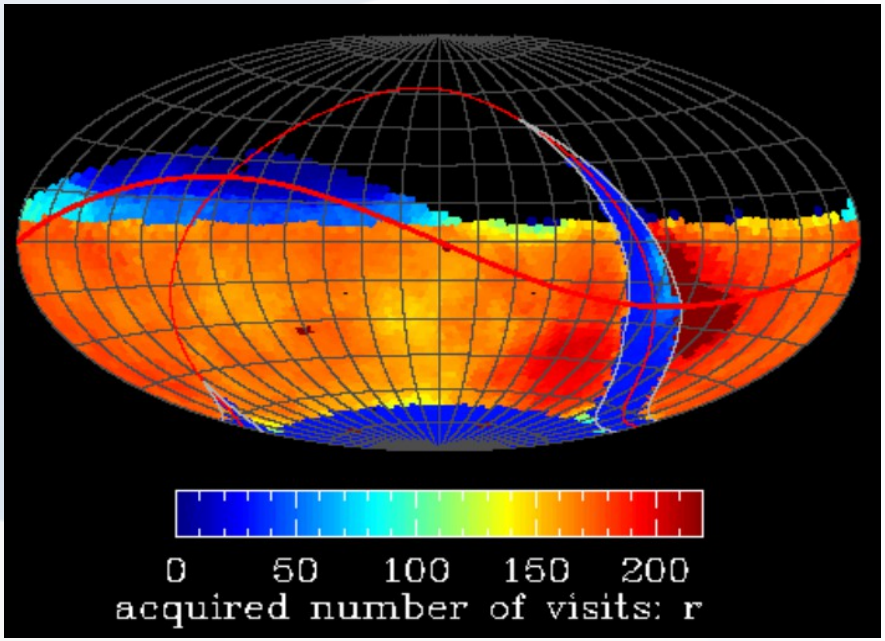
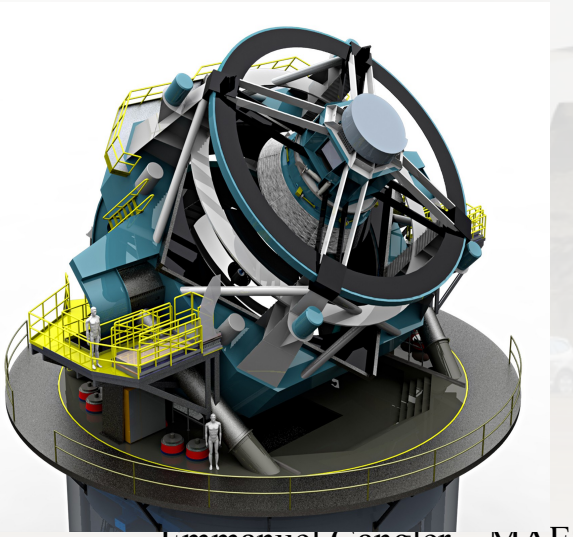
Big Data challenge posed by the

Emmanuel Gangler – UBP – Clermont-Ferrand (France)

capabilities :

- A stage-IV survey :
  - 8.4 m telescope
  - Cerro Pachon (Chili)
  - 3.2 Gpix 9.6□ FoV camera
  - 0.2 '' pixel
  - First light 2020





- All visible sky in 6 bands (ugrizy) (~20000□)
- 15 s exposure, 1 visit / 3 days
  r ~24 / visit
- During 10 years !
  - ~200 visits/band
- 30TB/day 100 PB/10 years

LSSTc institutions :

- **Institutional members:**

  Significantly contribute to LSST project and enabling science

  - 37 institutions, 5 being non-US:

    Republic of Chile

    
    FR (CNRS / IN2P3)

    
    UK (Oxford & Portsmouth Universities)

    
    CZ (Institute of Physics of the Academy)

- **International Contributors:**

  - Support for LSST operations (200 k$ / PI)
    - 25 institutions

      11 of them come from

      8 European countries

- All have **full data access rights**

European International contributors

Canary Islands
Instituto de Astrofisica de Canarias (IAC)

Croatia
Ruđer Bošković Institute (RBI)

France
IN2P3

Germany
Ludwig-Maximilians-Universität (LMU)
Max Planck Institute for Astrophysics (MPA)
Max Planck Institute for Astronomy (MPIA)

Hungary
Eotvos Lorand University (ELTE)
Konkoly Observatory

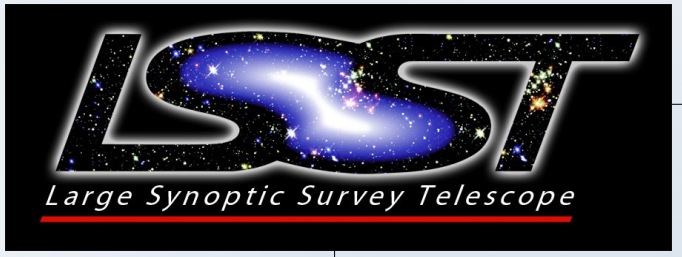
Serbia
Nano Center

Switzerland
Eidgenoessische Technische Hochschule Zuerich (Eth Zuerich)

United Kingdom
Science and Technology Facilities Council (STFC) - UK LSST Consortium

+ more coming (IT, ...)

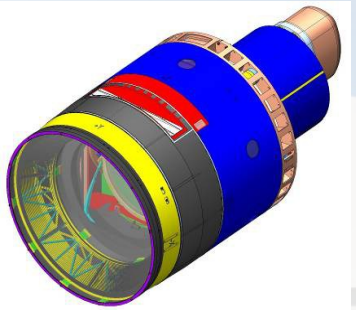
LSST project and Science:



L3

Telescope

Caméra



Data Management

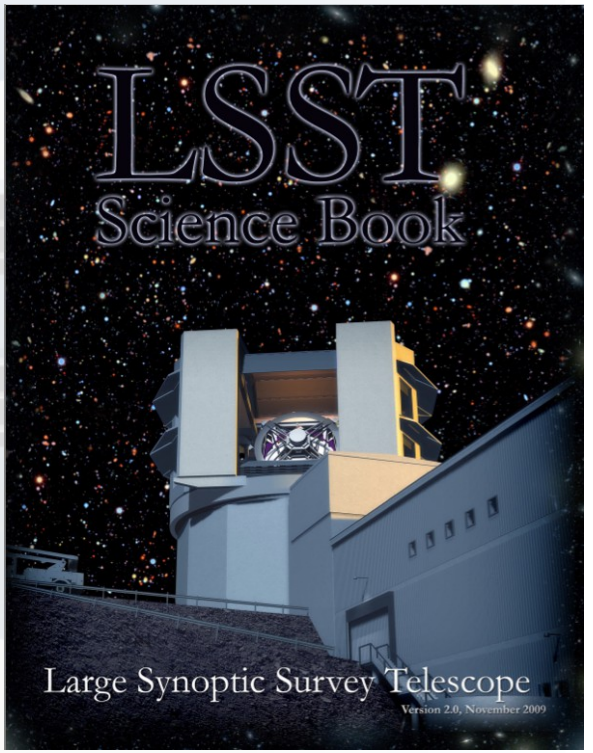
Outreach

L1 & L2

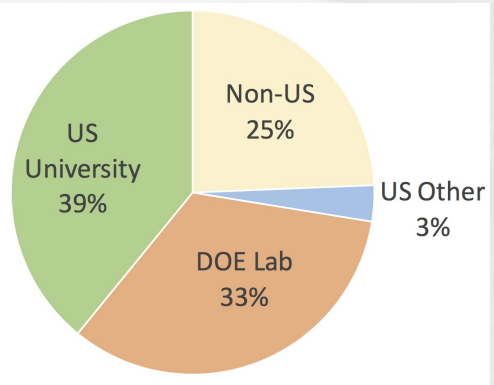
- **LSST covers 4 major scientific themes**
  - Dark Energy, Dark matter
  - Mapping Milky Way
  - Transient optical sky
  - Solar system
- **Scientific analysis is not part of the project**
  - Conducted by independent collaborations (**need data rights**)
  - With the help of LSSTc

LSST science

- **LSST is a world-wide project !**
  - Around *900 scientists* expected to have LSST data rights
  - ~450 from US
  - ~300 *from Europe* (9 countries)
- **9 science collaborations**
  - Galaxies (46 members)
    1 HR, 3 UK
  - Stars, Milky Way, and Local Volume (118 members)
    2 DE, 4 UK
  - Solar System (N/A)
  - Dark Energy (565 members)
    2 CZ, 1 ES, 65 FR, 70 UK
  - (Large scale structure/baryon oscillations)
  - (Strong lensing)
  - Active Galactic Nuclei (36 members)
    1 RS, 1 UK
  - Transient/Variable stars (>104 members)
    > 2 DE, 1 FR, 1 IL, 1 UK
  - ***Informatics and statistics*** (60 members)
    1UK



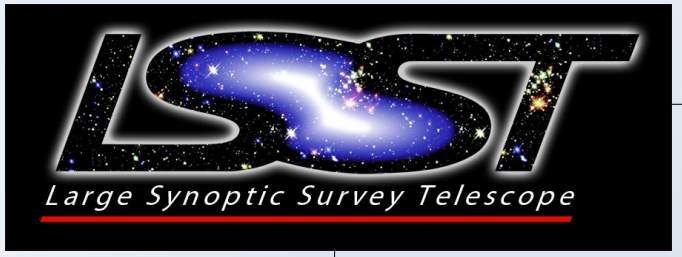
arXiv:09@12.0201



Numbers from 2016

+ ~45 FR joining ;
Those numbers are steadily growing !

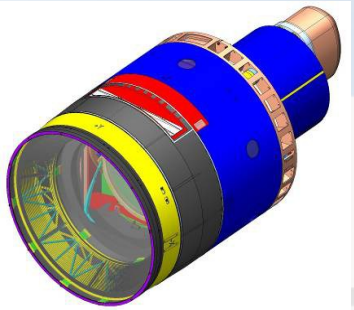
Challenges : building LSST



L3

Telescope

Caméra

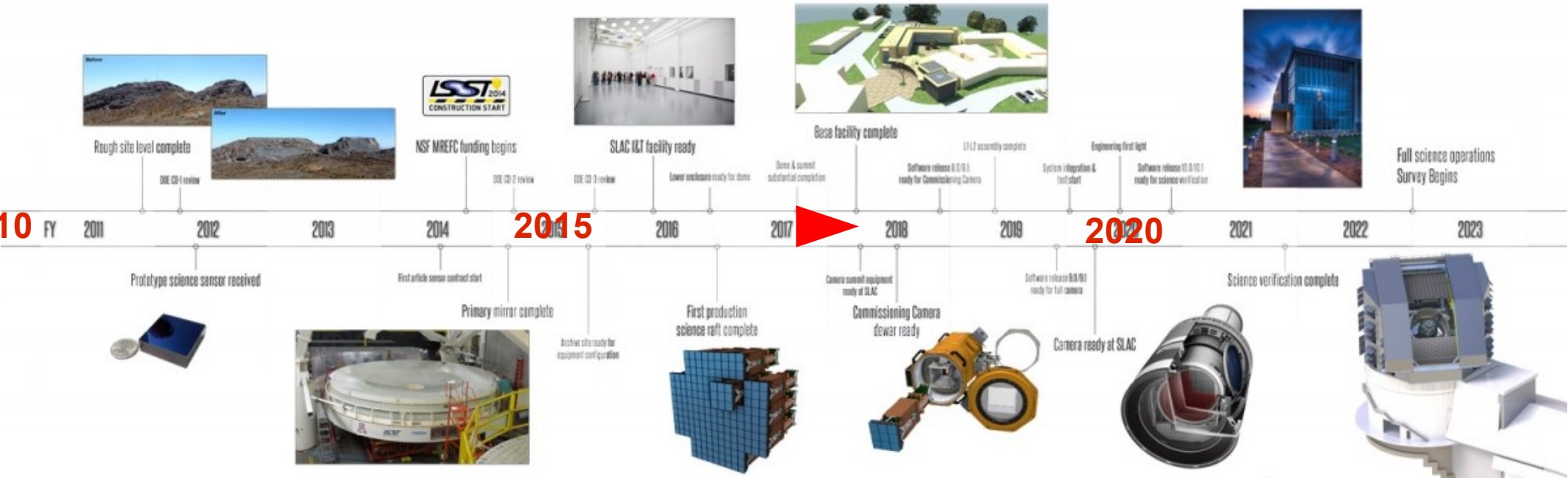


Data Management

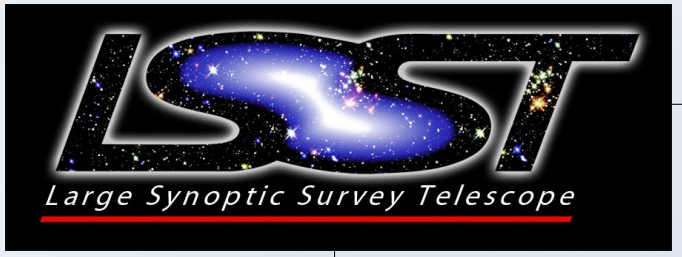
Outreach

L1 & L2

Challenges : building LSST



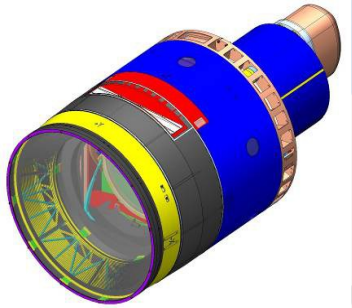
Challenges : building LSST



L3

Telescope

Caméra



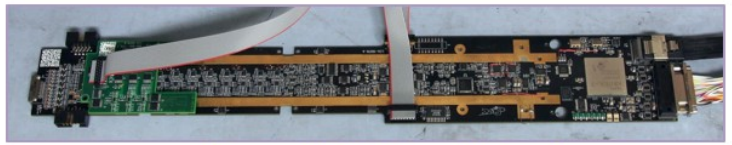
Data Management

Outreach

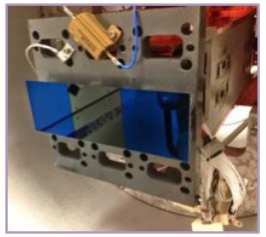
L1 & L2

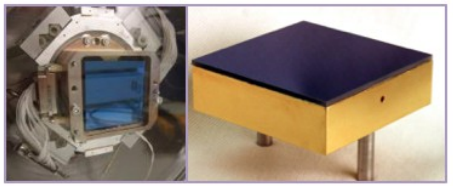
Most Camera Subsystems Have Been Prototyped

Raft Tower



Corner Raft Electronic Board

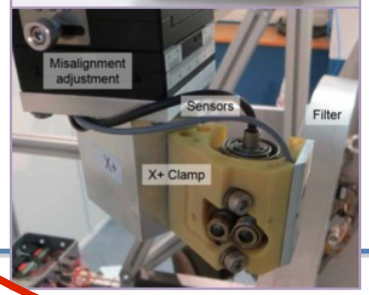




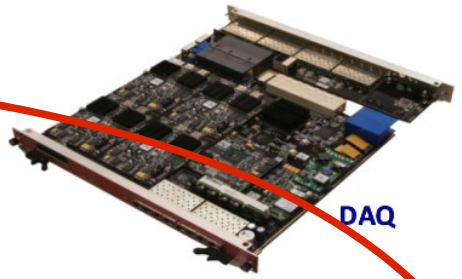
Prototype sensors

Refrigeration

Carousel Clamps



Preliminary test bench



DAQ

Autochanger

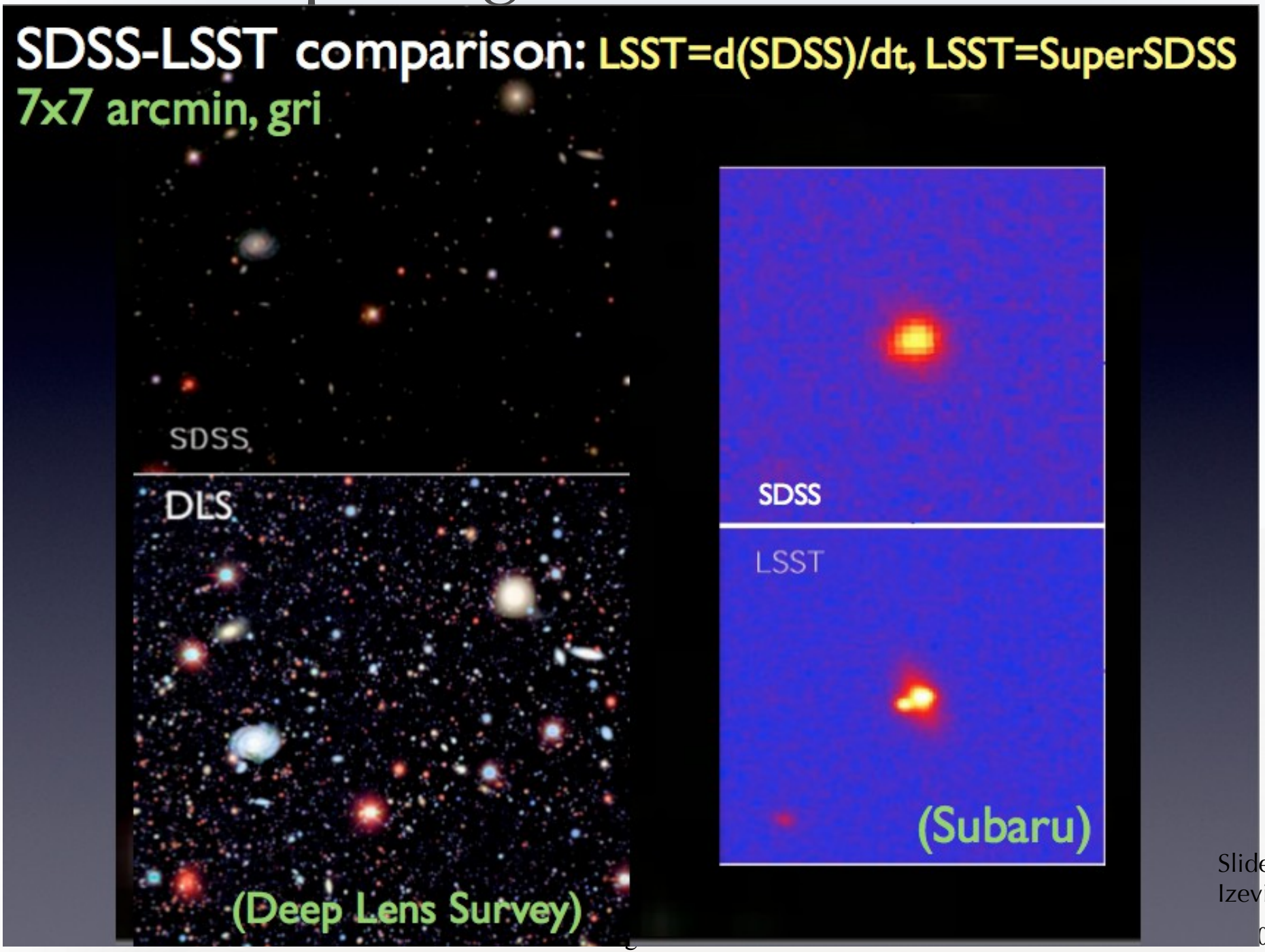
38

Sensor tests

ComCam

- **Sizeable hardware investment@ from FR**
  - Partner of LSST project since 2005

Comparing LSST data to SDSS



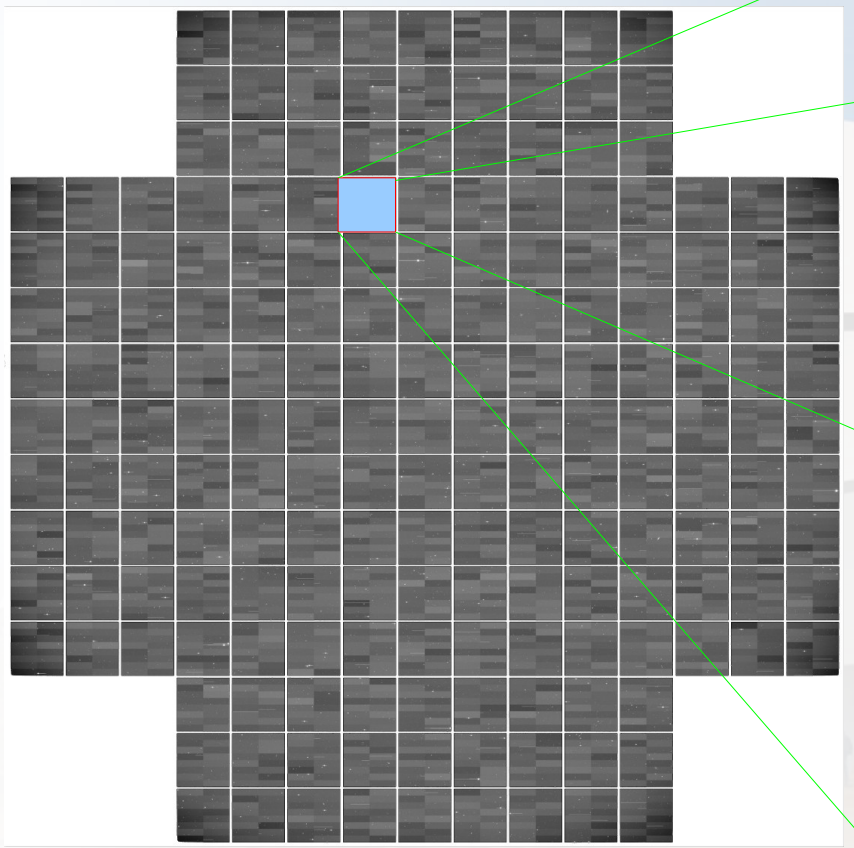
Slide from Ivezic

LSST data flow

Camera : 189 CCD (16 Mpix) read in parallel

- 3,2 G pixels !
- ~ 6 Gbyte / 17 seconds
- 15 TB / night

~ 1/1 000 000 000 of LSST data !



LSST data flow

Camera : 198 CCD (16 Mpix) read in parallel

- 3,2 G pixels !

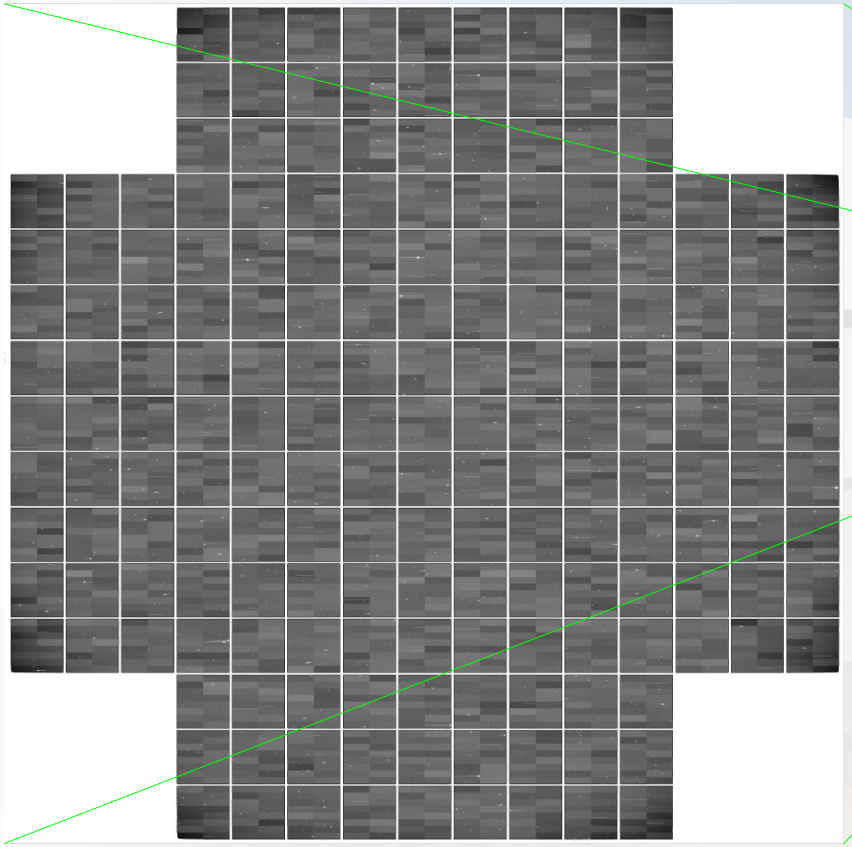
~ 6 Gbyte / 17 seconds

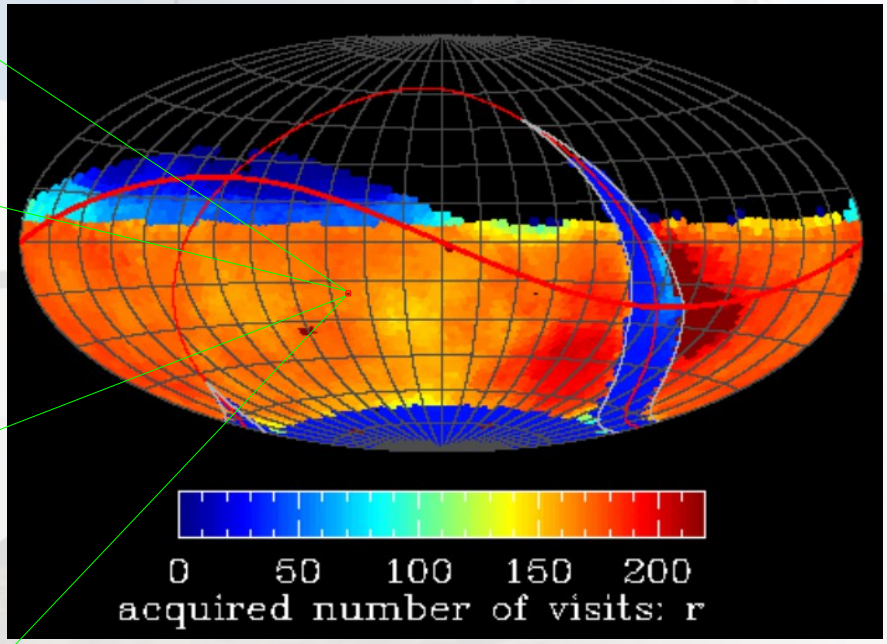
- 15 TB / night

During 10 years !

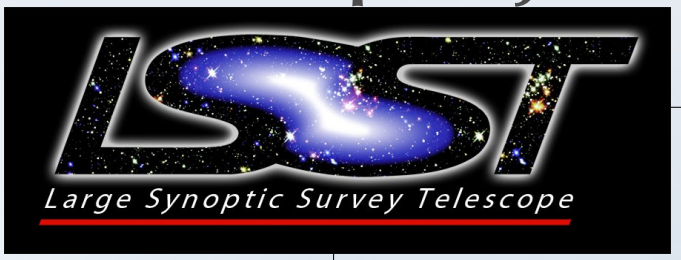
~ 1000 visits per field

- opens the time domain



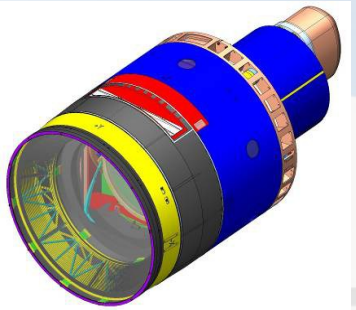


Data management is a pillar of the project :



Telescope

Caméra



Data Management

Outreach

« The data volumes […] of LSST are so large that **the limitation** on our ability to do science isn't the ability to **collect** the data, **it's the ability to understand** […] the data »

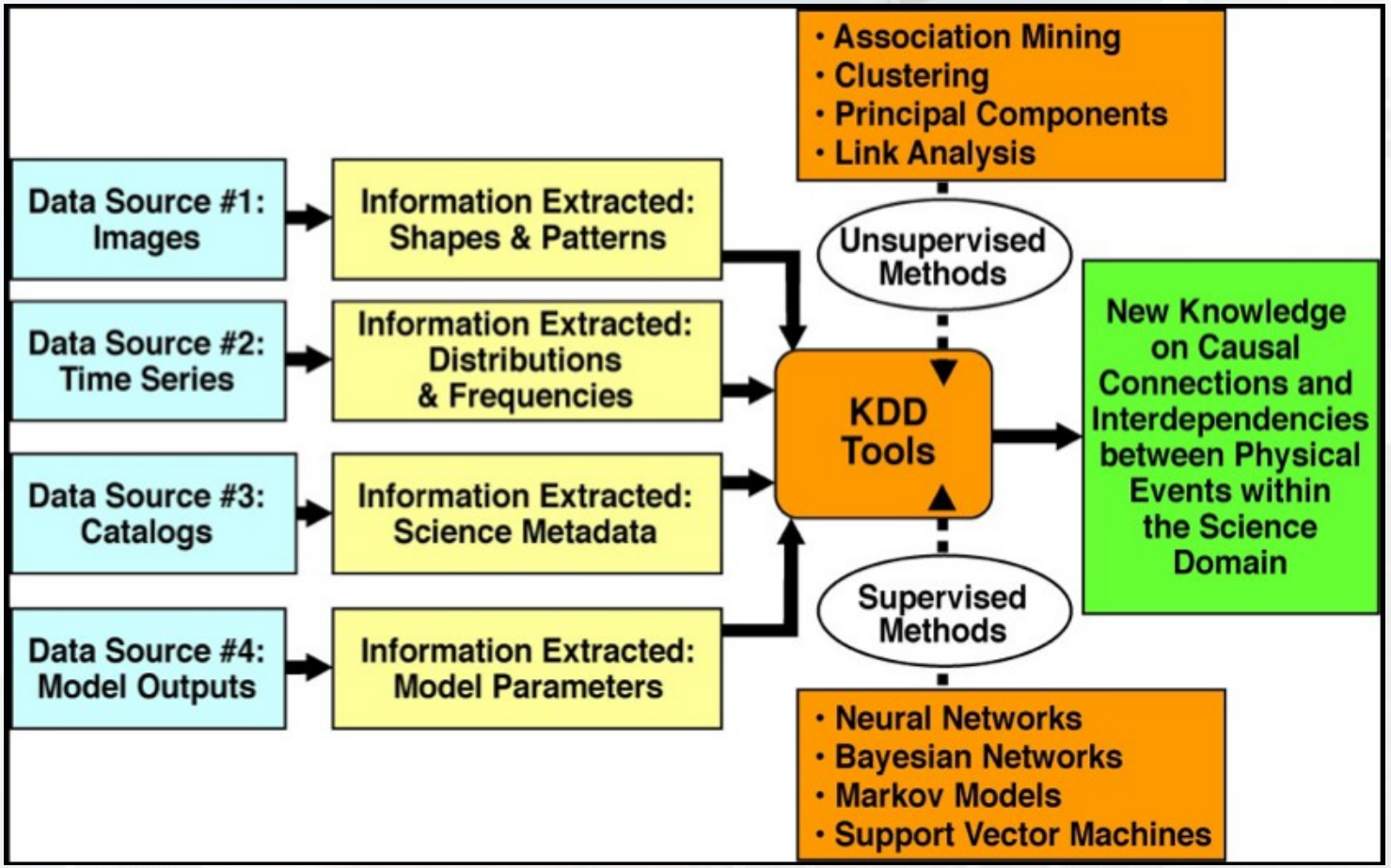
Andrew Conolly (U. Washington)

"How do you turn petabytes of data into scientific knowledge?"

Kirk Borne (George Mason U.)

How to handle astronomical big data ?

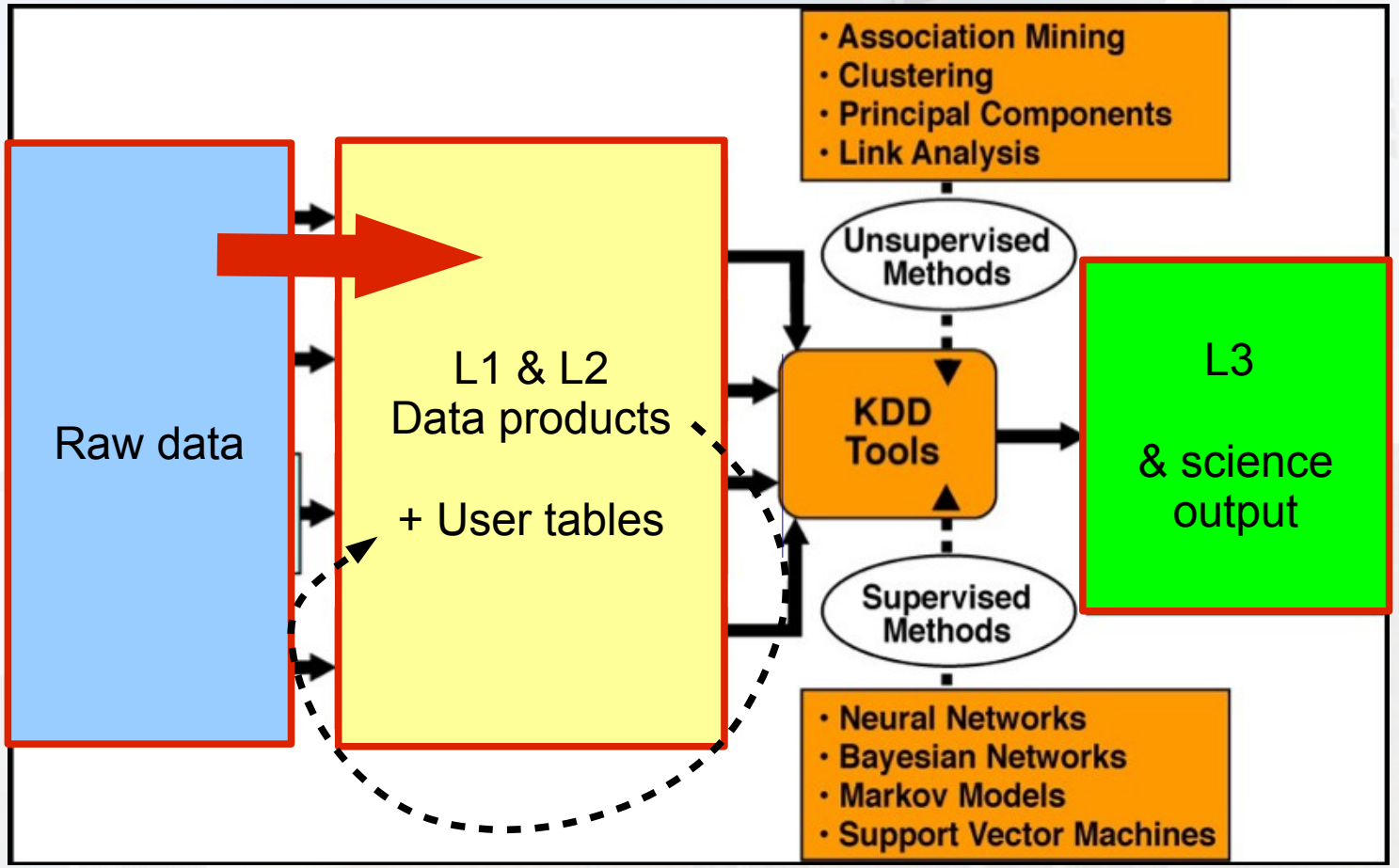
LSST is strongly inspired by Astroinformatics point of view :



Borne 2009

How to handle astronomical big data ?

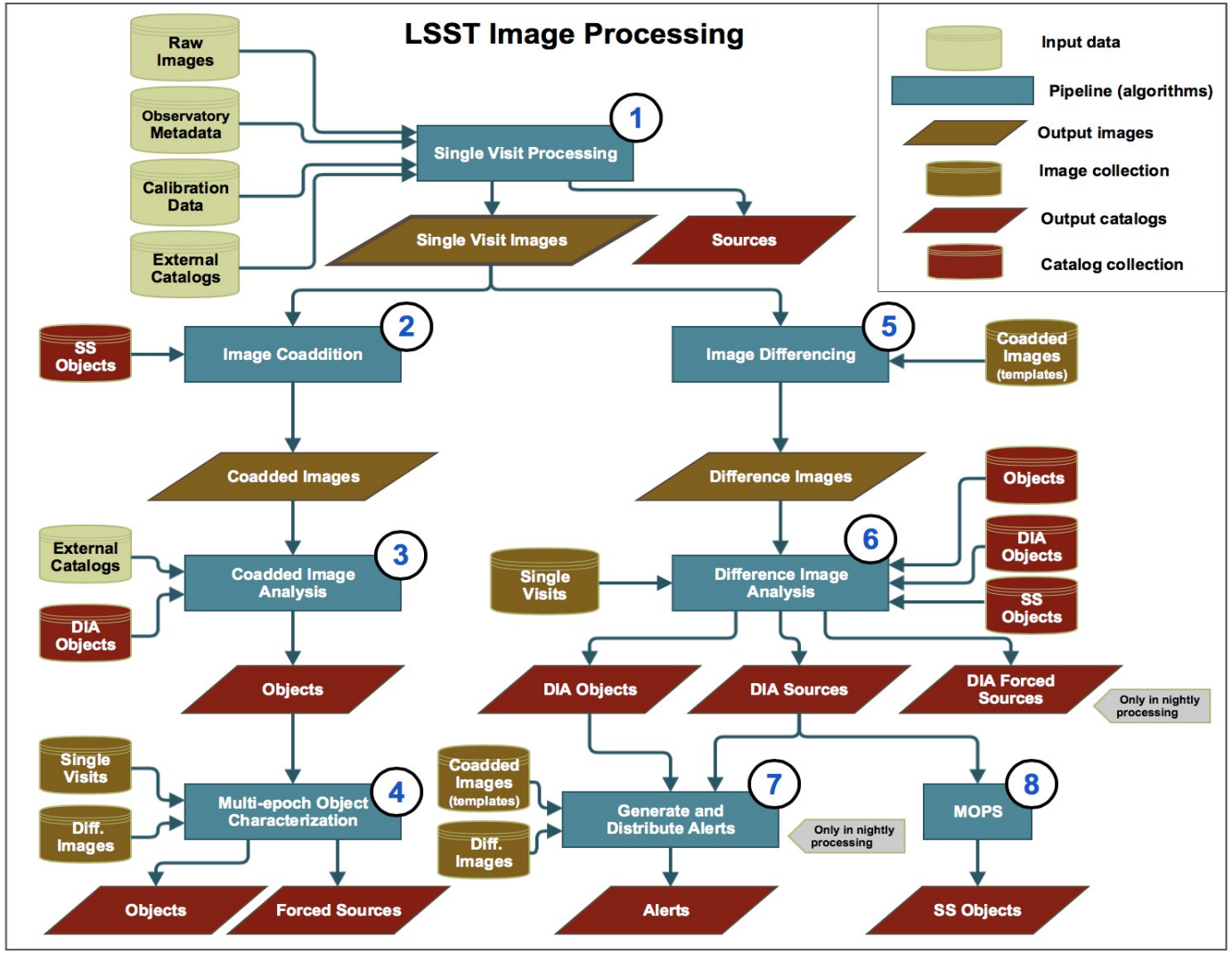
LSST is strongly inspired by Astroinformatics point of view :



Borne 2009

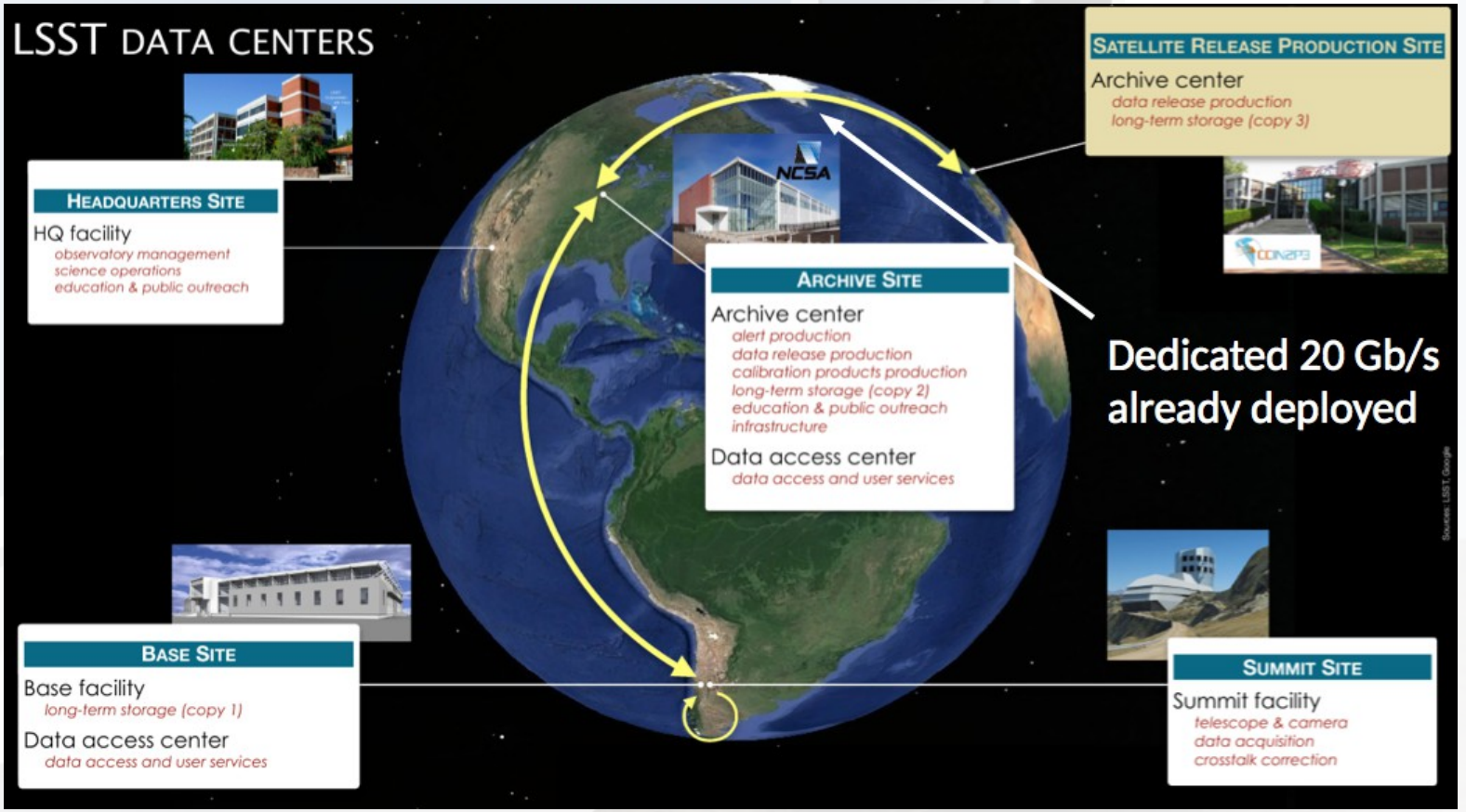
From Raw data to L1 & L2 products

- **Image Processing** pipeline by Data Management team
- Validation on precursor data (SDSS Stripe 82, CFHTLS, HSC)



From Raw data to L1 & L2 products

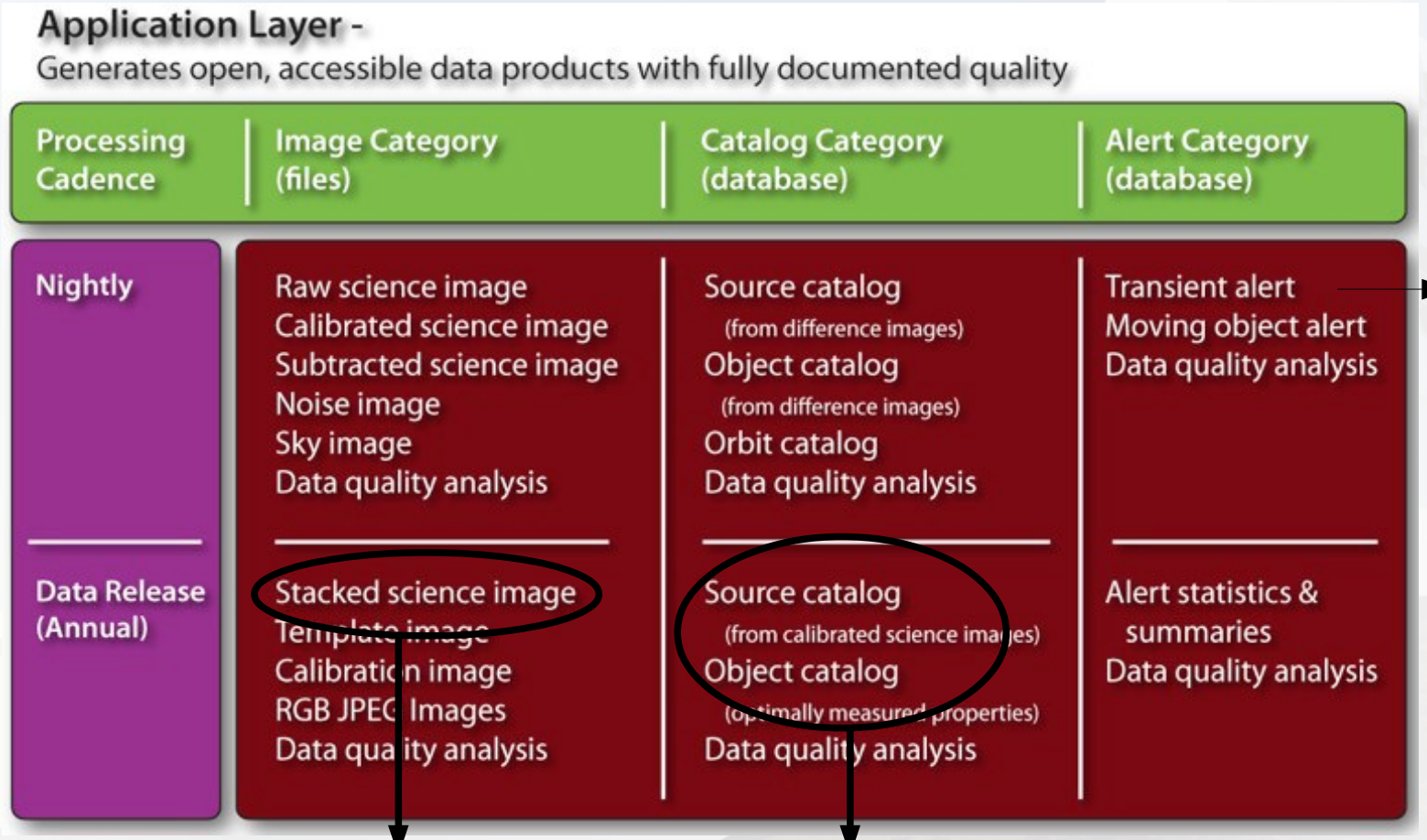
- **Embarrassingly parallel problem**
  - suited for HTC computing (I/O bound)



Per MoA, FR will:

- Process 50% of data up to L2 proucts
- Host an archive of LSST data
  - **There will be a full copy of LSST data in EU**

Available data:



Alerts : 2 M/day (within 60s.)

Static: 80 TB image
In 6 bands
(all visible sky)
Dynamic: 60 PB

Relation 1-Many
Object
Sources

Catalogs : 15 PB

Impact for database

Simulation 1
CCD 4k x 4k

- Huge data release by DR11:
  - 60 T rows
  - ~10 PB
  - 47. 10^9 objects (100 TB catalog)
  - 1 500. 10^9 objets extra (1.2 PB catalog)
  - 9 000. 10^9 detections "sources" (5 PB catalog)
  - 50 000. 10^9 measurements "forced sources" (2 PB catalog)
  - Nightly transient alerts: >2.10^6
- Processing paradigm : **characterize first, analyze later**
  - Data anlysis is NOT part of the project
- New subfield of astronomy: **Astroinformatics** (arXiv:0909.3892)

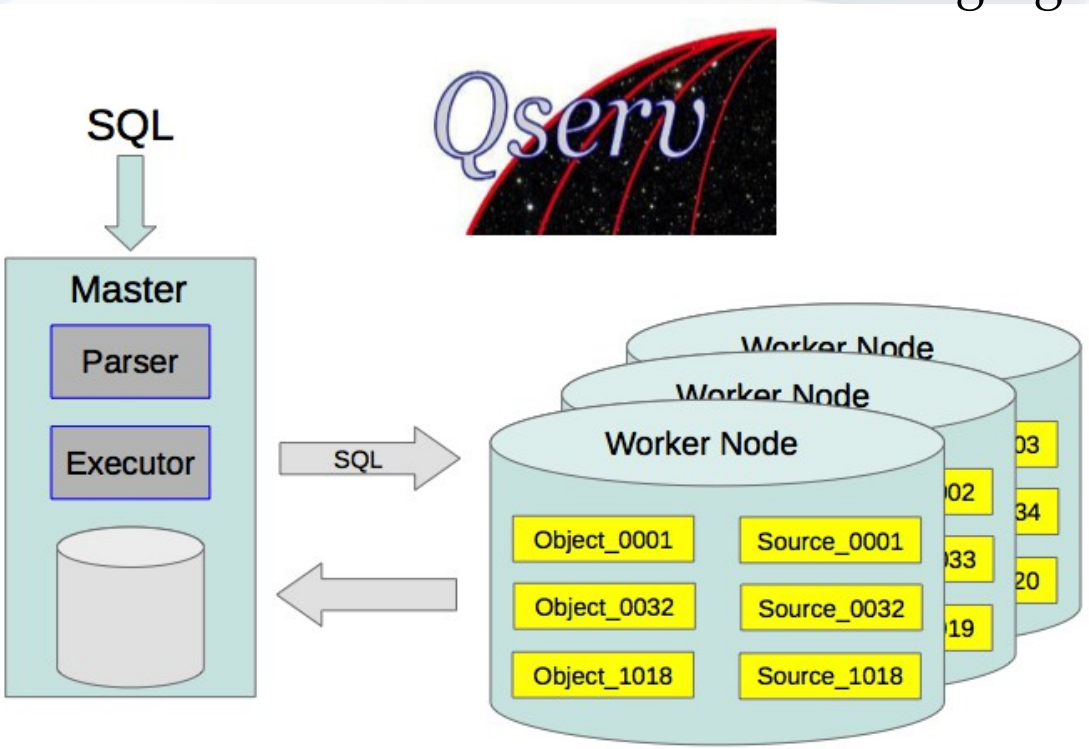
Qserv

Features :

- Shared-nothing MPP RDBMS
- *Spatial partitioning* with overlap
- *Shared scans*
- Replicated data
- Fixed-purpose dedicated hardware

Design

- SQL parser
  - Metadata DB
  - User defined function (geometry)
- Communication with xrootd
- MariaDB Backend
- Returns agregate results



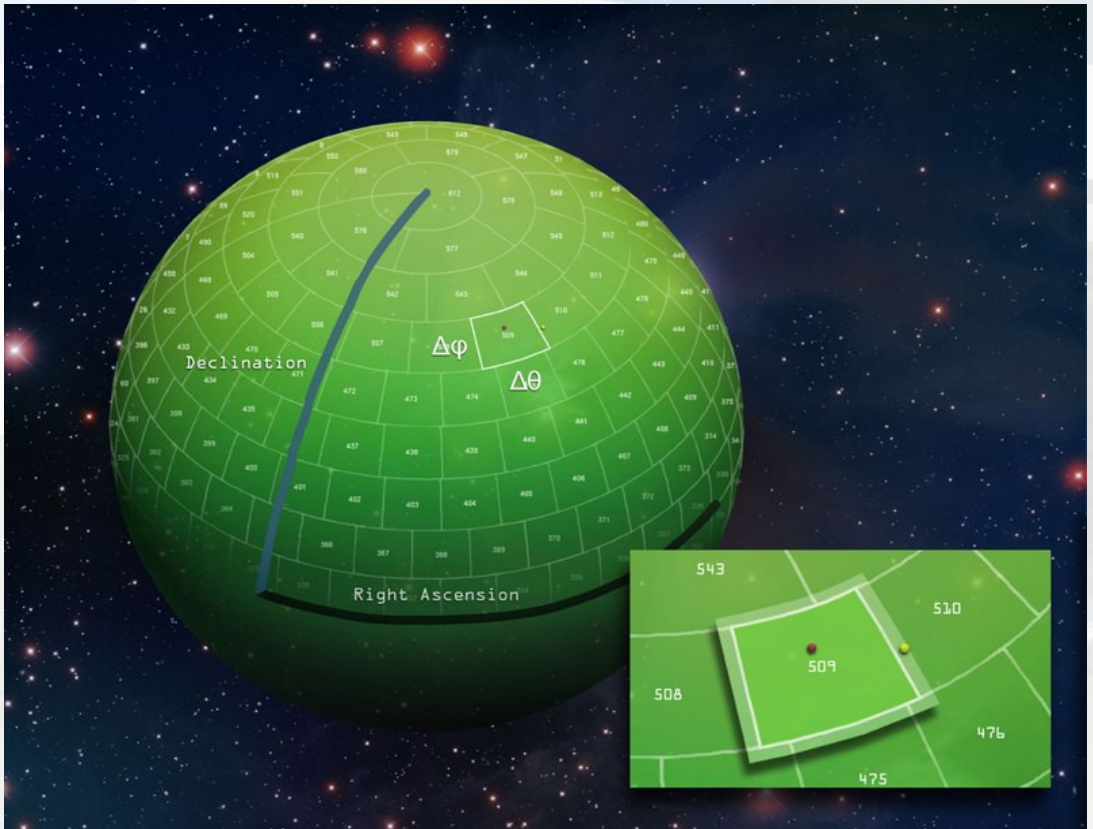
Qserv

Partitioning :

- Optimized for spatial joins on neighbors queries
- Spherical partitioning with overlap
- Two level
  - 2nd level materialized on the fly

Limitations

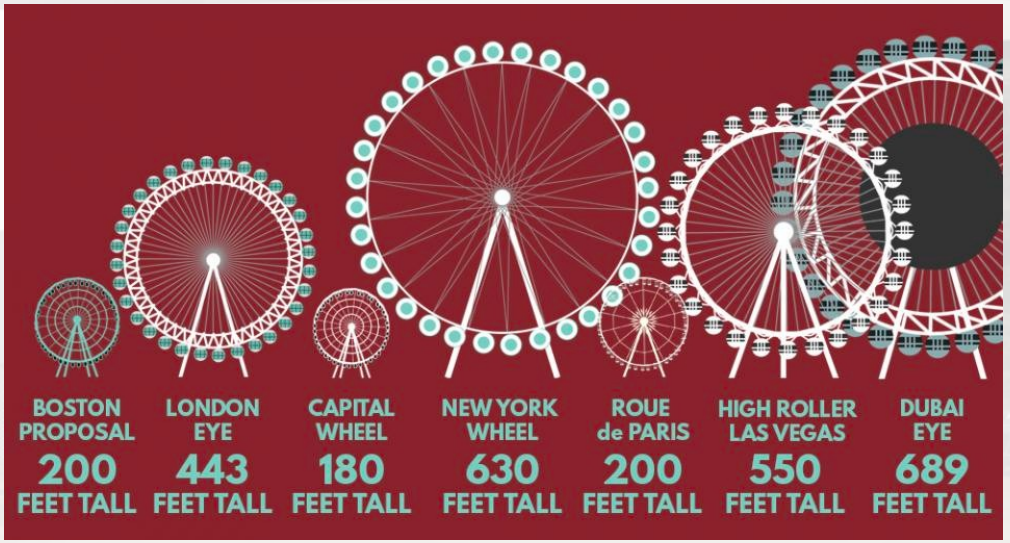
- Only subset of SQL
  - Spatial constraints
  - No subqueries
  - No function in ORDER BY clause
- Some queries can't be treated
  - (time, volume)



Qserv

Shared scan :

- Implementing Concurrency:
  - 100 simultaneous low volume queries (<0.5 GB @ 10/sec)
    - Eg : single object fetch, small spatial region
  - 50 simultaneous high volume queries (< 6GB @ 20/h)
    - Analytics and full scans
- *Continuous sequential scans*
- *Queries attached to appropriate running scan*
-



Qserv test platform

Target for production

~500 nodes clusters in 2 international data-centers

Running now

Development platform (CC-IN2P3)

400 cores, 800 GB memory
500 TB storage,
=> ~65 TB data set on 2*25 nodes

Prototype Data Access Center (NCSA)

500 cores, 4 TB memory
700 TB storage,
WISE data loaded

Thanks to a partnership with Dell we have deployed a Qserv test bench

The only test bench currently available in LSST for large scale tests

Lessons learned:

- **Qserv works well !**
  - Better than specs
- **Data deployment is not trivial**
  - Network configuration
  - Need 2-3x more disk space
- Large results stress master node
- Each scale increase raises new issues
  - **test platforms are vital**

Experimentations :

- Can cloud scale to big data ?
  - Openstack computation nodes + Ceph storage (cf. F. Gauet talk)
- Hive/HadoopDB (A. Mesmoudi et al.)
- MongoDB (C. Arnault)
  - Promising results on simple queries
  - Moving to Spark/Dataframe/GeoSpark

Optimization of aggregation queries

Aggregation in Massively Parallel Computing

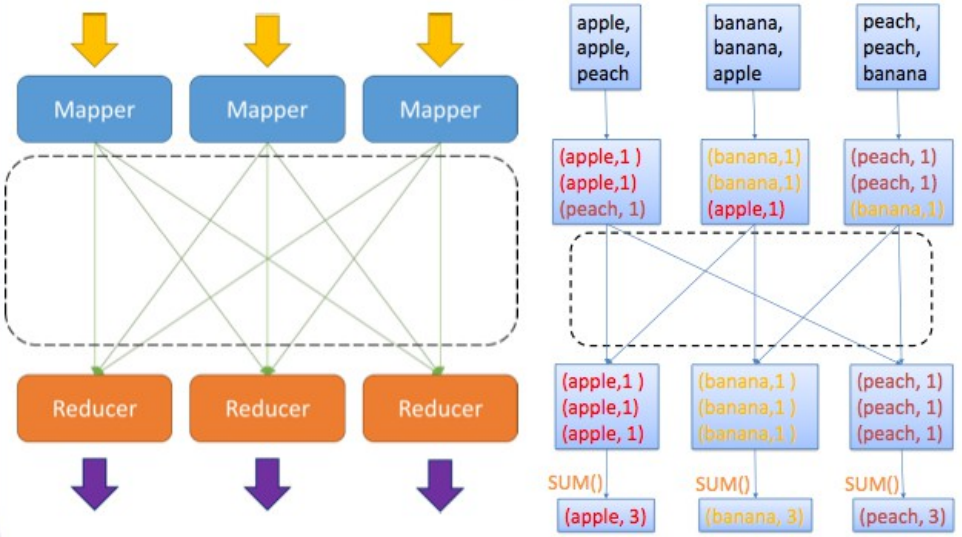
ZHANG Chao[12], Farouk Toumani[1], Emmanuel GANGLER[2]

[1] *Université Clermont Auvergne, LIMOS CNRS BP 10448, F-63000 Clermont-Ferrand*

[2] *Université Clermont Auvergne, LPC CNRS/IN2P3 BP 10448, F-63000 Clermont-Ferrand*

Motivation

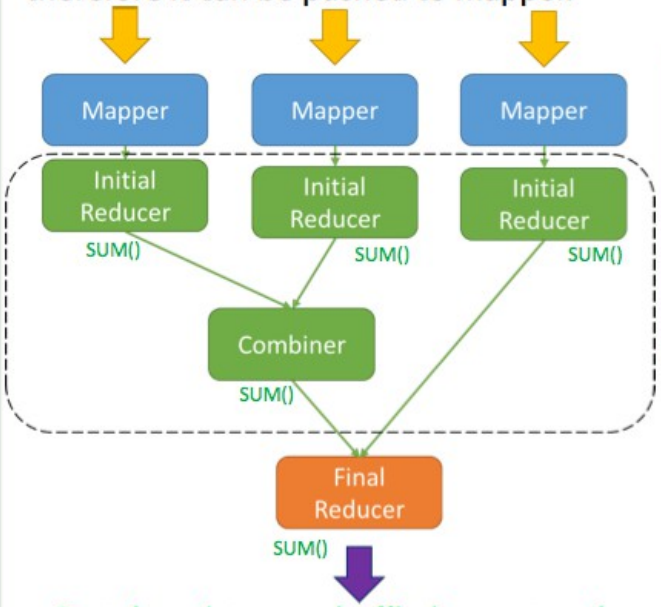
Word count example (naïve version)



Slow: all data are shuffled on network!!!

Word count example (optimized version)

SUM() is associative and commutative(AC), therefore it can be pushed to mapper.



AC is a strong condition, and not many aggregations are AC.

Question: given an arbitrary aggregation $\alpha()$, *how to process?*

Fast: less data are shuffled on network

KEY points

Main goal: generic algorithms to process arbitrary aggregation $\alpha()$ in parallel.

XLDB 2017 in Europe

Session and chairs:

Polystores

- **Patrick Valduriez, INRIA**: Senior researcher, head of Zenith research team.

Applications: earth and astronomy, neuroscience

- **Peter Baumann, Jacobs University**: Professor and head of the Large-Scale Scientific Information Systems research group.
- **Romulo Goncalves, Nederland eScience Center**: Expert in Databases, Data Structures, Distributed Computing.

Modern data management

- **Anastasia Ailamaki, EPFL**: Professor and Lab Director.
- **Mohand-Said Hacid, LIRIS**: Professor and Lab Director.

Scaling Cloud to Big Data

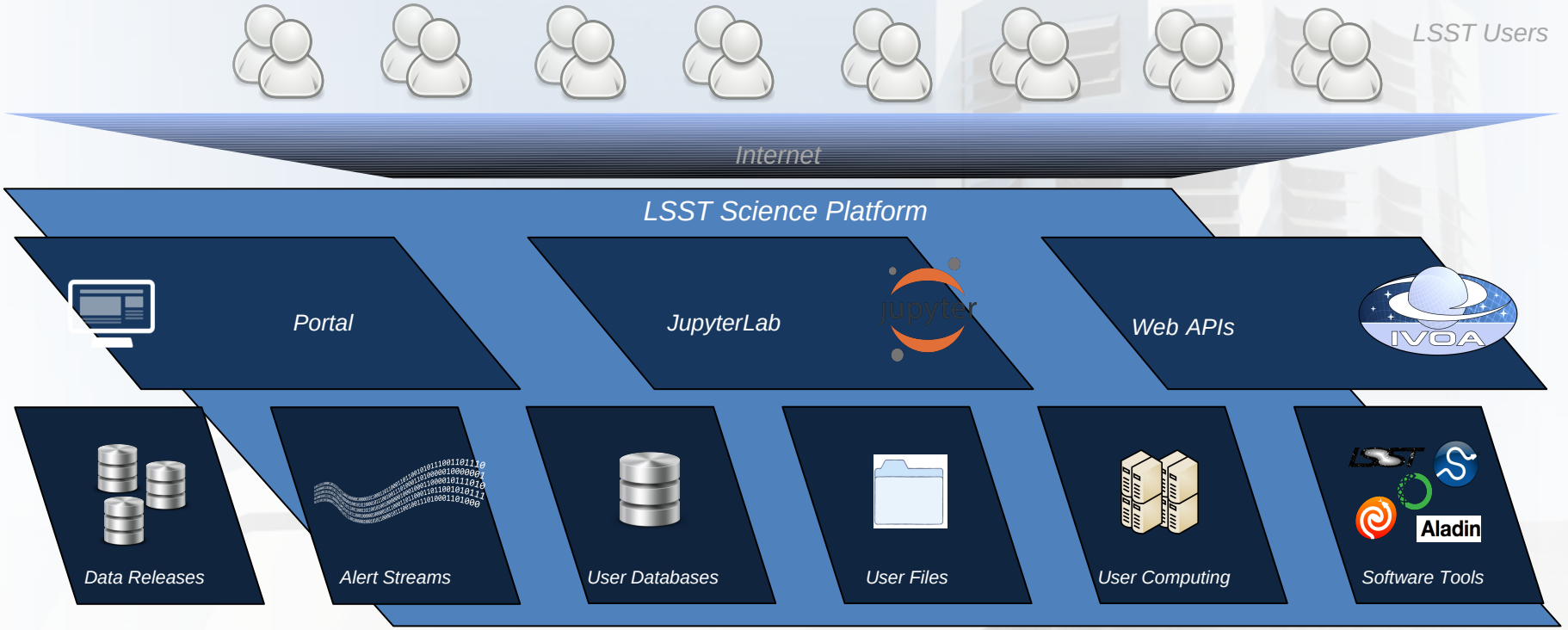
- **Dirk Duellmann, CERN**: Deputy leader of the data and storage services group in CERN's IT.
- **Yannick Legré, EGI**: Managing director.

1.5 days for main conference
~1 day for Hackaton

23

LSST science platform

Accessing LSST data and enabling LSST science



- Statistical analysis of a massive LSST dataset
  - **Portal:** browsing/vsualization of Pbyte-scale data
  - **JupyterLab:** user notebooks on DAC computing ressources
  - **Web APIs :** interface to VO, tools (eg TOPCAT…)

LSST Science cloud

Which knowledge to extract ?

- Some selected topics (Astroinfo...)
  - **Interpretation of spectral energy distribution (eg. photo-z)**
    - Machine learning vs. template fitting
    - Noise, posteriors pdf vs likelihood, optimal compression ...
  - **Spatial correlations**
    - Scalibility : can computations be done directly on DB ?
  - **Moving/Variable objects**
    - Irregular sampling, censored/sparse/missing data,
    - robustness, anomalies, characterization, classification
  - **New algorithms from LSST**
    - Automatic contamination/completeness tradeoff optimization
    - Robust detection of extremely rare events. Real-time
    - (un-)supervised clustering/classification of massive dataset (in db?)

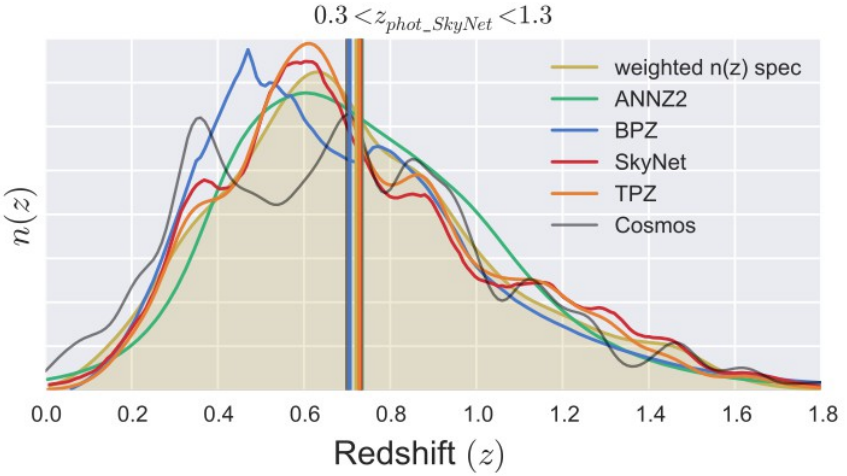
Example 1 : photo-z

Photo-z are critical for LSST

- **Which classifier ?**
  - Performance issue
  - Systematic uncertainty

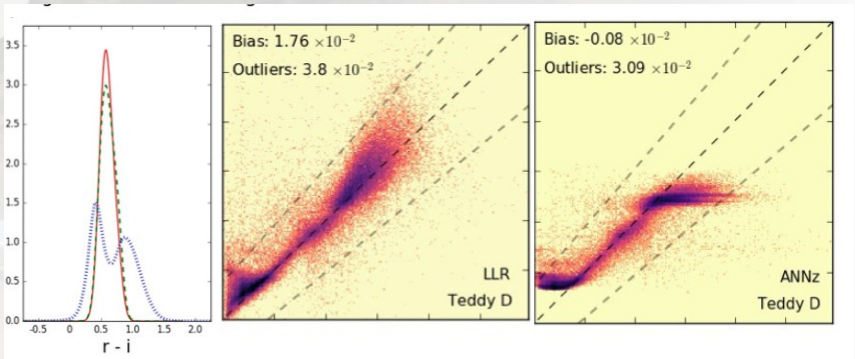
- **Training set**
  - **Small** spectroscopic sample
    - 50M spectra, 20 B galaxies
  - Different **coverage** in parameter space

Uncertainties in N(z) highly uncertain



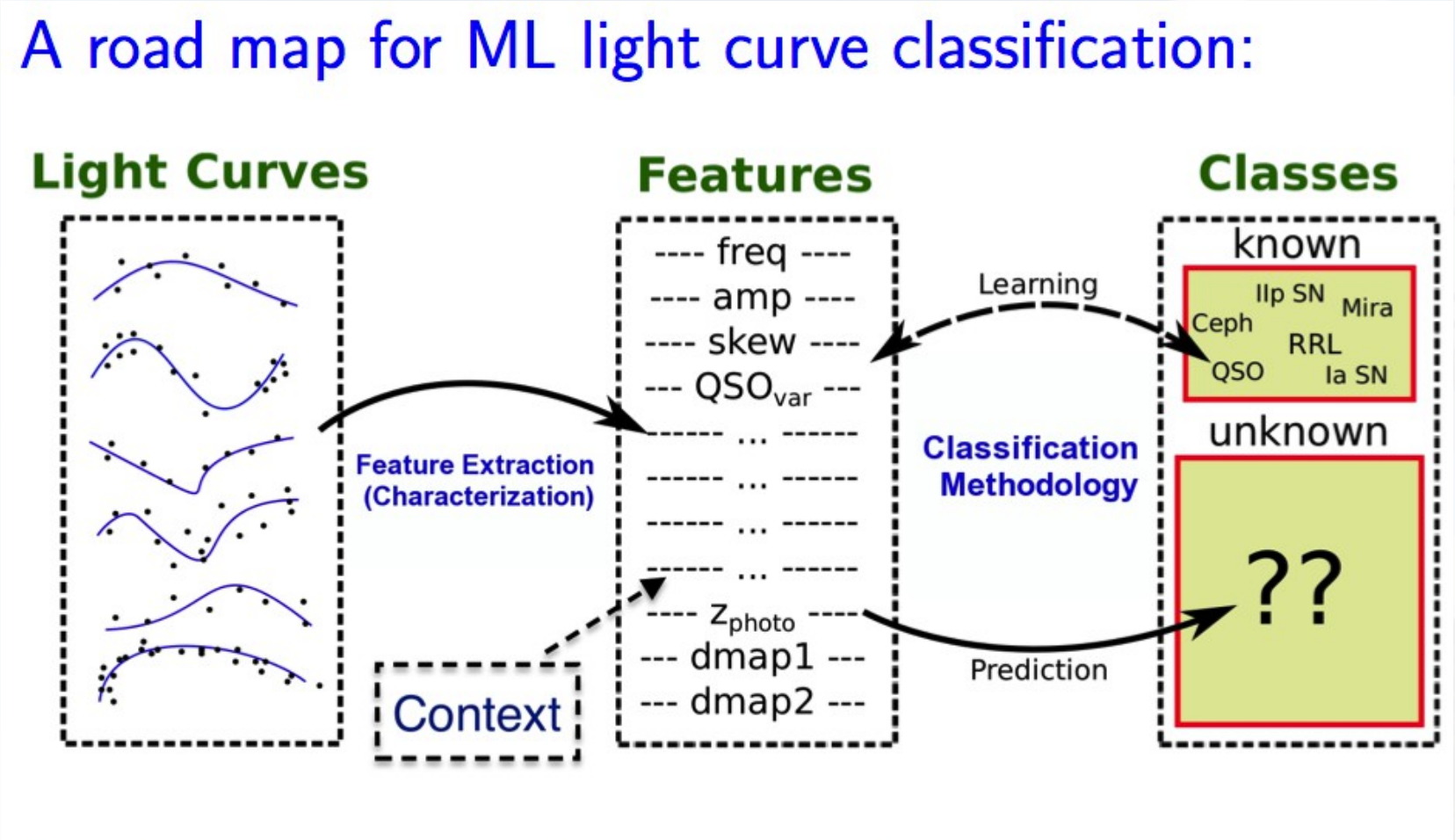
Redshift distributions for DES SV galaxies (Bonnett+ 2015)

Photo-z



Lin, C. A. *et at. - in prep – for the COIN collaboration*

Example 2 : time domain classification



Announcement : Plasticc

Photometric LSST Astronomical Time-Series Classification Challenge

Conclusions

- **LSST will provide unprecedented data**
  - Opens up time domain
  - … and a LOT of scientific opportunities
- **Proper knowledge on how to use these data needed**
  - Training of students
  - IT cross-disciplinary field : astroinformatics
    - ***New ideas are welcomed !***
- **Data access has to be organized**
  - A European DAC center is under study
    - Goal is to serve a broad community