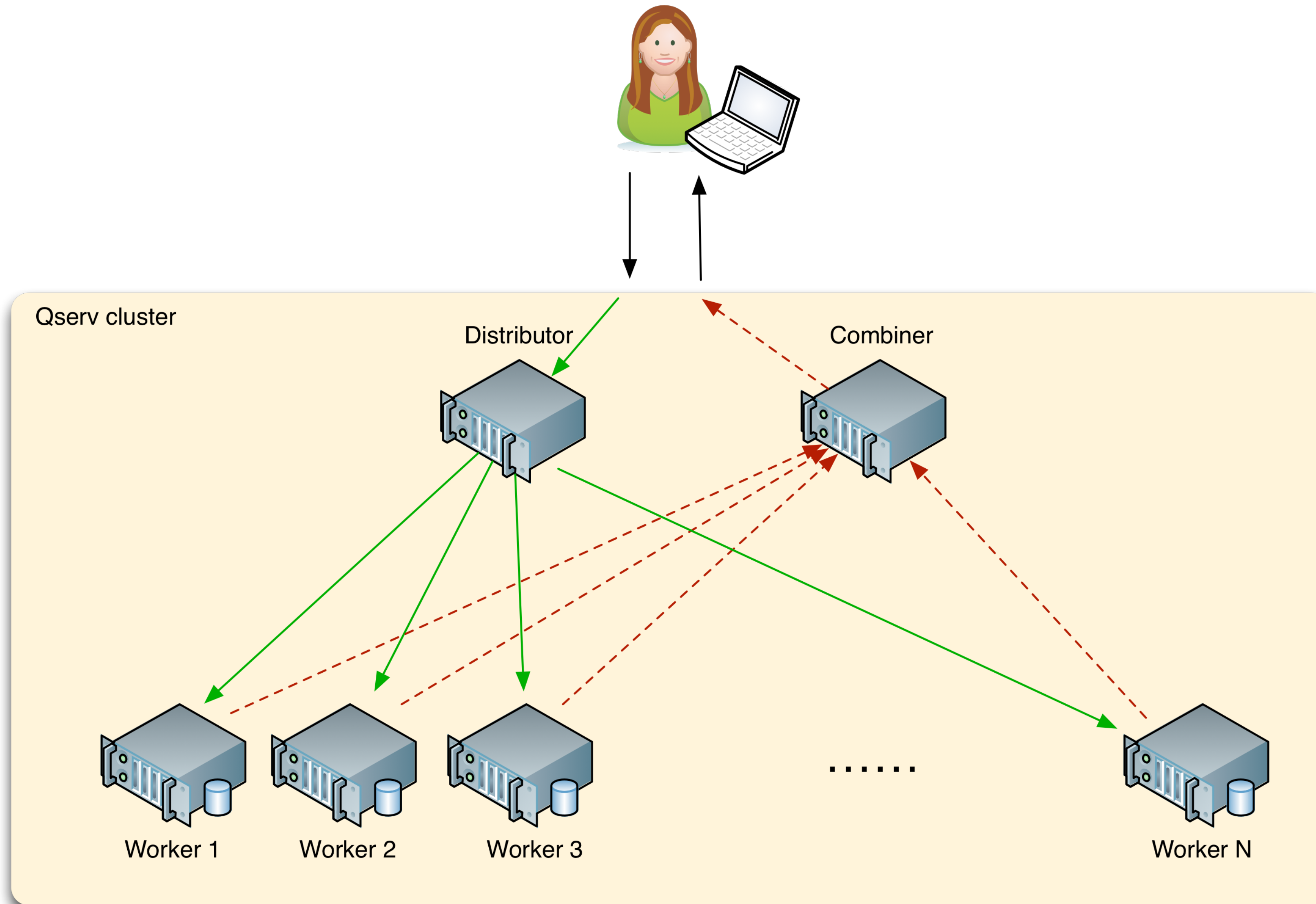
images: 515 PB, ~6M exposures

catalog: 83 PB



Source: LSST

LSST astronomical catalog architecture



user expresses queries in (extended) SQL

distributor interprets the query and dispatches translated requests to the relevant worker nodes

each **worker** performs the requested work against its own independent instance of MySQL with local data

combiner collects results and presents them to the user

LSST at CC-IN2P3

- Formal agreement signed between LSST Corp., NCSA and IN2P3 in March 2015

- Main roles of CC-IN2P3

satellite data release production under NCSA leadership

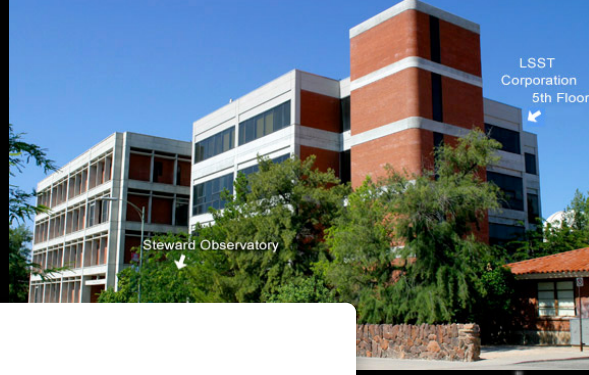
CC-IN2P3 to **process 50% of the data** and store the full dataset, both raw and derived data

both NCSA and CC-IN2P3 will exchange and validate the data produced by the other party

each site to host an **entire copy of every annual data release**

currently also working towards understanding the scope, requirements and services for a **data access center** (not part of the agreement)

LSST DATA CENTERS



HEADQUARTERS SITE

HQ facility
*observatory management
science operations
education & public outreach*



ARCHIVE SITE

Archive center
*alert production
data release production
calibration products production
long-term storage (copy 2)
education & public outreach
infrastructure*

Data access center
data access and user services

SATELLITE RELEASE PRODUCTION SITE

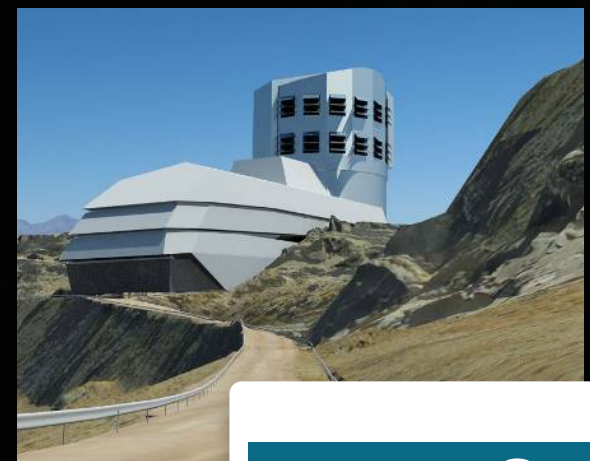
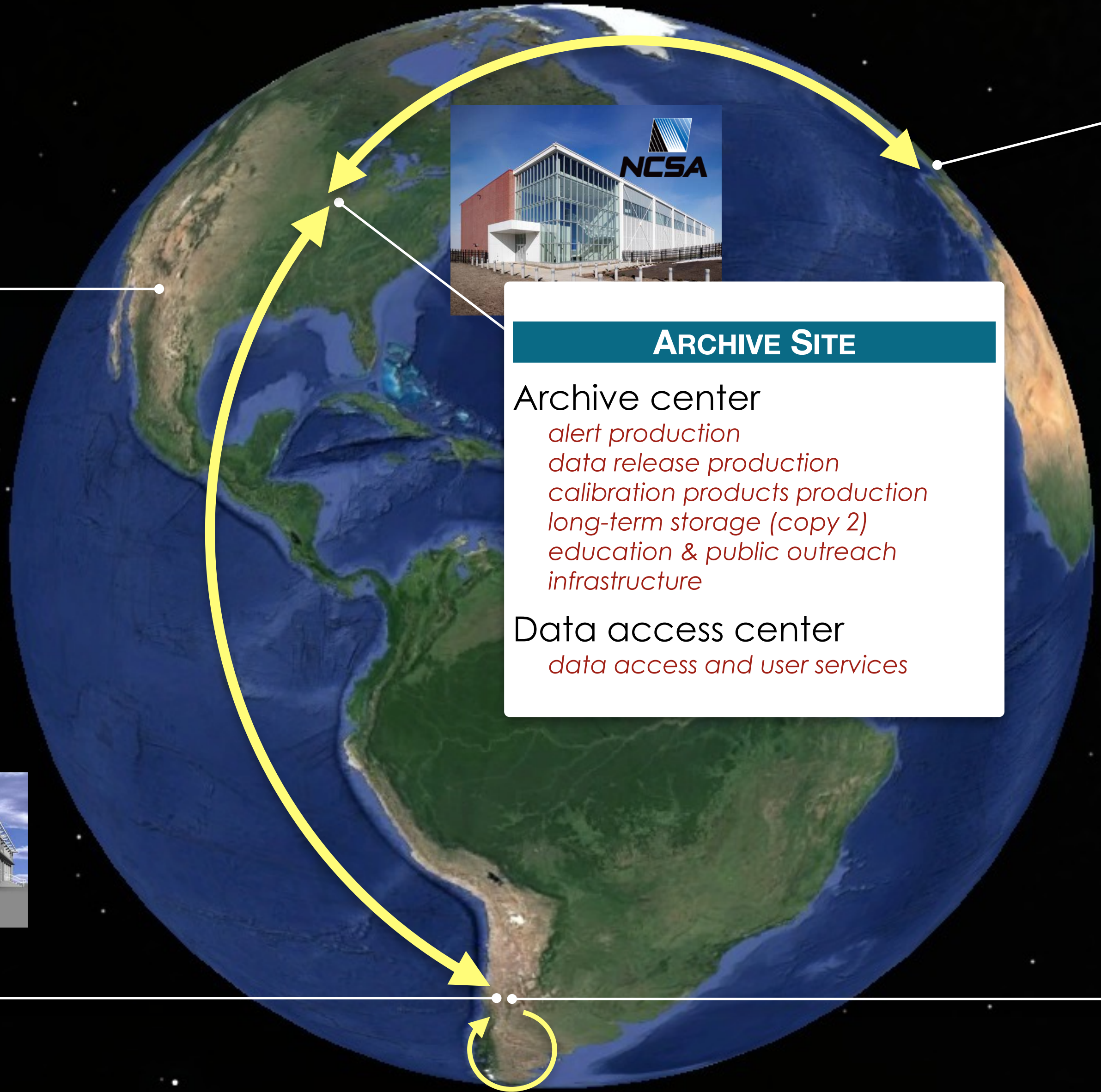
Archive center
*data release production
long-term storage (copy 3)*



BASE SITE

Base facility
long-term storage (copy 1)

Data access center
data access and user services



SUMMIT SITE

Summit facility
*telescope & camera
data acquisition
crosstalk correction*

LSST data management contributors



Stanford Linear
Accelerator Center



National Center for
Supercomputing Applications
University of Illinois at
Urbana-Champaign



Princeton University



Infrared Processing and
Analysis Center
California Institute of
Technology



IN2P3 / CNRS computing center

Ongoing activities

- [CC-IN2P3] LSST software distribution
both stable and weekly versions
local and cloud
- [LPNHE, LAPP, ...] Dataset repository at CC-IN2P3
- [LPNHE, CC-IN2P3] Experimentation with big memory compute nodes to understand what hardware components are relevant for future LSST data processing
fast local disks (SSD), large RAM, high bandwidth network card
- [LPC, CC-IN2P3] Operation of Qserv development cluster
50 nodes, 400 CPU cores, 800 GB of memory, 500 TB raw storage capacity
developers located at SLAC with contribution by F. Jammes (LPC Clermont)

Ongoing activities (cont.)

- [LAPP, LPC] Understanding mechanisms for Qserv data ingestion and query
- [LPC] Exploration of alternative infrastructure for Qserv based on distributed storage platform
- [LAL] Exploration of Apache Spark as a parallel data processing platform
- [CC-IN2P3] Prototyping a Python notebook-based data analysis platform *in line with the future project's Science Platform*
- [CC-IN2P3] Exploration of containers as packaging mechanisms for individual steps of the workflow for data release processing

Foreseen activities

- (Automated) Reprocessing of CFHT data with LSST software stack
- Processing of HSC data with LSST software stack
details to be discussed with contributors
- Ingestion of resulting objects into Qserv instance at CC-IN2P3
definition of data base schema, develop of bulk data ingestion tools
- Deployment of Science Platform components for visualisation and data exploration
development of high level tools for querying Qserv
- Prototyping LSST data backbone
data transport, data catalogue and file location services

Foreseen activities (cont.)

- To revisit the sizing model and deployment plan of computing resources for data release processing at CC-IN2P3
- To organise training sessions: generic Python, LSST software stack
- Documentation platform for LSST-France specifics

Upcoming event: lyon2017.lsst.fr

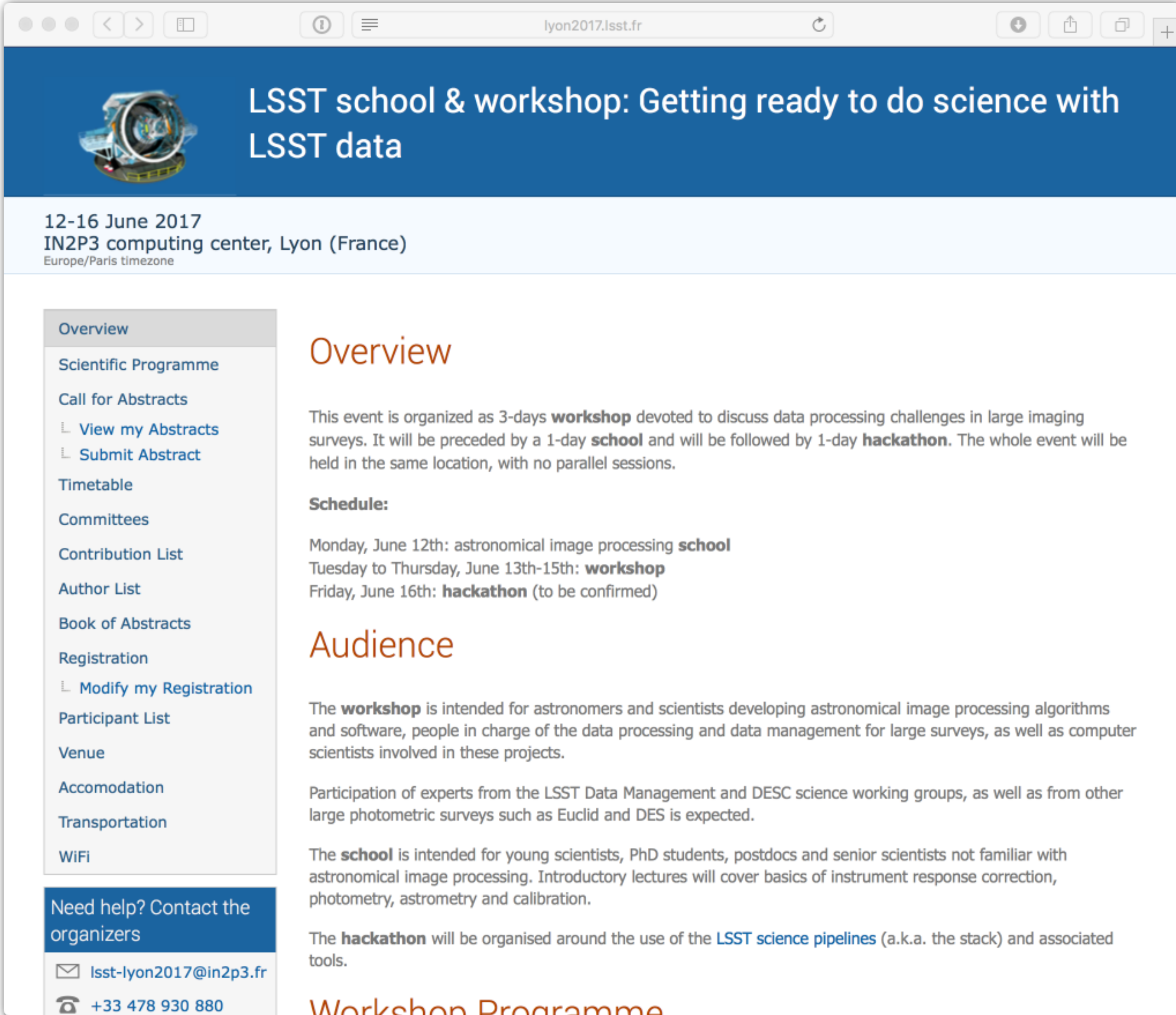
school + workshop + hackathon

school theme: introduction to image processing for astronomy

workshop theme: data processing challenges in large imaging surveys

hackathon theme: LSST software framework

no parallel sessions



The screenshot shows a web browser window with the URL lyon2017.lsst.fr. The page features a blue header with the LSST logo and the title "LSST school & workshop: Getting ready to do science with LSST data". Below the header, the event dates "12-16 June 2017" and location "IN2P3 computing center, Lyon (France)" are displayed. A left sidebar contains a navigation menu with items like "Overview", "Scientific Programme", "Call for Abstracts", "Timetable", and "Registration". The main content area is titled "Overview" and contains text describing the event's structure: a 1-day school, a 3-day workshop, and a 1-day hackathon. It also includes a "Schedule" section with dates for each activity and an "Audience" section detailing the target groups for the workshop, school, and hackathon. At the bottom, there is contact information for organizers, including an email address and a phone number.

Summary

Summary

- LSST-France intends to play a central role in LSST data release processing
- Ongoing activities in several fronts
- Expertise with required components being progressively acquired
- Reprocessing tasks with precursor data very useful to get insight into the whole process: from raw data to catalog query

QUESTIONS & COMMENTS

