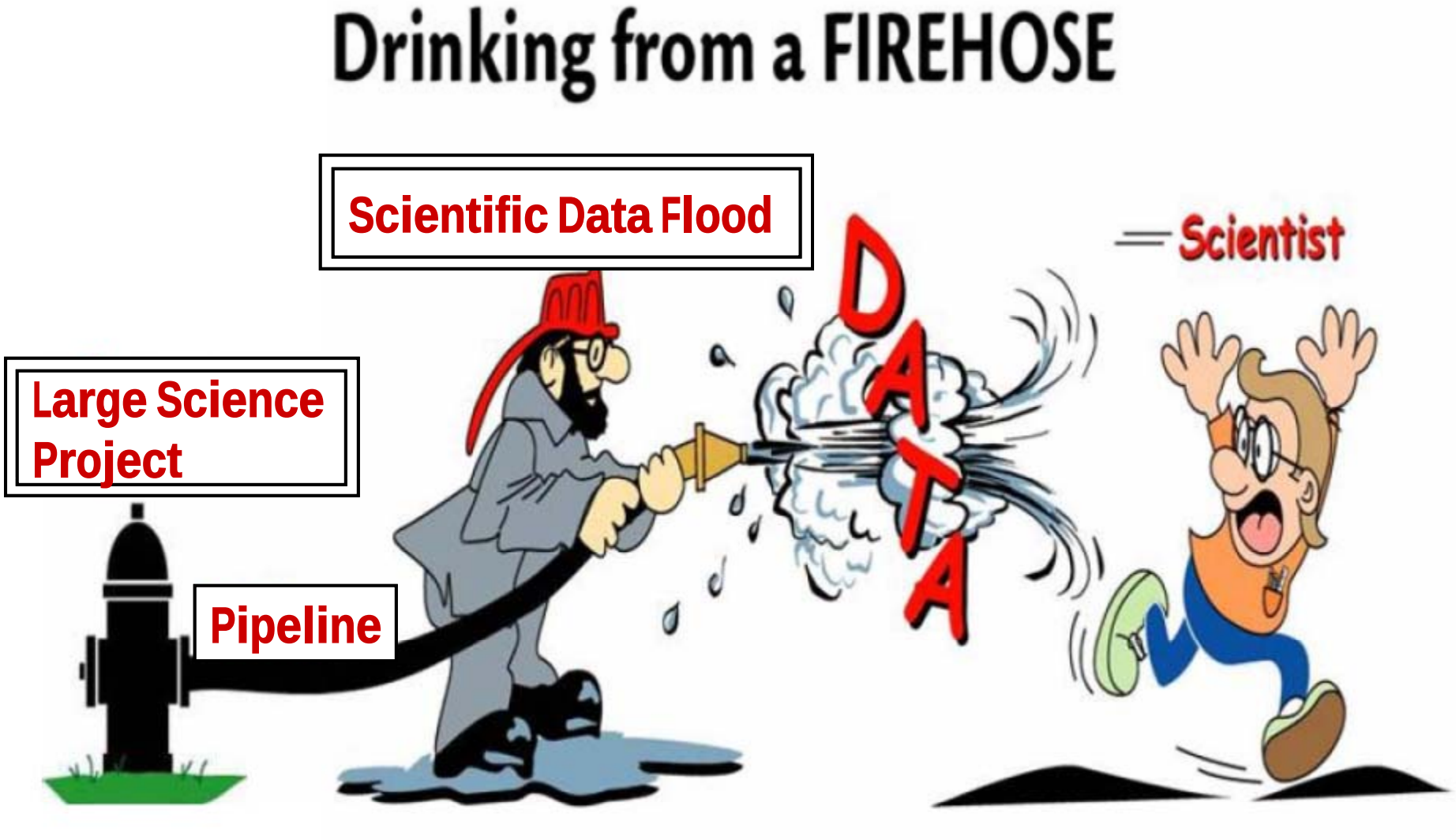
The Scientific Data Flood



The Dark Energy Survey

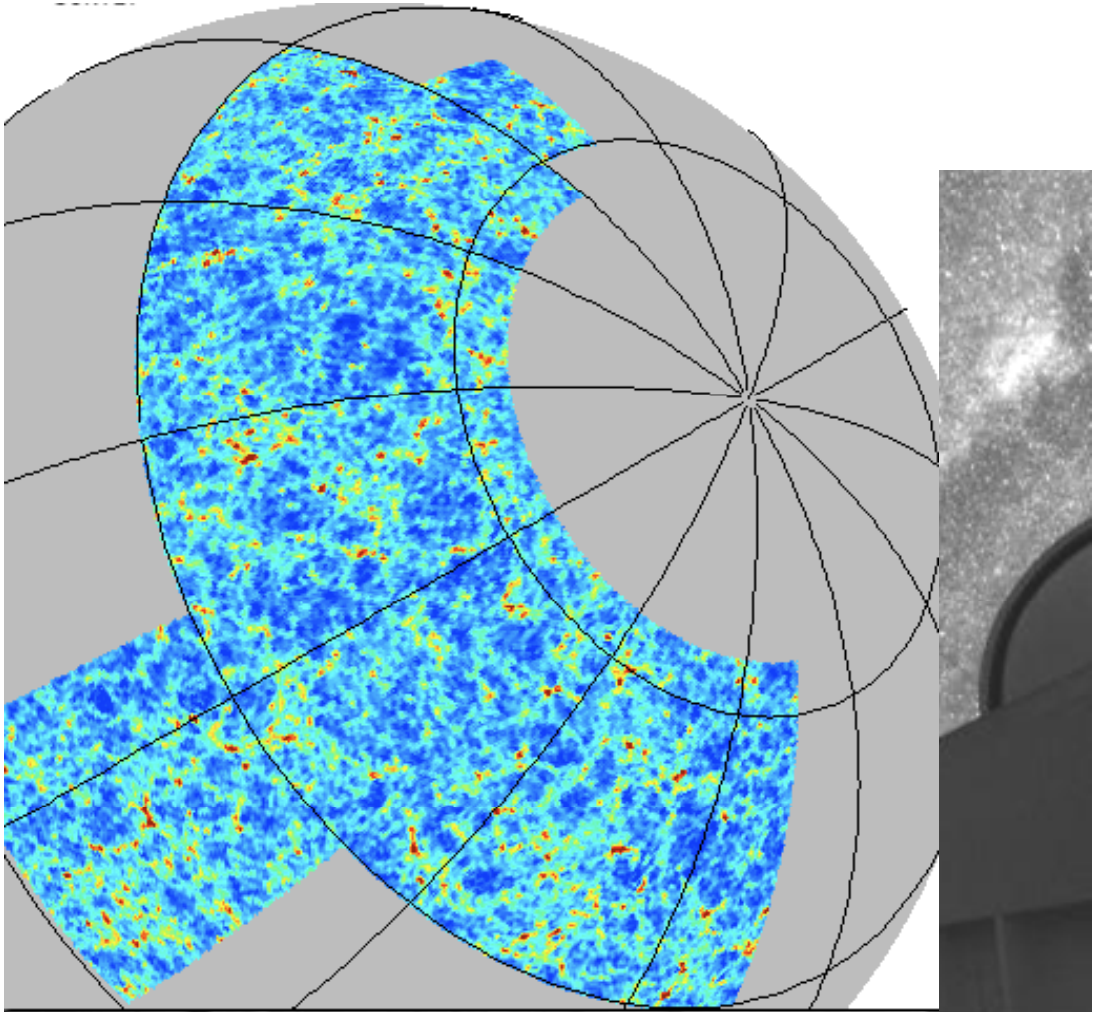
DARK ENERGY SURVEY

- Study Dark Energy using four complementary* techniques:
- I. Cluster Counts
- II. Weak Lensing
- III. Baryon Acoustic Oscillations
- IV. Supernovae

- Two multiband surveys:
- 5000 deg^2 *g, r, i, Z,Y* to *i~24*
- 9 deg^2 repeat (SNe)

- Build new 3 deg^2 camera and Data Management system
- DES 30% of 5 years of telescope
- Response to NOAO AO

- DES Forecast: FoM up by 4.6



*in systematics & in cosmological parameter degeneracies
*geometric+structure growth: test Dark Energy vs. Gravity

DARK ENERGY SURVEY

DES Data Management

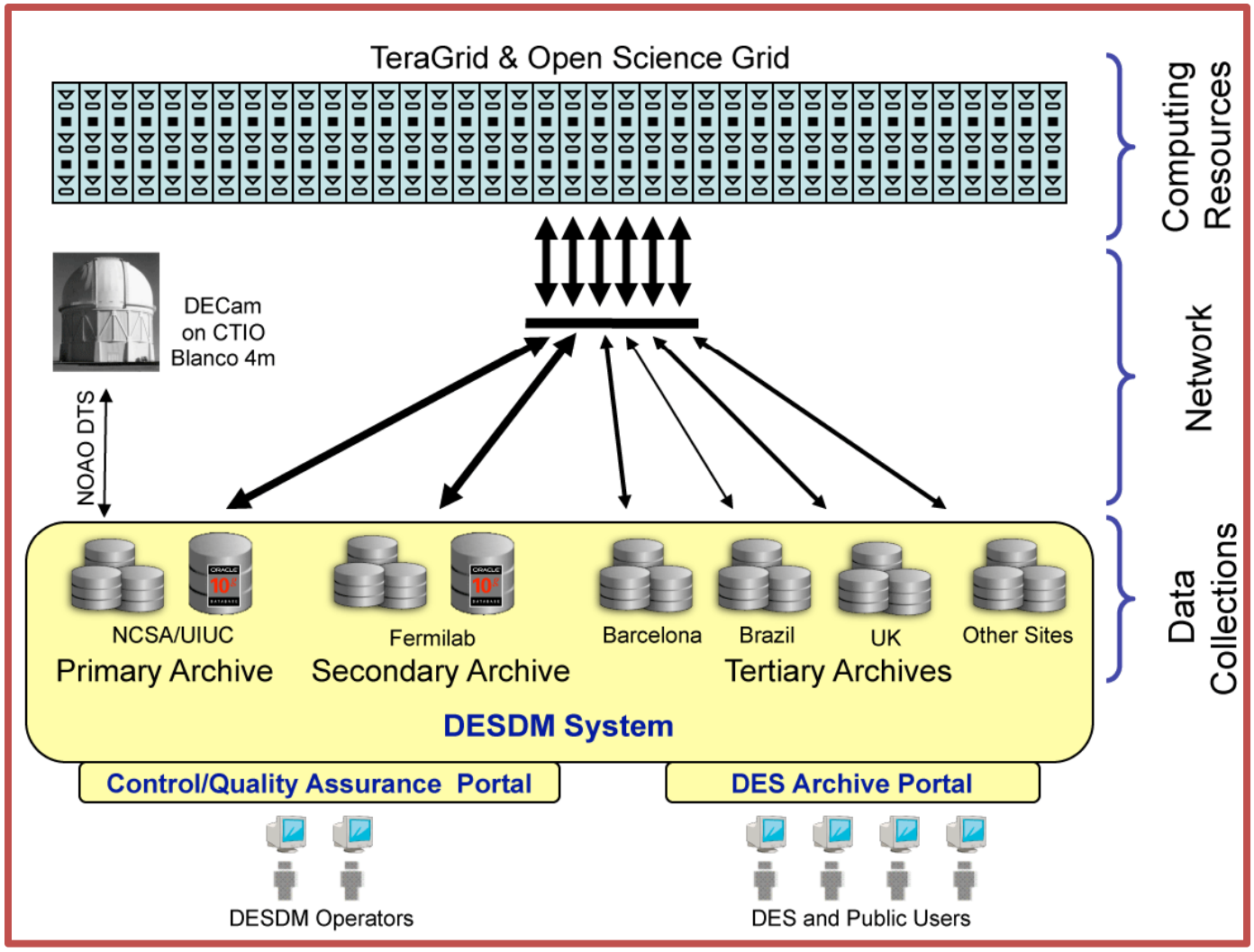
Expect ~2Pb final dataset, with ~100Tb in a relational database

LSST precursor

Transients – Supernovae

EPO – also drives data service needs

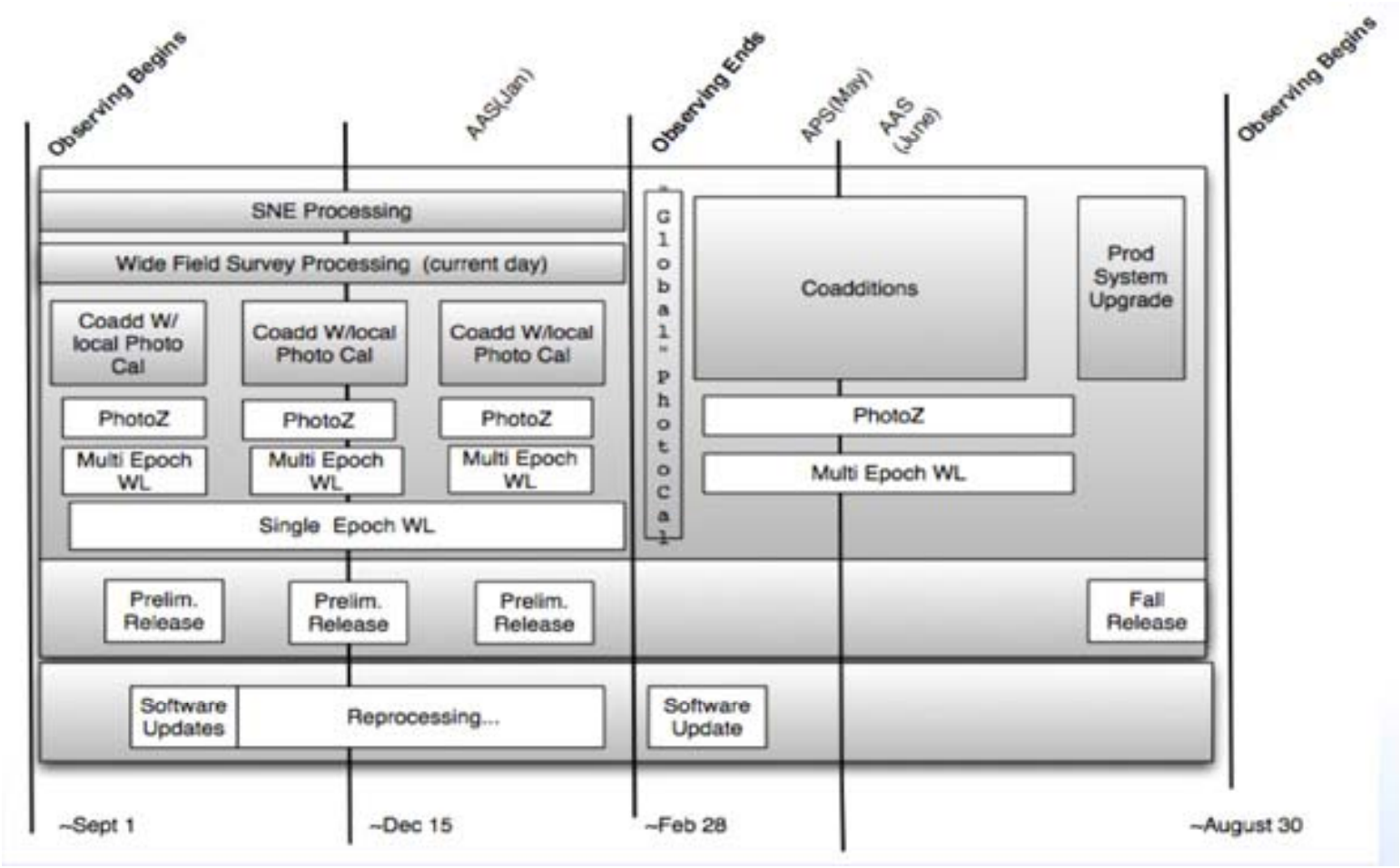
Commission late in CY2011

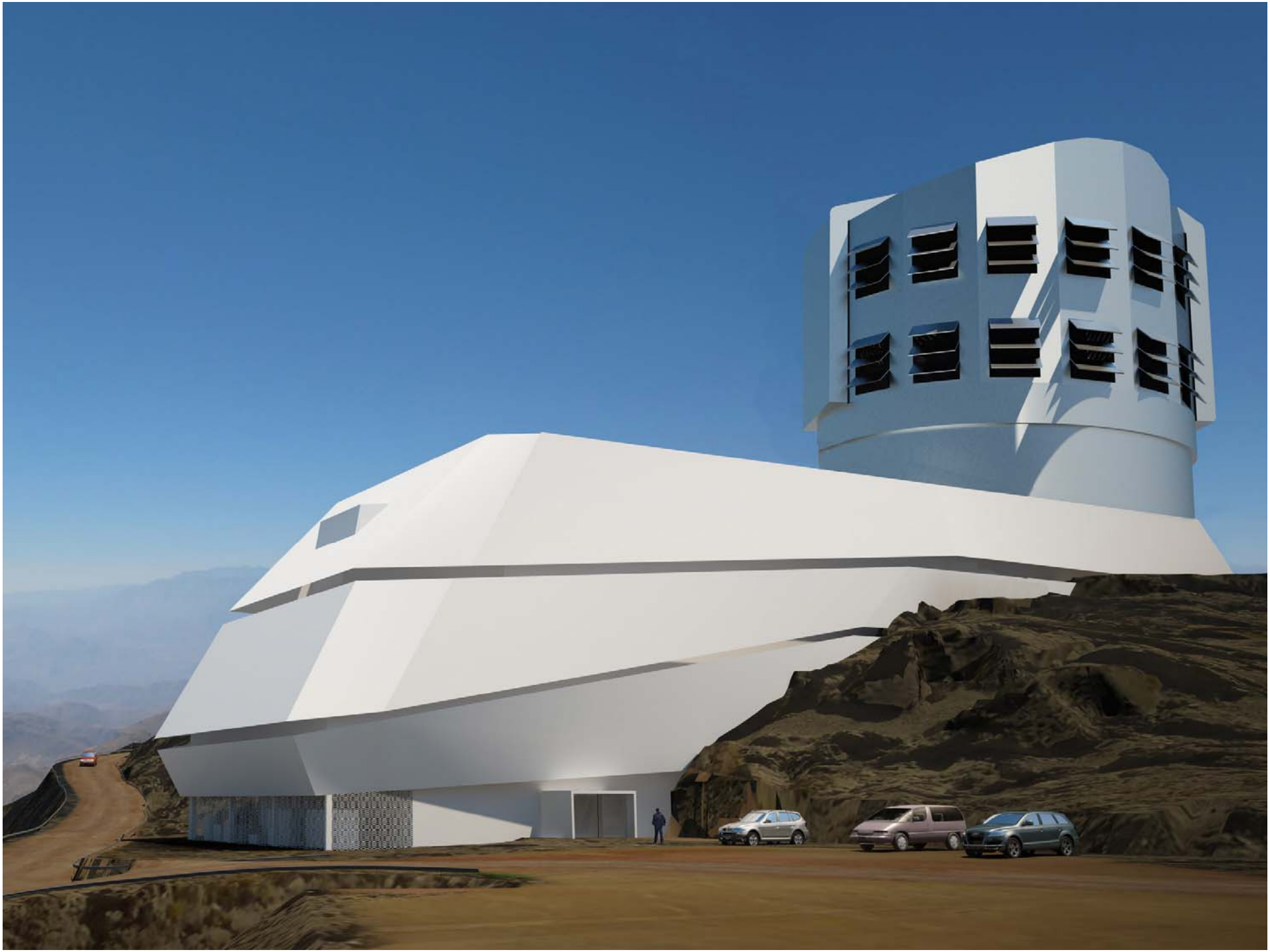


DARK ENERGY SURVEY

DES Data Management

DES itself runs in the southern spring, September to March, taking one third of the CTIO 4m Blanco telescope nights for five years



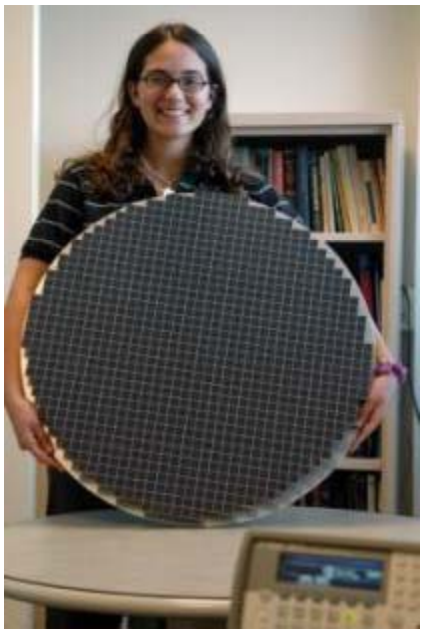


- **LSST (Large Synoptic Survey Telescope):**
  - Ten-year time series imaging of the night sky – mapping the Universe
  - **100,000 events each night** – *anything that goes bump in the night*
  - ***Cosmic Cinematography - The New Sky at http://www.lsst.org/***

Observing Strategy: One pair of images every 40 seconds for each spot on the sky, then continue across the sky continuously every night for 10 years, with time domain sampling in log(time) intervals (to capture dynamic range of transients).

Education and Public Outreach have been an integral and key feature of the project since the beginning – the EPO program includes formal & informal education, Citizen Science projects, Science Centers & Planetaria.

Data Challenges

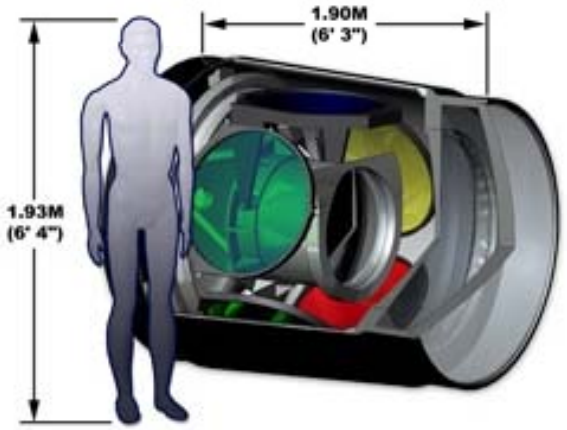


It's a large camera

201 CCDs @ 4096x4096 pixels <u>each</u>
= 3 Gigapixels = 6 GB per image, covering 10 sq.degrees
= ~3000 times the area of one Hubble Telescope image

Process

- Obtain one 6-GB sky image in 15 seconds
- Process that image in 5 seconds
- Obtain & process another co-located image within 20s
  (= 15-second exposure + 5-second processing & slew)
- Process the 100 million sources in each image pair, catalog all sources, and generate worldwide alerts within 60 seconds (e.g., incoming killer asteroid)
- Generate 50-100,000 alerts per night (VOEvent messages)
- Obtain 2000 images per night
- Produce ~40 Terabytes per night
- Move the data from South America to US daily
- Repeat this every day for 10 years
- Provide rapid DB access to worldwide community:
  - 100-200 Petabyte final 10-year image archive
  - 10-20 Petabyte final 10-year database catalog

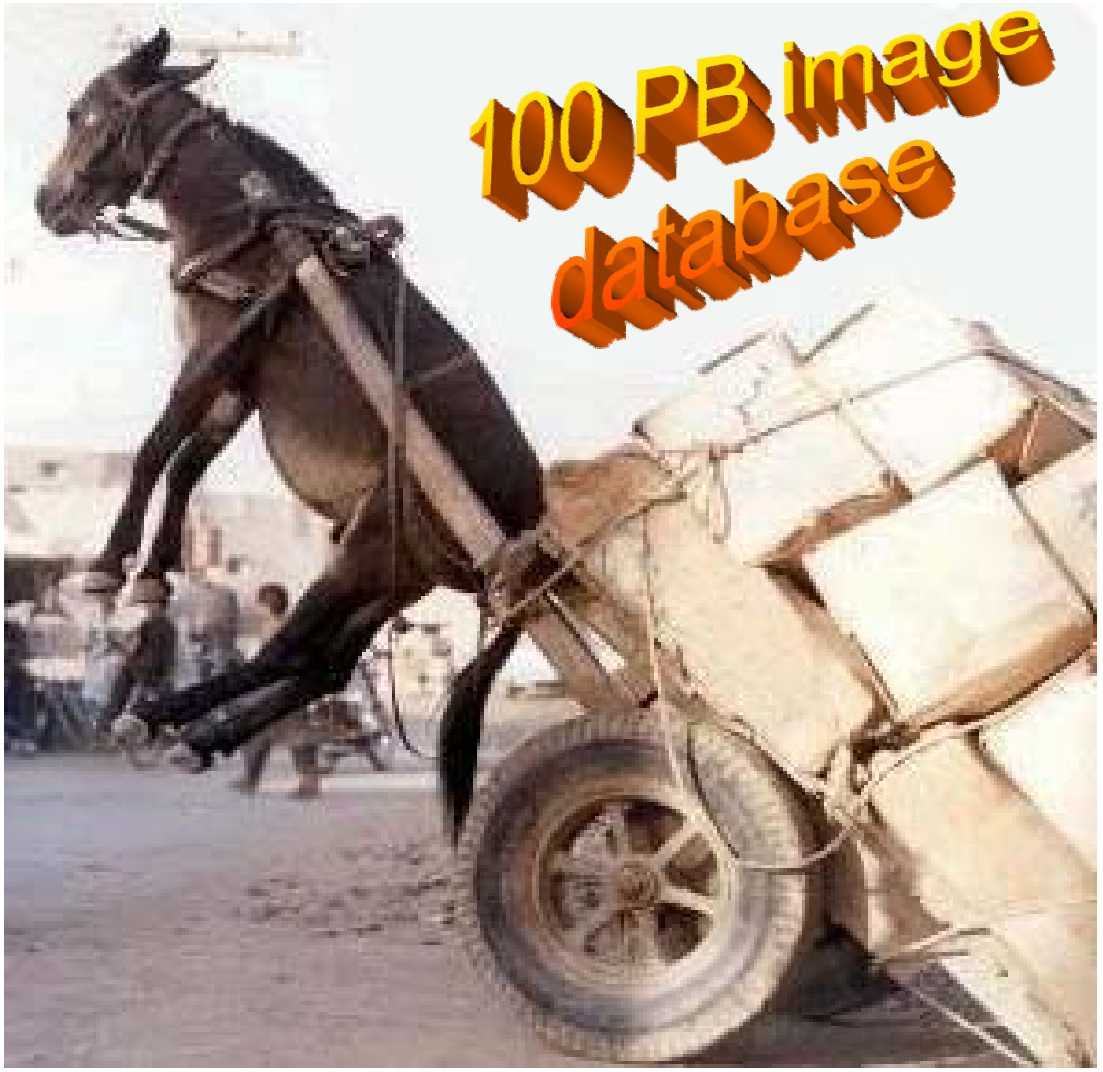


Data Management System

- **Science Requirements that drive Data Management**
  - **Very deep, time-dependent, wide field survey ➔ unprecedented data volumes (40Tb/night)**
  - **Controlled systematics ➔ advanced photometry, astrometry**
  - **Rapid transient alerting ➔ high performance, low latency**
- **Required system capabilities and performance**
  - **Highly distributed, scalable, and reliable**
  - **Evolvable scientifically and technically over the survey useful life**
- **Derived design and implementation features**
  - **Advanced scientific algorithms that scale to LSST volumes, rates**
  - **Layered architecture for extensibility and technology evolution**
  - **High-performance computing, storage, networking technologies**
  - **Rigorous formal development process**
- **Involves computer scientists, industry analysts, software developers, project managers, and others, from academic, commercial, and government enterprises.**

The LSST Data Challenge

- Massive data stream: ~2 Terabytes of image data per hour that must be mined in real time (for 10 years).
- Massive 20-Petabyte database: more than 50 billion objects need to be classified, and most will be monitored for important variations in real time.
- Massive event stream: knowledge extraction in real time for 100,000 events each night.



NSF/NASA - Virtual Astronomical Observatory

- Move beyond the very successful NVO (international standards, software, data mining support, user interfaces, etc.) to a real astronomical observatory

NATIONAL VIRTUAL OBSERVATORY

- "Just another observatory" where a wealth of detail is covered (hidden?) by the word "just"

- VAO, LLC preliminary certification to receive federal funding agreed by NSF business office; NSF award made, May 2010; NASA contribution funded through existing centers
- Strong link with Microsoft Research, World Wide Telescope product
- Research is being enabled – notable publications at http://www.us-vo.org/pubs/notablepubs.cfm

- **Problem of marketing, branding**

It takes a human to interpret a complex image

Citizen Science

- Exploit the cognitive abilities of **Human Computation**
- Novel mode of data collection:
  - Citizen Science = Volunteer Science
  - *e.g.*, VGI = Volunteer Geographic Information
  - *e.g.*, Galaxy Zoo @ http://www.galaxyzoo.org/
- Citizen science refers to the involvement of volunteer non-professionals in the research enterprise. The experience must be engaging, must work with real scientific data and information, **must address authentic science research questions** that are beyond the capacity of science teams and enterprises, and must involve the scientists.

Examples of Volunteer Science

- Audubon Bird Counts
- Project Budburst
- Stardust@Home
- VGI (Volunteer Geographic Information)
- Galaxy Zoo **(many refereed publications)**
- Zooniverse (buffet of Zoos)
- U-Science (semantic science 2.0)
  - includes Biodas.org, Wikiproteins, HPKB, AstroDAS
  - **U**biquitous, **U**ser-oriented, **U**ser-led, **U**niversal, **U**ntethered, **You**-centric **Science**