
Source: LSST

# The Large Synoptic Survey Telescope (LSST): big data for astrophysics research

fabio hernandez

CCIN2P3

*Visit of Chinese Academy of Sciences, Lyon, Sep 8th, 2017*

cnrs

# Contents

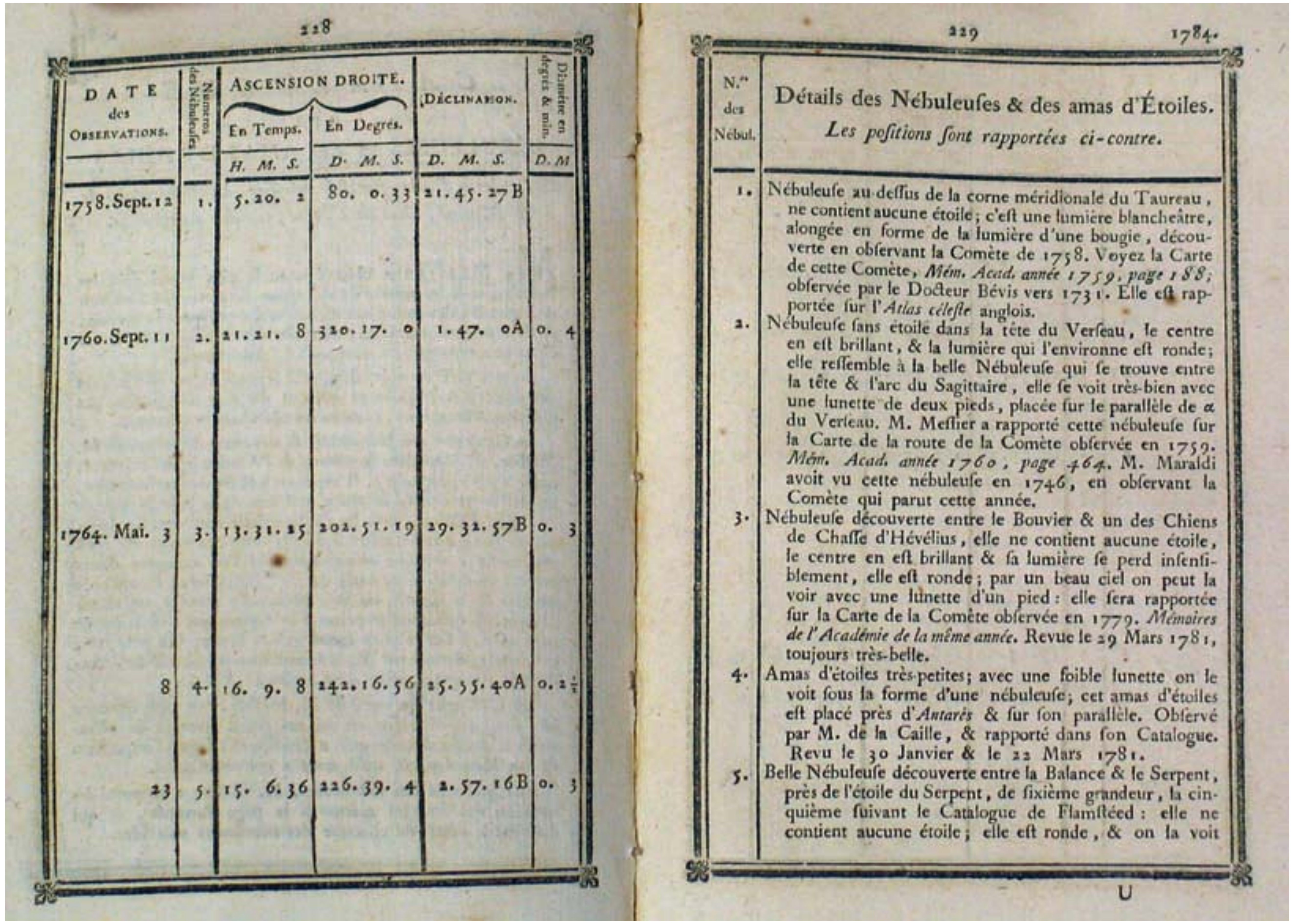

- Overview of the LSST project
- LSST data processing
- Summary

# Astronomical catalogs

# Astronomical catalog


Source: Wikipedia

Charles Messier
1730-1817



228

| DATE des Observations. | Numéros des Nébuleuses. | Ascension Droite. En Temps. H. M. S. | En Degrés. D. M. S. | Déclinaison. D. M. S. | Diamètre en degrés & min. D. M |
|---|---|---|---|---|---|
| 1758. Sept. 12 | 1. | 5. 20. 2 | 80. 0. 33 | 21. 45. 27 B | |
| 1760. Sept. 11 | 2. | 21. 21. 8 | 320. 17. 0 | 1. 47. 0 A | 0. 4 |
| 1764. Mai. 3 | 3. | 13. 31. 25 | 202. 51. 19 | 29. 32. 57 B | 0. 3 |
| 8 | 4. | 16. 9. 8 | 242. 16. 56 | 25. 55. 40 A | 0. 2½ |
| 23 | 5. | 15. 6. 36 | 226. 39. 4 | 2. 57. 16 B | 0. 3 |

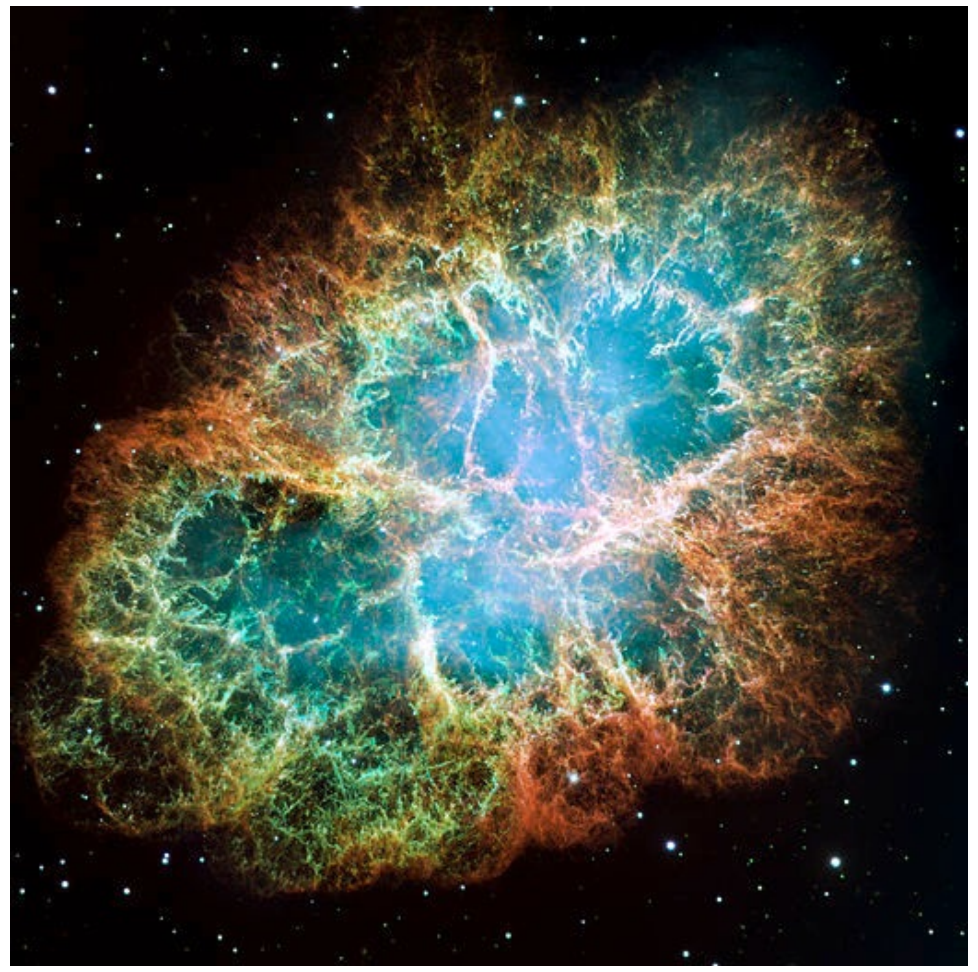
229 1784.

| N.os des Nébul. | Détails des Nébuleuſes & des amas d'Étoiles. *Les poſitions ſont rapportées ci-contre.* |
|---|---|
| 1. | Nébuleuſe au-deſſus de la corne méridionale du Taureau, ne contient aucune étoile; c'eſt une lumière blancheâtre, alongée en forme de la lumière d'une bougie, découverte en obſervant la Comète de 1758. Voyez la Carte de cette Comète, *Mém. Acad. année 1759, page 188;* obſervée par le Docteur Bévis vers 1731. Elle eſt rapportée ſur l'*Atlas céleſte* anglois. |
| 2. | Nébuleuſe ſans étoile dans la tête du Verſeau, le centre en eſt brillant, & la lumière qui l'environne eſt ronde; elle reſſemble à la belle Nébuleuſe qui ſe trouve entre la tête & l'arc du Sagittaire, elle ſe voit très-bien avec une lunette de deux pieds, placée ſur le parallèle de α du Verſeau. M. Meſſier a rapporté cette nébuleuſe ſur la Carte de la route de la Comète obſervée en 1759. *Mém. Acad. année 1760, page 464.* M. Maraldi avoit vu cette nébuleuſe en 1746, en obſervant la Comète qui parut cette année. |
| 3. | Nébuleuſe découverte entre le Bouvier & un des Chiens de Chaſſe d'Hévélius, elle ne contient aucune étoile, le centre en eſt brillant & ſa lumière ſe perd inſenſiblement, elle eſt ronde; par un beau ciel on peut la voir avec une lunette d'un pied: elle ſera rapportée ſur la Carte de la Comète obſervée en 1779. *Mémoires de l'Académie de la même année.* Revue le 29 Mars 1781, toujours très-belle. |
| 4. | Amas d'étoiles très-petites; avec une foible lunette on le voit ſous la forme d'une nébuleuſe; cet amas d'étoiles eſt placé près d'*Antarès* & ſur ſon parallèle. Obſervé par M. de la Caille, & rapporté dans ſon Catalogue. Revu le 30 Janvier & le 22 Mars 1781. |
| 5. | Belle Nébuleuſe découverte entre la Balance & le Serpent, près de l'étoile du Serpent, de ſixième grandeur, la cinquième ſuivant le Catalogue de Flamſtéed: elle ne contient aucune étoile; elle eſt ronde, & on la voit |

U

Source: Wikipedia

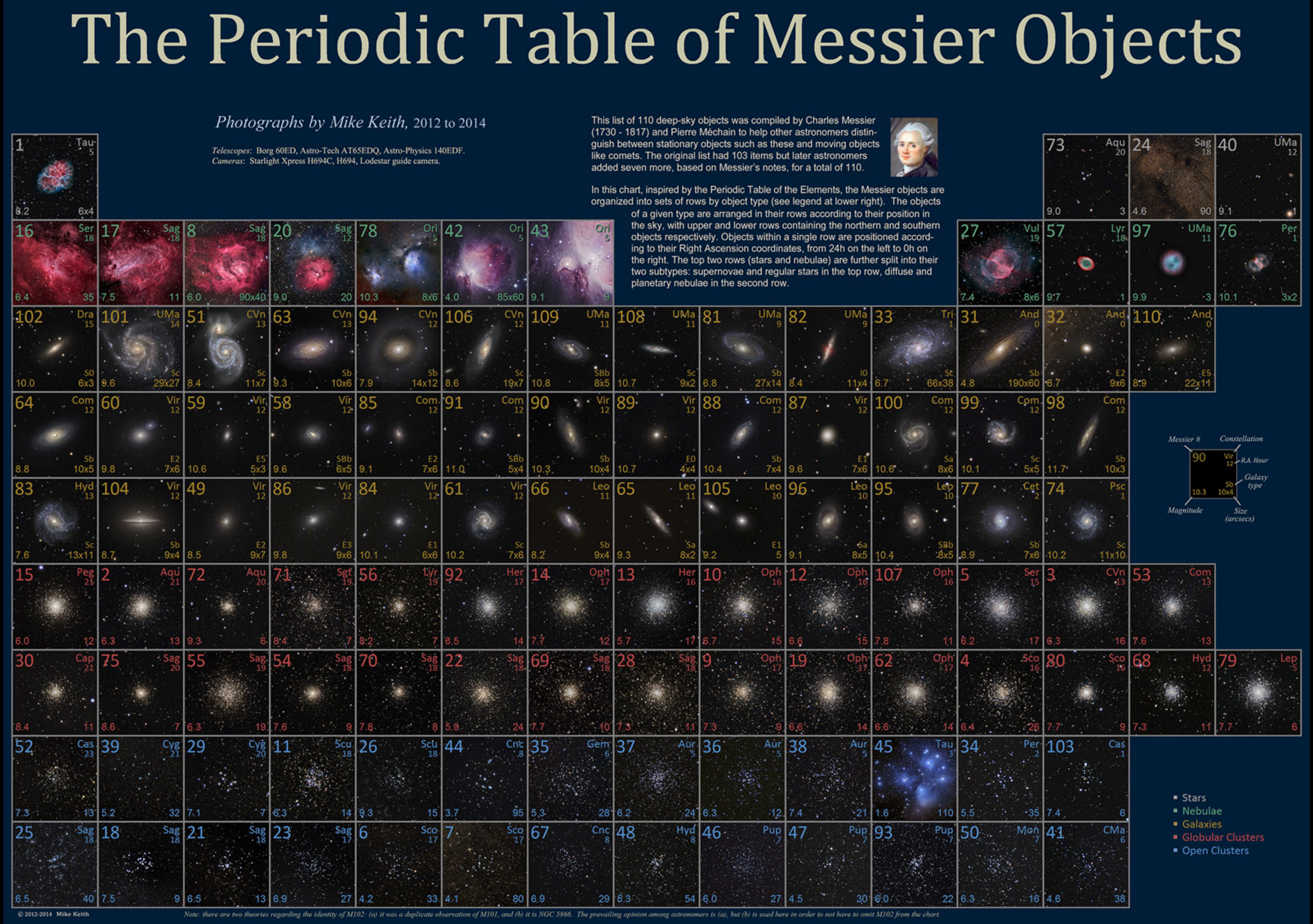
First 5 entries (out of 110) of the Messier catalog (1781)


Source: Wikipedia

M1 Crab Nebula


Source: Wikipedia

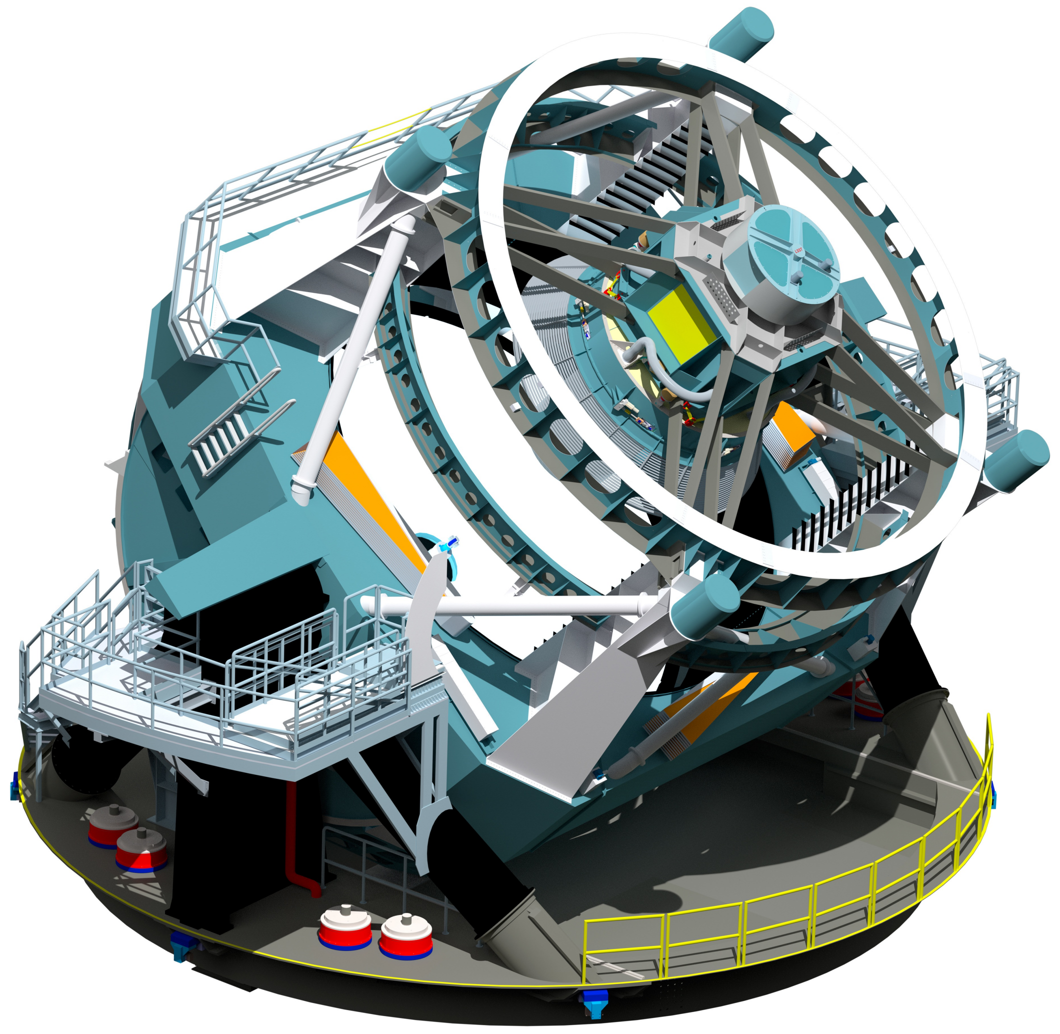
M31 Andromeda galaxy

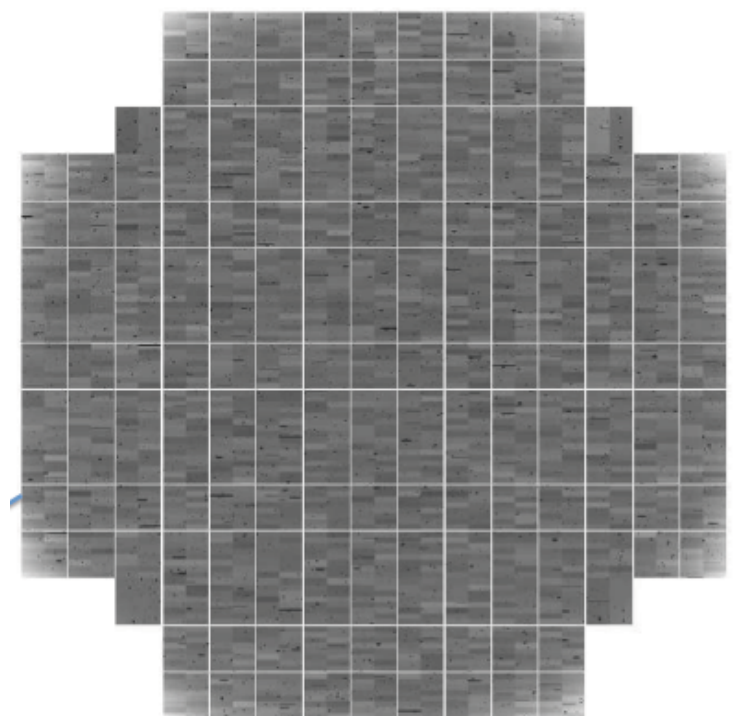



Source: Mike Keith

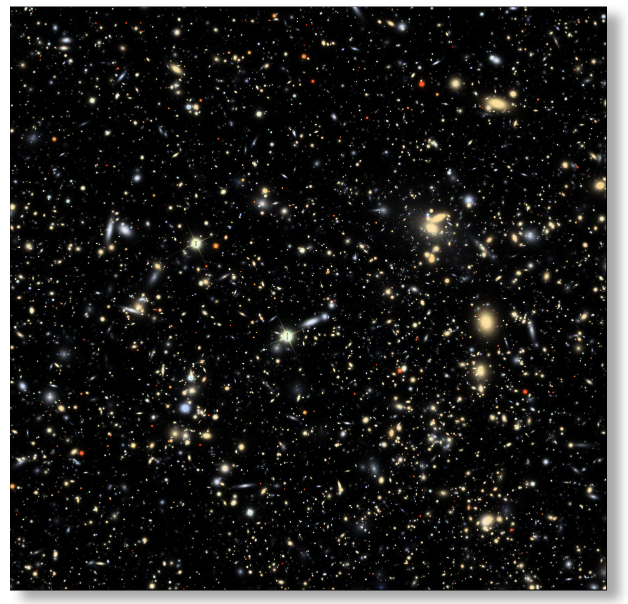
# Large Synoptic Survey Telescope

# LSST project overview

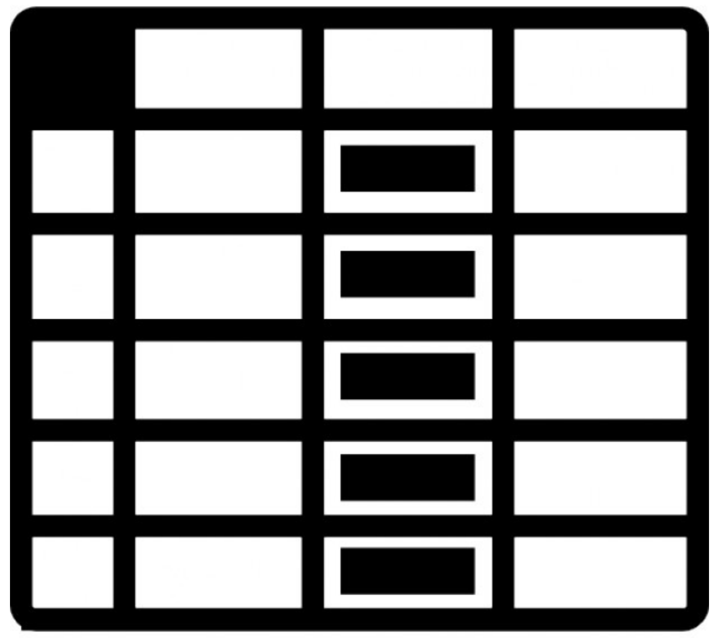




raw data



processed image



catalog
(stars, galaxies, objects, sources, transients, etc.)

LSST will build a catalog of 20 billion galaxies and 17 billion stars and their associated physical properties

# LSST project overview (cont.)

- Large Synoptic Survey Telescope

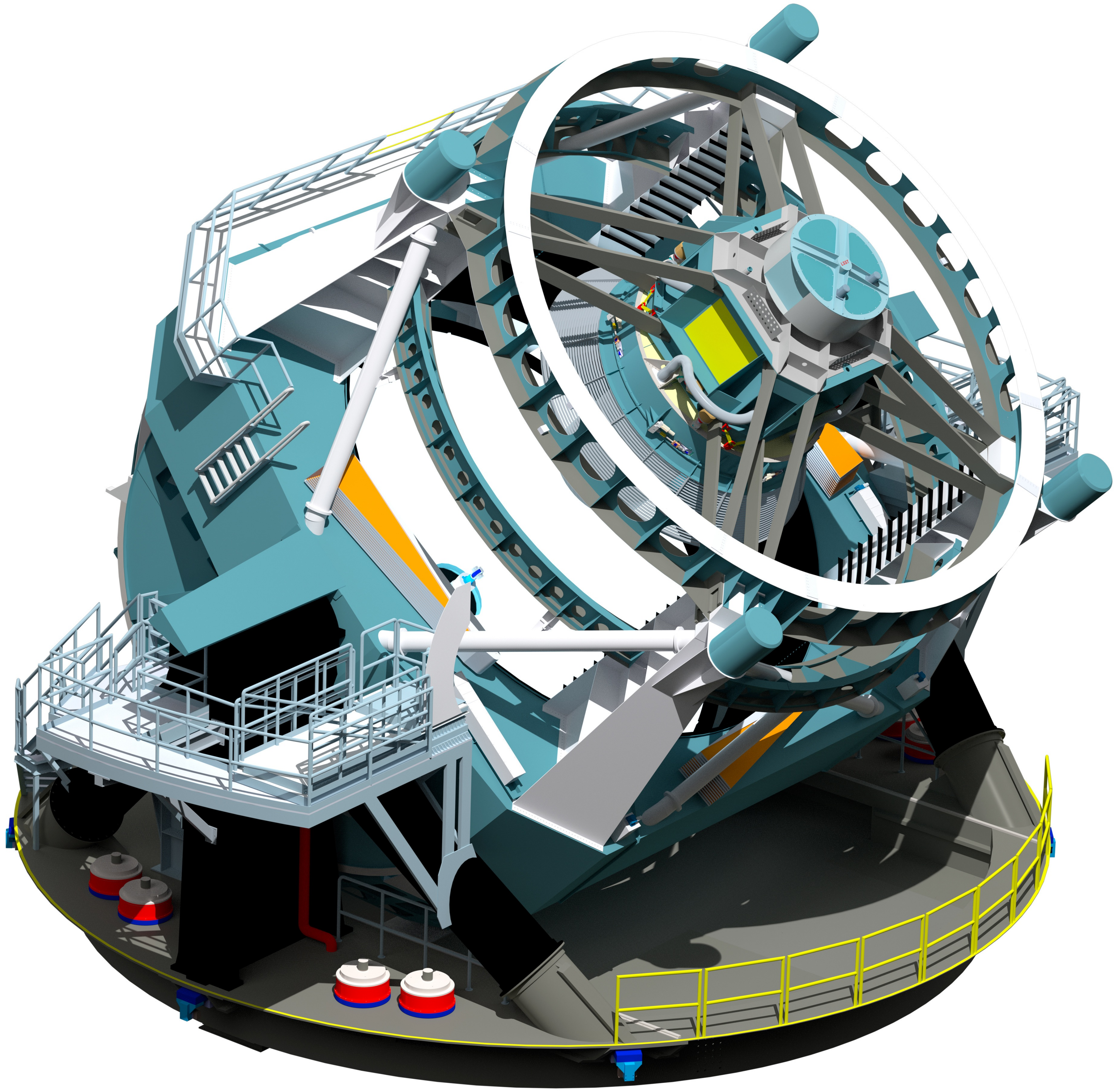
  *large aperture, wide-field, ground-based **survey telescope***

- Characteristics

  ***large aperture** to image faint astronomical objects across the sky with short exposures*

  ***wide field of view** to observe large areas of the sky at once*

  ***compact** so it can move quickly between images*



# LSST project overview (cont.)

- Science themes

  *determining the nature of **dark energy** and **dark matter***

  *taking an inventory of the **solar system***

  *exploring the **transient** optical sky*

  *mapping the structure and evolution of the **Milky Way***

- Principle of operations

  *90% of the observing time of the telescope devoted to a **deep-wide-fast survey***

  ***one complete visit of the southern hemisphere sky every 3-4 nights**, from 2022 for **10 years***

  *each patch of the sky to be visited about 1000 times*

  *43% of the celestial sphere will be covered by this survey*

# LSST project overview (cont.)

- Deliverable

***the science-enabling, ultimate deliverable of the project will be the fully reduced data***

*the scientific exploitation of the collected data will be performed by the scientific community*

- Schedule

*construction: 2014 - 2022*

*operations: 2022 - 2032*

*first light: 2020*

- Open data and open source software

*complete cumulative data set (images and catalogs), open to the scientific community of the participating countries, once per year, with no proprietary period*

*alerts of detected variable sources (transients) made available for world-wide distribution within 60 seconds of observation, published via standard protocols*

*software:* https://github.com/lsst

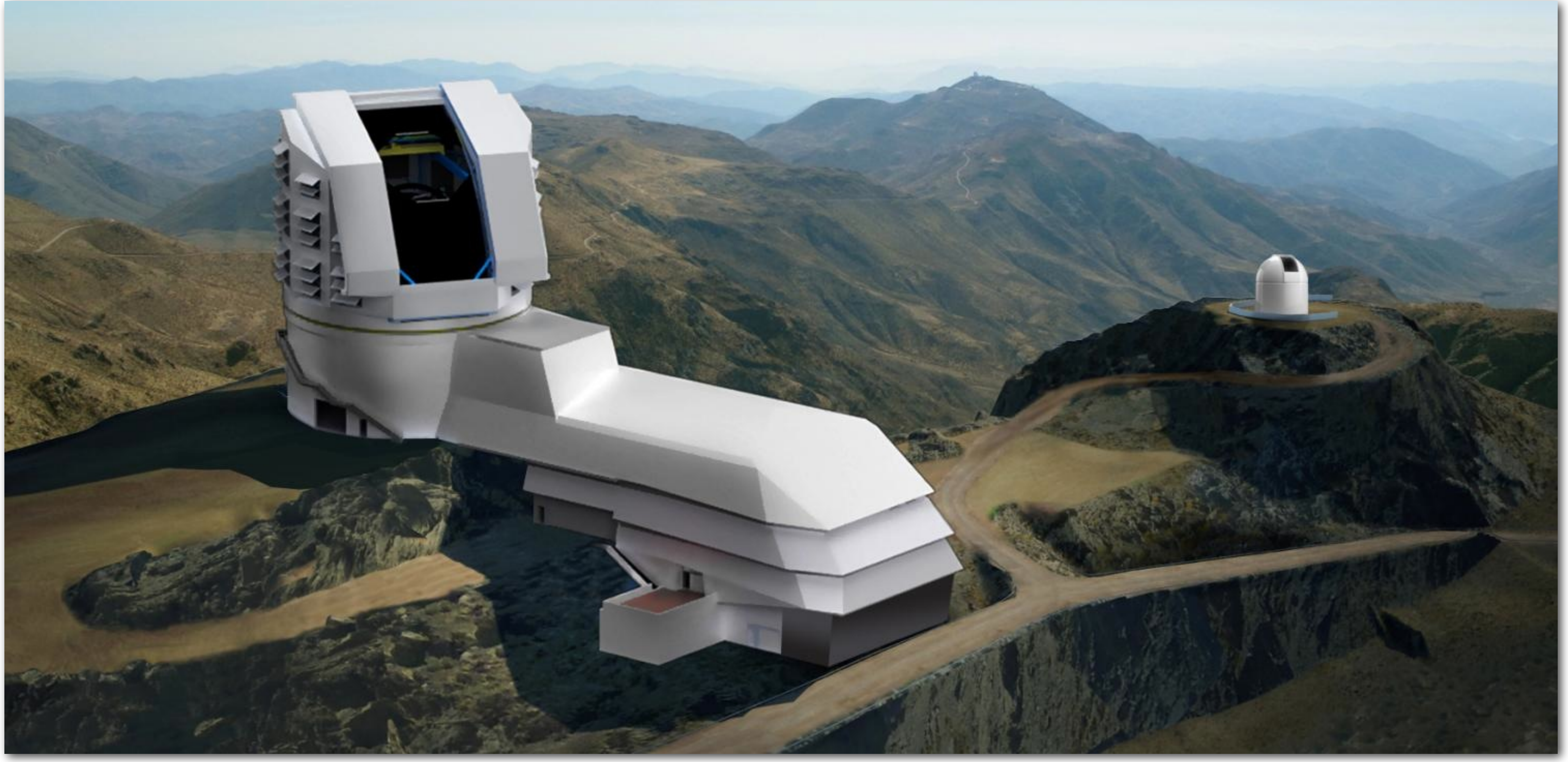
# LSST project: funding and budget

2014-2022 — Construction phase budget: US$ 671M

About 20% of the construction budget goes to the *Data Management* subsystem

2019-2032 — Operations phase budget: US$ 41M/year

# Observatory



Source: LSST

2647 m a.s.l.

500 km North of Santiago,
90 km East of La Serena

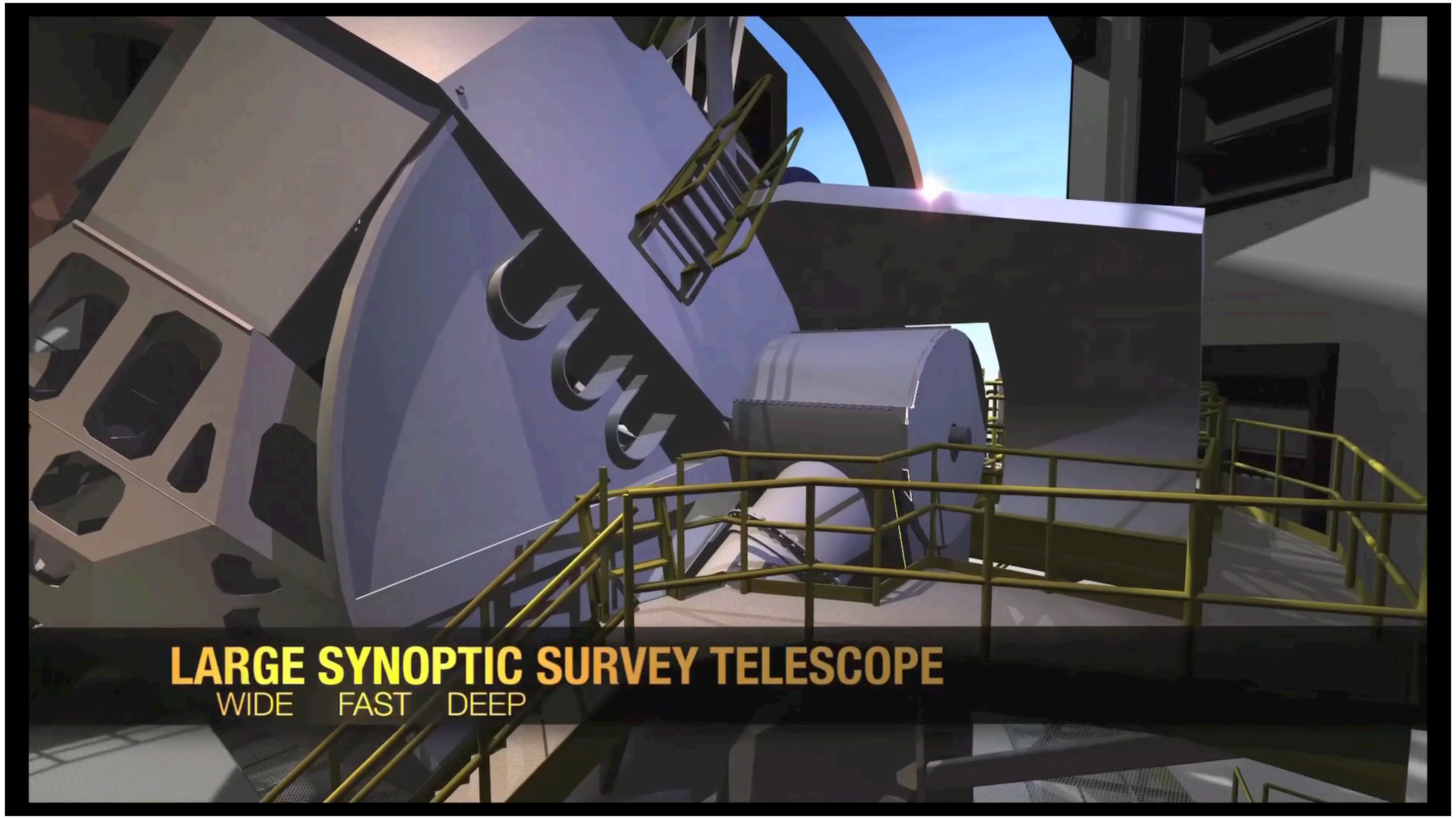
excellent observation conditions (stable air, clear sky, dark nights)

stable Chilean infrastructure

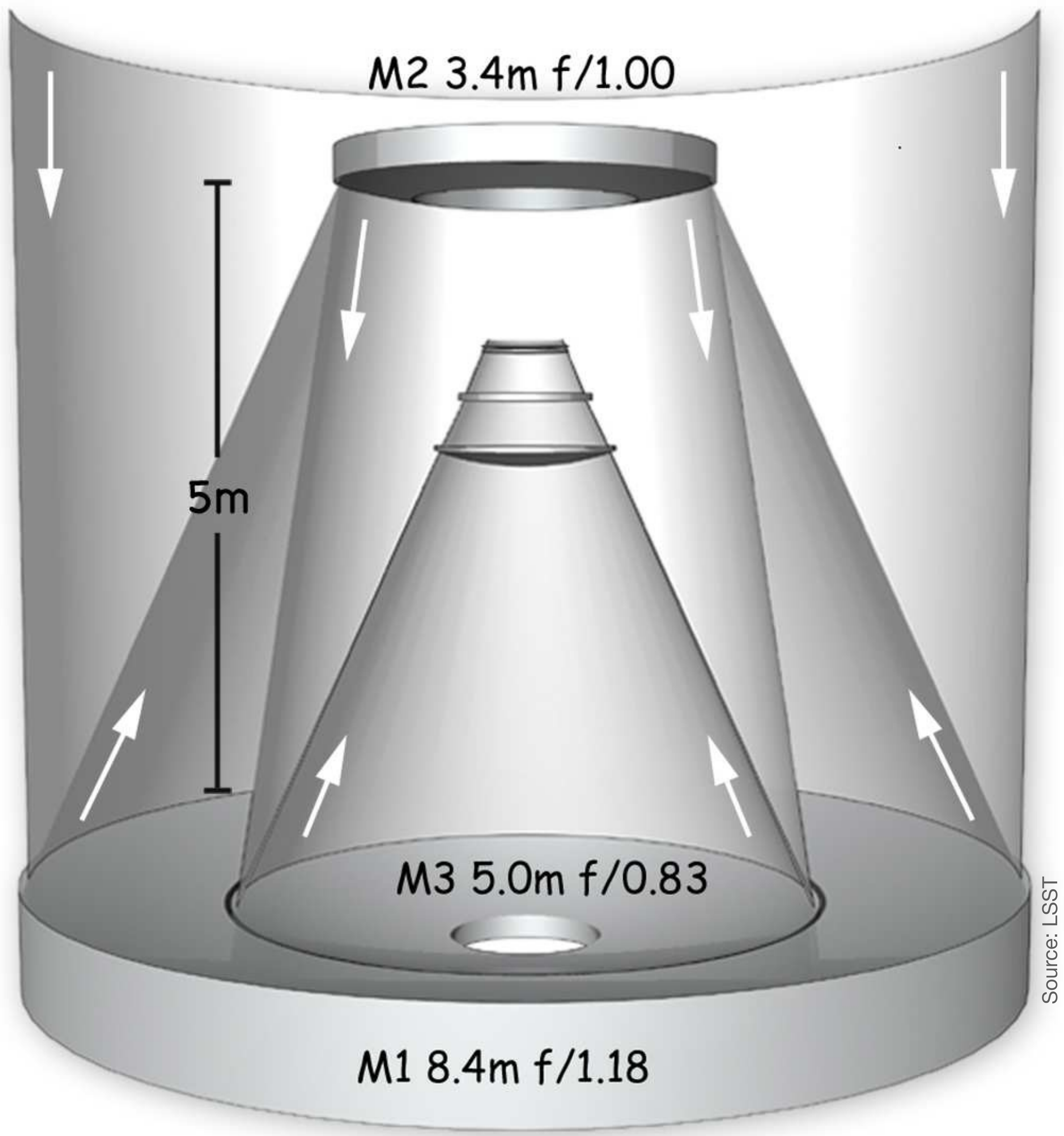
# ongoing civil engineering works, Aug 2017


Source: LSST

# Telescope



Source: LSST



Source: LSST

mobile structure, 350 ton

to be repositioned about 3 million times over the 10 years of operations

main mirror ⌀ 8.4 m

(effective aperture 6.5 m)

f/1.234

# Telescope (cont.)


Source: LSST

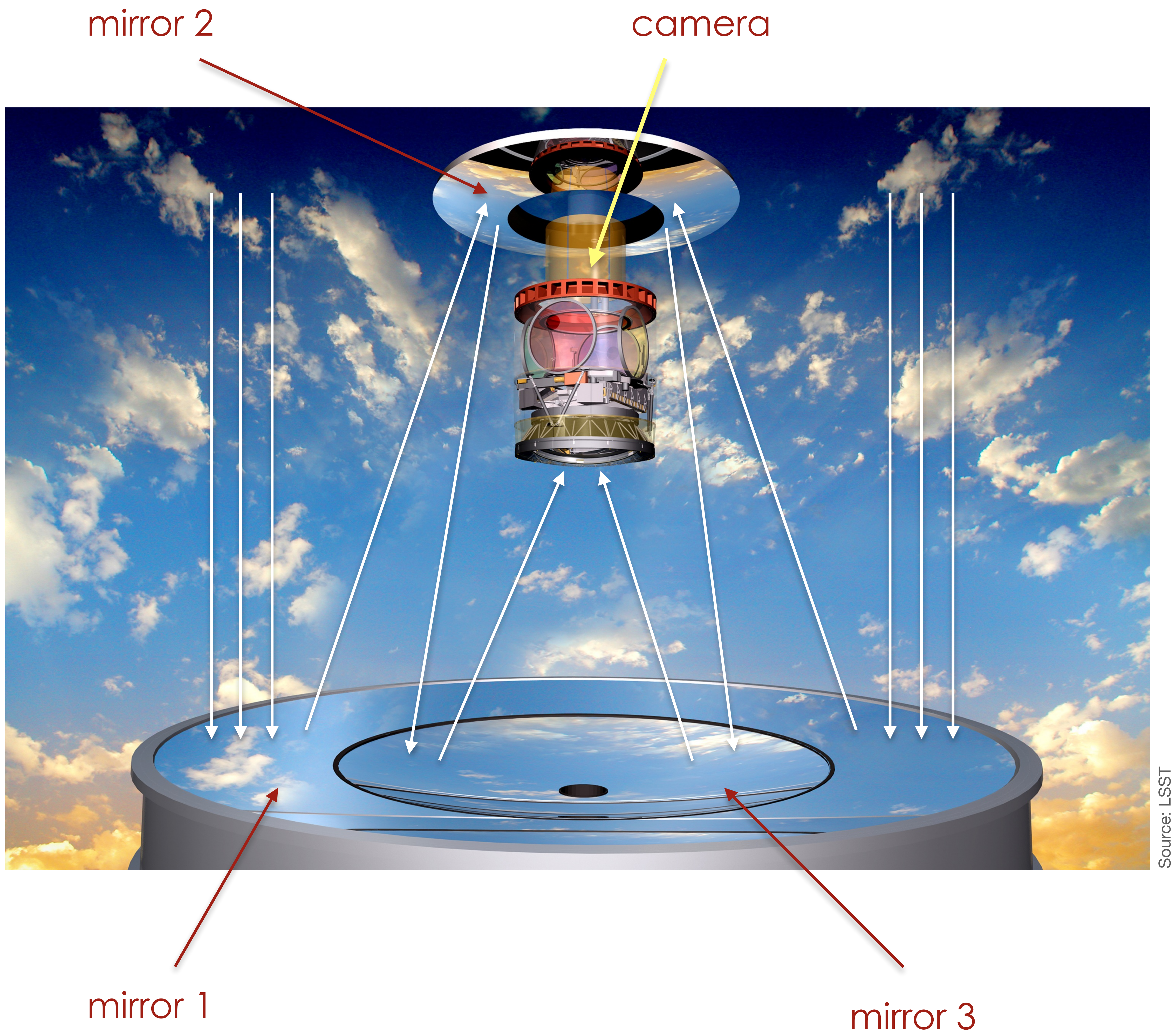
M1M3 mirrors, built on a single piece of glass

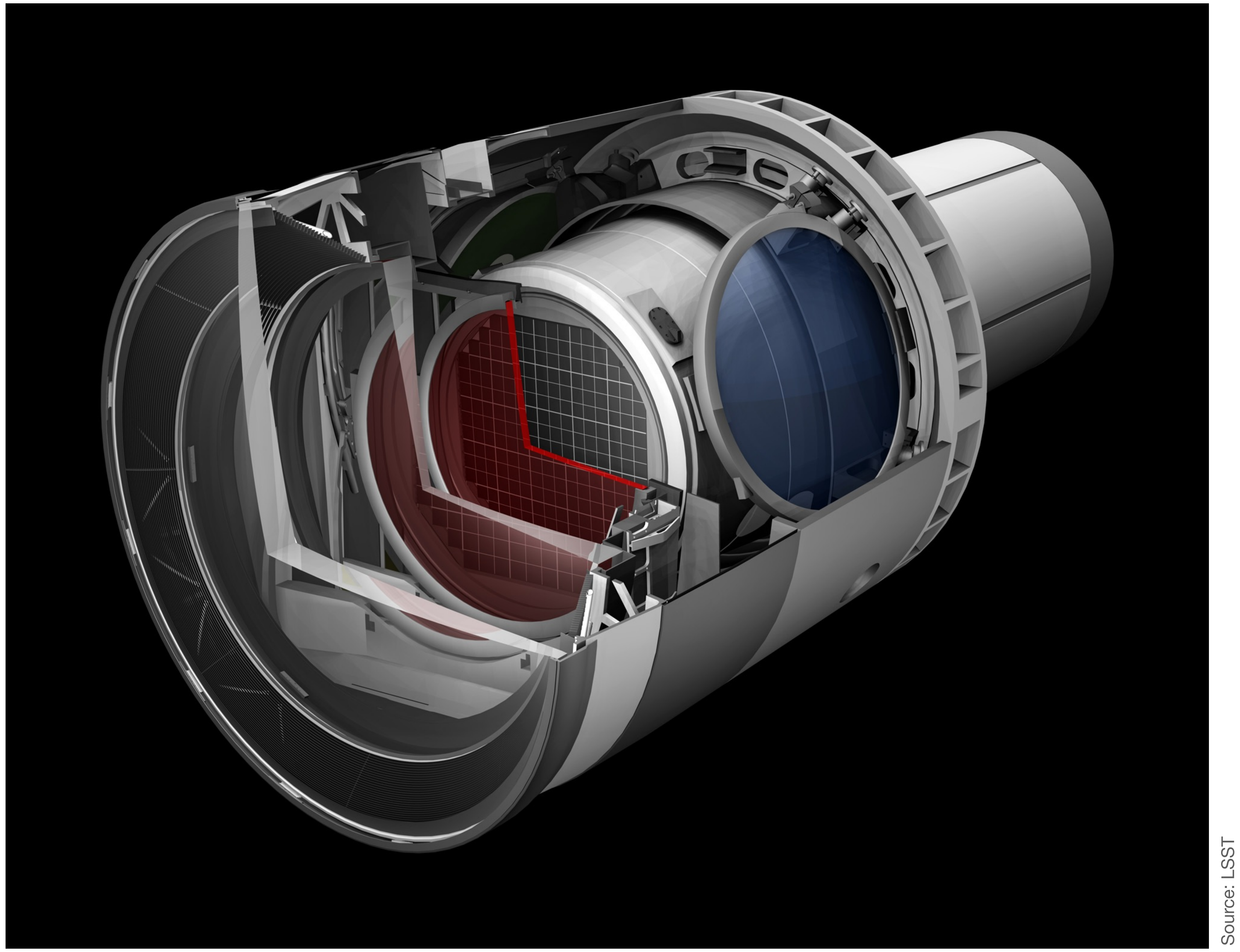
University of Arizona Tucson

16 ton

20M US$

# Camera




⌀ 1.65 m | 3.7 m long | 3 ton

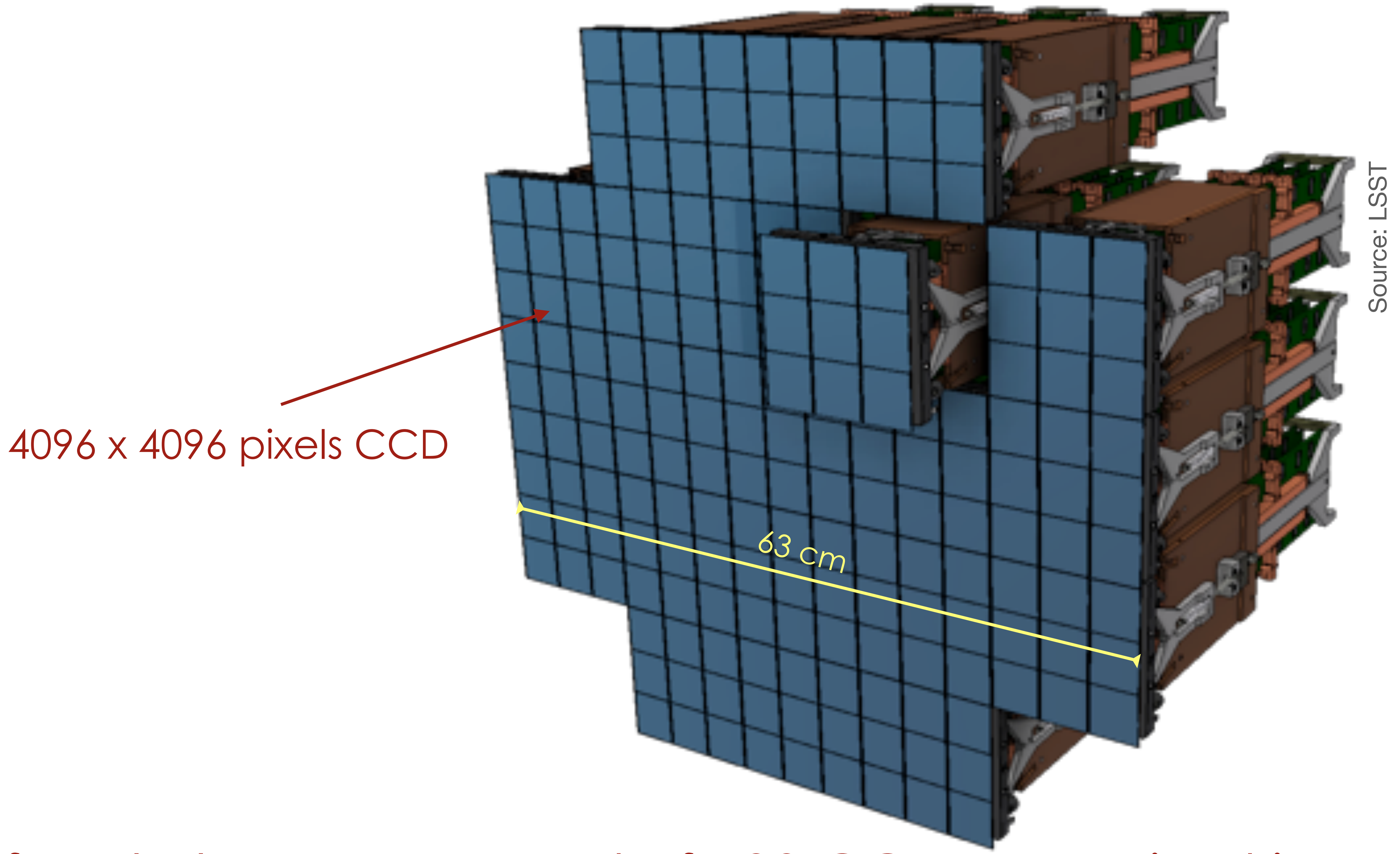
**3.2 G pixels** | 3 lenses | 5 embedded filters

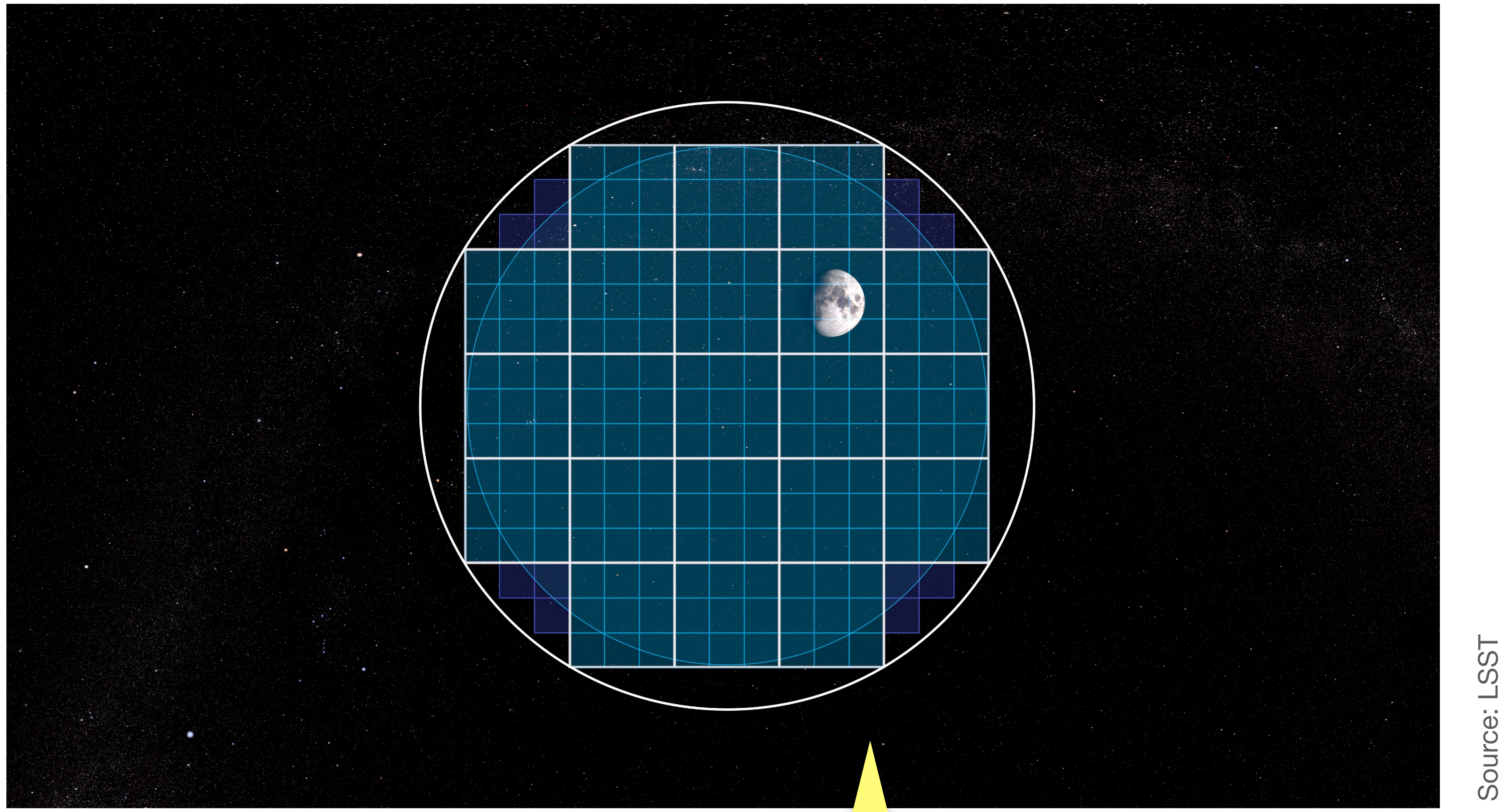
3.5° field of view | 9.6 degree$^2$

focal plane and electronics in cryostat at 173K

# Camera (cont.)

## the largest imager ever built for astronomy

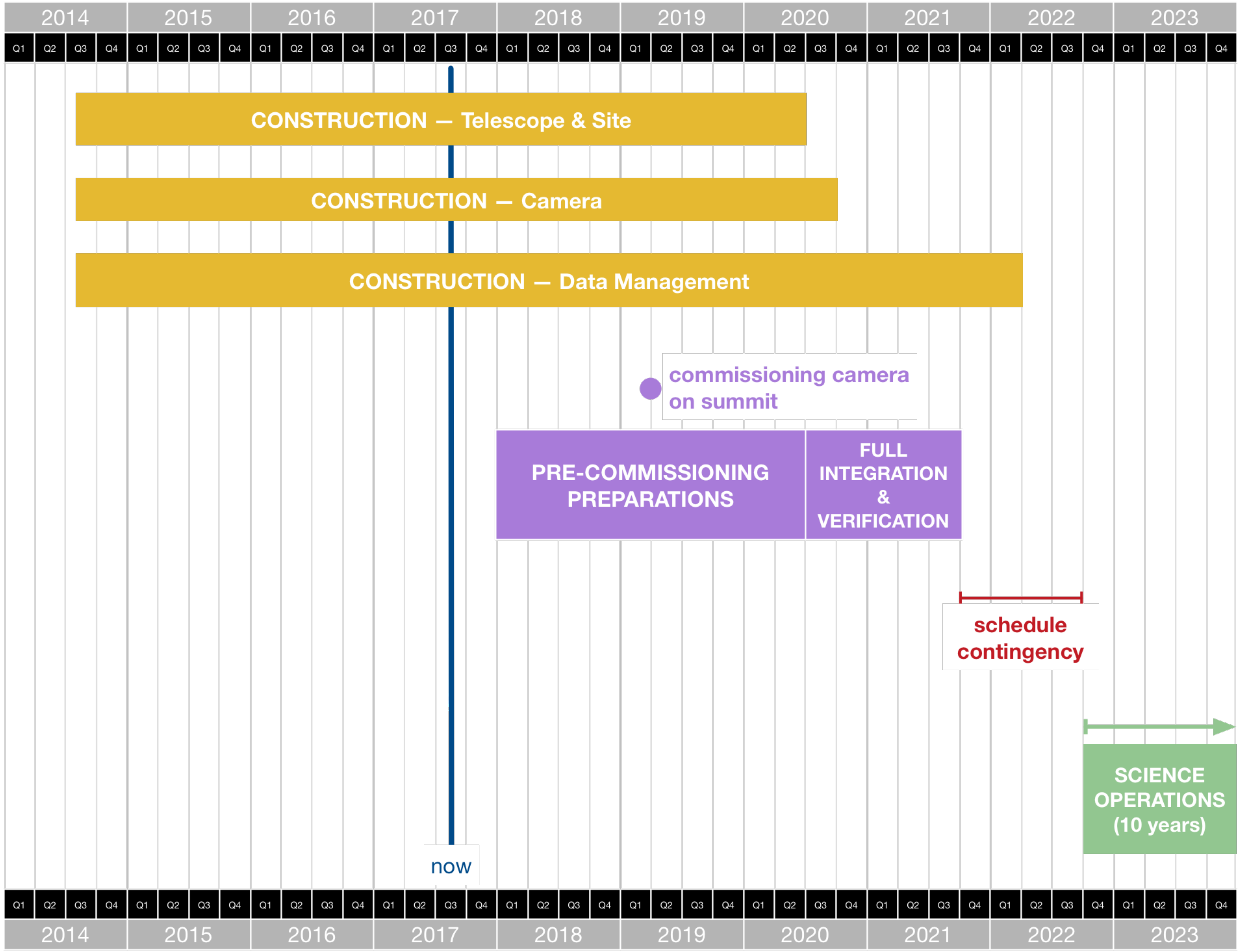



sensors can image an area equivalent to 48 full moons

focal plane composed of 189 CCDs organised in autonomous modules of 3 x 3
15 seconds exposures, 2 seconds read-out time
wavelengths: six bands, ultraviolet to near infrared (320 - 1050 nm)
0.2 arcsec per pixel | 10μm pixels

# Macro schedule




Adapted from V. Krabbendam

# LSST data processing

# 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**

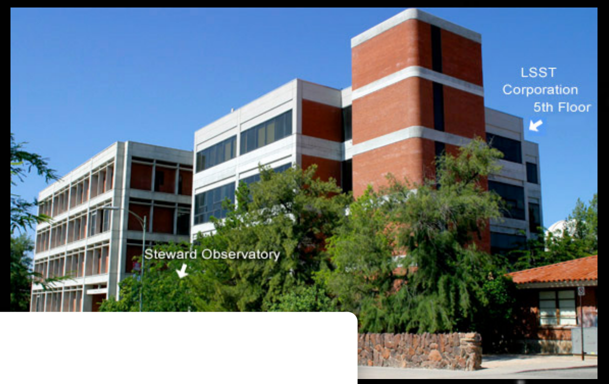
**On demand**

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

**Software** and APIs enabling development of analysis code

Source: LSST

# 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*

Sources: LSST, Google

# 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

UNIVERSITY of WASHINGTON

National Optical Astronomy Observatory

IN2P3 / CNRS computing center

# Data volume

- Raw data

  *7.2 GB per image*

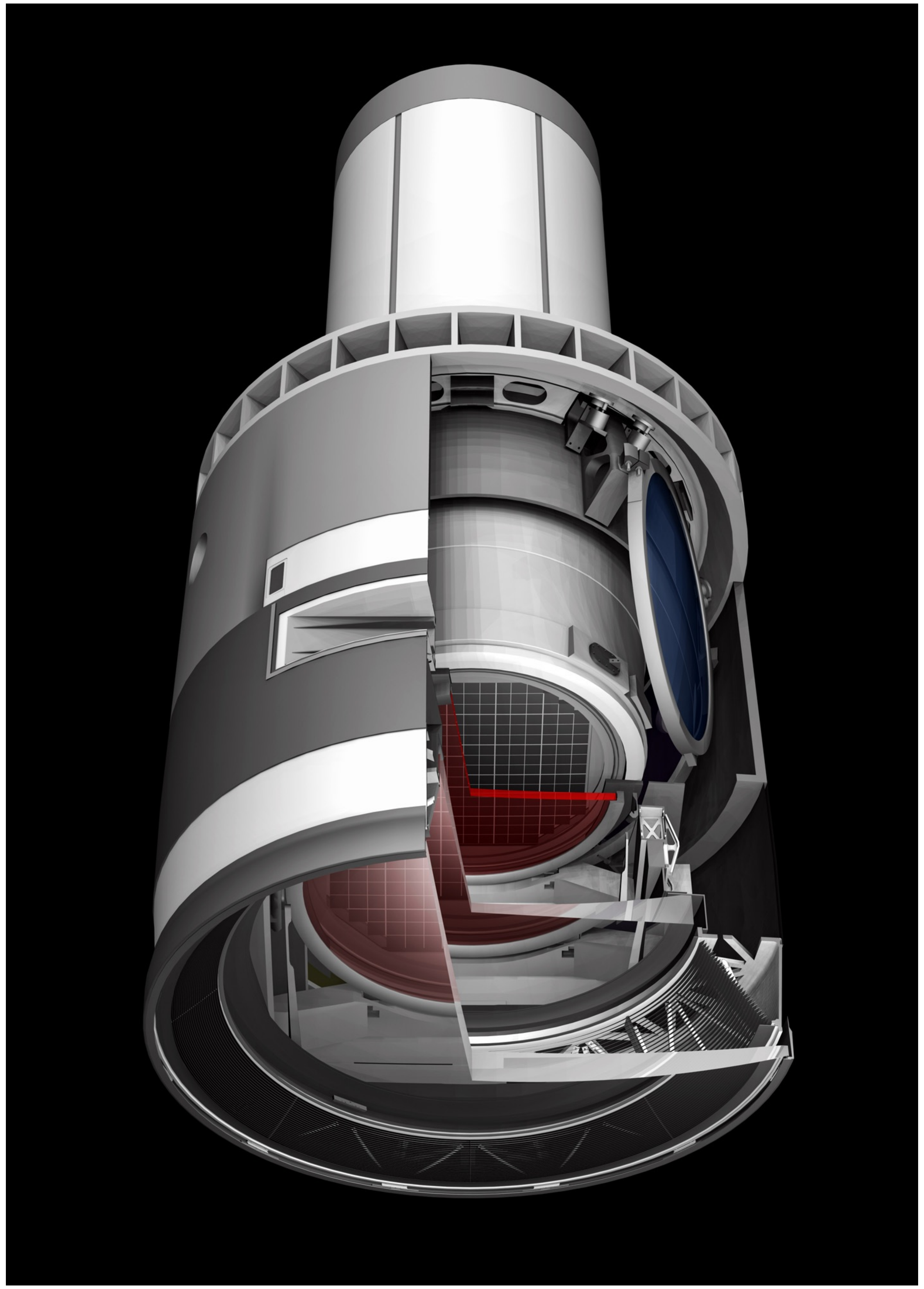
  *2000 science images + 450 calibration images per night, 300 nights per year*

  *15 TB per night, 4.5 PB per year*

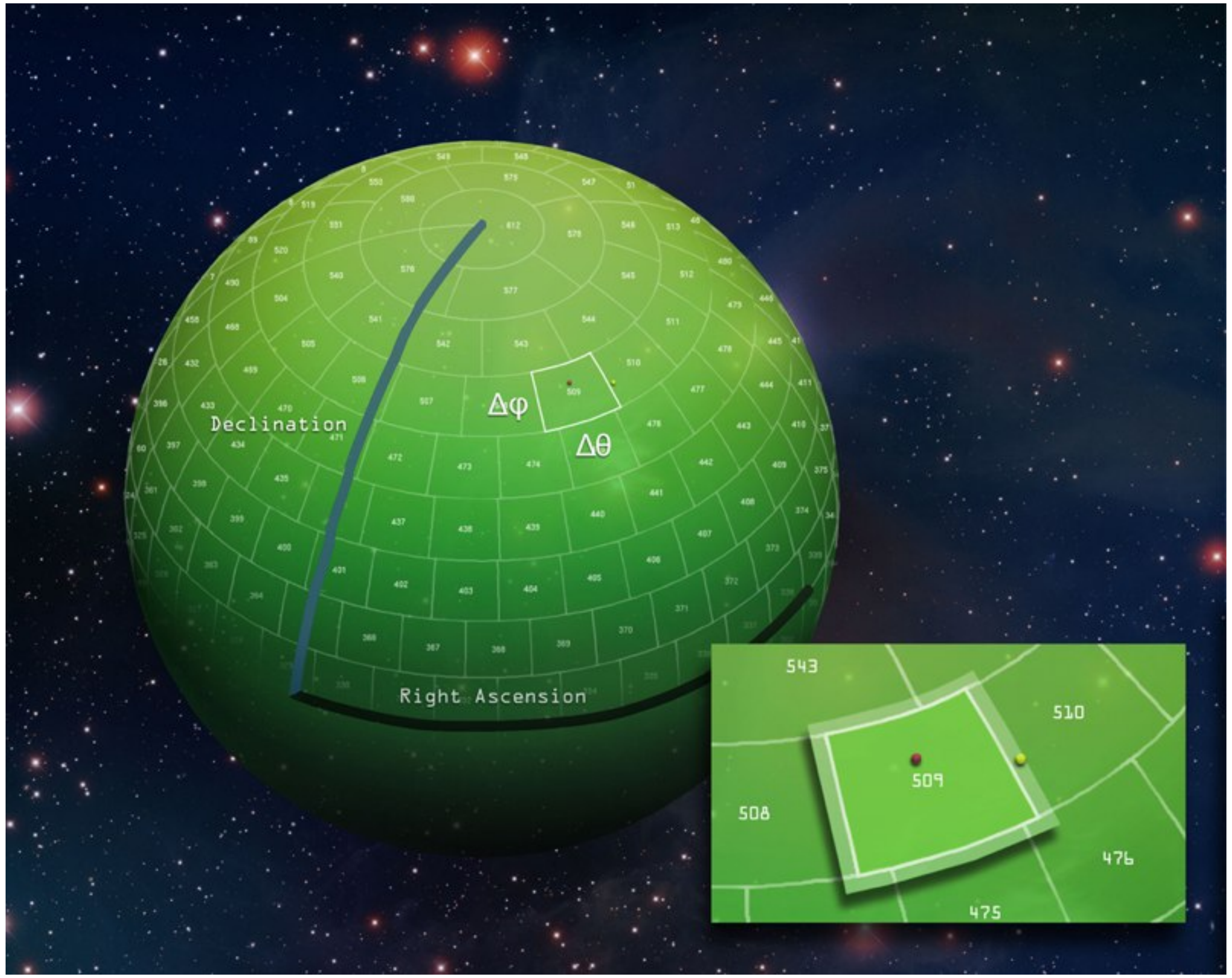
- Aggregated data over 10 years of operations, including derived data

  *images: ~6M exposures, 515 PB*

  *catalog: 83 PB*


Source: LSST

# LSST astronomical catalog




Source: LSST

### *Examples of queries*

SLAC

- *find all the observed galaxies in a given region of the sky*
- *find all objects similar to the colors of a quasar with redshift in a given range*
- *find all galaxies brighter than a given magnitude*
- *find all objects within x arc seconds of one another that have similar colors*

### *Qserv*

*custom, distributed relational database*

*spatial partitioning by sky coordinates, with overlaps*

***very high number of rows: ~37 trillion***

# LSST at CC-IN2P3

- Main roles

  ***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***

# Challenges

- Data volume

  *500+ PB of image data, 80+ PB of astronomical catalog*

- Knowledge extraction in real time

  *1-2 billion objects monitored nightly for important variations*

  *3 TB per hour must be mined in real time for emitting alerts*

- Database for astronomical catalog

  *to support simultaneous, large spectrum of queries from users over a catalog of ~40 billion rows and 7 to 7000+ columns*

- Data analysis

  *automated detection and extraction of physical properties for a large number of celestial objects (~40 billion) present in a large number of images (100 to 1000) using several models (~200)*

# International contributions

| Construction | Operations | | |
|---|---|---|---|
| United States | Argentina | Hungary | South Africa |
| Chile | Australia | India | Spain |
| France | Brazil | Italy | Sweden |
| | Canada | Korea | Switzerland |
| | **China** | New Zealand | Taiwan |
| | Croatia | Poland | United Kingdom |
| | Denmark | Serbia | |
| | Germany | Slovenia | |

*LSST-China consortium*

Dr. ZHAO Gang (NAOC)
MoA signed in Jan 2014

25 countries
39 research institutions

LSST all hands meeting, Aug 2017

# Conclusions

# Summary

- LSST aims to produce the most complete astronomical catalog as well as a digital color movie of the Universe

  *the project is set to deliver a processed dataset, science papers will be produced by the scientific community*

- Processing LSST data is challenging

  *significant data volume, large number of objects, sophisticated algorithms, time constraints*

- IN2P3 intends to make a major contribution to the LSST data processing effort

  *in addition to the significant contribution to the camera subsystem*

  *scientists will have the possibility to exploit the data for their research in the best possible conditions*

- High-visibility, multi-institution, long term, highly distributed project

# QUESTIONS & COMMENTS

# Acknowledgements

# Acknowledgements

This presentation includes material extracted from several sources, among them:

*[lsst.org](lsst.org)*

*Mario Juric*

*Victor Krabbendam*

*Andrew Connolly*

*Chris Smith*

*Michael Strauss*

*Jim Bosch*

*Željko Ivezić*

# Backup Slides

# Who I am

- Senior research engineer

  *computer science background*

  *25+ years involved in high energy physics research*

- Currently

  *LSST project leader at CC-IN2P3, in charge of preparation of data processing infrastructure*

- Previously

  *senior visiting scientist at CAS' Institute of High Energy Physics (IHEP), Beijing (China)*

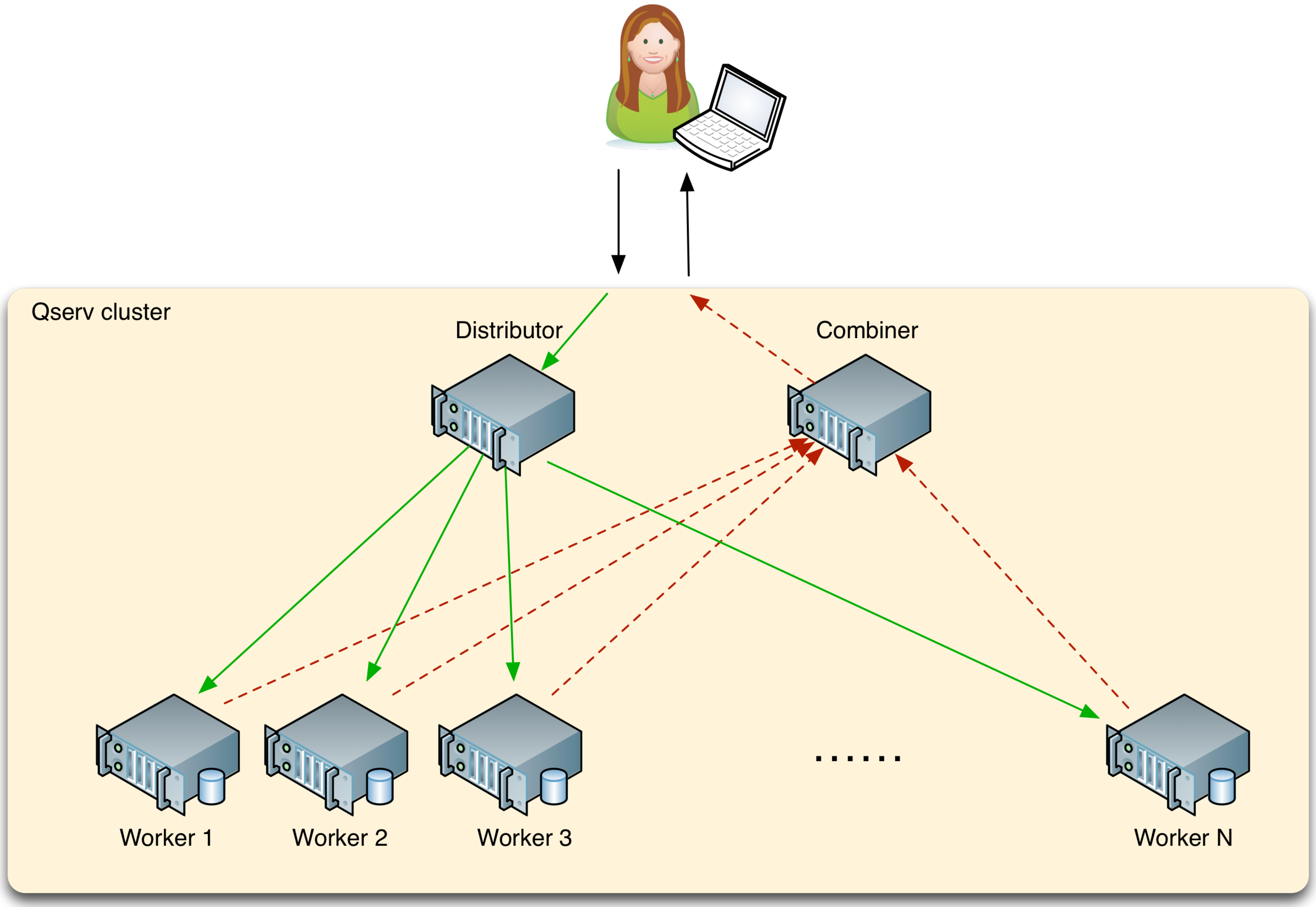
  *international technical expert at the office of science of the Embassy of France in China*

  *technical leader of French contribution to CERN's Large Hadron Collider computing grid*

  *CC-IN2P3 deputy director*

  *leader of engineering teams doing grid computing, data center operations, storage & systems administration*

# 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*