

Kako automatski obraditi 100 PB astronomskih slika?

Željko Ivezić

University of Washington, Sveučilište u Zagrebu

FER & IEEE, Zagreb, 16.XII 2010



Što je LSST? Large Synoptic Survey Telescope

- LSST je sistem za brzo snimanje neba: digitalni film u boji
- Glavni ciljevi: kozmologija, opasni asteroidi
- Najveća kamera na svijetu: 3200 Megapix
- Svaku noć preko 20 TB podataka, 2018-2027
- Najveća javna baza podataka: preko 100,000 TB
- Cijena: oko milijardu USD



Teme predavanja:

- Zašto nam treba LSST?
- LSST dizajn
- Softverski izazovi



Moderne promatračke metode u astronomiji i astrofizici:

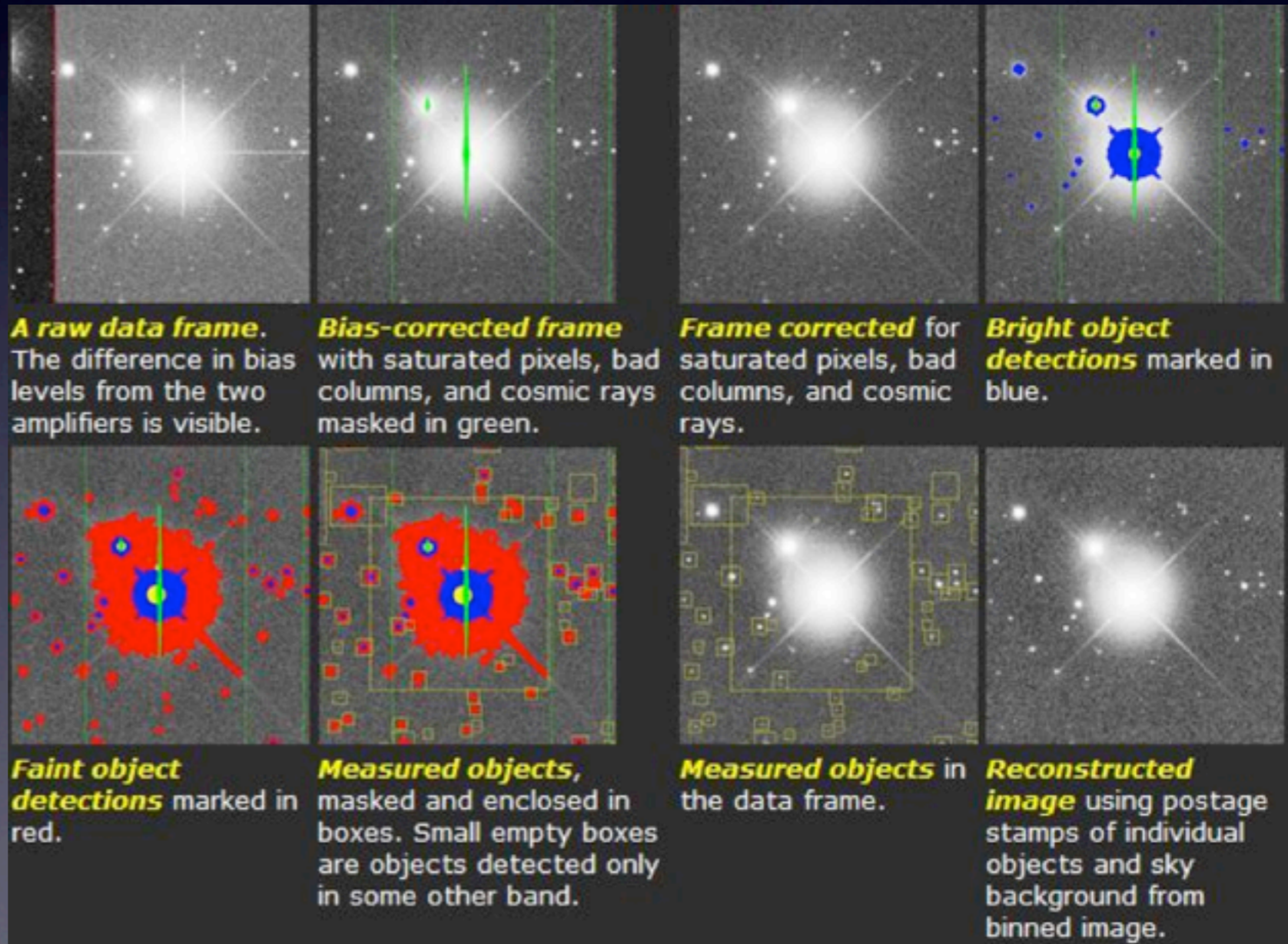
- **Veliki teleskopi (~10m):** tamni objekti, spektroskopija
- **Teleskopi iznad atmosfere:** visoka rezolucija (HST) i ne-optička valna područja (X, radio, IR)
- **Pregledi (mapiranje) neba:**
 - prodor digitalne tehnologije za senzore (CCD: charge-coupled device), za obradu podataka, i za distribuciju podataka i drugih informacija

Što je mapa neba?

Zašto praviti mapu neba?

- **Mapa neba:**

- popis svih vidljivih objekata (zvijezde, galaksije...)
- popis mjerenih parametara (sjaj, veličina, boja...)



Osnovni koraci u procesiranju astronomskih slika (primjer: Sloan Digital Sky Survey):

Što je mapa neba?

Zašto praviti mapu neba?

- **Mapa neba:**
 - popis svih vidljivih objekata (zvijezde, galaksije...)
 - popis mjerenih parametara (sjaj, veličina, boja...)
- **Zašto praviti mapu neba?**
 - Otkrivanje novih objekata: “jel’ to novi asteroid, ili nekaj već poznato?”
 - Klasifikacija objekata: “kaj se sve vidi na nebu?”
 - Statistički pregled objekata: “a kol’ko ima kvazara?”
 - Potraga za neobičnim objektima: “je li to NLO?”
 - Kozmološka mjerenja: “koliko se brzo širi Svemir?”

Kratka povijest mapiranja neba

- **Hiparkos**

- prije oko 3,000 godina
- sve zvijezde vidljive golim okom iz Grčke: oko 3,000
- glavni izvor promatračkih astronomskih podataka idućih 2,500 godina

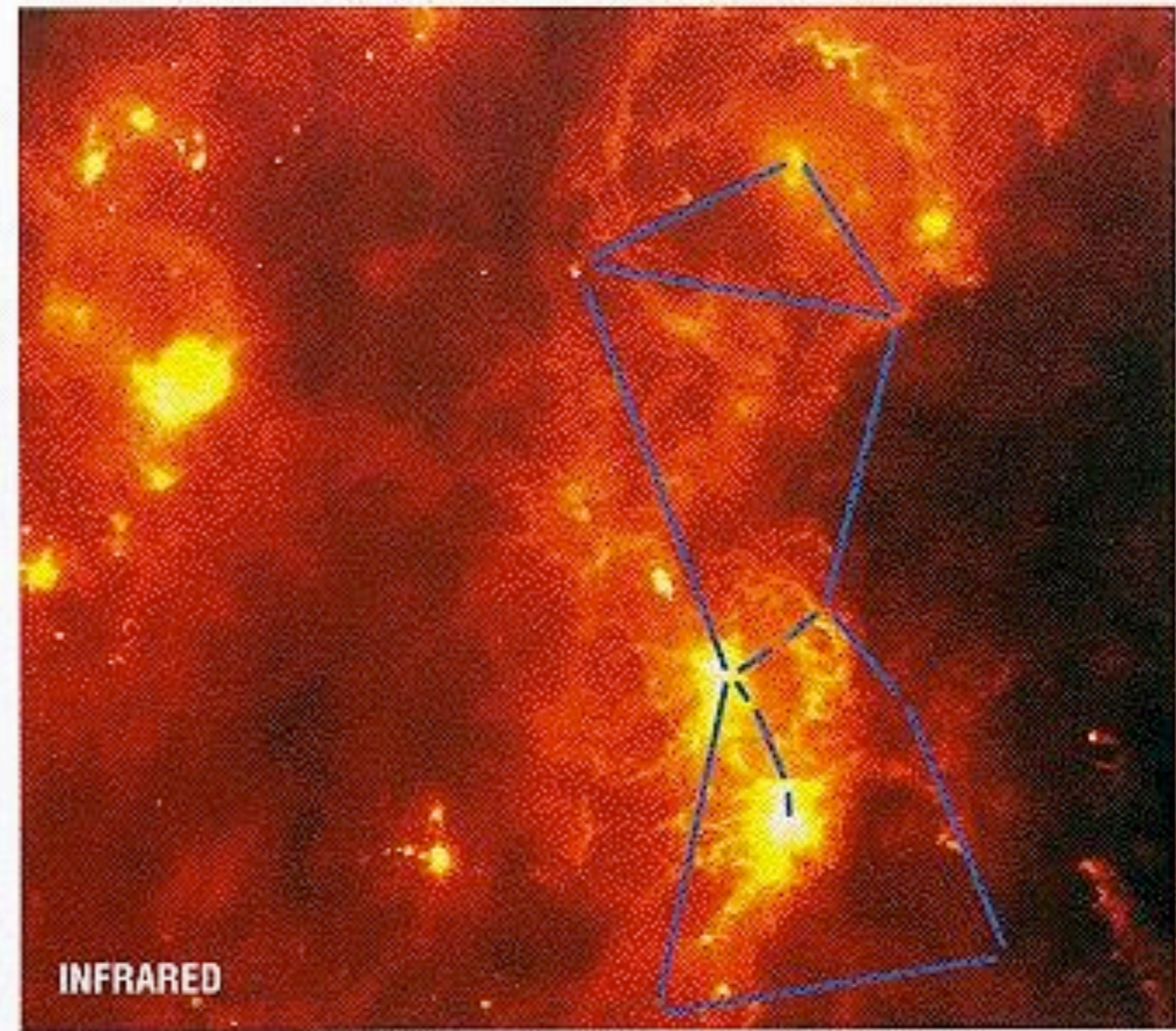
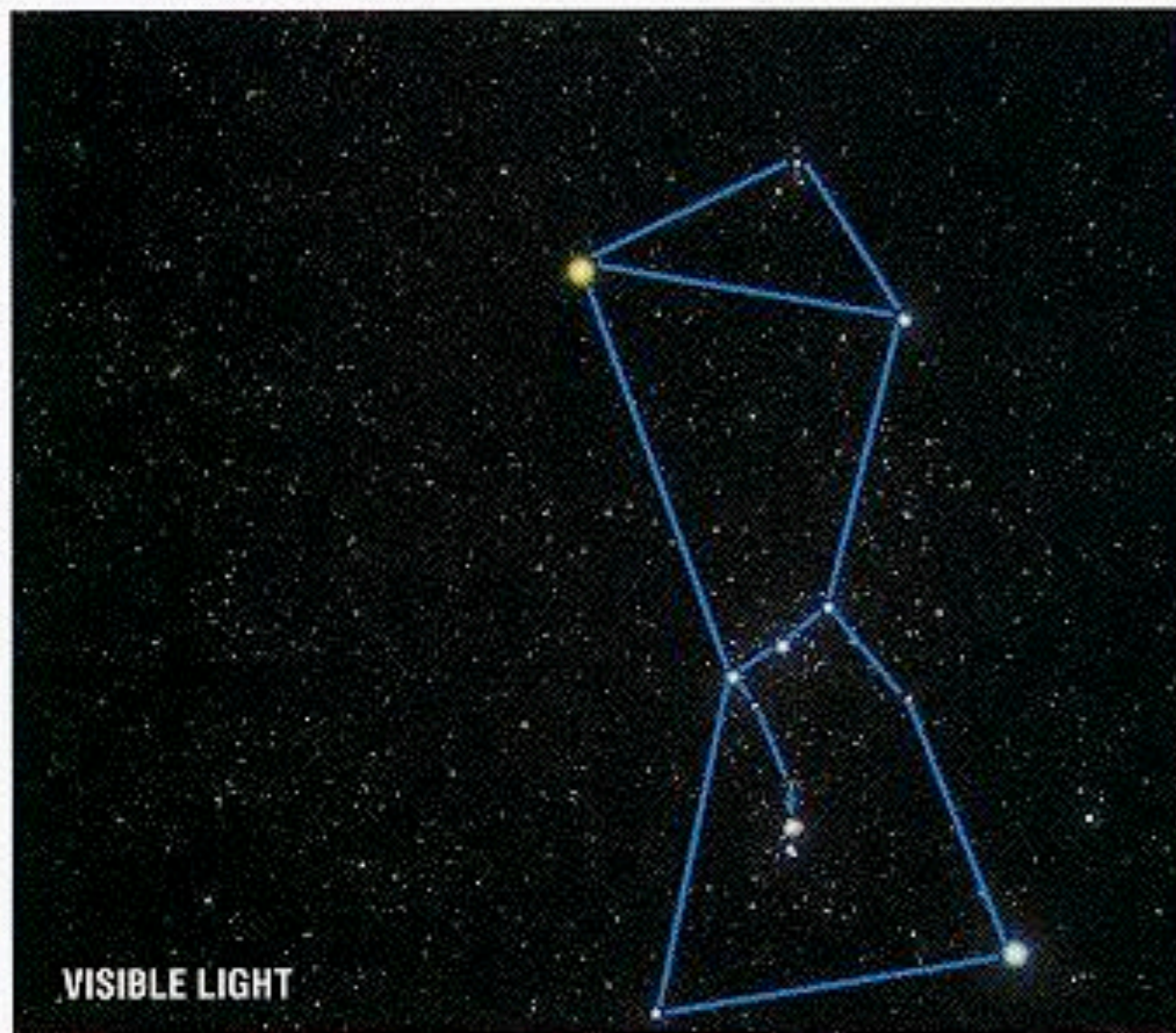
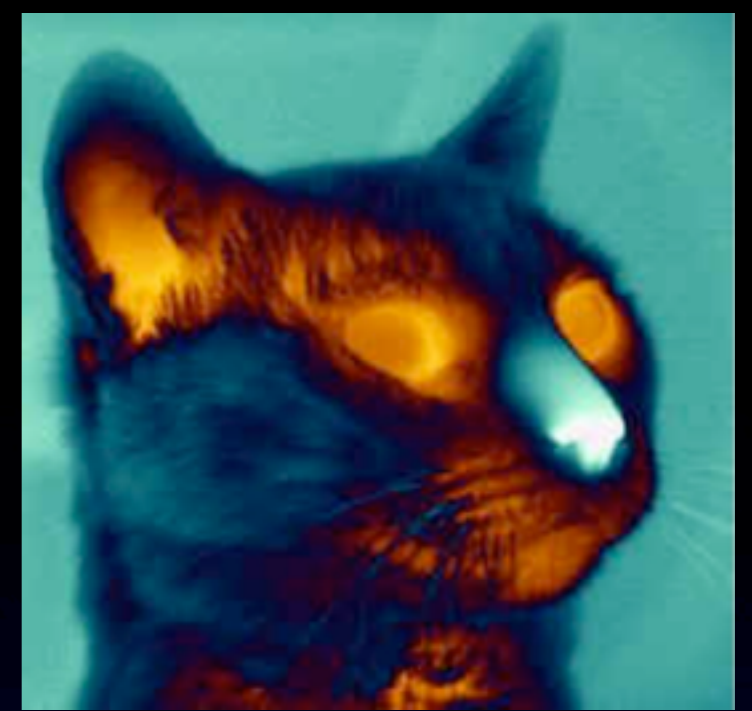
- **Tycho Brahe**

- Srednji vijek, puno točniji od Hiparkusa
- Još uvijek bez teleskopa: oko 3,000 zvijezda
- Glavni rezultati: Keplerovi zakoni kretanja planeta, Newtonova teorija gravitacije

Moderno mapiranje neba

- **Palomar Observatory Sky Survey**
(National Geographic Sky Survey):
 - optičko područje
 - 1950-1955 (druga faza '80tih)
 - oko 1,000 fotografija (cijelo nebo)
- **Druge valne duljine:**
 - X zrake
 - ultraljubičasto područje
 - infracrveno područje
 - radio astronomija

Optičke valne duljine otkrivaju samo mali dio stvarnosti...

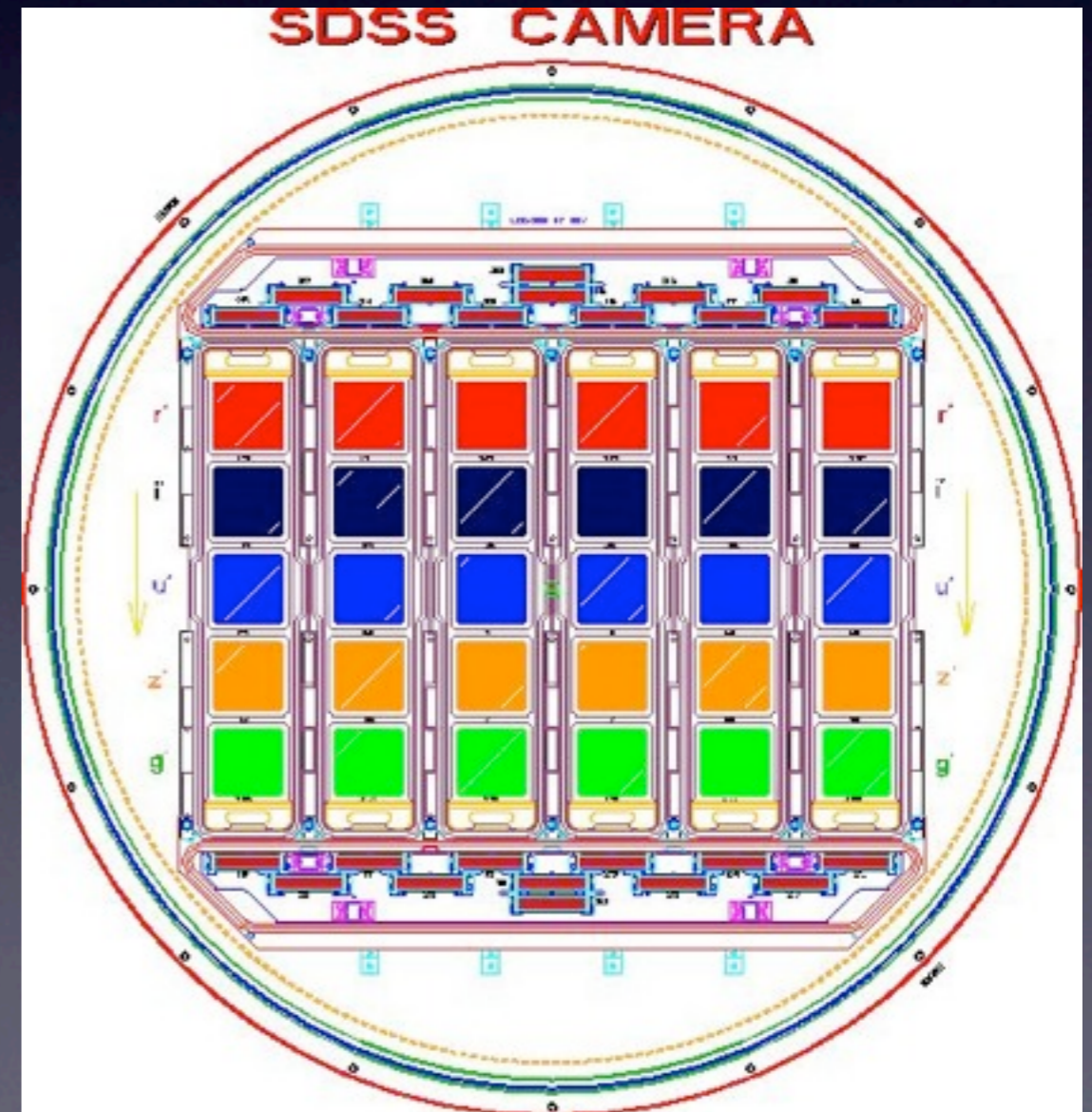


Zviježđe Orion: vidljivo svjetlo

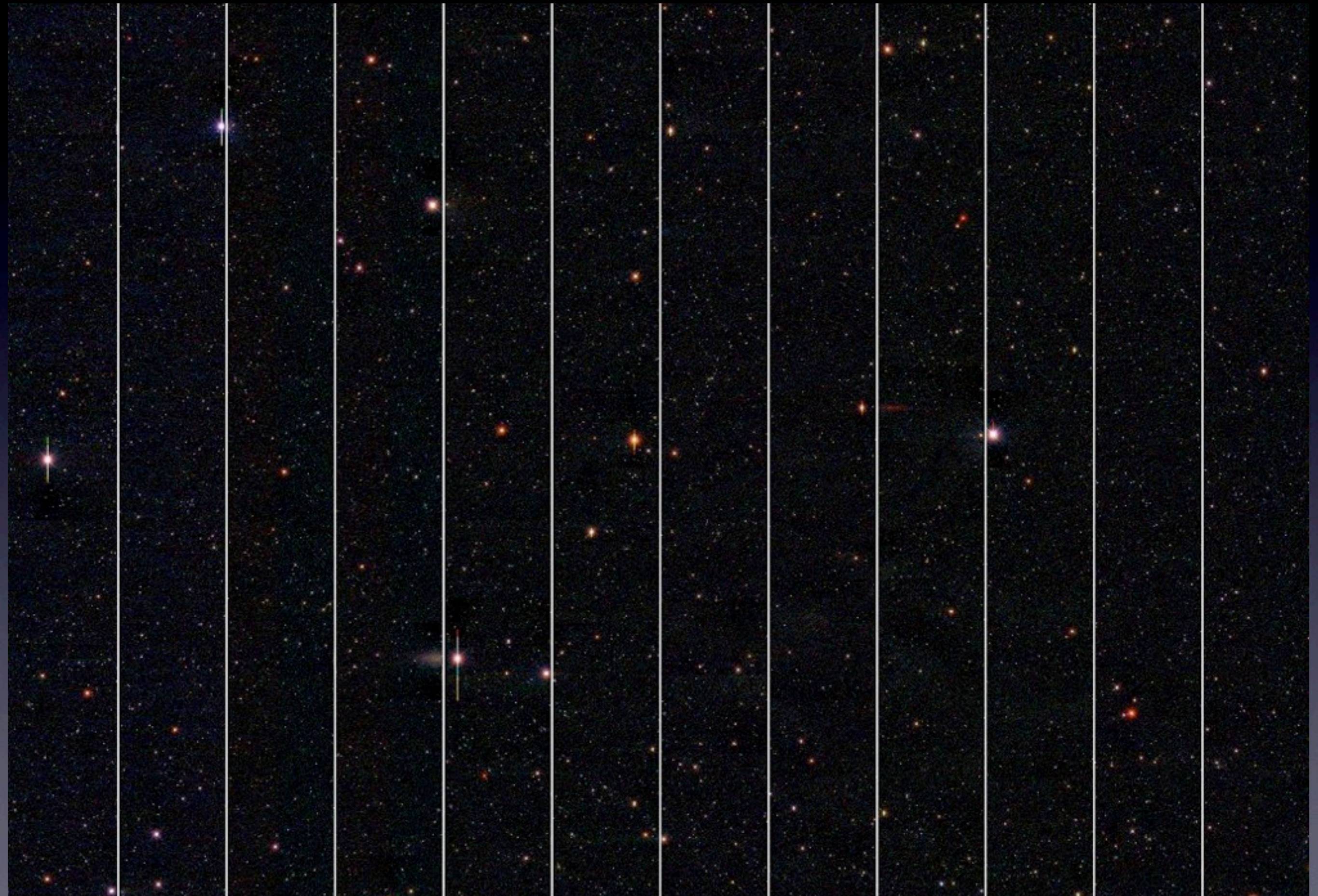
infracrveno svjetlo

Zadnjih 10 godina: Sloan Digital Sky Survey

- Prvi digitalni pregled neba (CCD kamera)
- Precizni podaci za oko 400,000,000 objekata
- Revolucija u radu astronoma: javne baze podataka



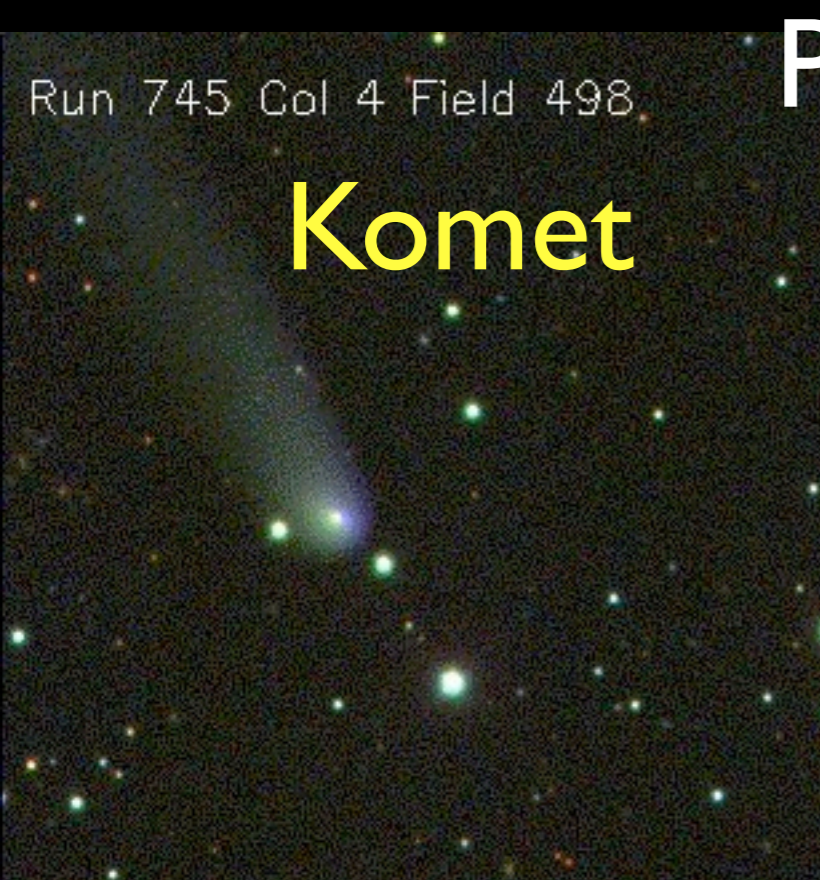
Primjer SDSS snimanja neba



Run 745 Col 4 Field 498

Primjeri SDSS slika

Komet



Patuljasta galaksija



Spiralna galaksija



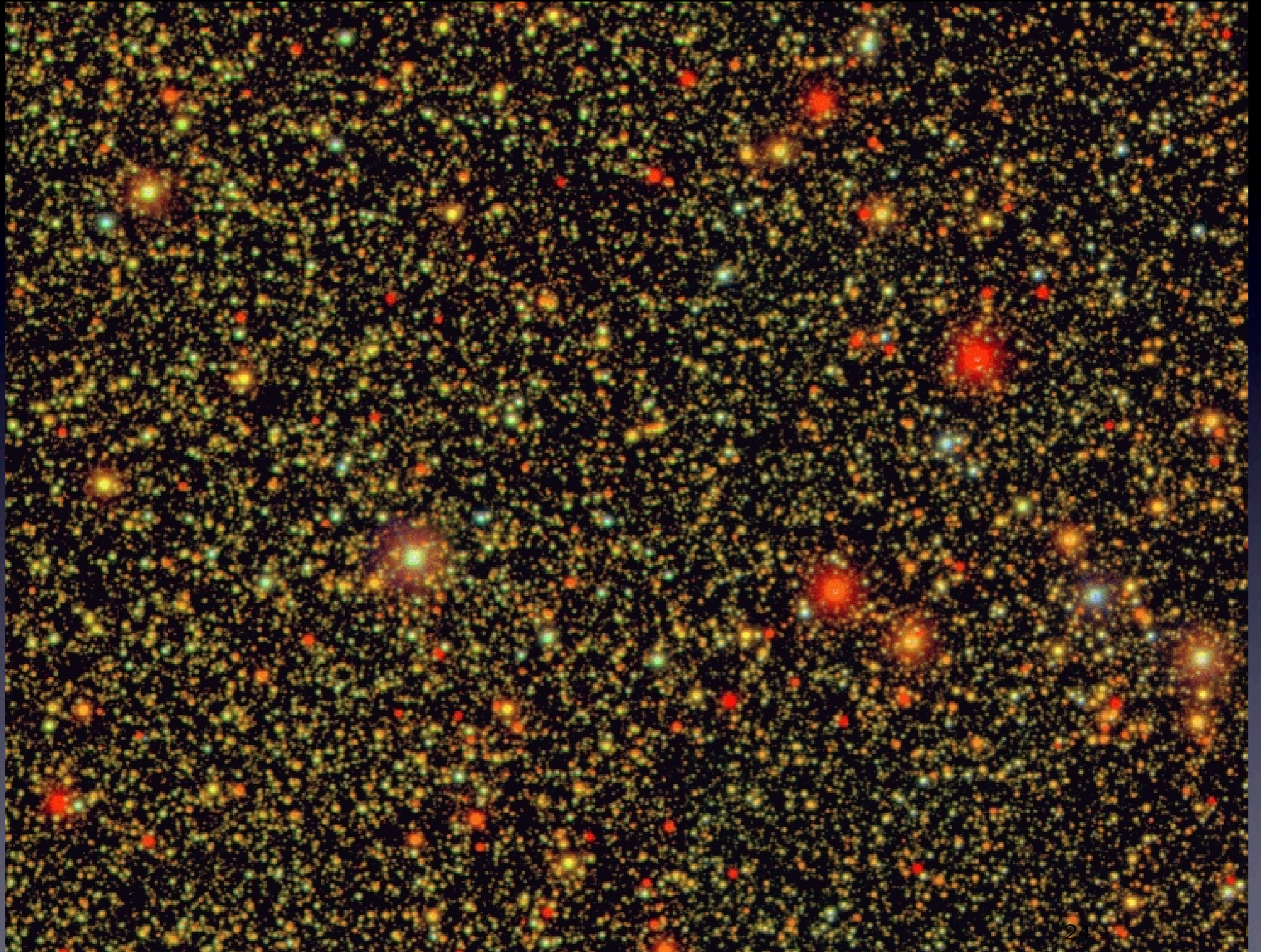
Maglica



Spiralne galaksije



SDSS pogled kroz ravninu Mlječnog Puta



Astronomija "od doma"

SDSS SkyServer DR7

http://cas.sdss.org/astro/en/

Address Book Apple Customize Links Yahoo! Free Hotmail Google Maps Windows YouTube Wi

Sloan Digital Sky Survey / SkyServer

Home Tools SQL Search Schema Finding Chart Download Projects DR7 DAS Site Search Help

Welcome to the **DR7 Catalog Archive Server** site providing public access to SDSS data for professional astronomers.

The following databases are available


BestDR7	[Default] The best version photo (imaging), spectro and tiling data
TargDR7	The version of the data from which spectroscopic targets were chosen

News
This site contains data from **Data Release 7 (DR7)**. Please see the **Site News** page, **What's New in DR7** page and the **Known Problems** page for more information.

To run a query on one of the DR7 DBs other than BESTDR7, name the database explicitly in the query:

```
SELECT TOP 100 * FROM TARGDR7..PhotoObj WHERE r<17 and r-i>2
```

SDSS is supported by



Search Tools

- Radial Search
- Rectangular Search
- SQL Search
- Imaging Query
- Spectro Query
- Object Cross-ID
- Get images
- Emacs Interface
- sqlcl

Advanced Tools

- Finding Chart
- Navigate
- Image List
- Explore
- CasJobs
- CasJobs CL tool
- VO Services
- Open SkyQuery
- Spectrum Services

Links

- About the SDSS
- About the SkyServer
- SDSS Project Website
- Data Archive Server
- Public SkyServer
- Famous places
- Images of RC3 Galaxies
- Educational Projects
- NVO Website

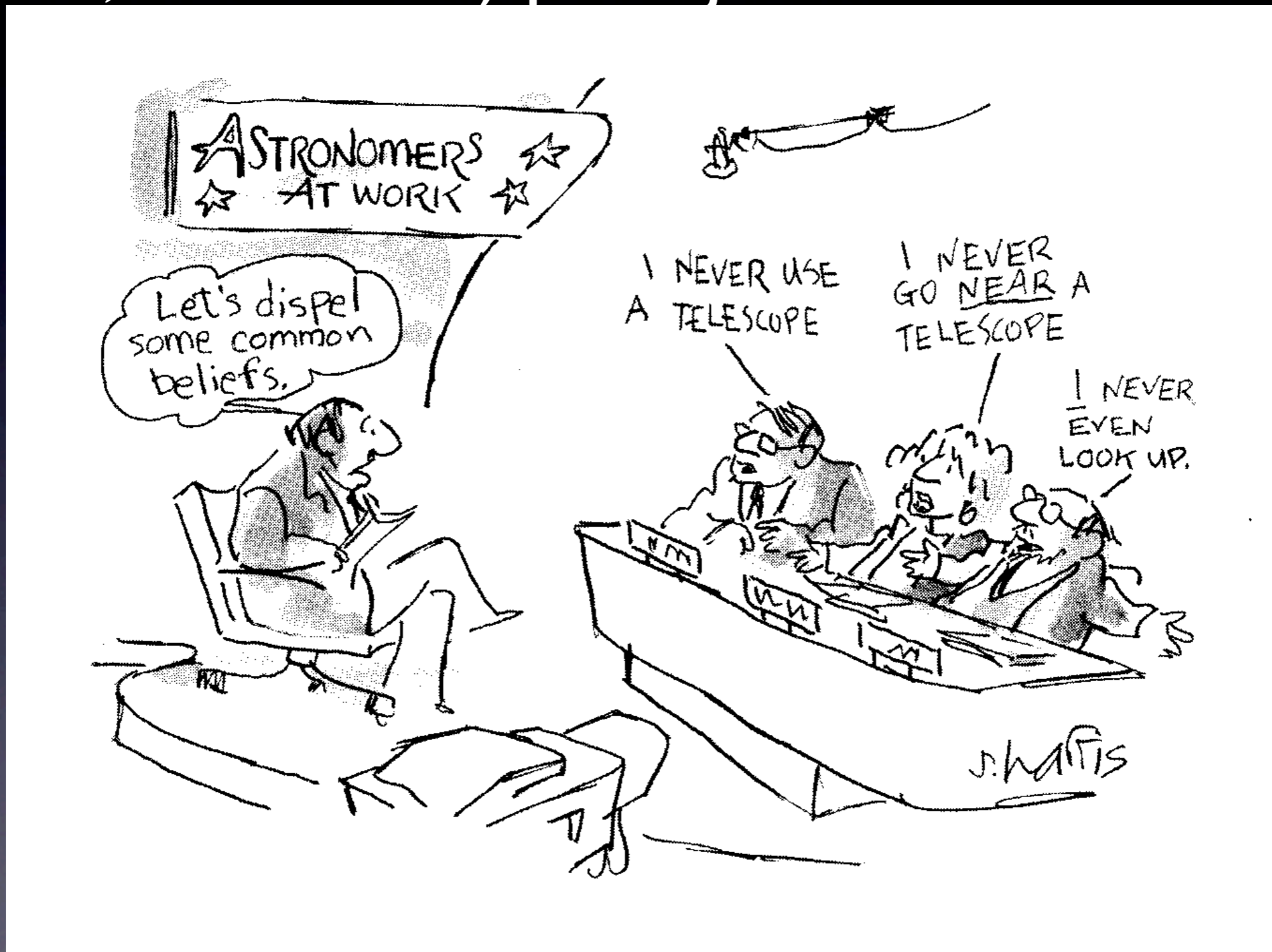
Help and Tutorials

- Archive Intro
- FAQ
- Glossary
- Algorithms
- Table Descriptions
- Schema Browser
- Sample SQL Queries
- Details of SDSS Data
- Skyserver.org

Powered by **Microsoft**

Site Traffic
Privacy Policy

Uz teorijsku i promatračku astronomiju, analiza masivnih baza podataka (stotine TB, uskoro PB) postaje novi način rada

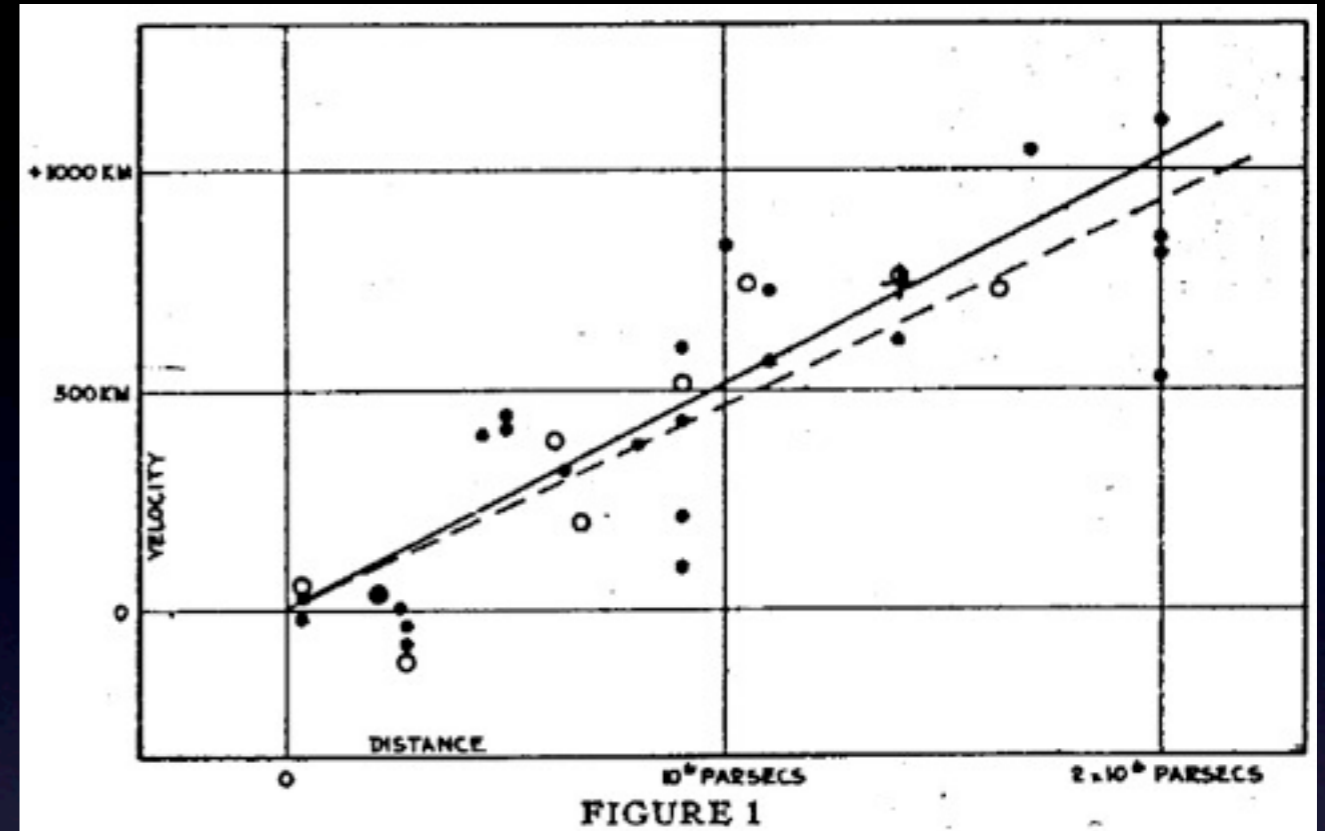


Zašto nam treba LSST ako već imamo SDSS?

Stare kozmološke zagonetke

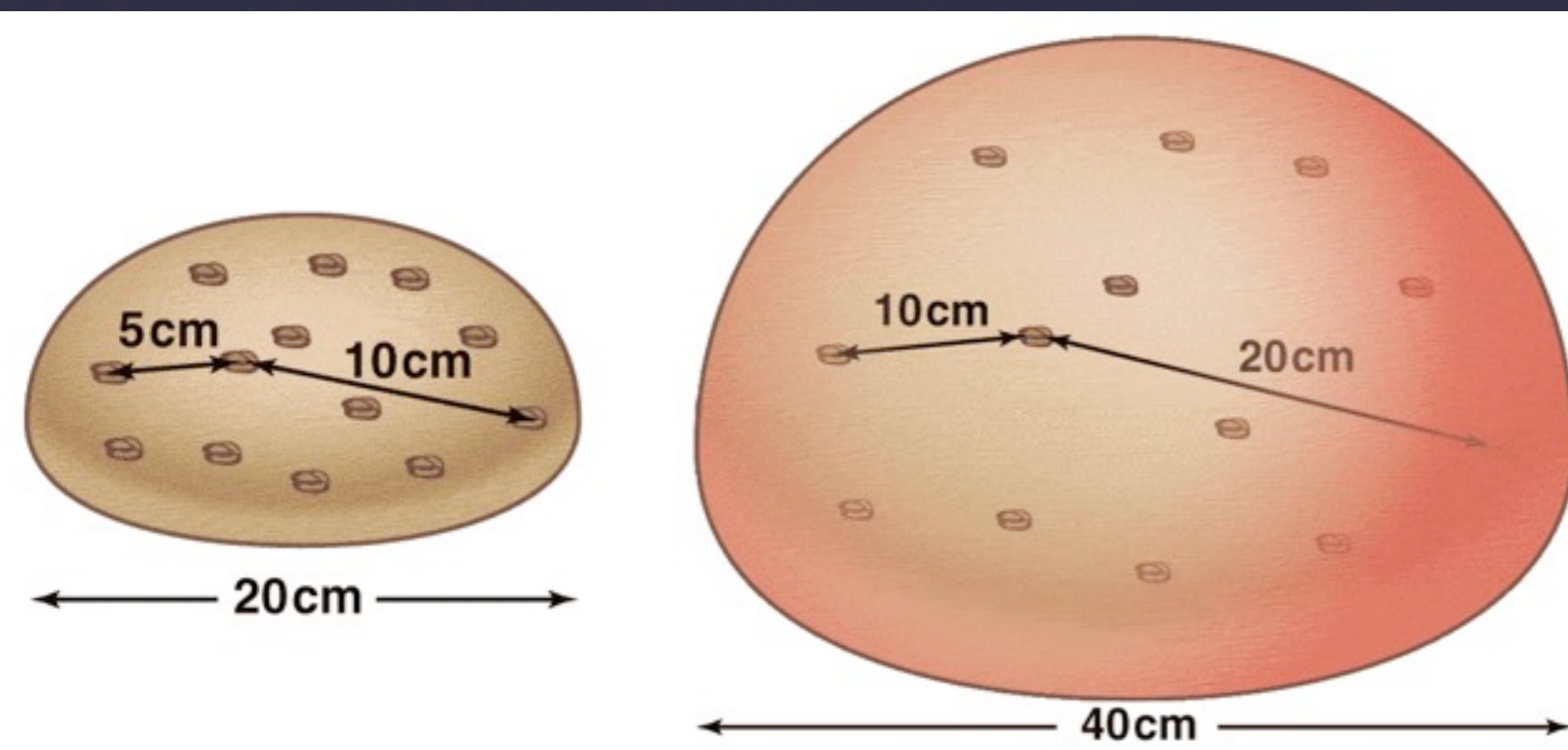


brzina udaljavanja



Edwin Hubble (1929): Svemir se širi!

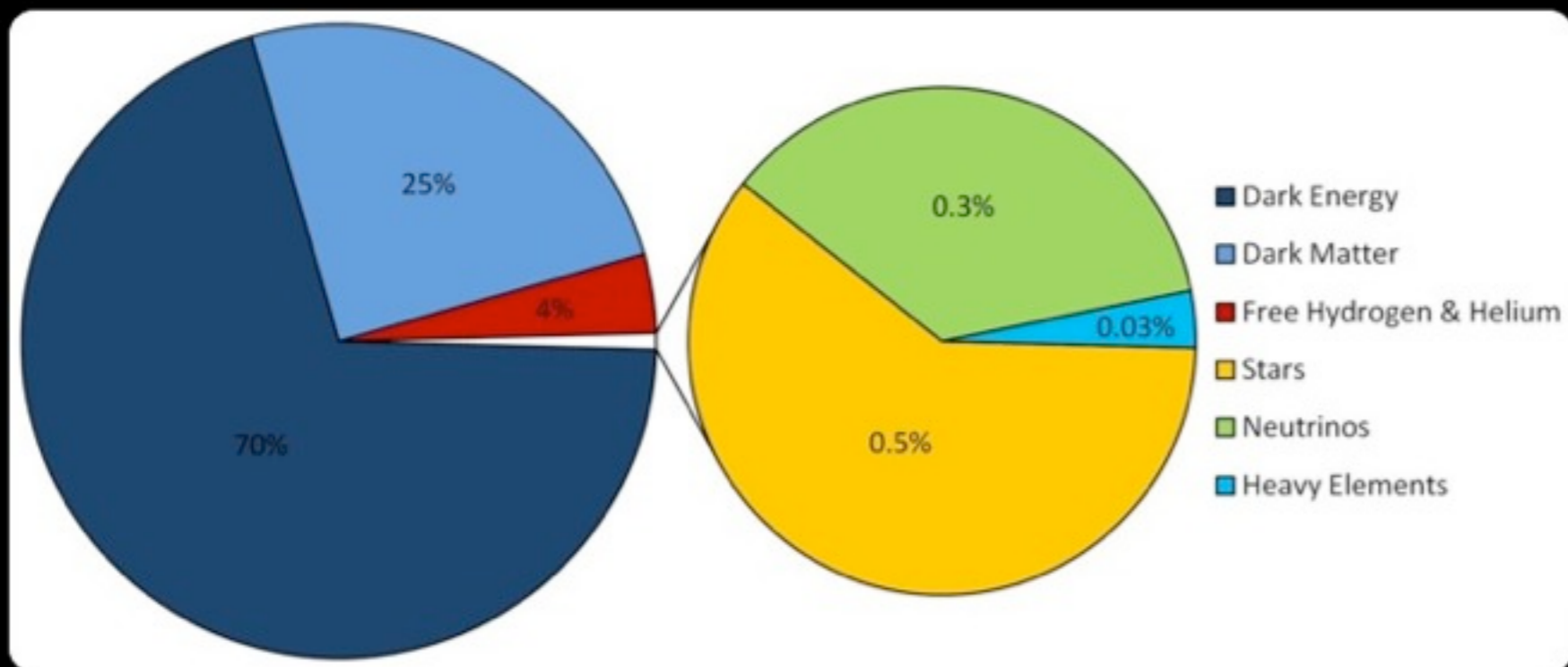
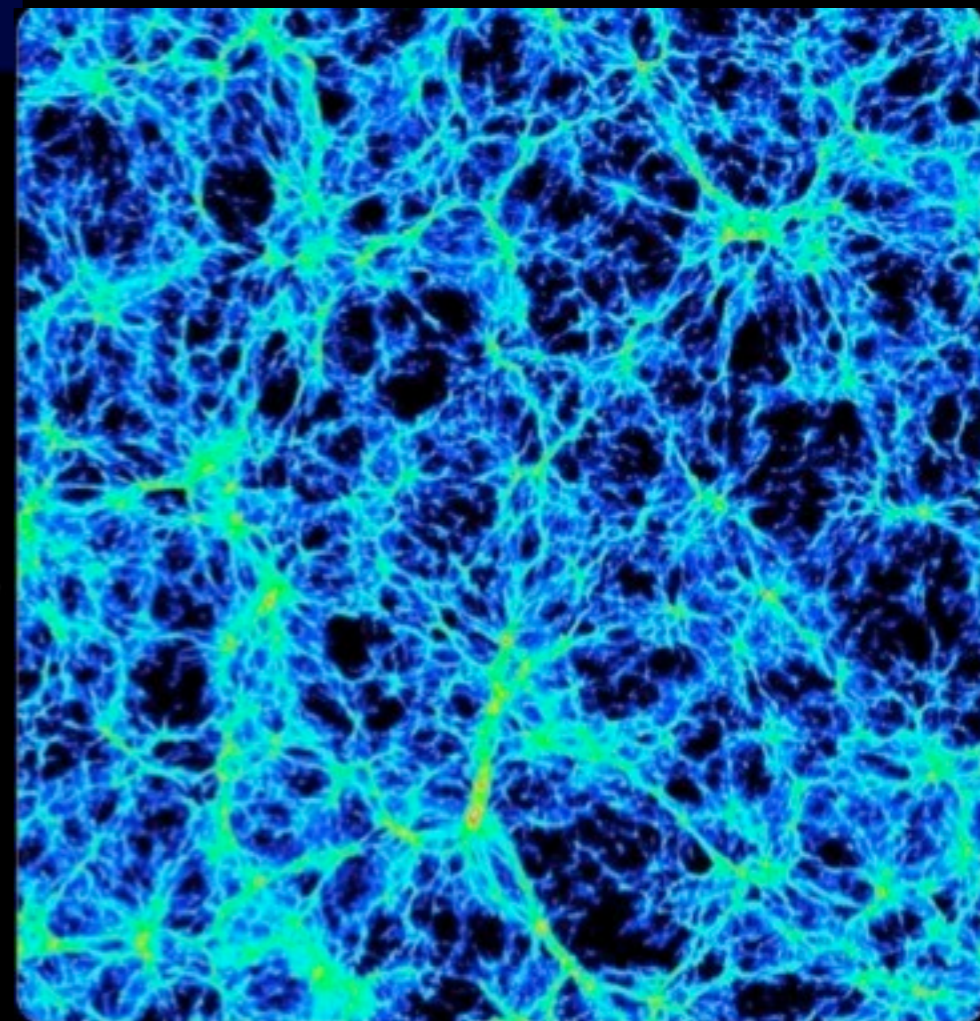
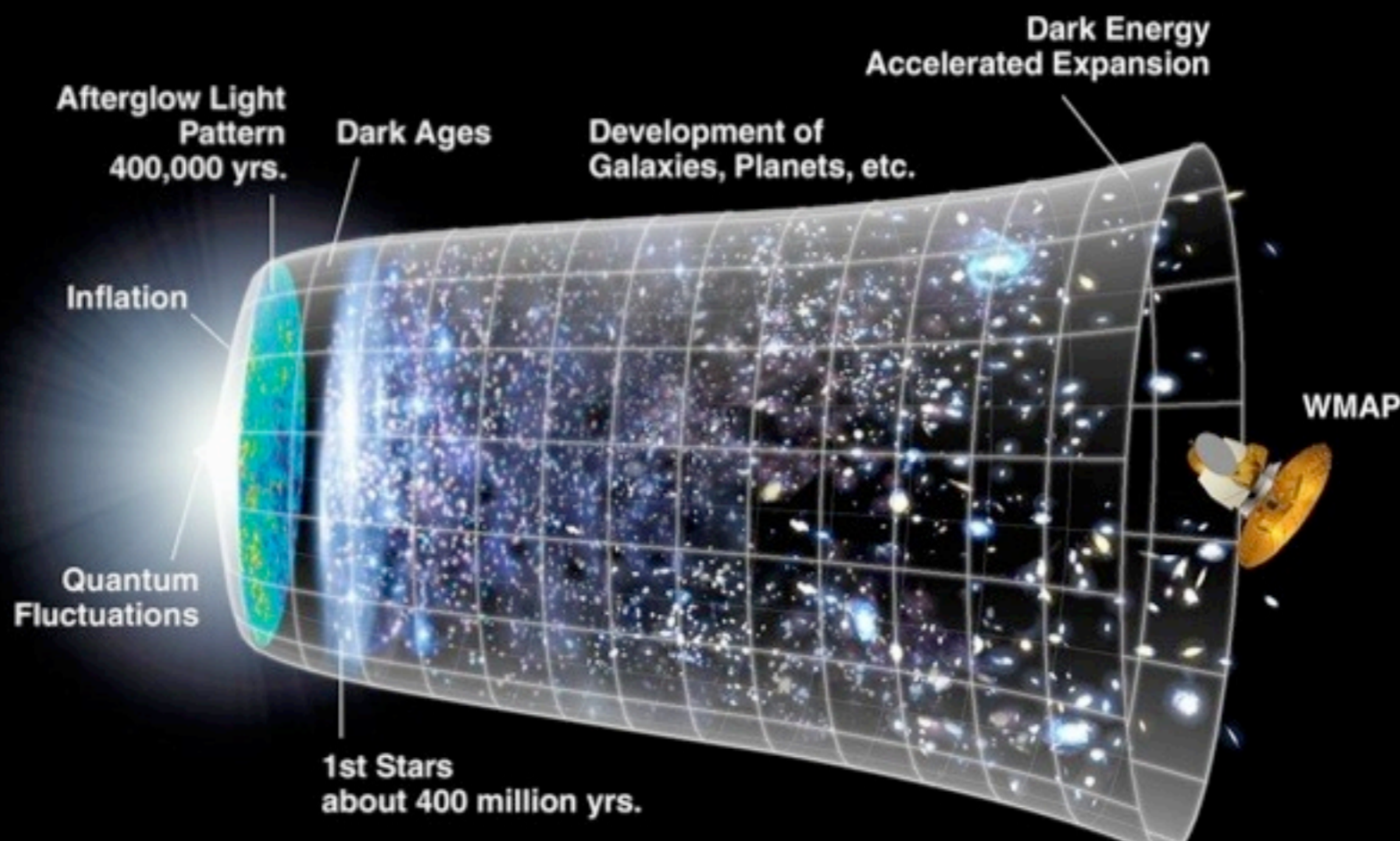
udaljenost



**Svemir se širi;
mislilo se kako to
širenje mora
usporavati zbog
djelovanja
gravitacije, ali...**

Nove kozmološke zagonetke

Λ CDM: The 6-parameter Theory of the Universe



Moderni model širenja Svemira objašnjava sva promatranja, ali mora postulirati tamnu tvar i tamnu energiju (no moguće je da opis gravitacije nije točan)

Moderna kozmološka mjerenja

- Supernove (SNe): lako im je odrediti udaljenost
- Gravitacijske leće: raspodjela tvari (obične i tamne)
- Prostorna raspodjela galaksija (statistika)
- Kozmičko pozadinsko zračenje

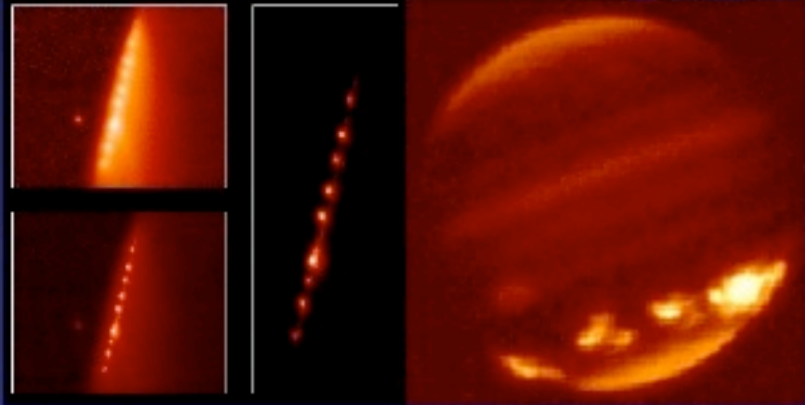
Za precizna mjerenja širenja Svemira i stvaranja strukture potrebni su uzorci od nekoliko milijardi galaksija, te stalna promatranja da bi se otkrile SNe

Slična su promatranja potrebna za otkrivanje opasnih asteroida, te za mnogo drugih grana astrofizike (npr. proučavanje Mliječnog Puta, kvazara, itd): potreba za “novim i boljim SDSSom”: razlog za potrošiti milijardu \$

Potruga za opasnim asteroidima

Vjerojatnost udara asteroida u Zemlju nije zanemariva

NASA ima mandat od Kongresa SAD za pronalaženje 90% asteroida većih od 140m do 2020



Tunguska
(1908)

Shoemaker-Levy 9
(1994)



Krater Barringer u Arizoni: udarac 40m objekta prije 50,000 god.

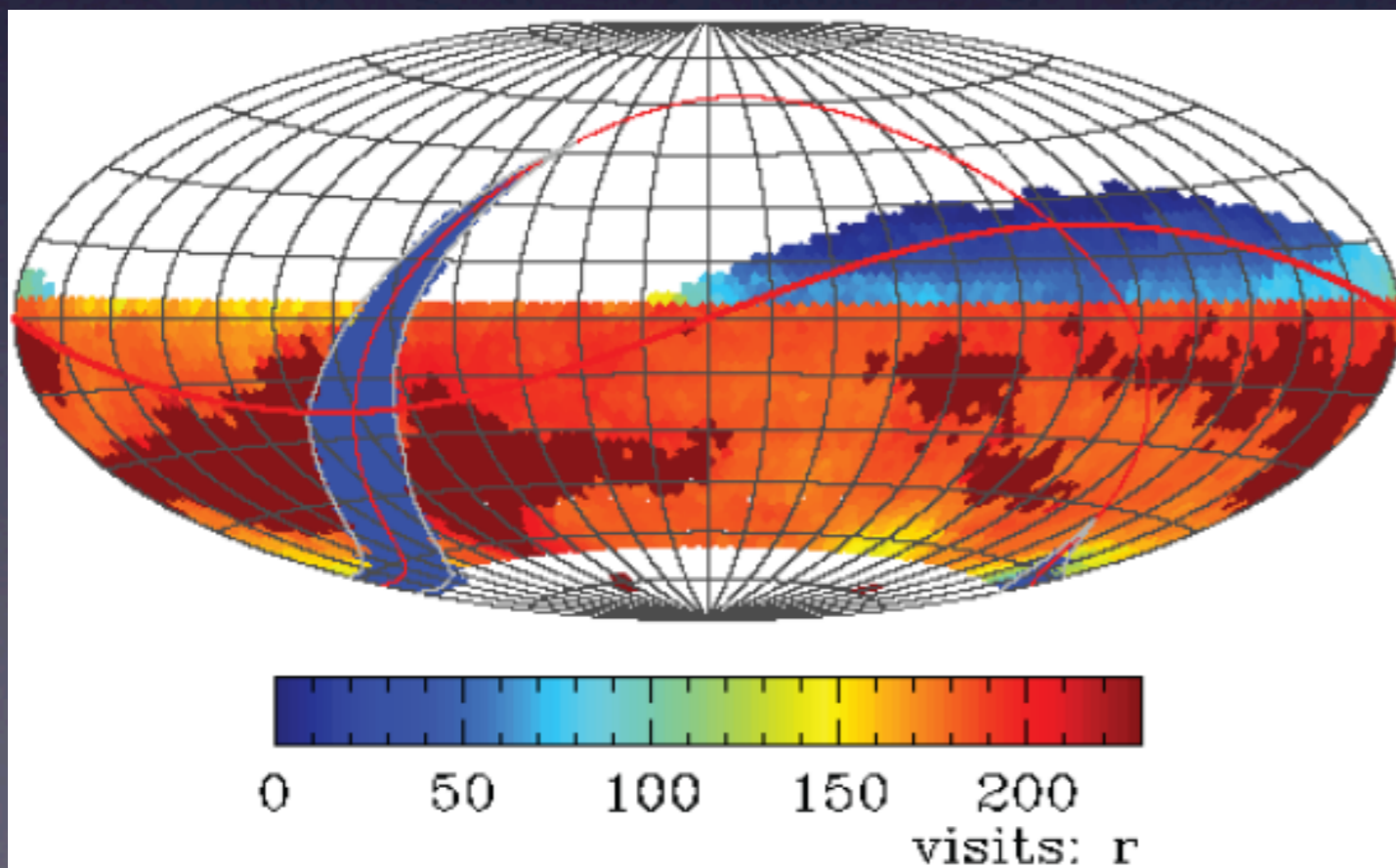


Osnovni koncepti za LSST

- Zrcalo velikog promjera (barem 6m) da bi se moglo koristiti kratke ekspozicije (30 s)
- Agilan teleskop (5 sekundi između eksp.)
- Veliko vidno polje da bi se moglo “pokriti” cijelo nebo sa malim brojem slika ($\sim 1,000$)
- Male optičke deformacije
- Kamera sa 3000 Mpix (zbog rezolucije)
- Sofisticirani software za obradu podataka (20,000 GB/dan, oko 20 milijardi objekata)

Osnovna ideja LSSTa: uniformni pregled neba

- 90% vremena će biti utrošeno na uniformni pregled neba: svake tri noći cijeli dostupni dio neba će biti snimljen dva puta
- nakon 10 godina, pola cijelog neba će biti mjereno oko 1000 puta (u 6 filtera)
- biti će oko 500 milijuna slika od 16 Mpix, sa mjerenjima za 20 milijardi objekata: preko 100,000,000 GB podataka (>100 PB)



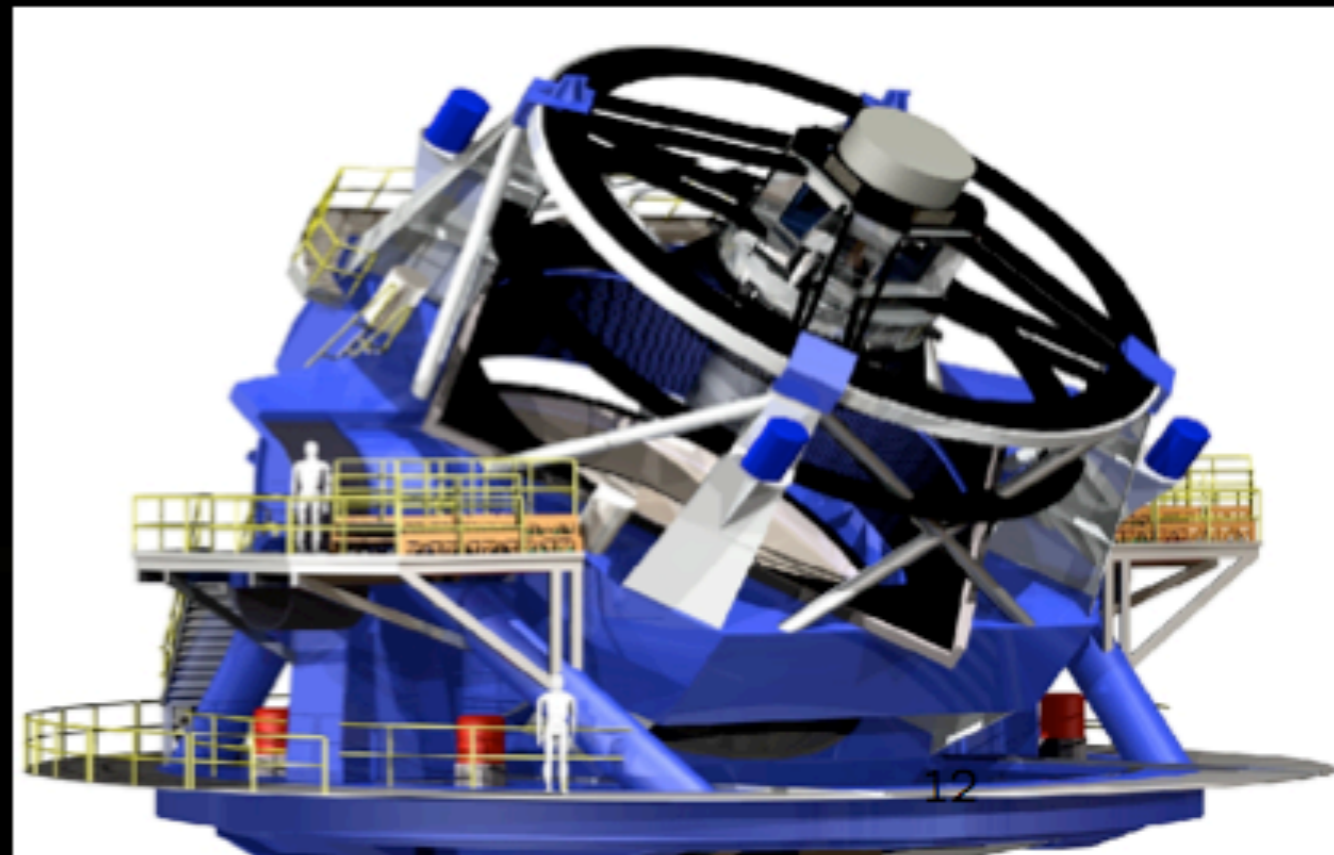
Simulacija 10 godina rada LSSTa: broj promatranja u jednom od filtera (r)

Usporedba SDSS-LSST



SDSS: one US Library of Congress worth of data

LSST: one SDSS per night, or all the words ever printed!

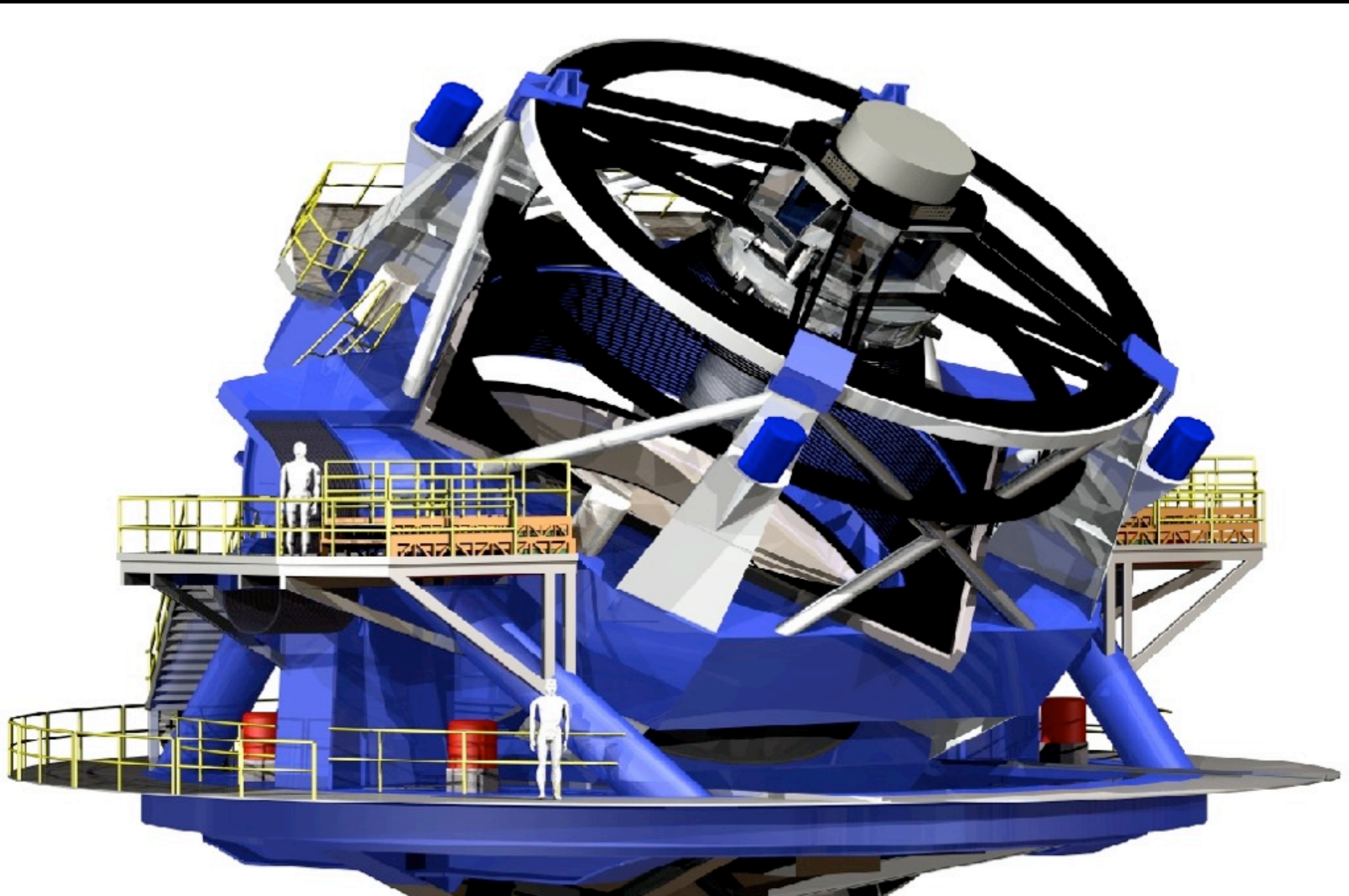


LSST Observatorij

LSST sistem:
Teleskop
Kamera
Softver



LSST Teleskop



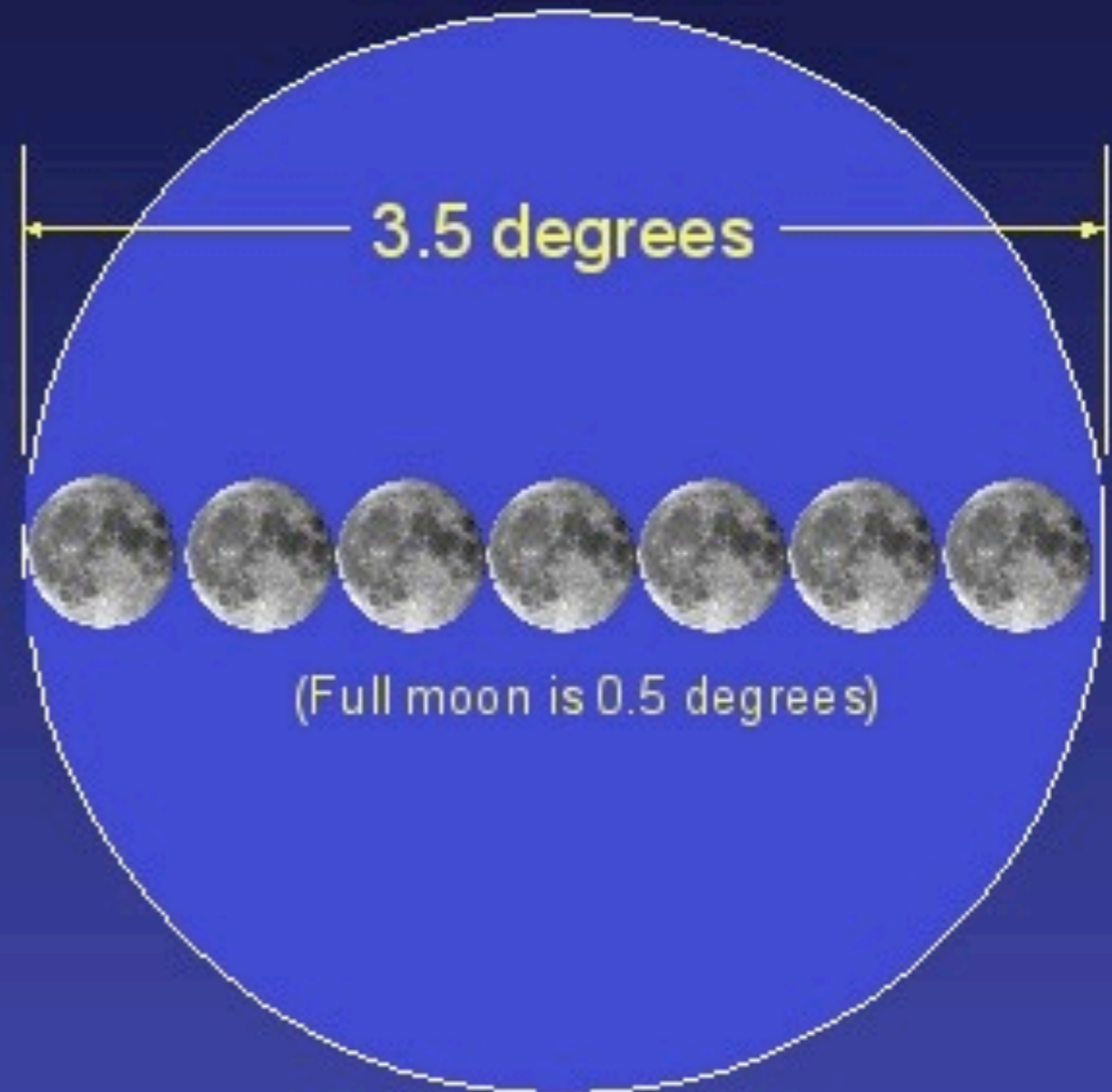
Usporedba vidnog polja Gemini-LSST

Primary Mirror Diameter

Field of View



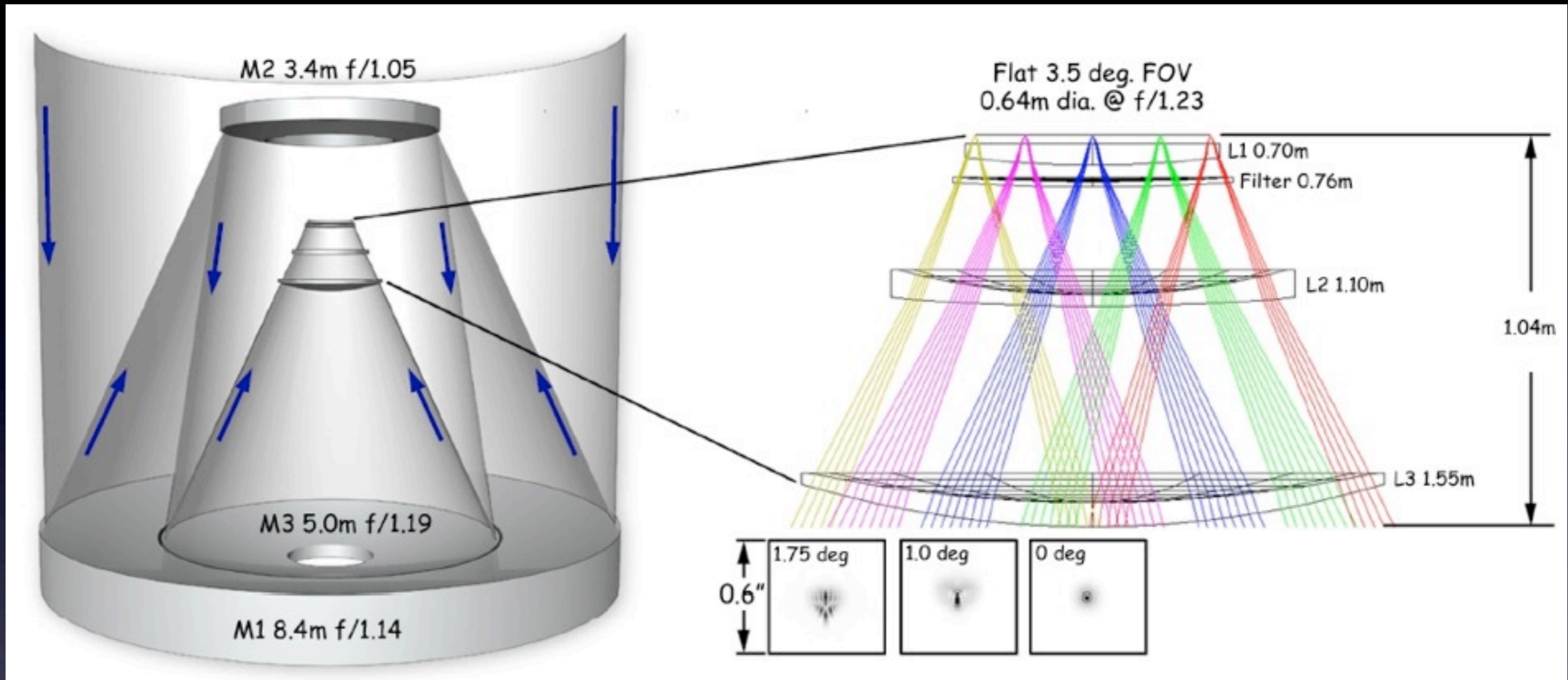
Gemini South Telescope



LSST



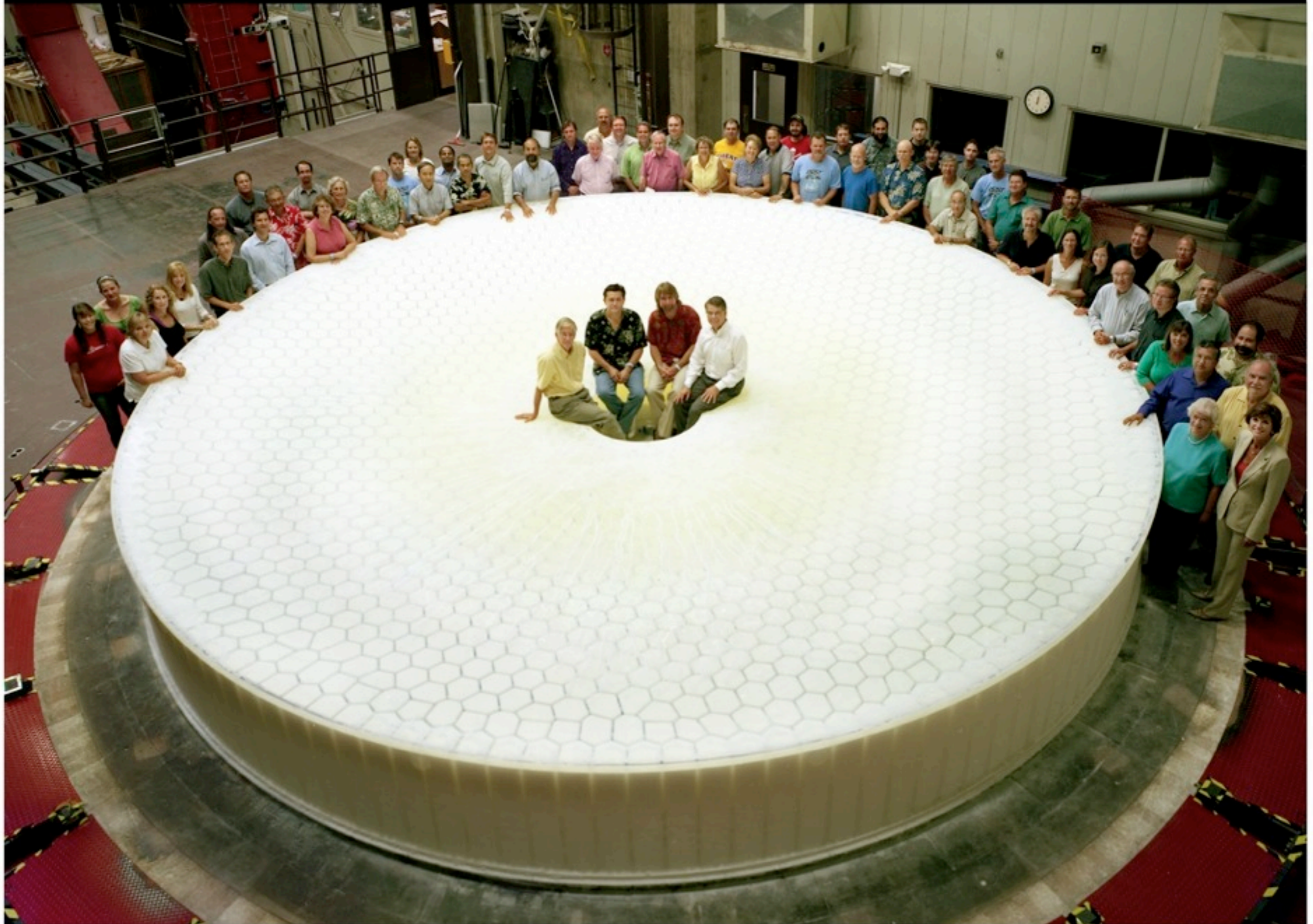
Optički dizajn za LSST teleskop



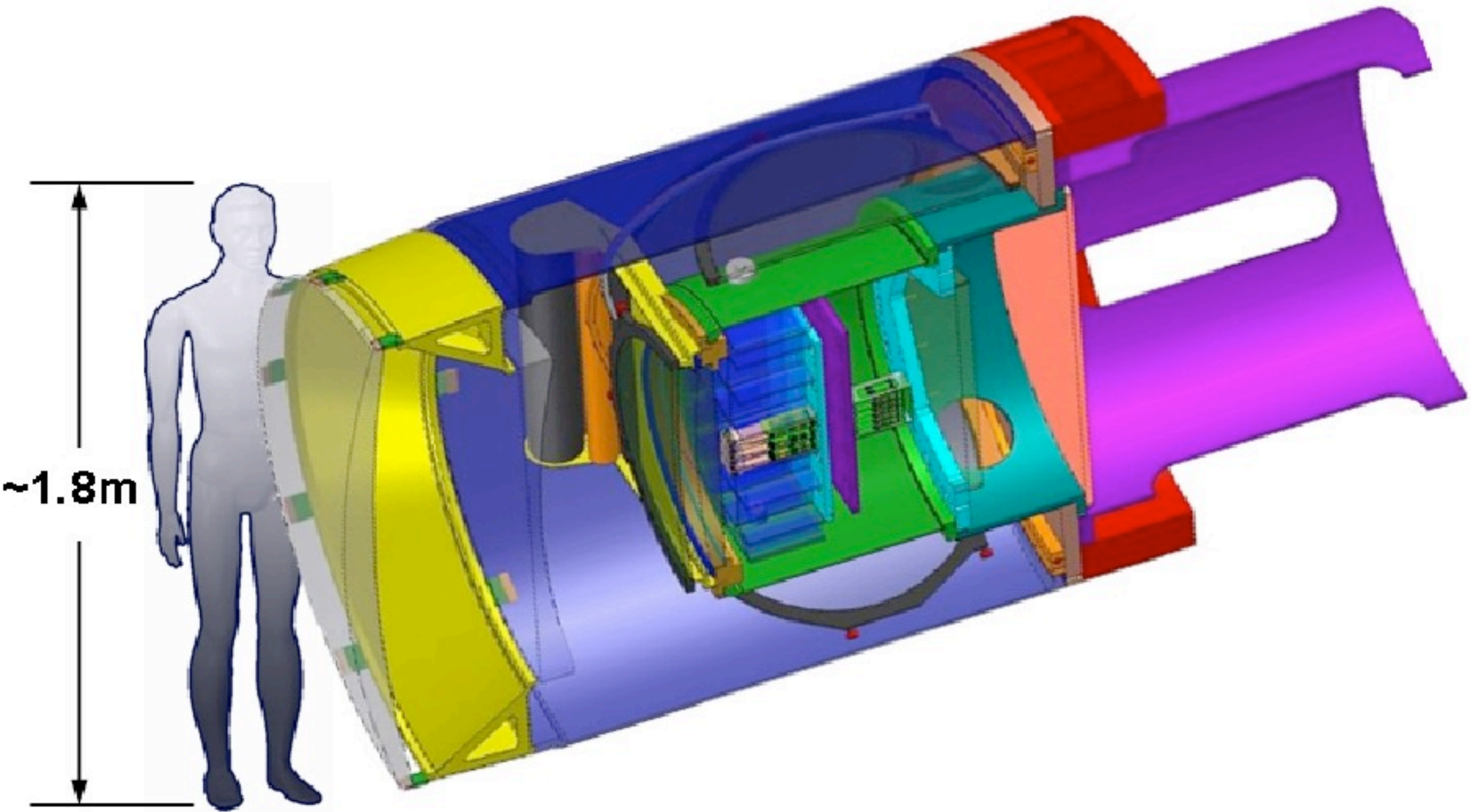
“Klasični” teleskopi: dva zrcala (ograničeno vidno polje)
LSST: tri zrcala koja daju veliko vidno polje sa malim deformacijama slike (Paul-Baker sistem)



Large Synoptic Survey Telescope



LSST kamera



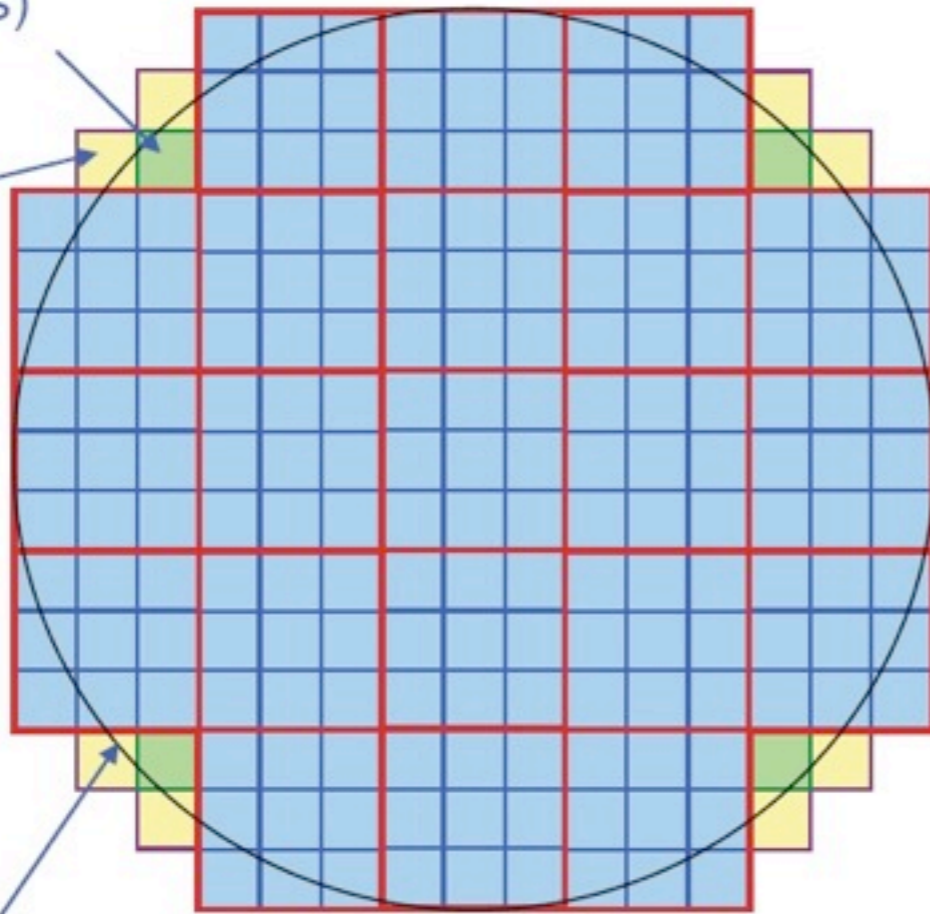
Najveća astronomska kamera: 2800 kg, 3200 Megapixela

LSST kamera

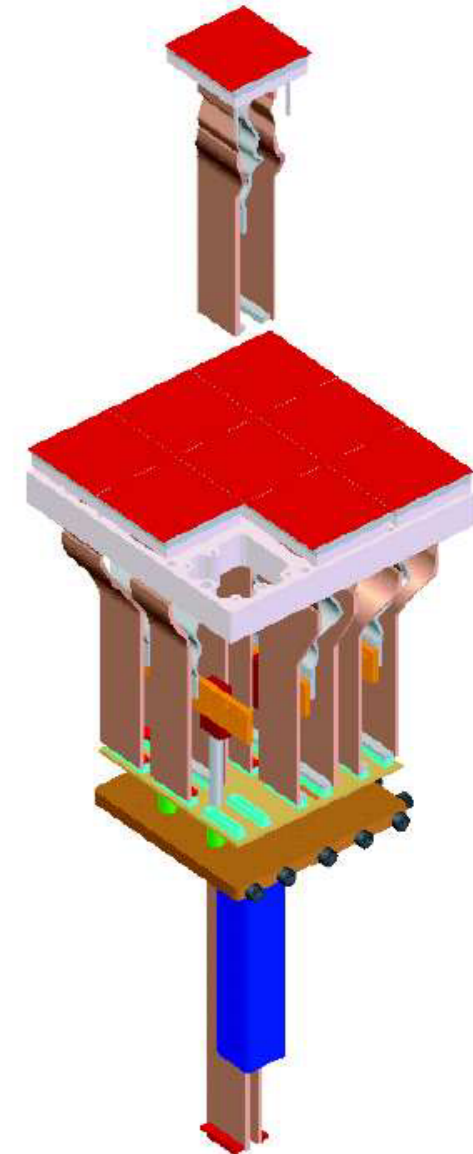
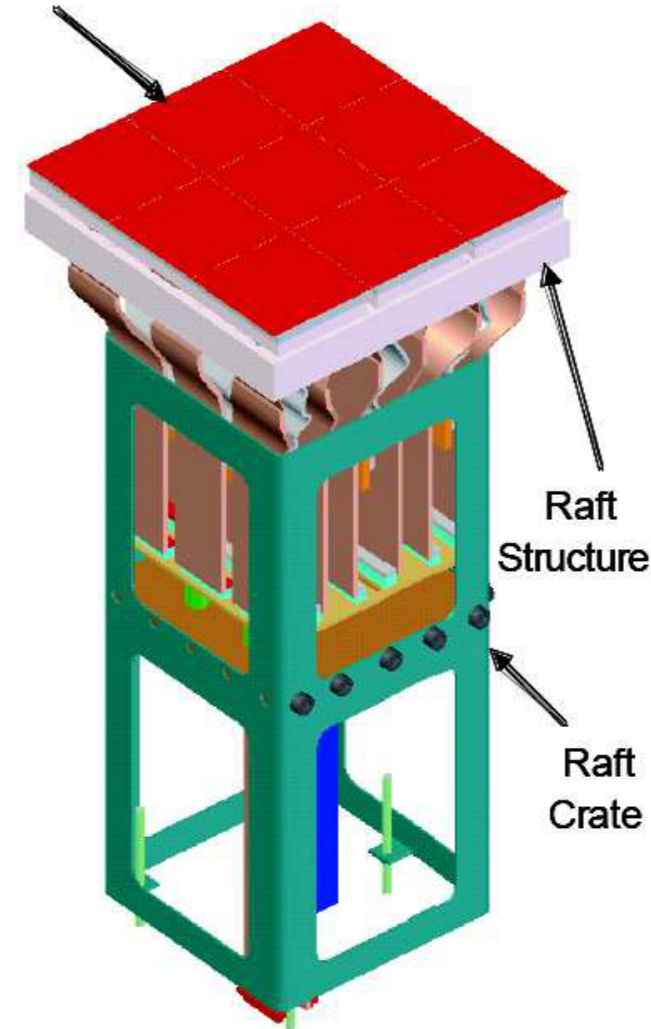
Wavefront Sensors
(4 locations)

Guide Sensors
(8 locations)

3.5 degree Field
of View (634 mm diameter)



Imaging Sensors

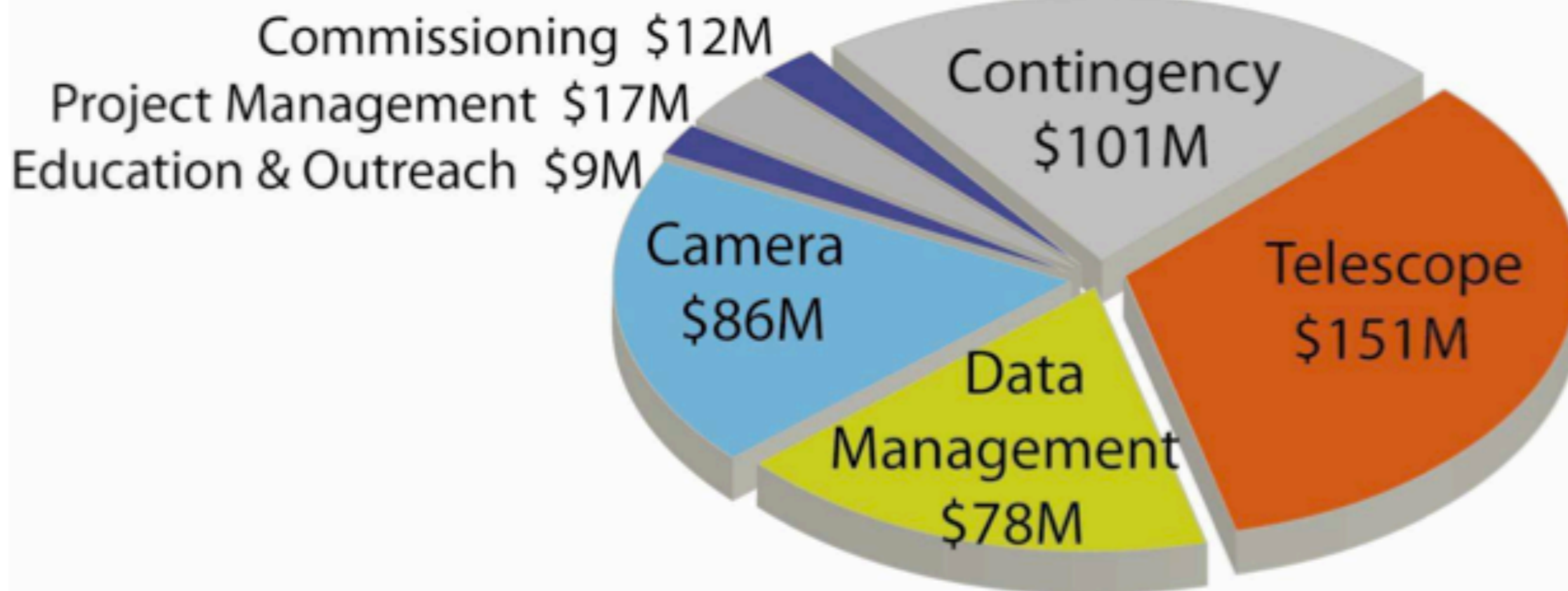


Modularni dizajn: 3200 Megapix = 189 x 16 Megapix CCDs
9 CCDova imaju zajedničku elektroniku: raft = kamera
Raft sa problemima se može zamijeniti tokom dana

Kako potrošiti milijardu dolara?

Pola na konstrukciju, pola na 10 godina rada.

Total Project Cost: 455M 2009USD

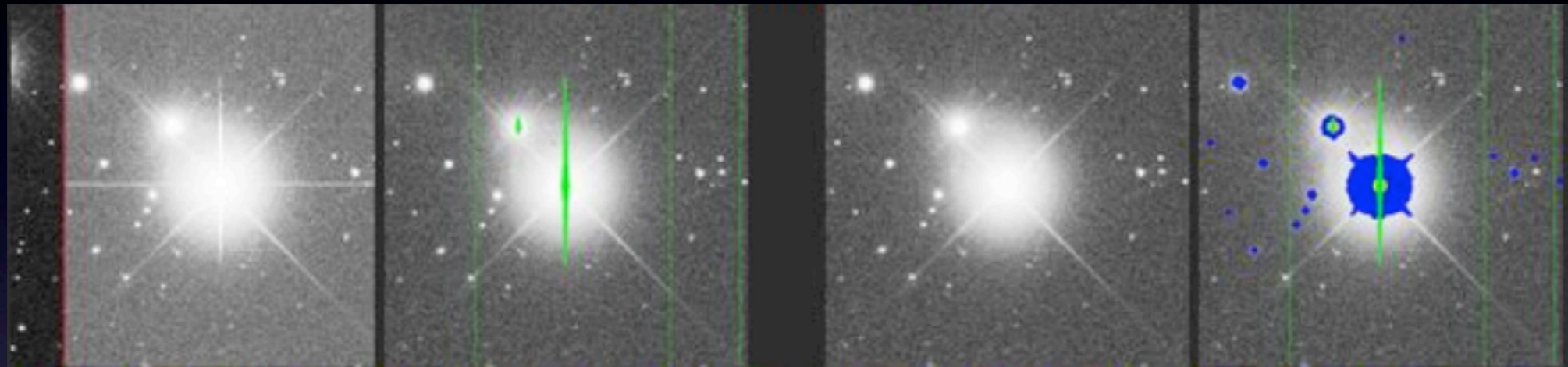


LSST Construction Component Cost

Početak rada: oko 2018 (zeleno svjetlo: kolovoz 2010)

Što LSST softver treba raditi?

- **Obrada slika: korekcije za instrumentalne efekte, detekcija izvora, mjerenje parametara**



A raw data frame.

The difference in bias levels from the two amplifiers is visible.

Bias-corrected frame

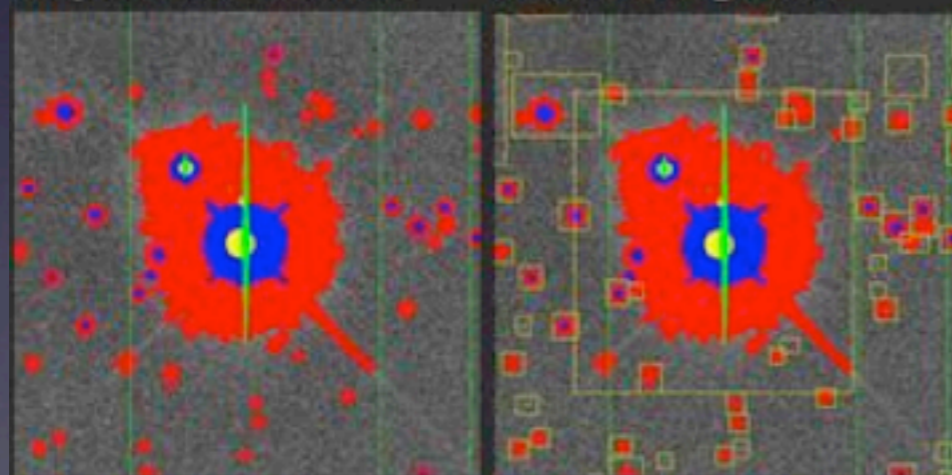
with saturated pixels, bad columns, and cosmic rays masked in green.

Frame corrected

for saturated pixels, bad columns, and cosmic rays.

Bright object

detections marked in blue.

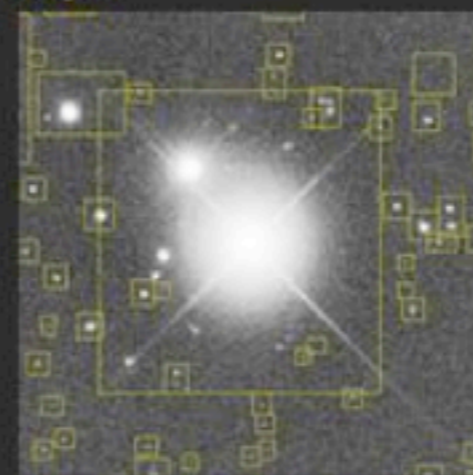


Faint object

detections marked in red.

Measured objects,

masked and enclosed in boxes. Small empty boxes are objects detected only in some other band.



Measured objects

in the data frame.



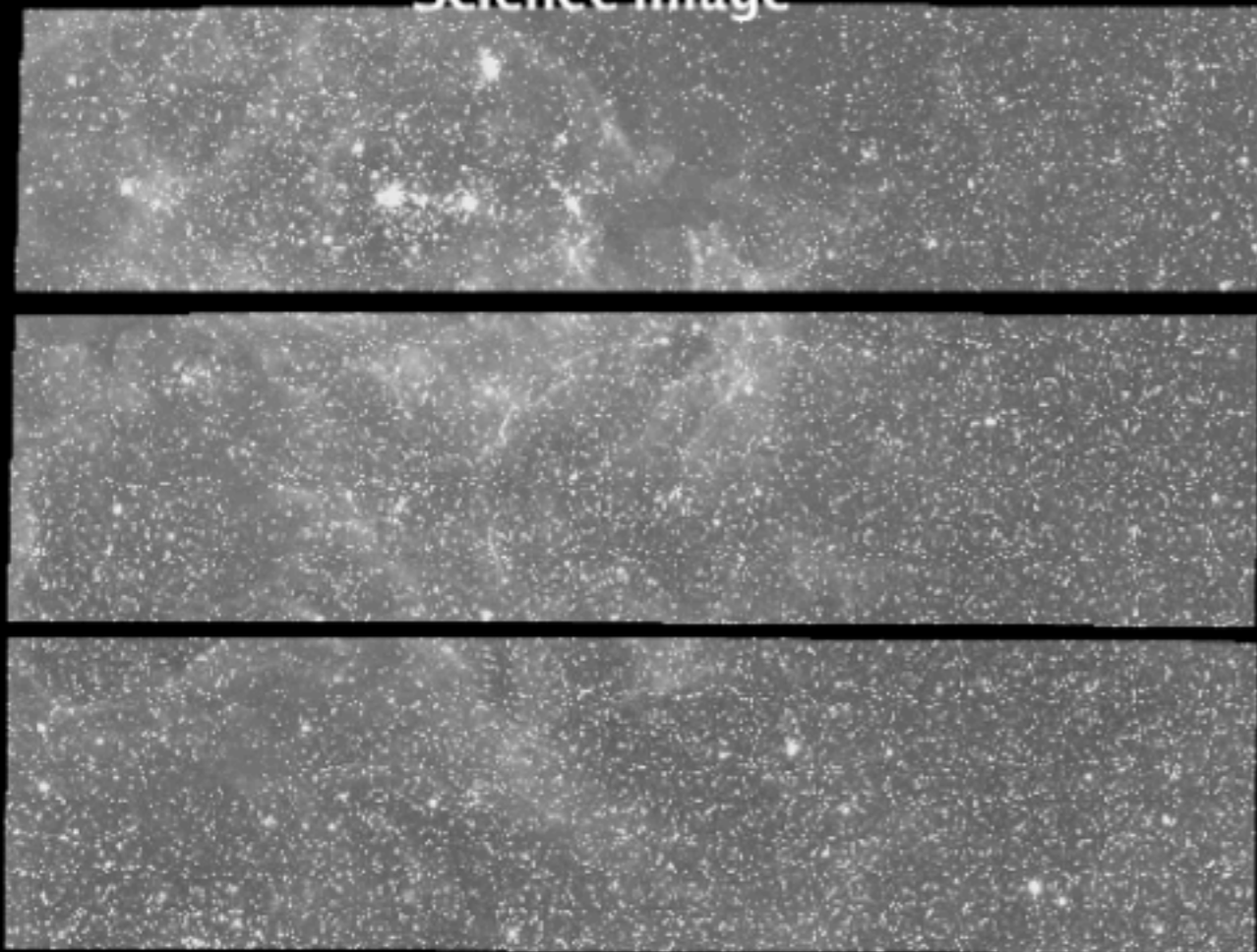
Reconstructed

image using postage stamps of individual objects and sky background from binned image.

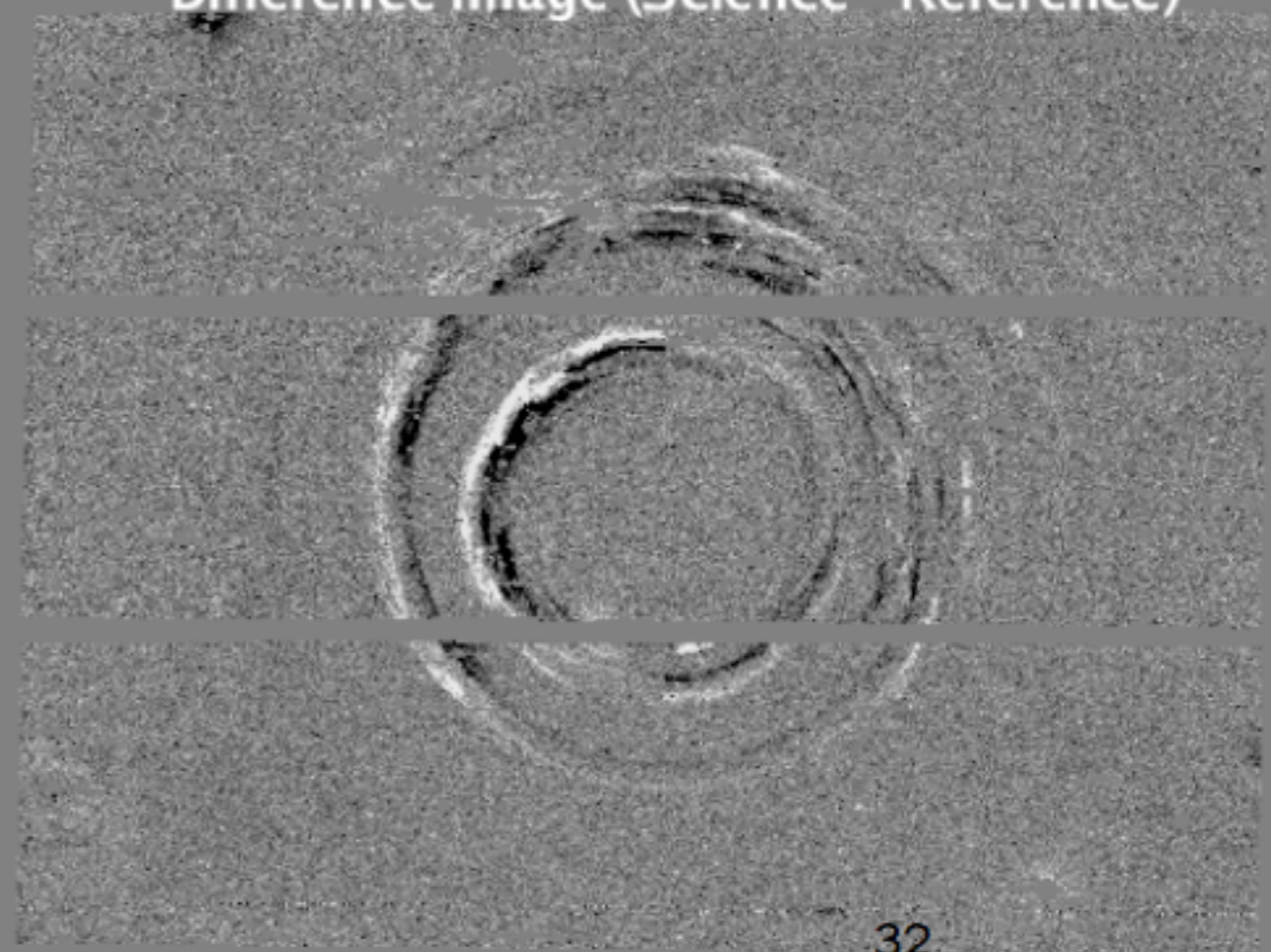
Što LSST softver treba raditi?

- **Oduzimanje slika:** dvije 3200 Mpix slike se moraju oduzeti svakih 40 sec, pronaći i izmjeriti svi objekti, te na kraju svi podaci staviti na web (oko 100,000 tranzienata)

Science Image



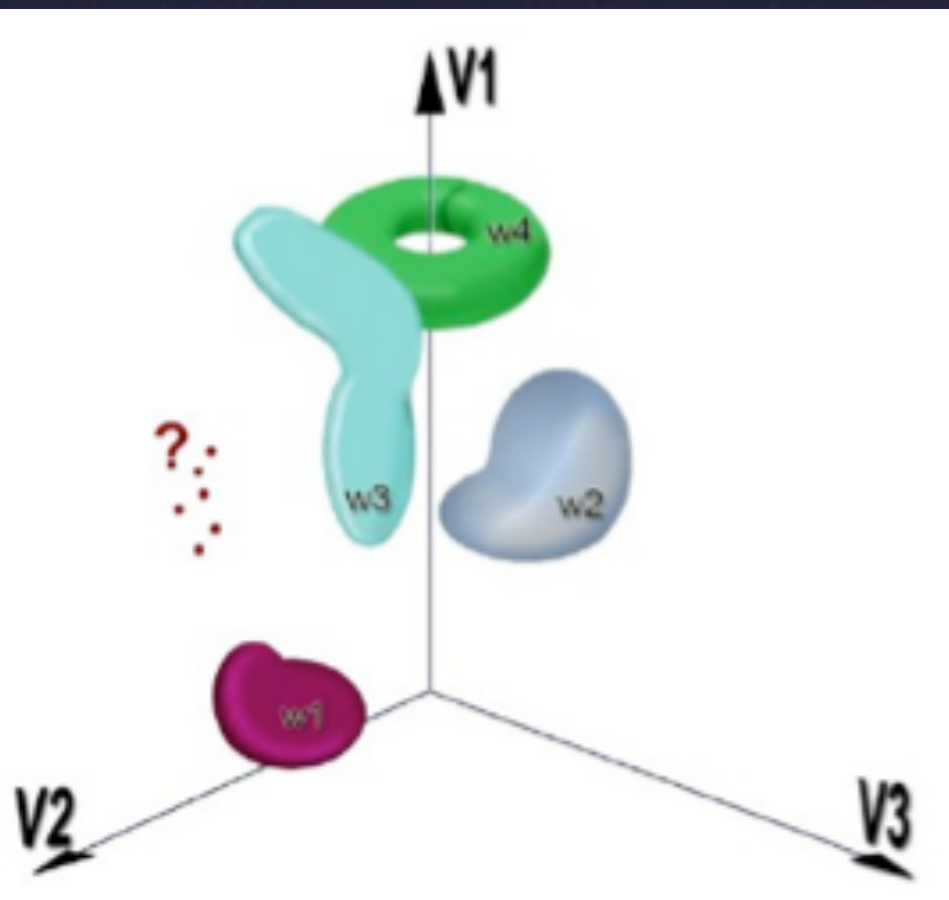
Difference Image (Science - Reference)



Što LSST softver treba raditi?

- **Baza podataka i alati za analiziranje:** za svaki od oko 20 milijardi objekata biti će oko 1000 mjerenja (svako mjerenje ima nekoliko desetaka mjerenih parametara)

Data mining and knowledge discovery



- 10,000-D prostor s 20 milijardi točaka
 - Karakterizacija poznatih objekata
 - Klasifikacija novih populacija
 - Otkrivanje “neobičnih” objekata
- Clustering, classification, outliers

LSST Software

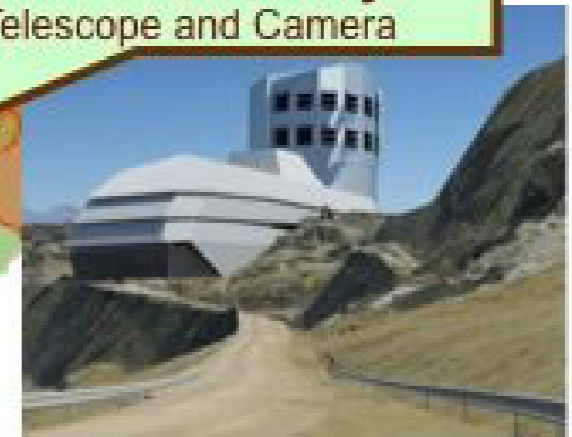
TBD Site
Science Centers
Data Quality Analysis

Archive Site
Archive Center
Nightly Reprocessing
Data Release Production
Long-term Storage (copy 2)
Data Access Center
Data Access and User
Services
120 – 330 TFLOPS
35 – 250 PB

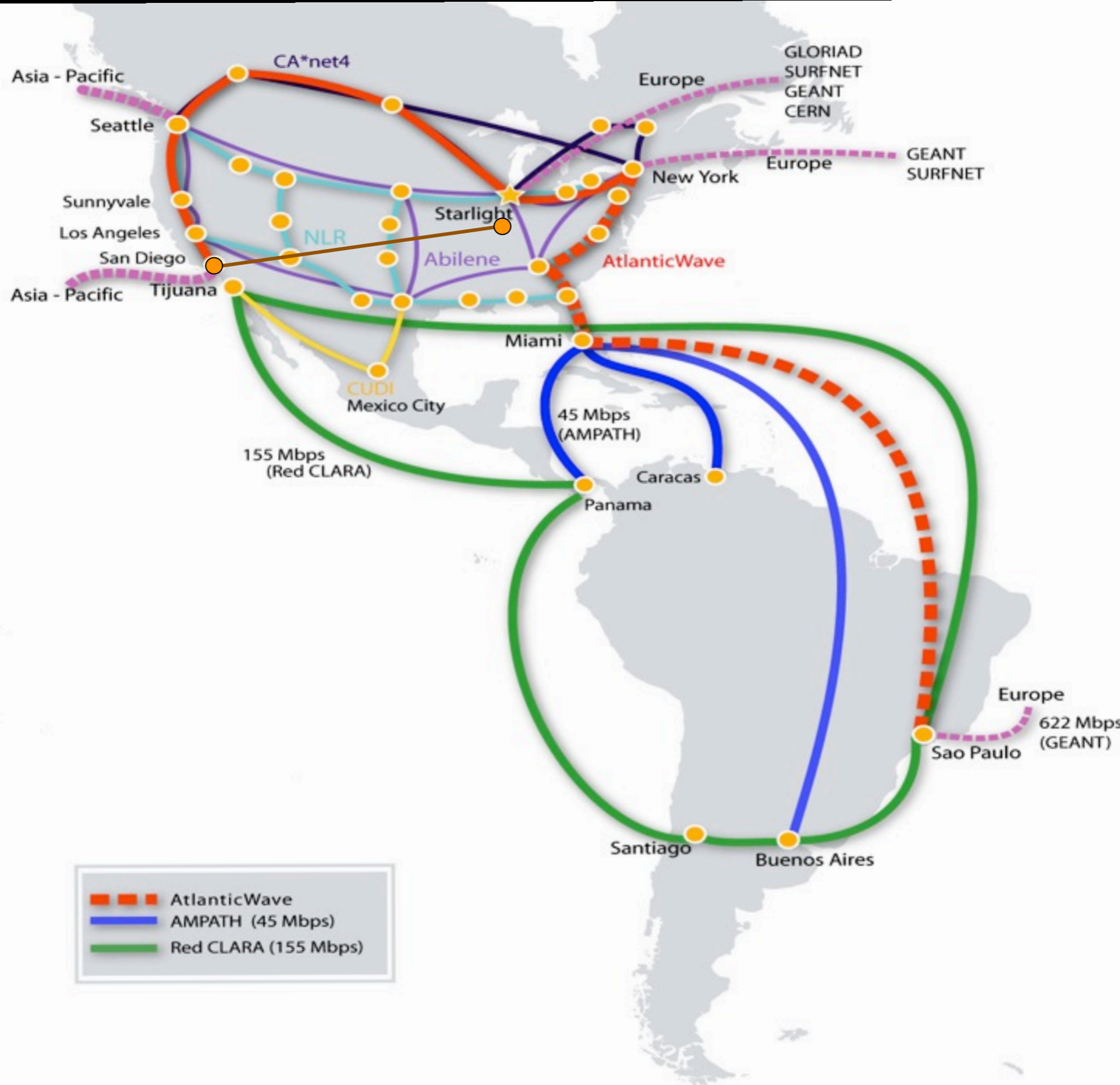
Headquarters Site
Headquarters Facility
Observatory Management
Science Operations
Education and Public Outreach

Base Site
Base Facility
Data Access Center
Alert Production
Long-term storage (copy 1)
60 TFLOPS
35 – 250 PB

Summit Site
Summit Facility
Telescope and Camera



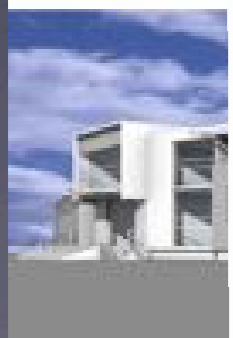
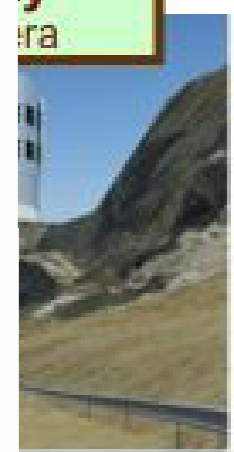
LSST Software



He



te
y
era



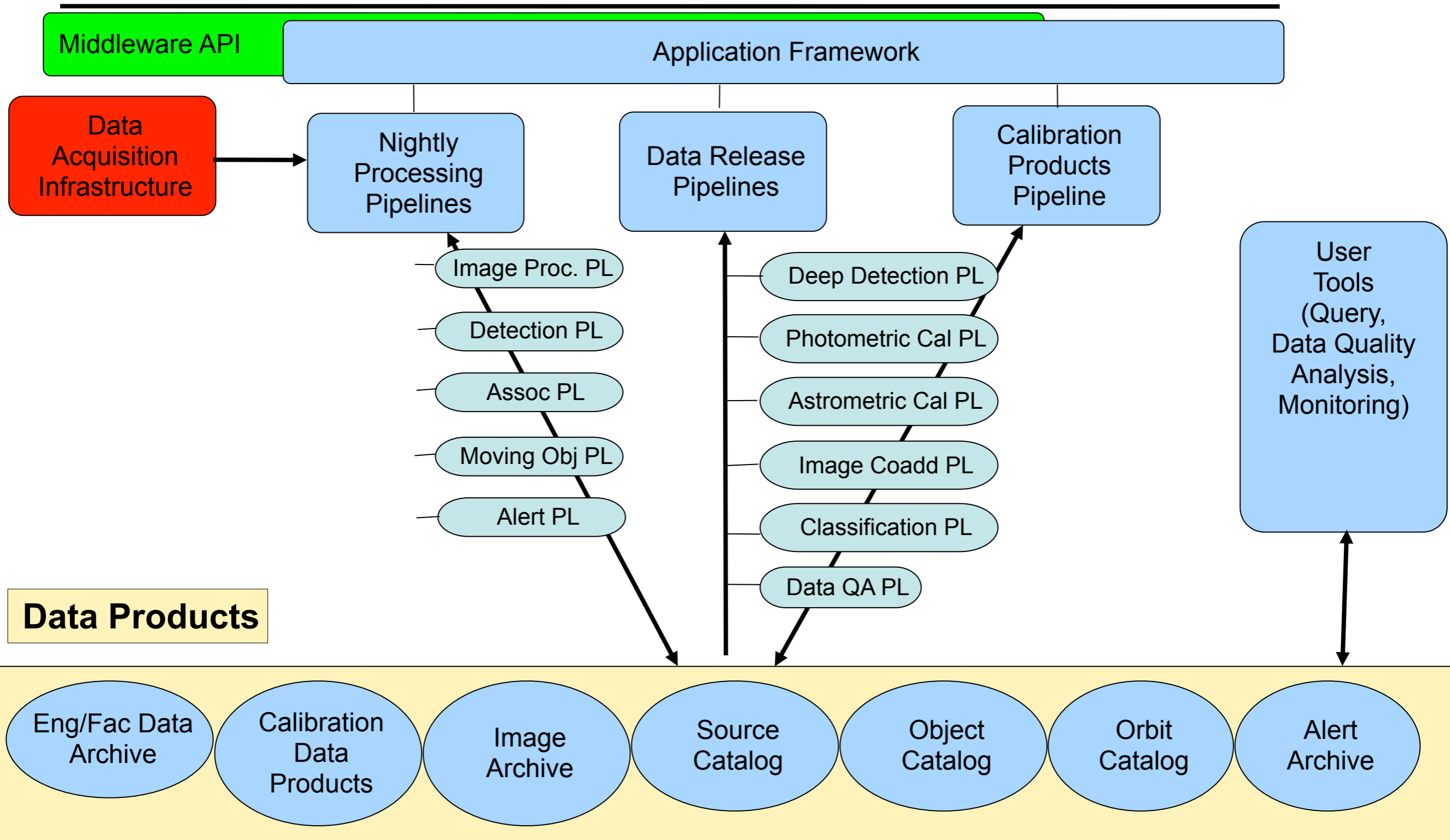
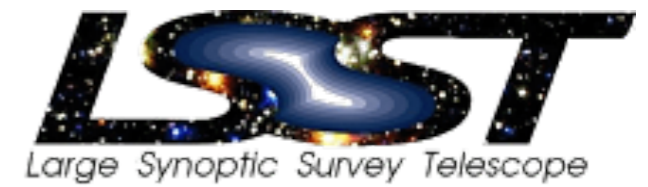
LSST Softver - problemi:

- 20 TB podataka koje kontinuirano treba obrađivati svaki dan
- 20 milijardi objekata sa 1000 mjerenja tokom 10 godina
- Potreba za novim pristupom obradi i analizi podataka:
software, software,
software!



Application Layer - pipelines process

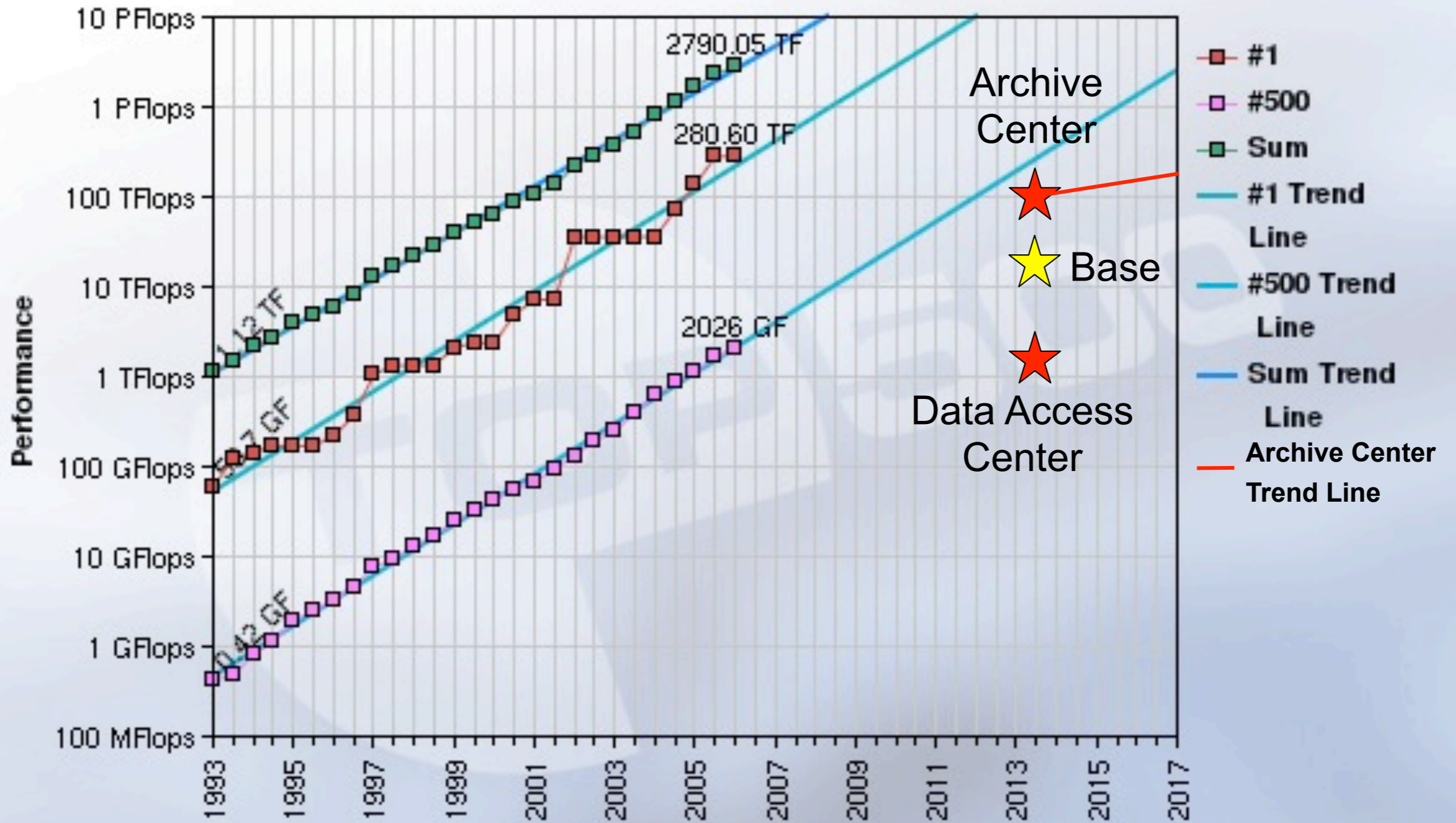
raw data to products



Usporedba LSST softvera sa drugima

- **Complexities we have to deal with in DM**
 - Very high data volumes (transfer, ingest, and especially query)
 - Advances in scale of algorithms for photometry, astrometry, PSF estimation, moving object detection, shape measurement of faint galaxies
 - Provenance recording and reprocessing
 - Evolution of algorithms and technology
- **Complexities we DON'T have to deal with in DM**
 - Tens of thousands of simultaneous users (e.g. online stores)
 - Fusion of remote sensing data from many sources (e.g. earthquake prediction systems)
 - Millisecond or faster time constraints (e.g. flight control systems)
 - Very deeply nested multi-level transactions (e.g. banking OLTP systems)
 - Severe operating environment-driven hardware limitations (e.g. space-borne instruments)
 - Processing that is highly coupled across entire data set with large amount of inter-process communication (e.g. geophysics 3D Kirchhoff migration)

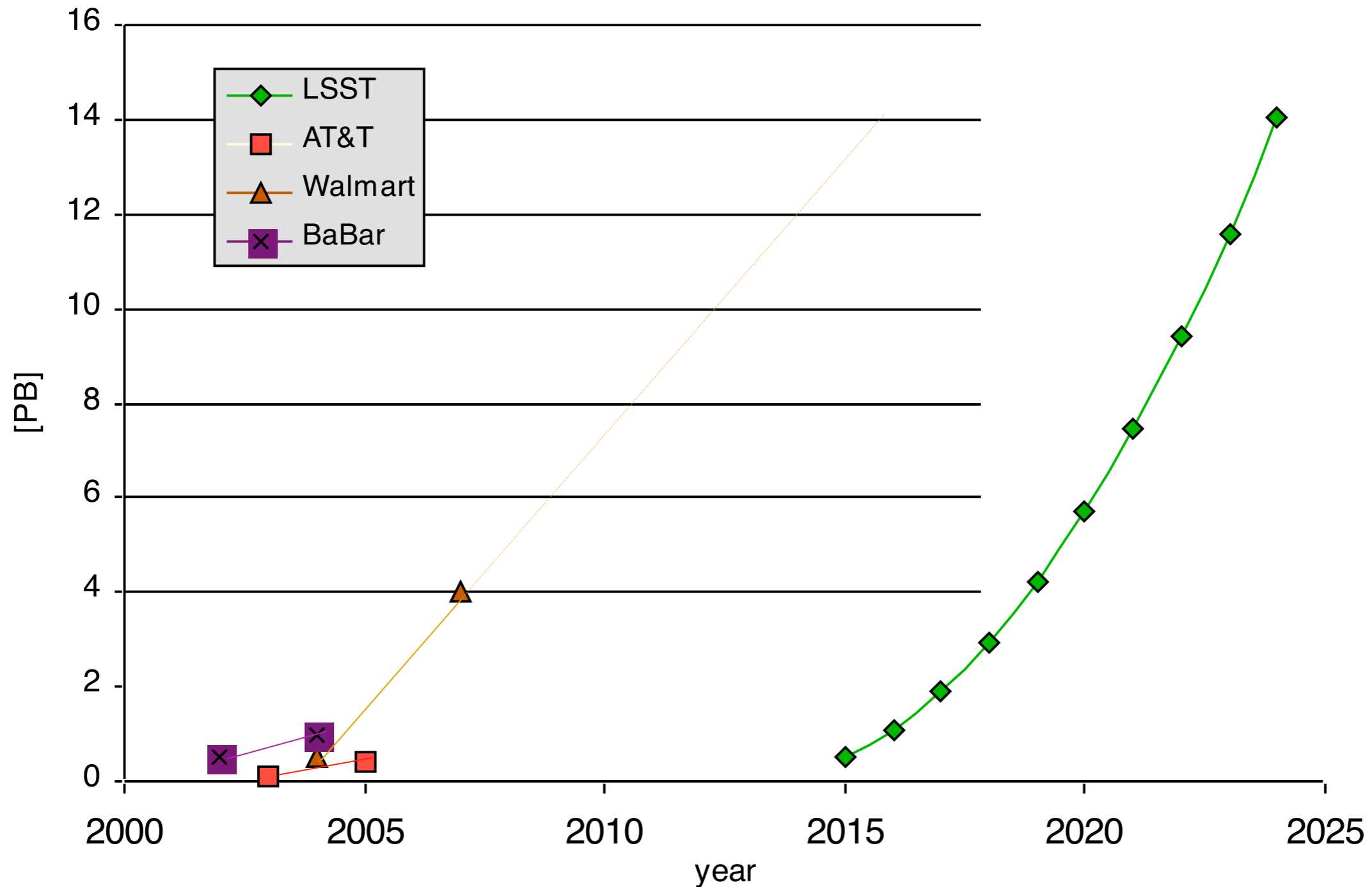
Computing needs show moderate growth



28/06/2006

<http://www.top500.org/>

Large RDBMS Systems - Data Volumes



** All numbers based on publicly available data*

LSST Softver: ukratko

- Vjerojatno najrizičniji dio sistema
- Mora biti brz, točan i robustan (20 TB/dan)
- Oko 5-10 milijuna linija novoga koda (~1000 FTE years)
- Uglavnom C++/python (fleksibilnost & devel brzina)
- Multi-D multi-attribute db od >100 PB
- Kolaboracija astronoma, fizičara i profesionalnih programera
- Suradnja sa Hrvatskom (GPU klaster grupe Dejana Vinkovića u Splitu)

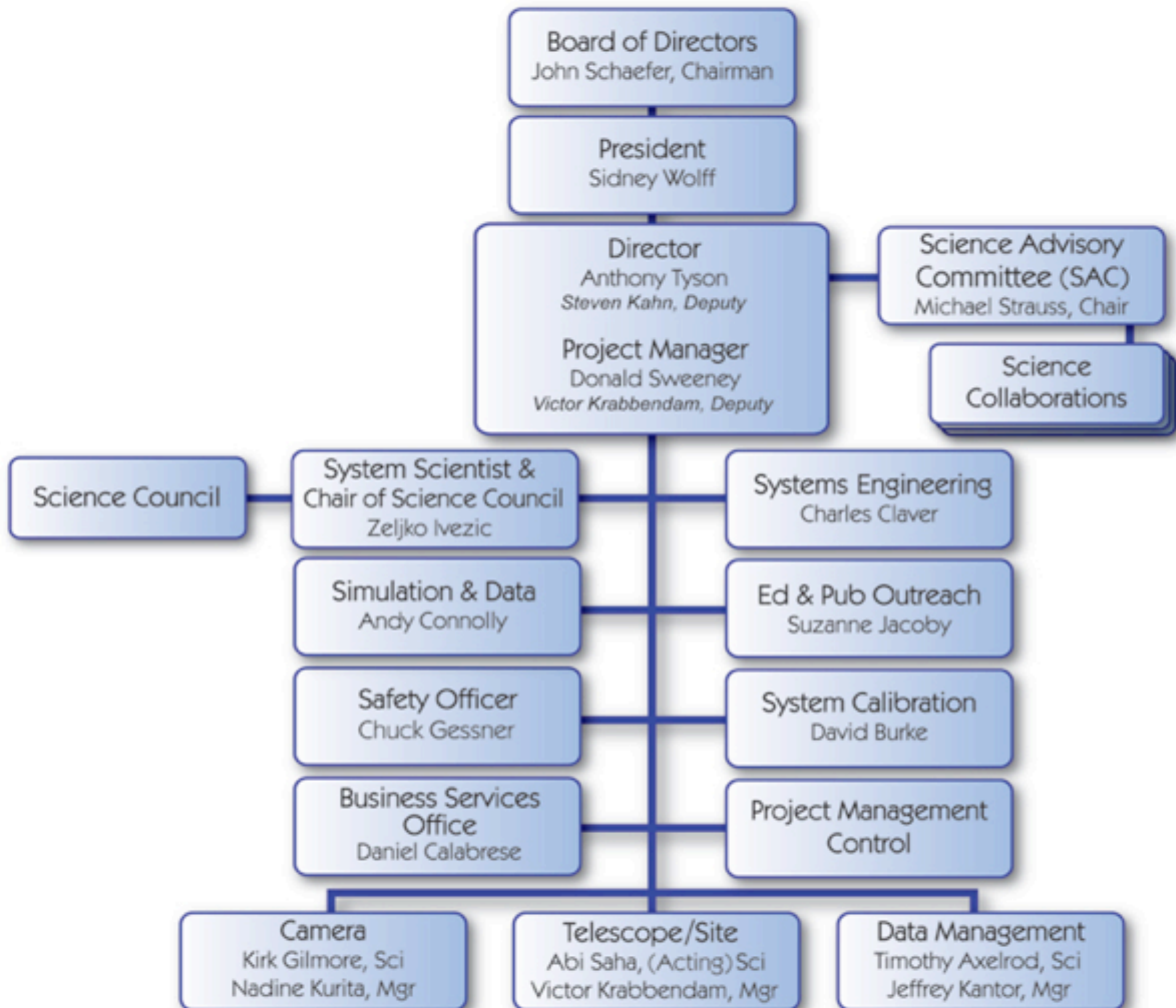
LSST All Hands Meeting at NCSA

LSST kolaboracijski sastanak 2009



Project Management Org Chart

SA



List

[cameracontrol](#)

Description

discussion group for camera control

[LSST-agn](#)

LSST AGN

[LSST-calibration](#)

LSST system calibrations.

[LSST-Calypso](#)

Calypso Group

[LSST-camera](#)

LSST Camera

[LSST-cameramgrs](#)

LSST cameramgrs

[LSST-camerasteer](#)

LSST camerasteer

[LSST-ChangeControlBoard](#)

[no description available]

[LSST-cosmology](#)

Cosmology related work

[LSST-data](#)

LSST Data Management

[LSST-datascience](#)

LSST-DataScience

[LSST-DeepDrill](#)

[no description available]

[Lsst-dm-db](#)

[no description available]

[Lsst-dmscientist](#)

[no description available]

[Lsst-epo-oab](#)

EPO Outreach Advisory Board

[LSST-FriendsOnly](#)

Auto-enrollment

[LSST-galaxies](#)

LSST Galaxies

[Lsst-galaxy-clusters](#)

the mailing list for the galaxy clusters working group

[Lsst-imagesim](#)

LSST Image Simulation

[LSST-ImSimData](#)

ImSim Data Management for LSST.

[LSST-lss](#)

LSST LSS

[LSST-ManagementMeetingAttendee](#)

[no description available]

[LSST-MembershipCommittee](#)

[no description available]

[LSST-milkyway](#)

LSST Milkyway

[LSST-Operations-Infrastructure](#)

Computing Infrastructure for Operations (not DCs)

[Lsst-opsim](#)

Operations Simulator

[LSST-PDR-Team](#)

PDR Preparation Team

[LSST-ProjectAndScience](#)

Auto-enrollment

[LSST-ProjectOnly](#)

Everyone working on LSST project

[LSST-sc](#)

LSST Science Council Mailing List

[LSST-science-working-group](#)

LSST-Science-Working-Group

[LSST-ScienceOnly](#)

Everyone who is a member of a science collaboration

[LSST-solarsystem](#)

Solar system science collaboration

[LSST-stellar pops](#)

LSST Stellarpops

[LSST-stronglens](#)

LSST Strong Lensing Science Collaboration

[LSST-supernovae](#)

LSST Supernovae

[LSST-Systems-Engineering](#)

Systems Engineering Group

LSST Ukratko

- SDSS je sakupio količinu podataka (20 TB) jednaku svim knjigama u Kongresnoj Knjižnici SAD - LSST će toliko sakupiti svaku noć. Ukupna količina LSST podataka (60,000 TB) biti će veća nego sve riječi do sada tiskane u cijelom svijetu. Trebalo bi oko 3 milijuna HDTV za prikazati LSSTovu mapu neba.
- SDSS je napravio prvu digitalnu mapu neba - LSST će napraviti prvi digitalni film neba. Trebalo bi 11 mjeseci za to “pogledati”.
- LSST će popisati oko 20 milijardi zvijezda, galaksija i drugih objekata: po prvi put će biti više astronomskih objekata nego živih ljudi na Zemlji. (Želite li svoju zvijezdu?)

Glavni ciljevi LSST:

- 1) tamna energija ili pogrešna gravitacija?
- 2) opasni asteroidi
- 3) promjenljivi Svemir

Više informacija na www.lsst.org

