



Open Sky, Open Data

*The Community Science and Data Center
at NSF's National Optical Astronomy
Observatory*

Adam S. Bolton

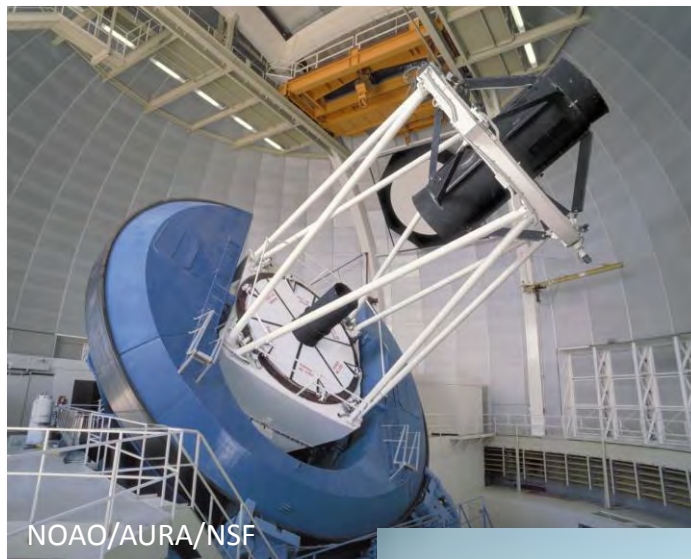
NOAO Acting Deputy Director

NOAO Associate Director for CSDC



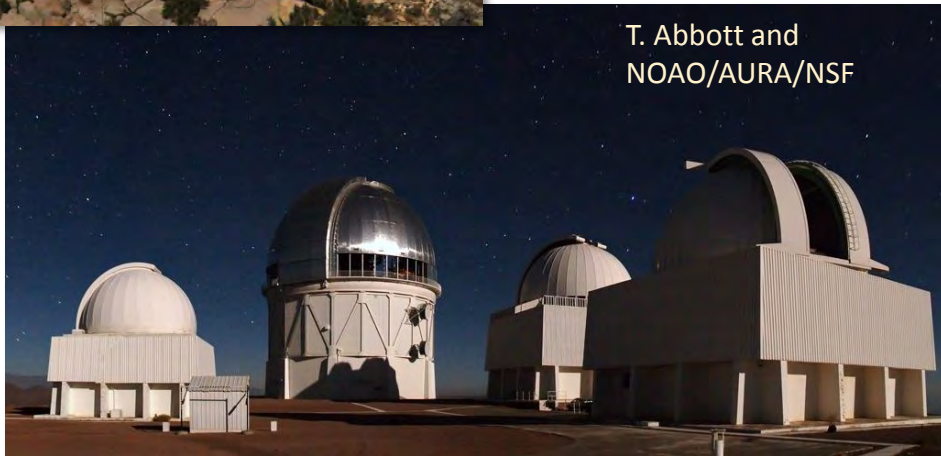
Open access to world-class astronomy research facilities

P. Marenfeld/NOAO/AURA/NSF and E. Acosta/LSST/AURA/NSF



NOAO/AURA/NSF

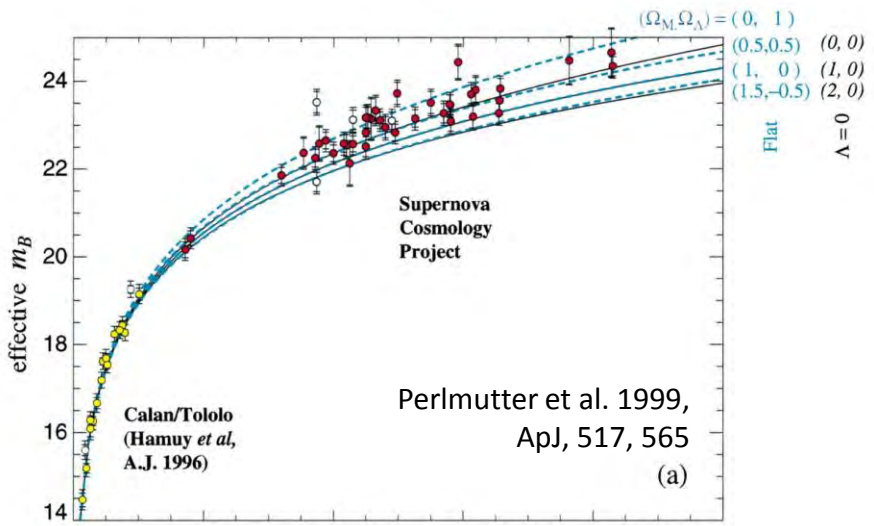
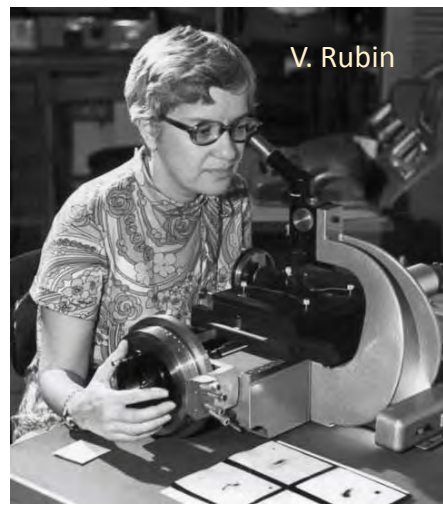
T. Abbott and
NOAO/AURA/NSF



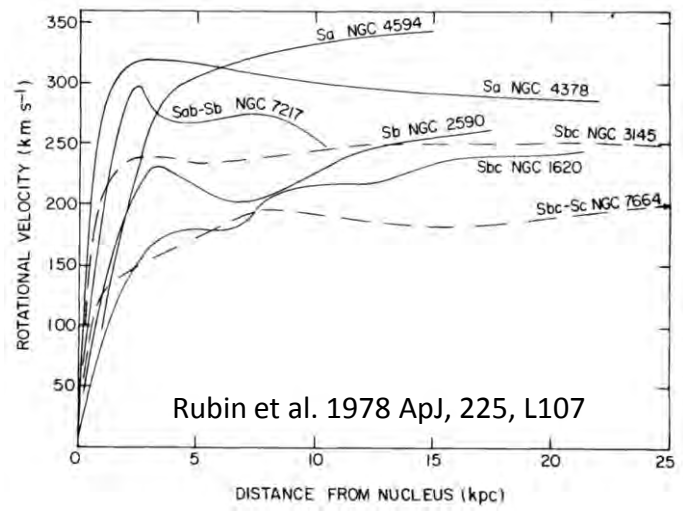
M. Urzúa Zuñiga/Gemini Observatory



NOAO in the history books: Dark Matter and Dark Energy



(also Riess et al. 1998, AJ, 116, 1009)



Nobel Foundation / U. Montan

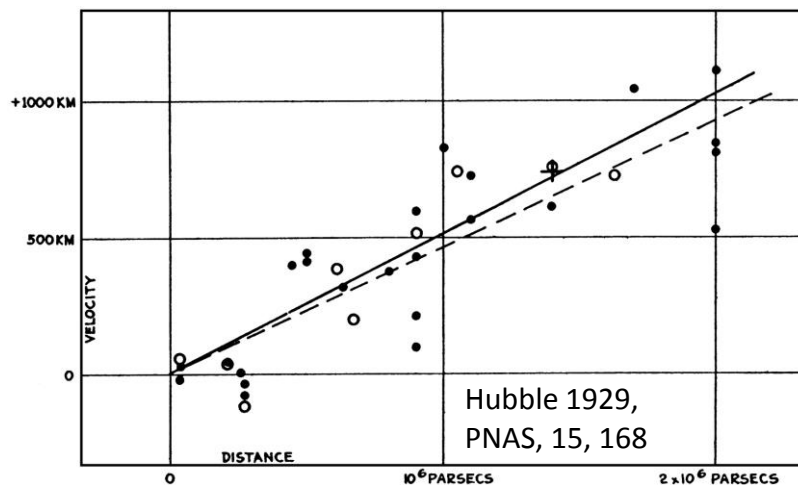


“Classical” astronomy

- PI-driven
- Targeted observations
- Accessible to relatively few

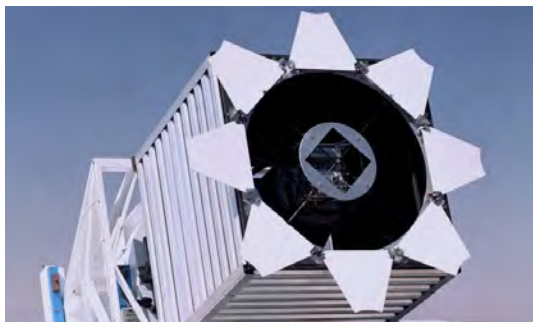


Margaret Bourke-White—The Life Picture Collection/Getty Images

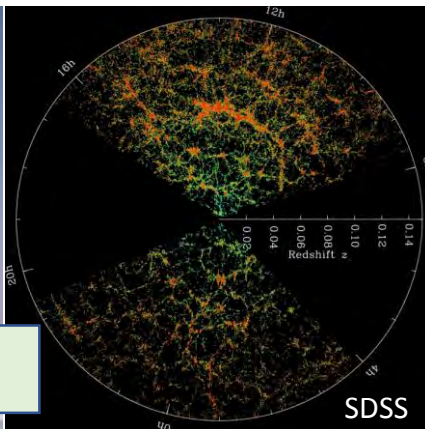




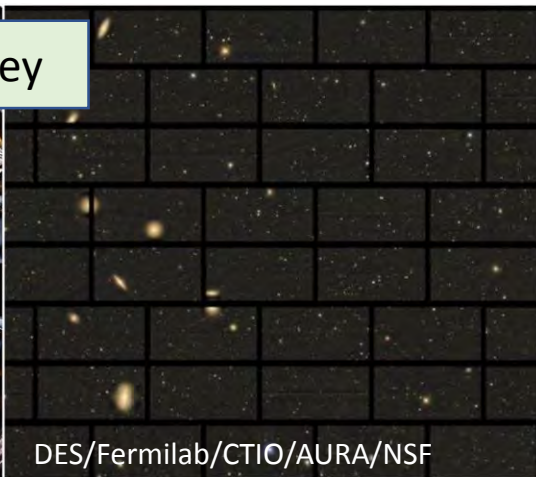
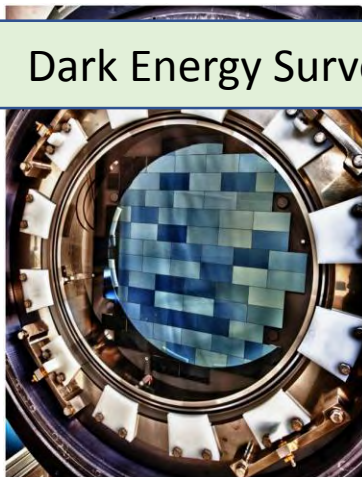
Survey-scale astronomy: a data-intensive revolution



Sloan Digital Sky Survey



Dark Energy Survey



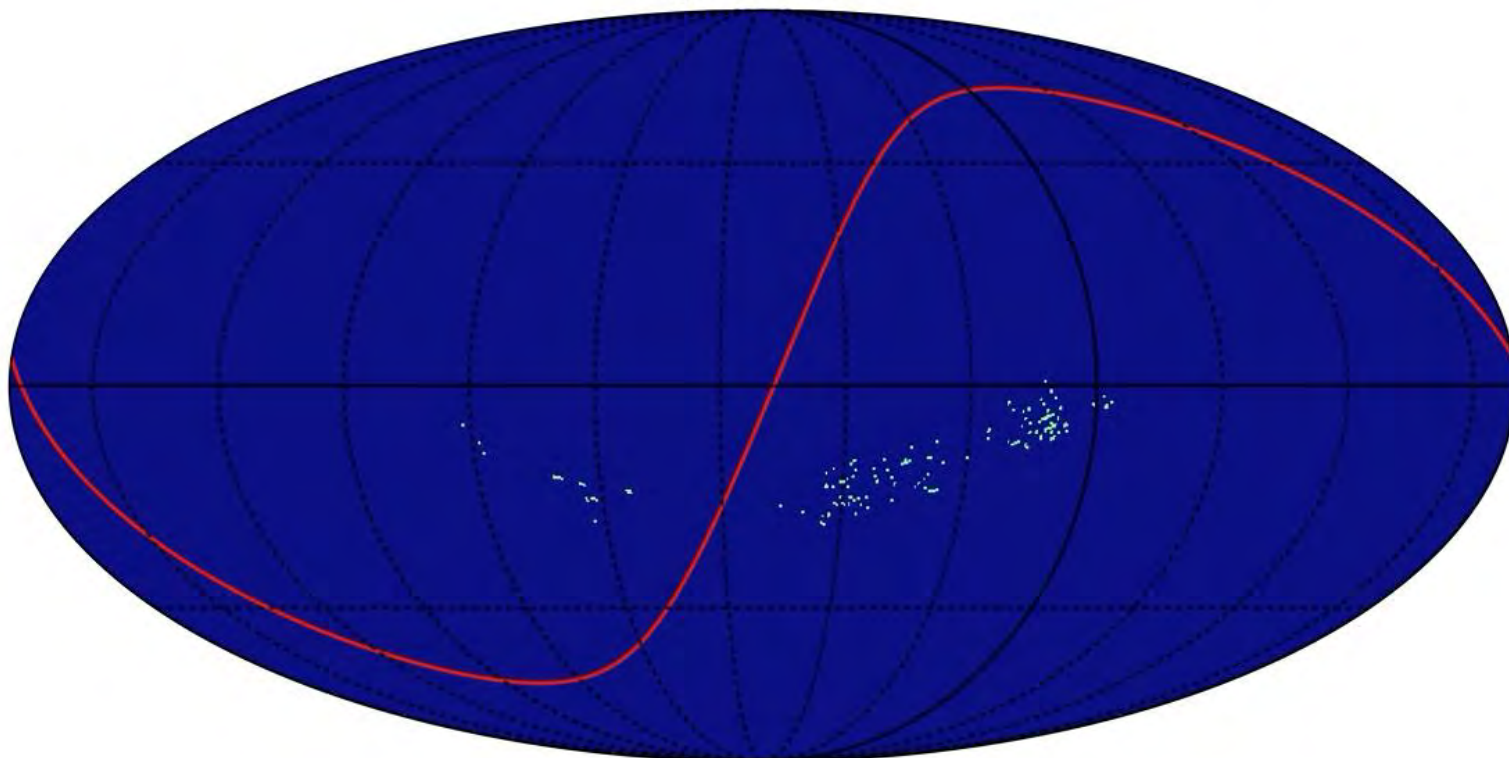
- Astronomy + Physics collaboration
- Purpose-built survey facilities
- Systematic surveys of large areas
- Large collaborative teams
- Large homogeneous data sets
- ***Diverse archival research opportunities***



The archival phase transition

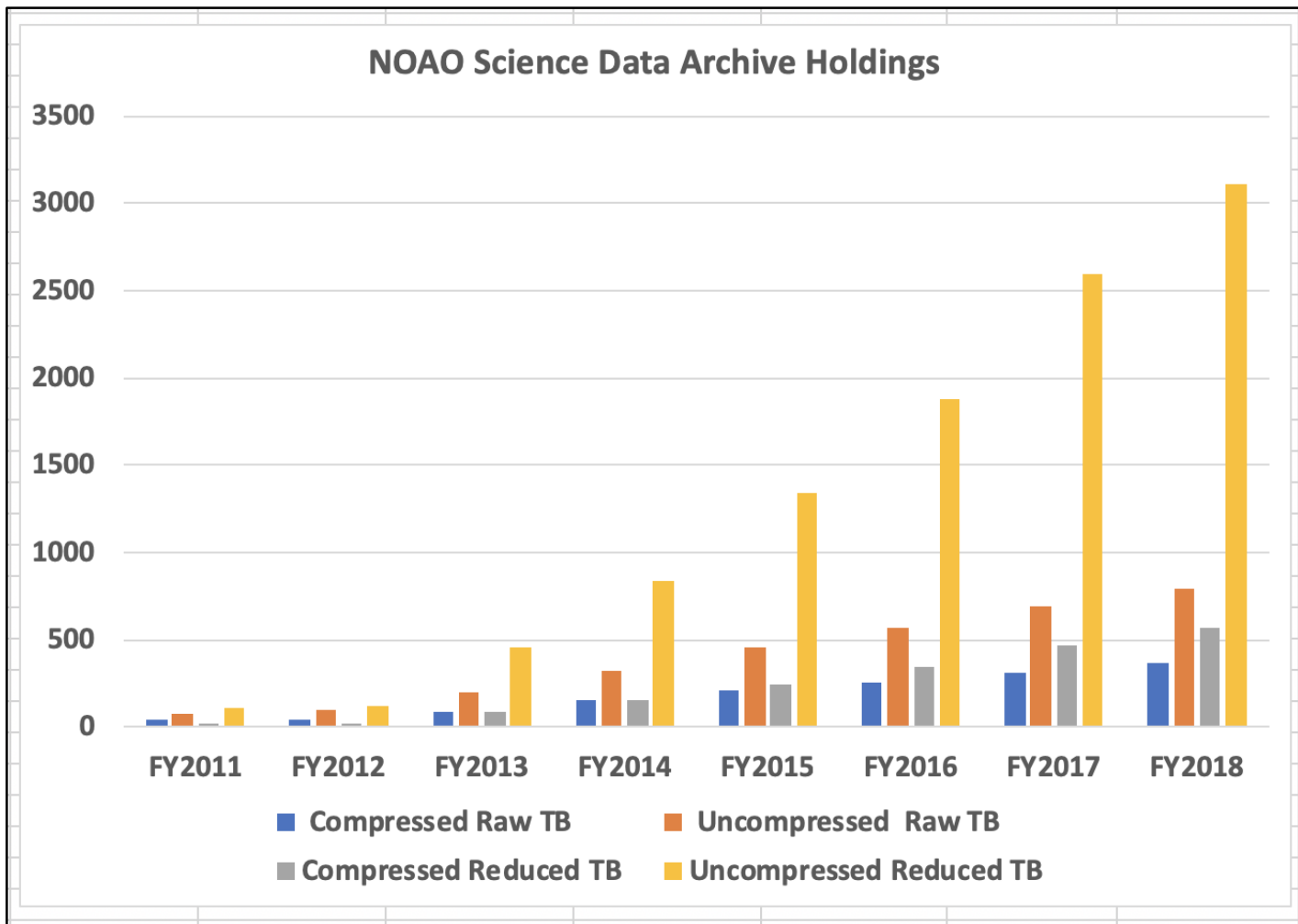
Animation: K. Olsen & S. McManus

August 11, 2004





(half-)Gigapixel imaging, petascale data archives

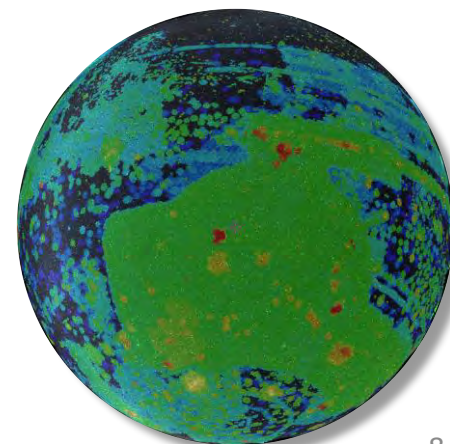




NOAO's Community Science and Data Center (CSDC)

The “third mountaintop”: enabling community science in the era of data-intensive astronomy

- *Enable science now with the data sets of today*
- *Prepare the community for science with the data sets of tomorrow*
- *Leverage existing frameworks and technologies*
- *Realize the potential for data-intensive astronomy to drive an “inclusion revolution” (see D. Norman, ASP AstroBeat 162, July 2018)*





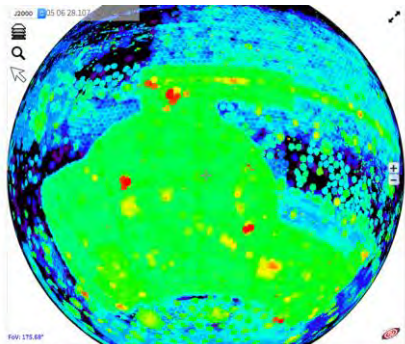
NOAO Data Lab: open-access science with big survey data sets

The screenshot shows the NOAO Data Lab website interface. At the top, the title "NOAO Data Lab" is centered. On the right, there are links for "Login" and "Sign up", and a "Service Status" indicator with five green checkmarks. A navigation menu includes "About", "Quick Start", "Tools", "Survey Data", "Docs/Help", and "Events". The main content area features a large visualization of the "All-sky NOAO Source Catalog" (2.9 billion objects, 34 billion measurements) shown as a colorful map of the sky with a grid. Below the visualization, the text "All-sky NOAO Source Catalog (2.9 billion objects, 34 billion measurements)" is displayed. At the bottom, there are logos for NSF, NOAO, and AURA, along with social media icons for GitHub and Twitter. A text block states: "The NOAO Data Lab is operated by the National Optical Astronomy Observatory, the national center for ground-based nighttime astronomy in the United States operated by the Association of Universities for Research in Astronomy (AURA) under cooperative agreement with the National Science Foundation." Below this, it says "For further information, contact datalab@noao.edu." A prominent green box at the bottom right contains the text "Visit <https://datalab.noao.edu>".



NOAO Data Lab: open-access science with big survey data sets

Discover data



Query catalogs

datalab.noao.edu/tap

Column Information Query interface Virtual Storage Job Status

- allwise
- dad_dr1
- dad_dr2
- decaps_dr1
- des_dr1
- des_sva1
- gaia_dr1
- gaia_dr2
- ivoa
- ivoa_calibrated
- ivoa_coadd
- ivoa_des_dr1
- ivoa_ls_dr3
- ivoa_ls_dr4
- ivoa_ls_dr5
- ivoa_raw

```
* required field.
All queries must be written in ADQL.
To issue direct SQL database queries, use the queryClient module or datalab command-line.

SELECT ring256,count(ring256) as nb FROM ls_dr3.tractor_primary GROUP BY ring256
-- In this query we return the number of objects from Legacy Survey DR3, but aggregate the count over the ring256 column. Ring256 stores the Healpix index (RING scheme, NSIDE=256), so the returned result is a Healpix map that can then be visualized with the healpy module.
```

Query:

Crossmatch catalogs

X-match Service

Xmatch Table Management

Step 1: Pick the 1st table:

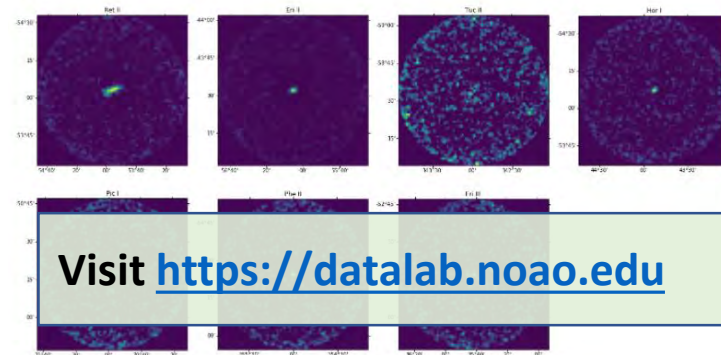
Step 2: Pick the 2nd table: Datalab table User table

Step 3: Result: Table Name
 Radius: 5 arcseconds
 Find all matches Nearest neighbor
 Download results to your computer only

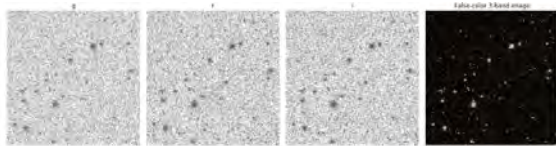
Step 4:

Analyze, visualize, and discover in Jupyter

Ret II: Computing differential convolution
 Eri II: Computing differential convolution
 Tuc II: Computing differential convolution
 Hor I: Computing differential convolution
 Pic I: Computing differential convolution
 Phe II: Computing differential convolution
 Eri III: Computing differential convolution
 Done.



Make Image cutouts



RA:159.815_DEC:0.653 3 images found.

Thumbnail	Instrument_name	obs_bandpass	exptime	prodtype	proctype	date_obs	All items
	DECam	g	0	Image	Stack	2014-02-18T05:29:42.550	<input type="checkbox"/>
	DECam	r	0	Image	Stack	2014-12-30T06:12:39.397	<input type="checkbox"/>
	DECam	z	0	Image	Stack	2013-04-01T03:34:19.201	<input type="checkbox"/>

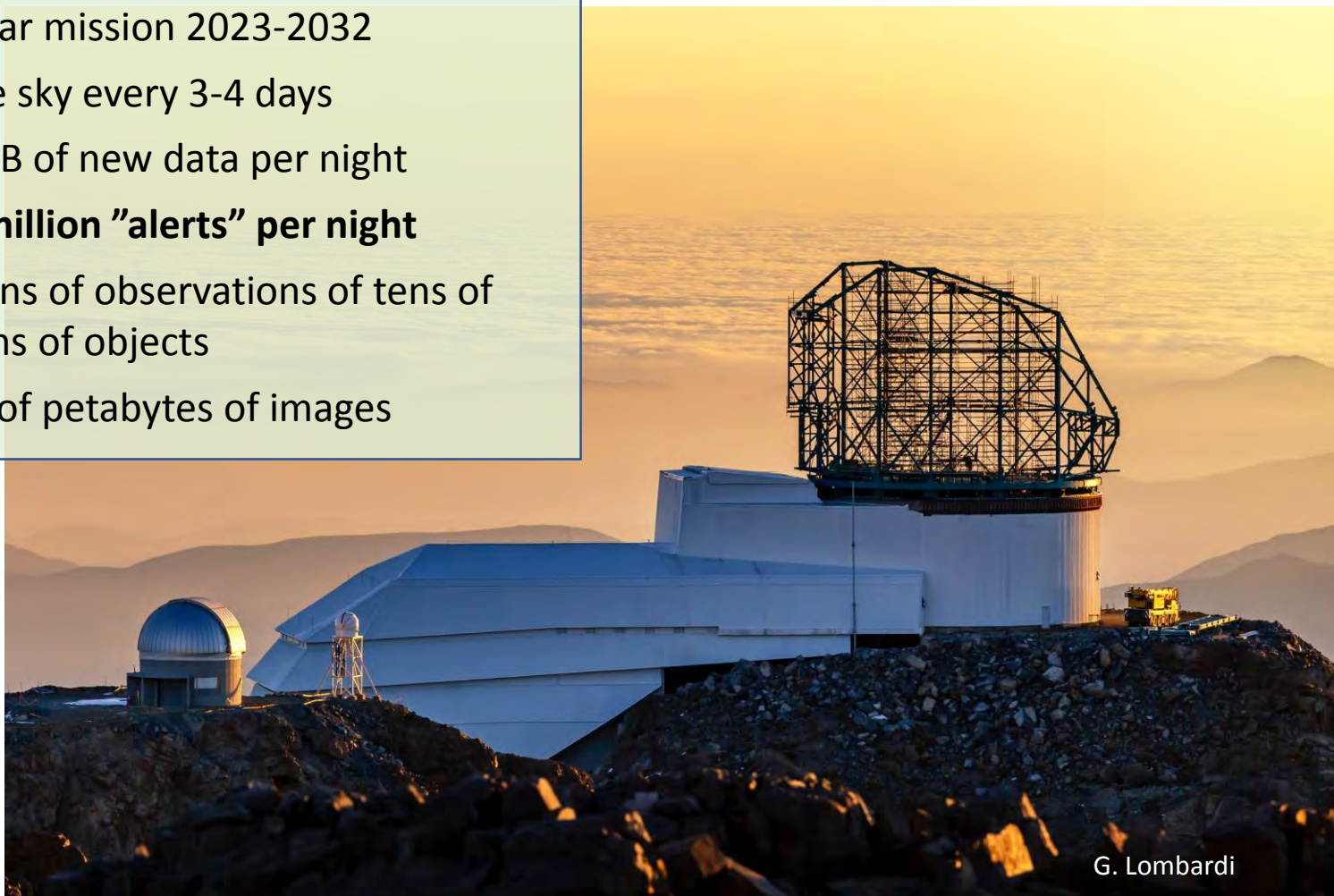
Visit <https://datalab.noao.edu>





The future: Large Synoptic Survey Telescope (LSST)

- 10 year mission 2023-2032
- Entire sky every 3-4 days
- ~20 TB of new data per night
- **~10 million "alerts" per night**
- Trillions of observations of tens of billions of objects
- 100s of petabytes of images



G. Lombardi



ANTARES: an “event broker” for the LSST era (and today)

Data-intensive science challenge:

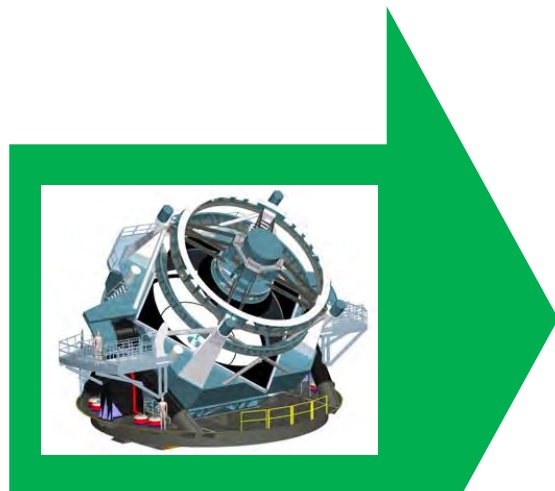
- Flexible public “event brokers” needed to enable science with LSST alert stream
(Elmegreen et al. 2015, Najita & Willman et al. 2016)

NOAO/CSDC solution:

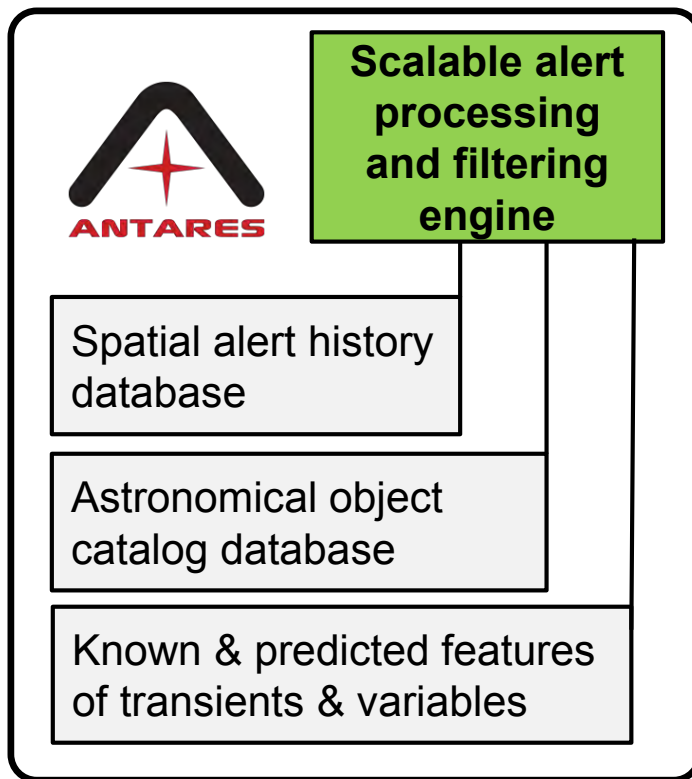
- ANTARES: the Arizona-NOAO Temporal Analysis and Response to Events System
- Collaboration between NOAO and U. of Arizona Department of Computer Science
- A “software instrument” for time-domain astronomy



ANTARES schematic



LSST and other surveys: $\sim 10^7$ alerts per night

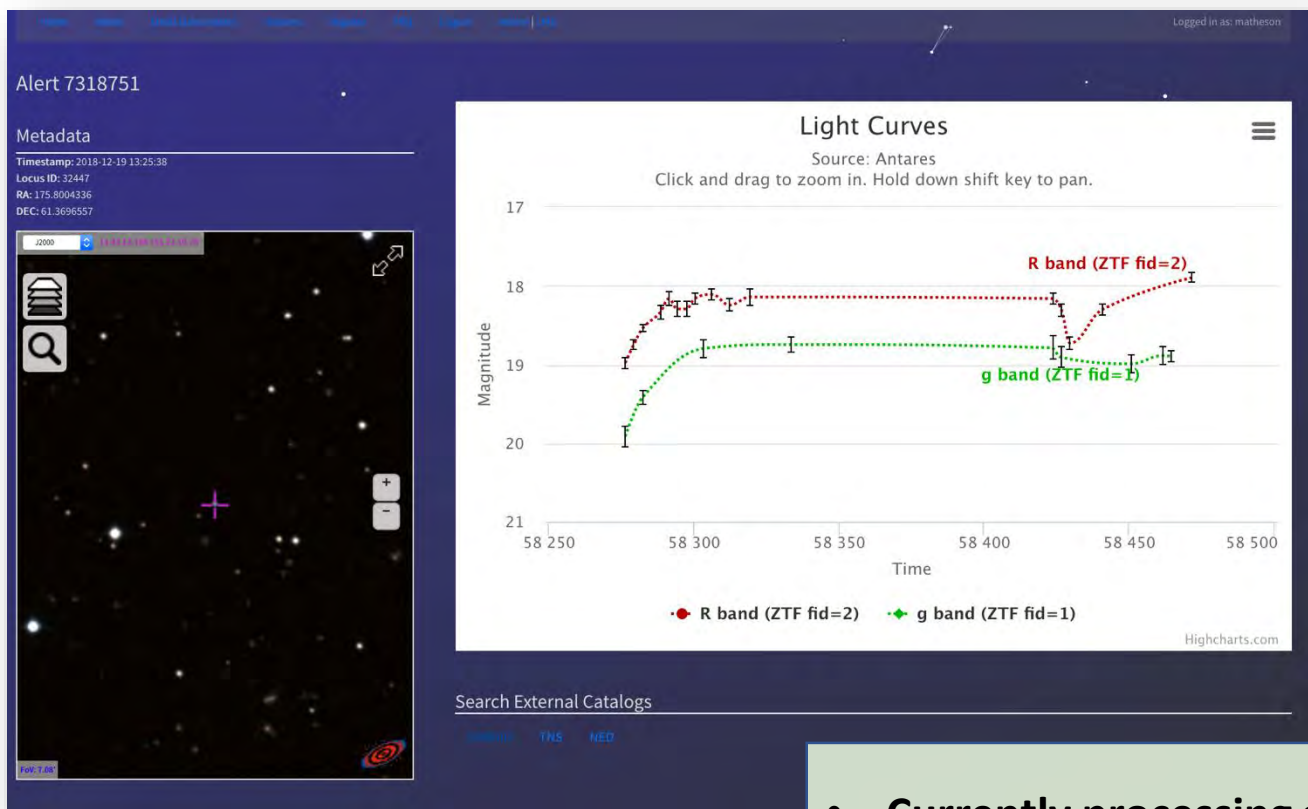


Filtered and value-added alert stream ("the rarest of the rare")





ANTARES: open for business as of December 2018

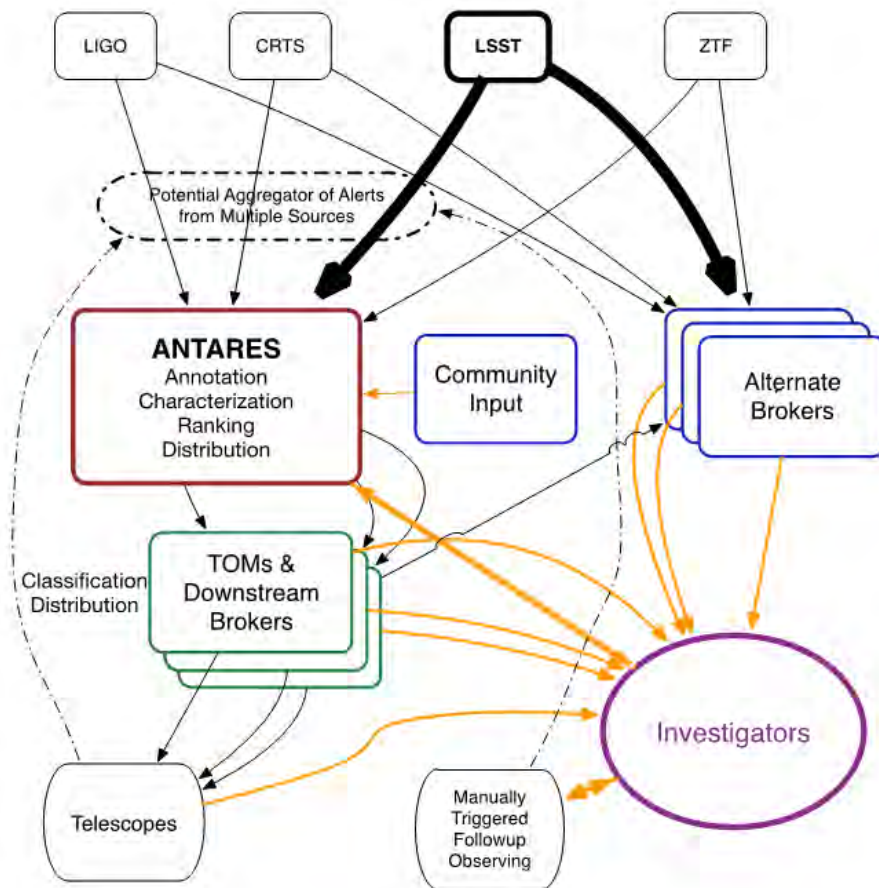


- Currently processing alerts from the Zwicky Transient Facility
- Visit <https://antares.noao.edu>



ANTARES: open for business as of December 2018

- Ingests real-time streams from alert-producing sky surveys
- Processes alerts through customizable filters to select the rarest or most interesting events for individual astronomers
- Adds value to alert packets (spatial history, catalog association, characterization)
- Forwards filtered streams to astronomers through chosen interface (web portal, Slack channel, or direct API connection)
- Retains alert database for offline query & data mining



Visit <https://antares.noao.edu>



Top five challenges / opportunities

- Evolving the core competencies of a National Observatory in the era of data-intensive astronomy
- Building (and scaling) effective mixed teams of scientists and engineers
- Managing agile, user-driven software projects in a waterfall-oriented culture
- Achieving collaboration and synergistic development across silos
- Leveraging commercial and open-source solutions



Backup Slides



NOAO Data Lab: open-access science with big survey data sets

- Data-set discovery capabilities
- Queries of large public survey databases
- Extraction of image cutouts from NOAO Science Data Archive
- Survey file archive service (NOAO Surveys + SDSS)
- Crossmatch service
- User storage and “MyDB”
- Web interface, command line, and Jupyter notebook support

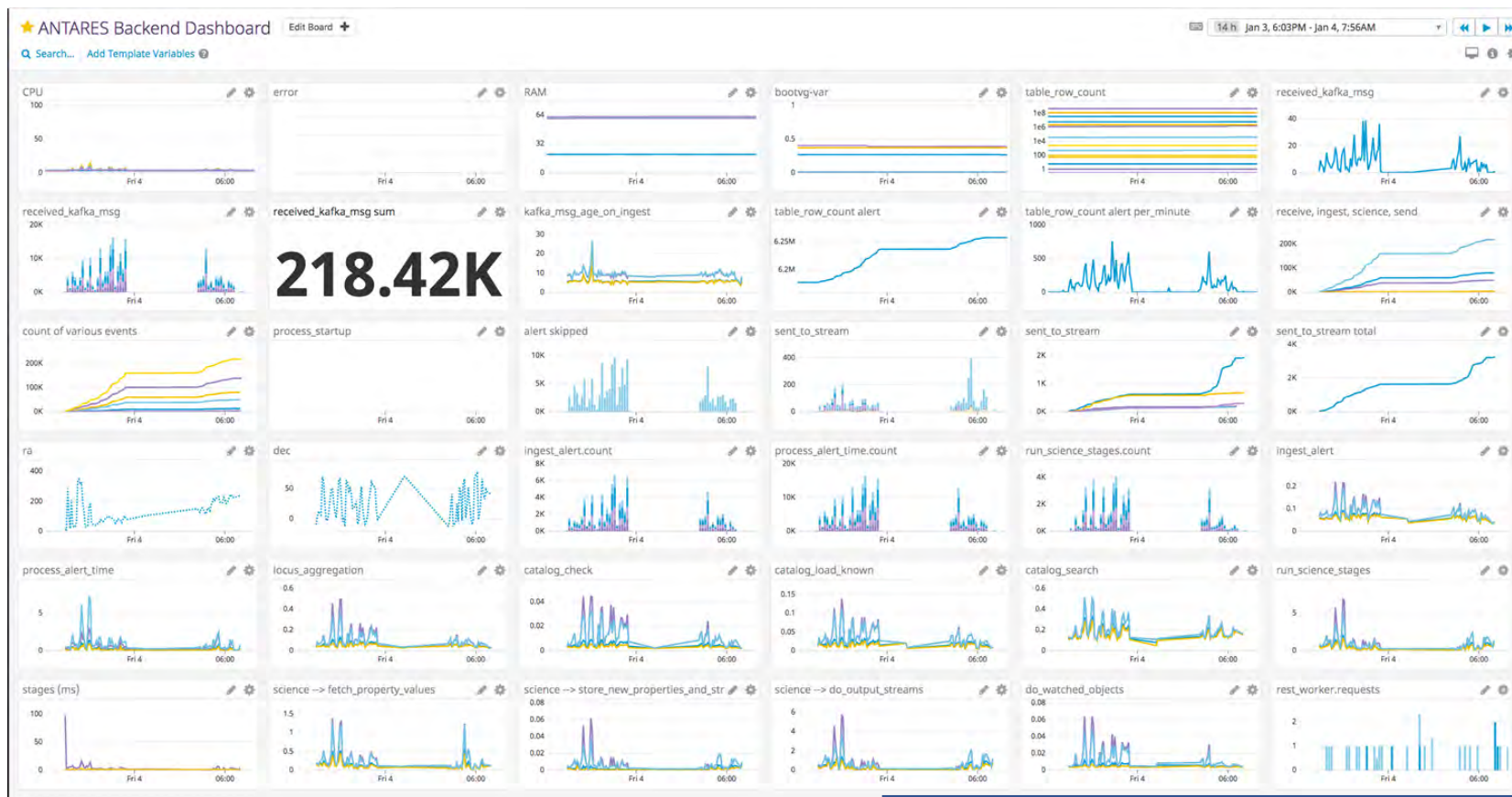
Statistics:

- ~700 registered users
- ~400 visits per month
- Catalogs from 16 distinct surveys, multiple data releases
- Database with ~150 billion rows of catalog data, ~50 TB
- ~20 billion rows fetched in January 2019
- ~200 million files via file service

Visit <https://datalab.noao.edu>



ANTARES back-end dashboard



Visit <https://antares.noao.edu>