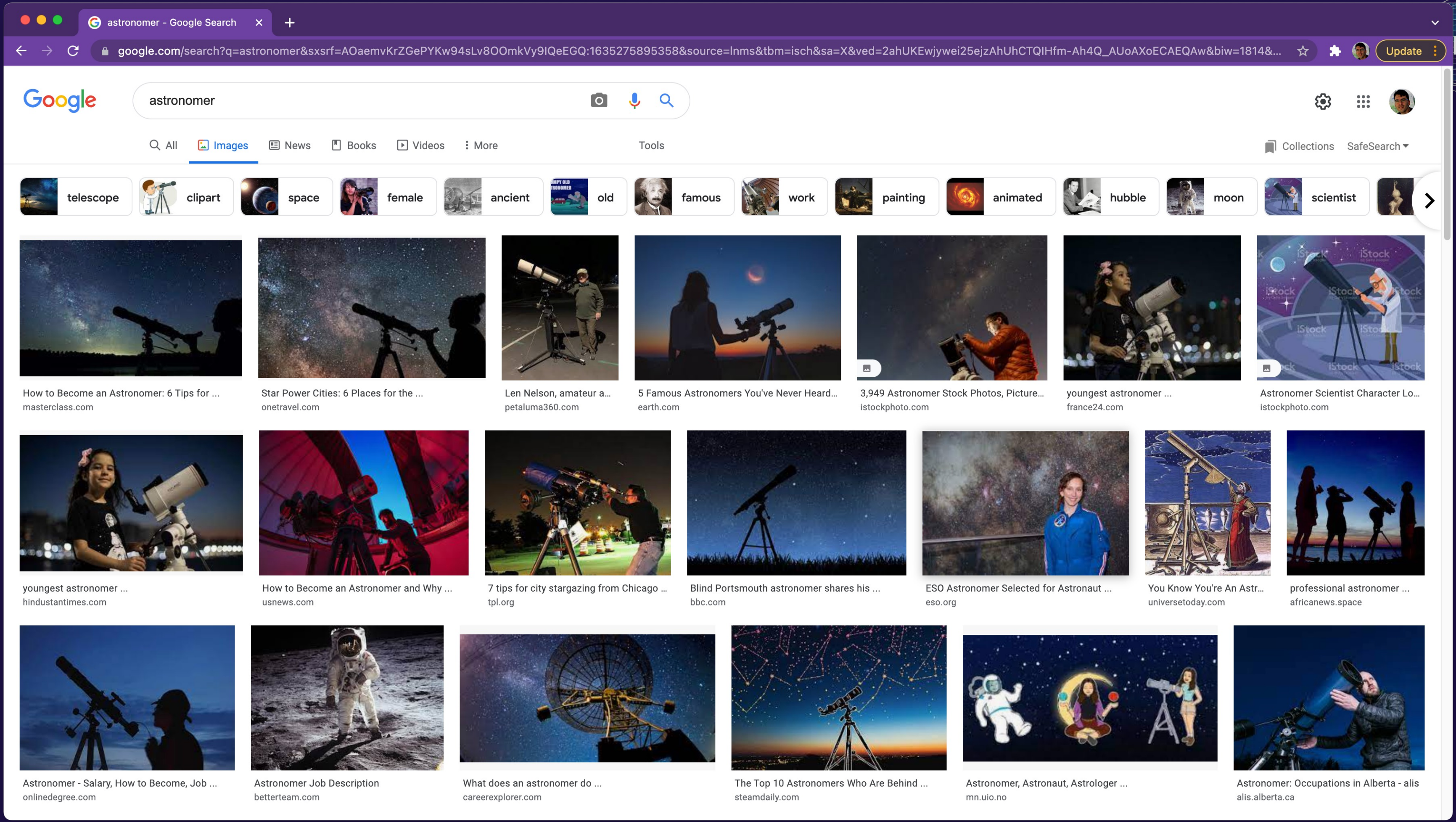
Downloading and making sense of the Sky: Large Sky Surveys and AI

Mario Juric
Director, DiRAC Institute, University of Washington

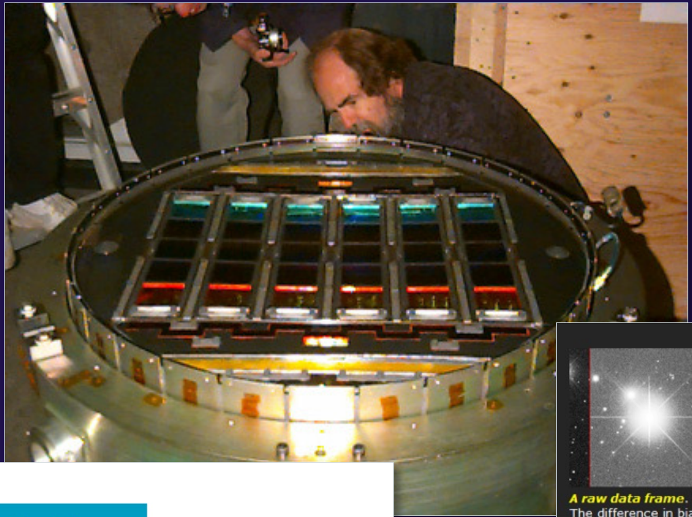
Department of Astronomy, University of Washington
LINCC Frameworks Project

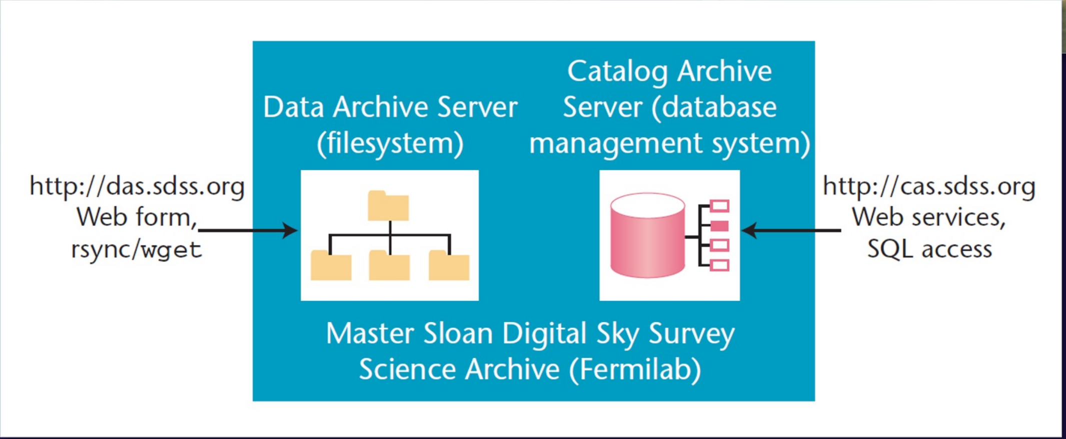
AST-2003196

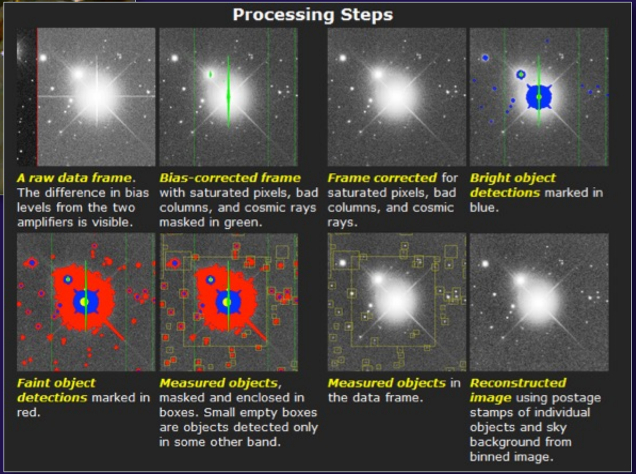

astronomer - Google Search
google.com/search?q=astronomer&sxsrf=AOaemvKrZGePYKw94sLv8OOmkVy9IQeEGQ:1635275895358&source=lnms&tbm=isch&sa=X&ved=2ahUKEwjywei25ejzAhUhCTQIHfm-Ah4Q_AUoAXoECAEQAw&biw=1814&...
Update
Google
astronomer
All
Images
News
Books
Videos
More
Tools
Collections
SafeSearch
telescope
clipart
space
female
ancient
old
famous
work
painting
animated
hubble
moon
scientist
How to Become an Astronomer: 6 Tips for ...
masterclass.com
Star Power Cities: 6 Places for the ...
onetravel.com
Len Nelson, amateur a...
petaluma360.com
5 Famous Astronomers You've Never Heard...
earth.com
3,949 Astronomer Stock Photos, Picture...
istockphoto.com
youngest astronomer ...
france24.com
Astronomer Scientist Character Lo...
istockphoto.com
youngest astronomer ...
hindustantimes.com
How to Become an Astronomer and Why ...
usnews.com
7 tips for city stargazing from Chicago ...
tpl.org
Blind Portsmouth astronomer shares his ...
bbc.com
ESO Astronomer Selected for Astronaut ...
eso.org
You Know You're An Astr...
universetoday.com
professional astronomer ...
africanews.space
Astronomer - Salary, How to Become, Job ...
onlinedegree.com
Astronomer Job Description
betterteam.com
What does an astronomer do ...
careerexplorer.com
The Top 10 Astronomers Who Are Behind ...
steamdaily.com
Astronomer, Astronaut, Astrologer ...
mn.uio.no
Astronomer: Occupations in Alberta - alis
alis.alberta.ca

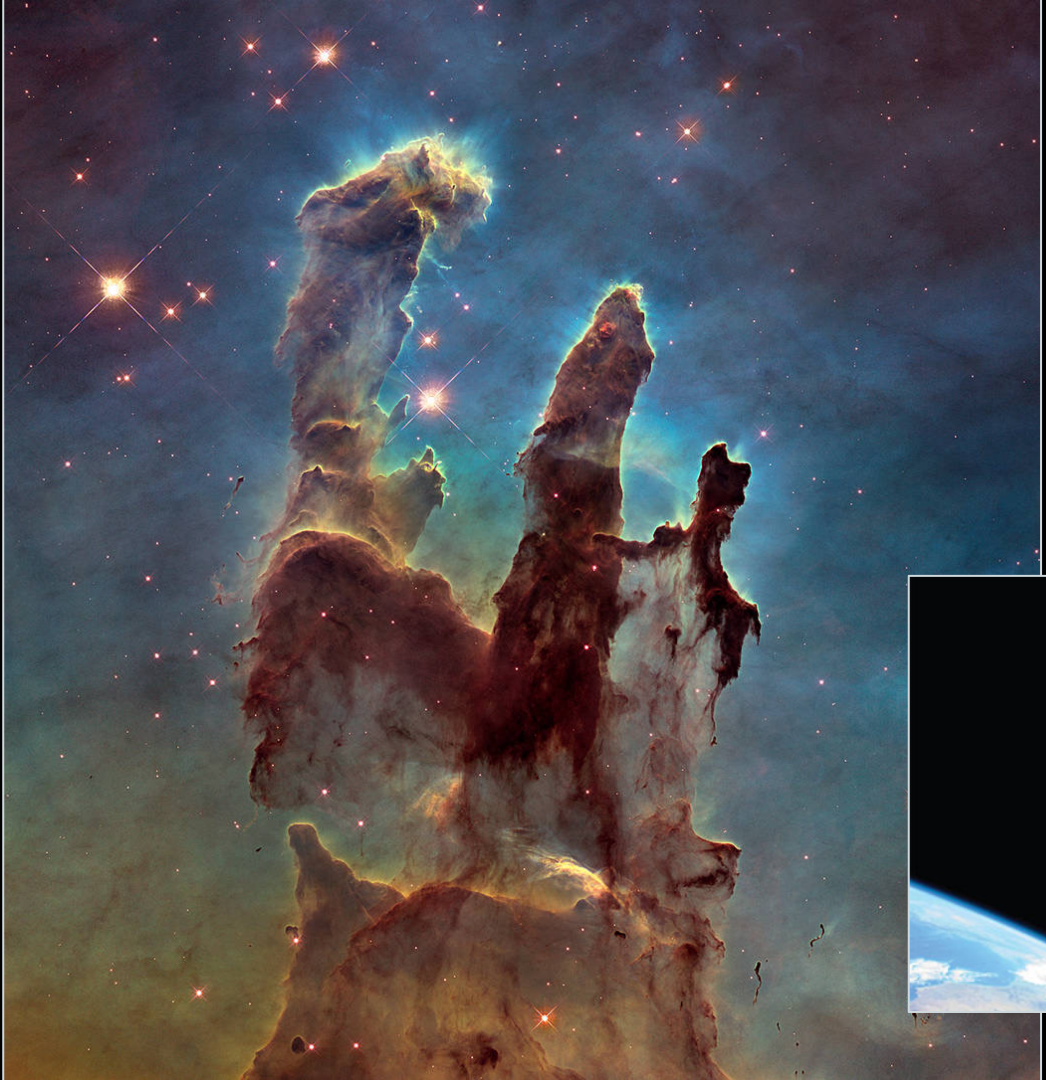
What people *think* astronomers do

What we *(increasingly)* do



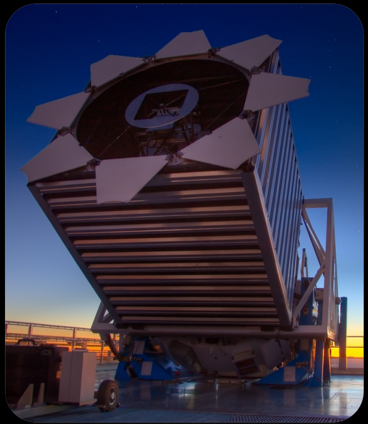






"Pillars of Creation" in the Eable Nebula

… "active star-forming region within the nebula, harboring newborn stars".

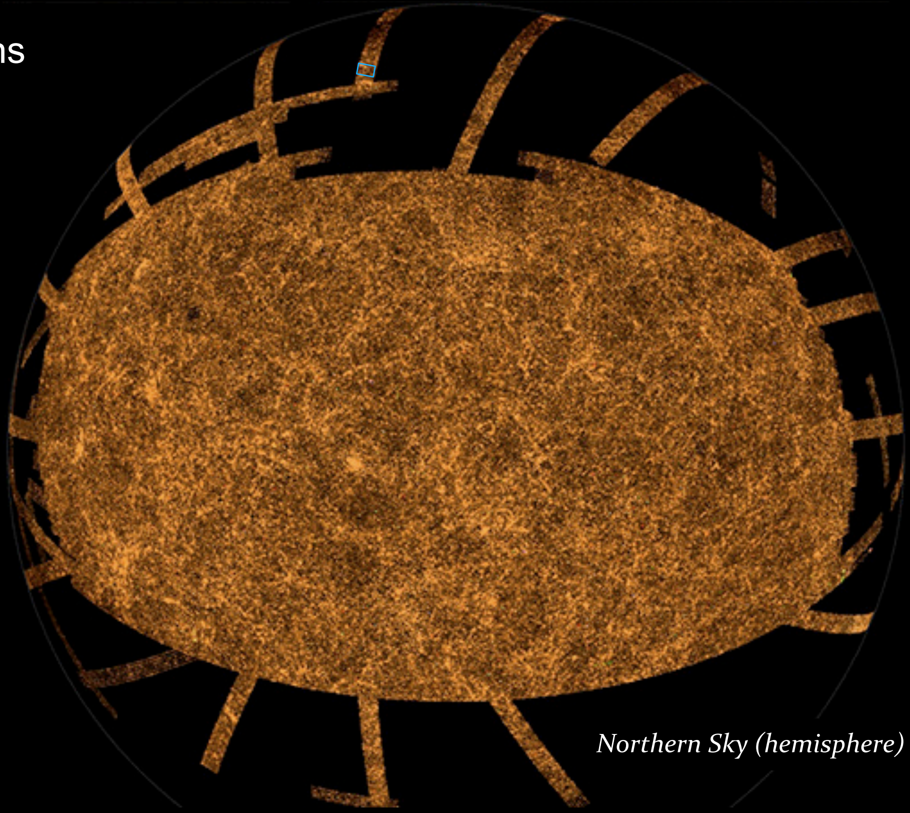


10 years of SDSS

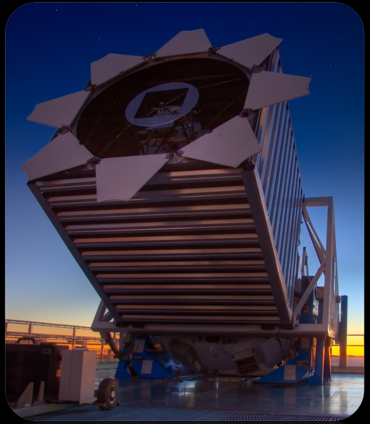
Over 1.2 billion observations of stars and galaxies

Approximately 20 TB of raw imaging data

(1998 – 2009)



Northern Sky (hemisphere)

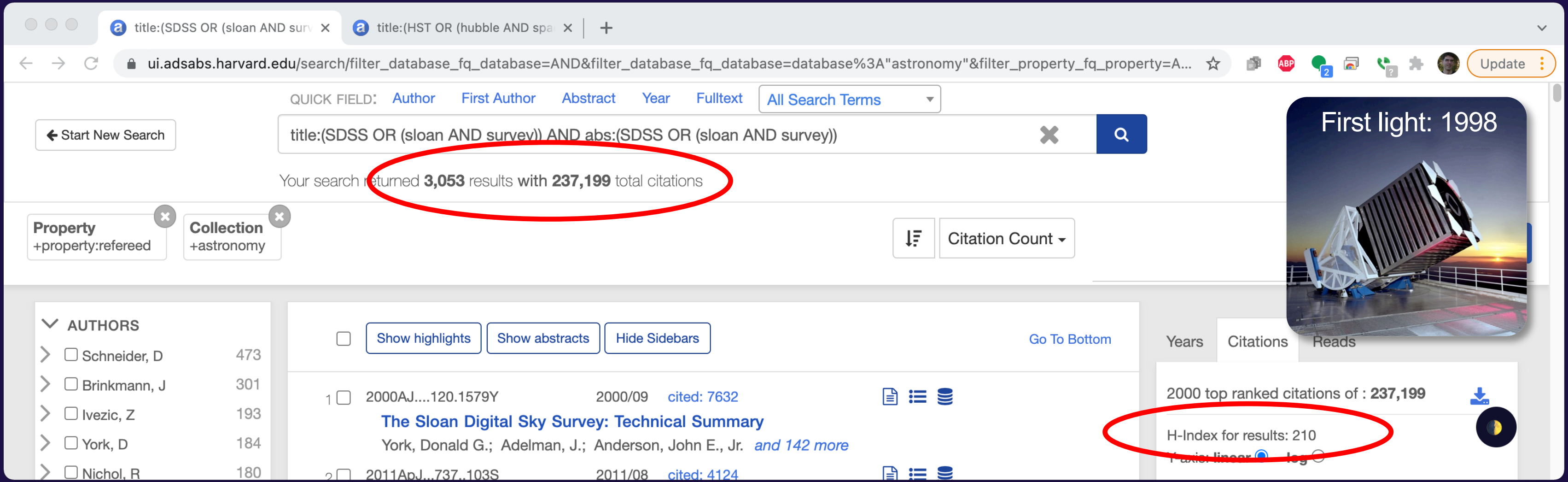


Data in large databases

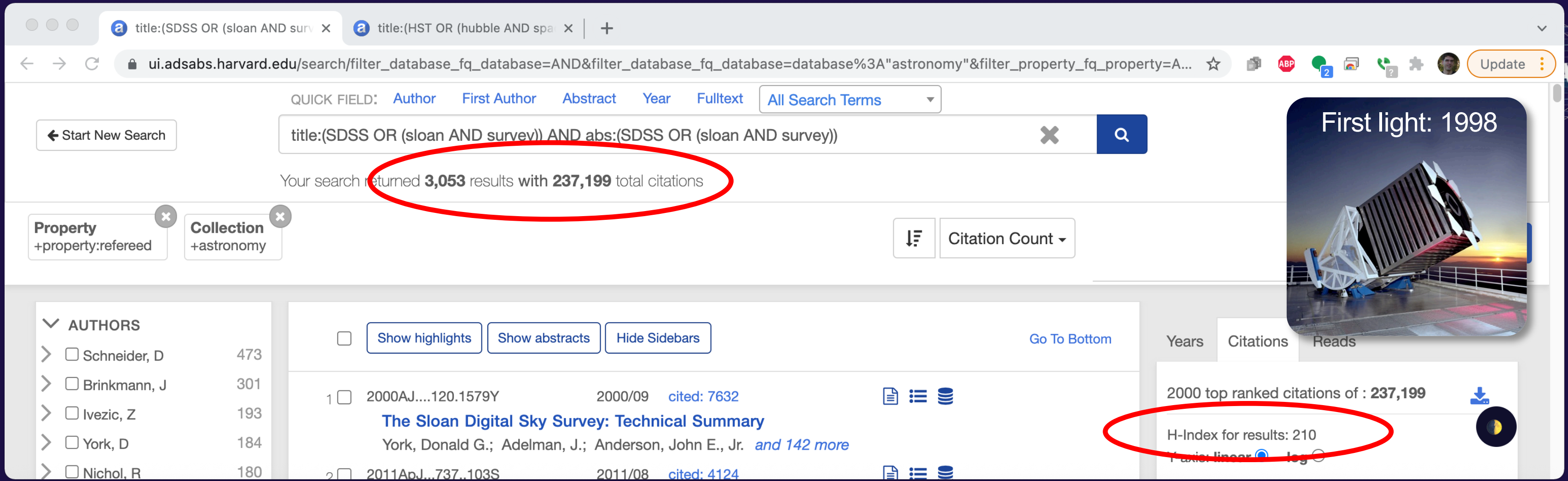
Started by U.W. and Princeton (ARC)

How big of a deal was SDSS?

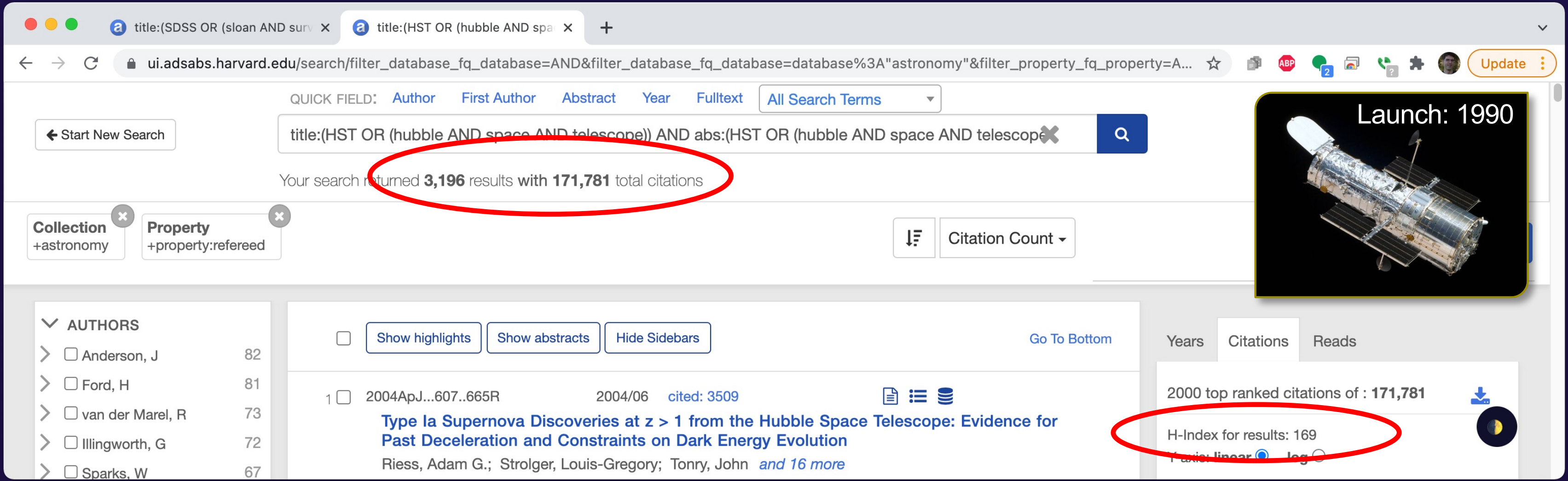
Sloan Digital Sky Survey



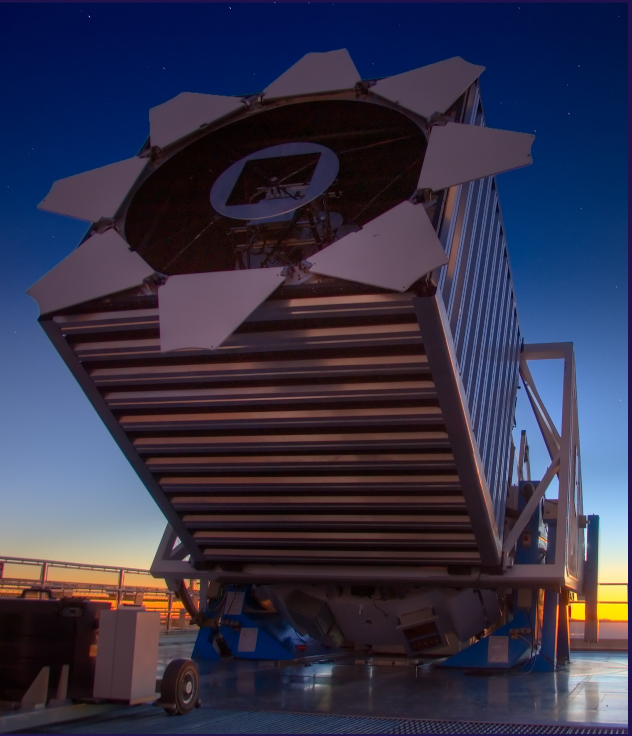
Sloan Digital Sky Survey



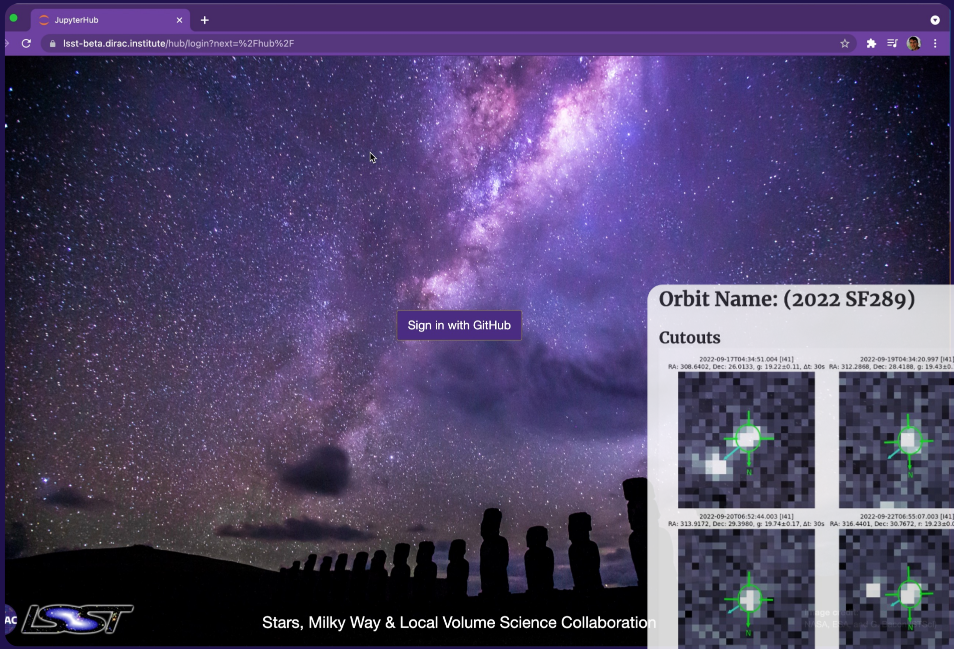
Hubble Space Telescope

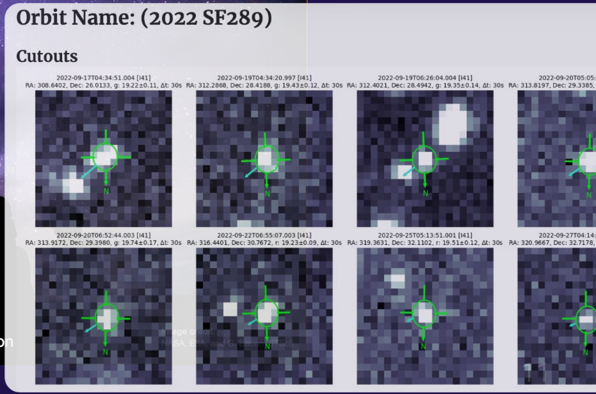


What we *(increasingly)* do.



Most of us don't go to telescopes: we devise algorithms to analyze huge databases to draw inferences about the Universe. We employ machine learning and data science. *Our day-to-day work has more in common with someone at Microsoft or Google, than with our predecessors 20 years ago.*





This is a *dramatic* shift in what an astronomer is, and what skills they need to be successful.

Vision

A Universe Understood through Data-Intensive Discovery

Mission

Develop advanced astronomical datasets, algorithms, and software, and use them to explore and understand the universe.

People

Bring together talent in astronomy, software, and AI, incubate big ideas, seed the next generation of leaders.

Launched in 2017

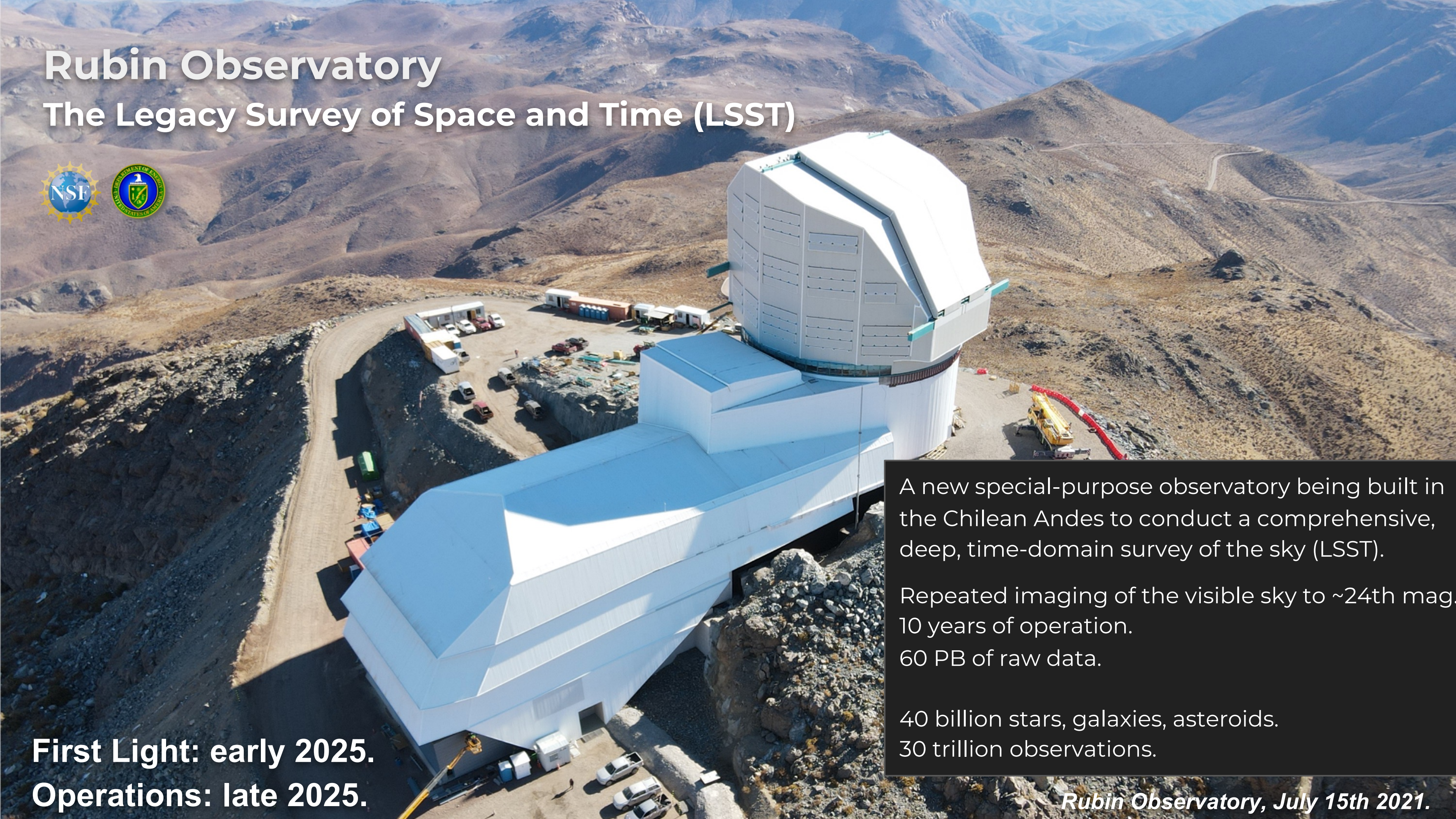
DiRAC Today:

- 7 faculty members
- 17 software eng. and researchers
- 6 postdocs
- 11 graduate students
- 1 administrative staff

DiRAC @ Rubin Annual Meeting 2022

Rubin Observatory

The Legacy Survey of Space and Time (LSST)



A new special-purpose observatory being built in the Chilean Andes to conduct a comprehensive, deep, time-domain survey of the sky (LSST).

Repeated imaging of the visible sky to ~24th mag.
10 years of operation.
60 PB of raw data.

40 billion stars, galaxies, asteroids.
30 trillion observations.

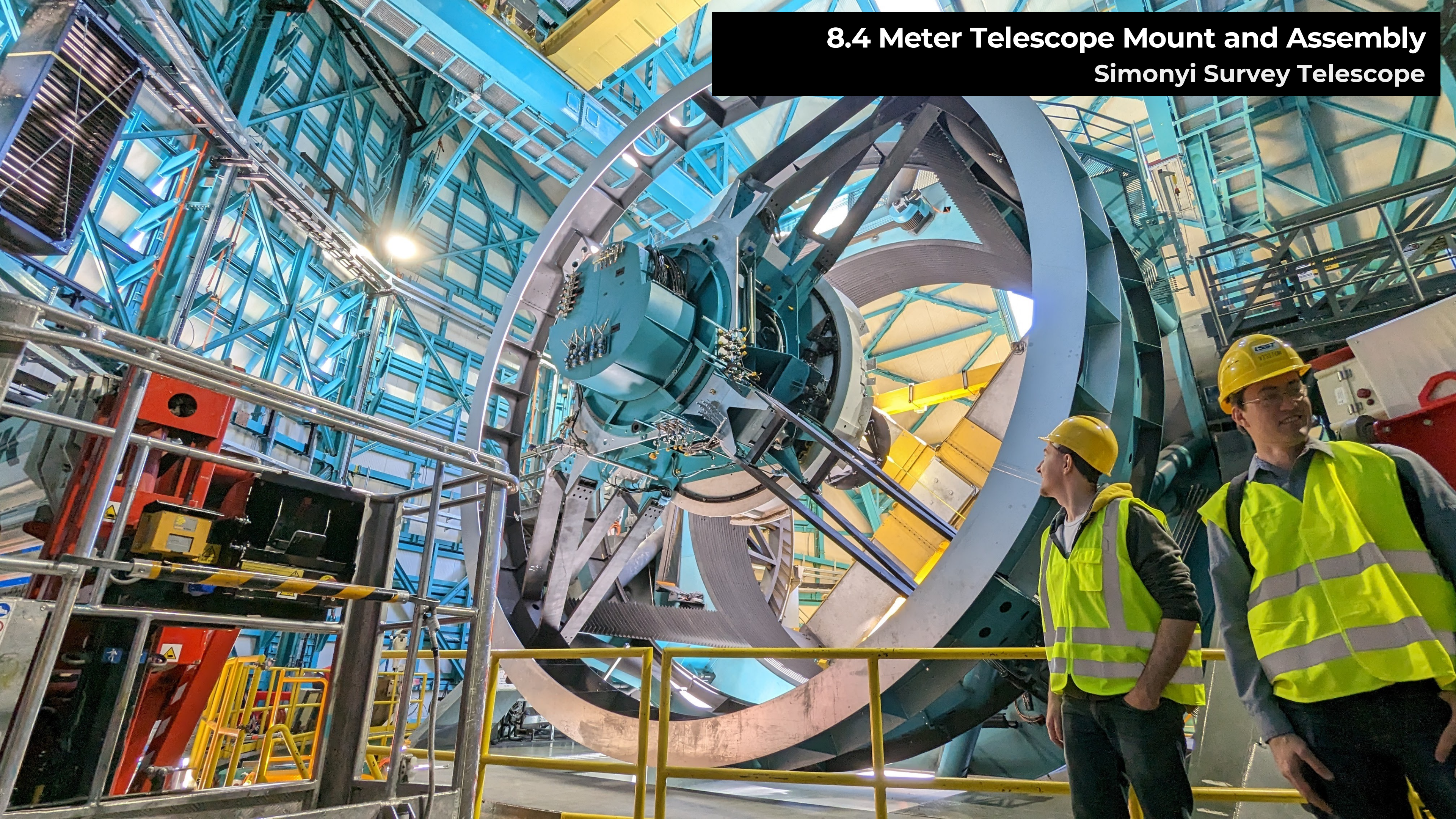
First Light: early 2025.
Operations: late 2025.

Rubin Observatory, July 15th 2021.

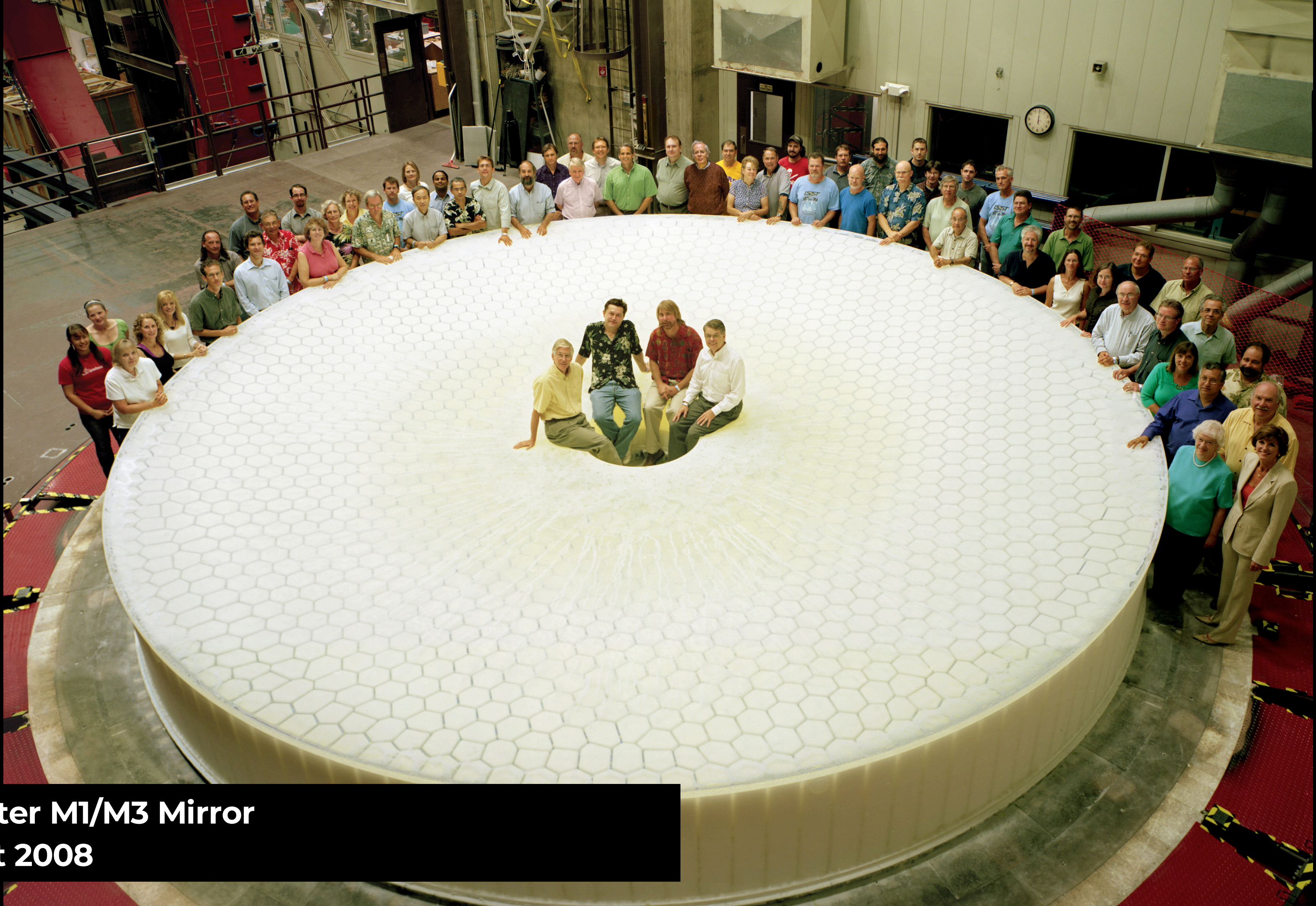

Rubin Observatory, March 15, 2023.
Cerro Pachon, Chile

8.4 Meter Telescope Mount and Assembly

Simonyi Survey Telescope



8.4 Meter Telescope Mount and Assembly

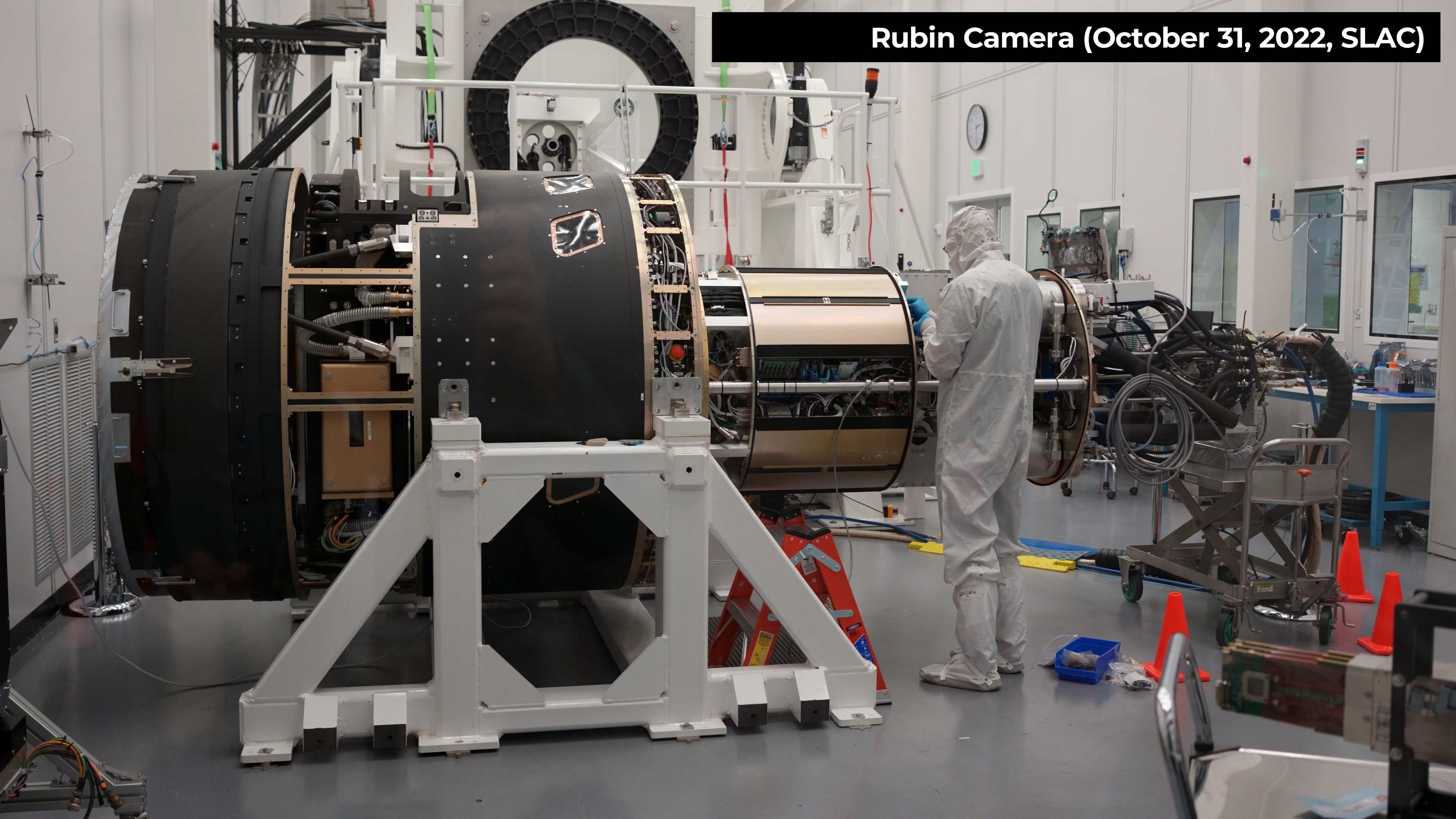


8.4 Meter M1/M3 Mirror
August 2008

Polishing completed in 2015

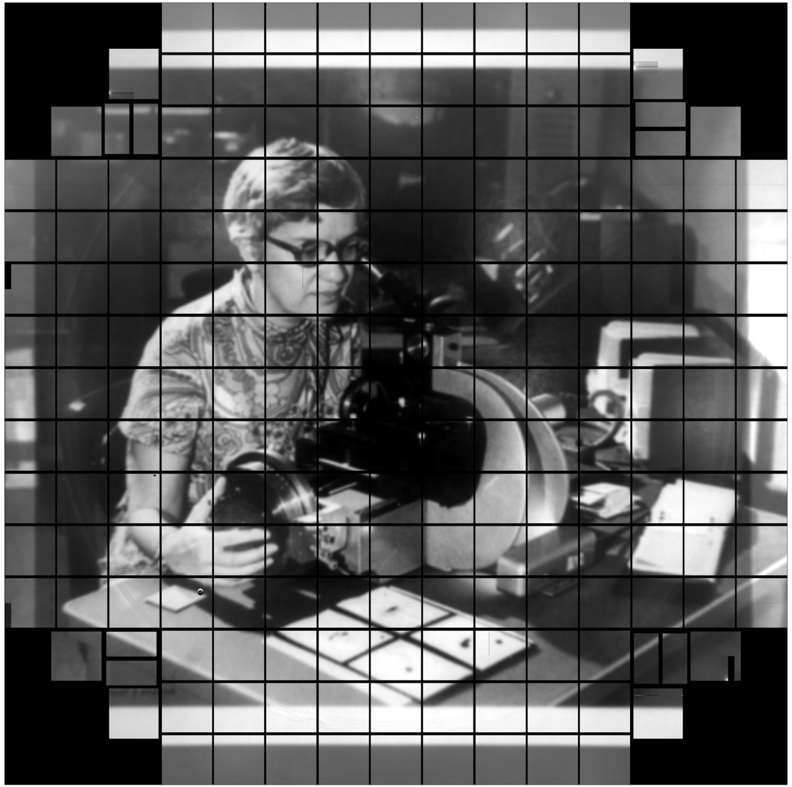

April 27, 2024.
VON ARDENNE
SKYVA 8.4
KONECRANES

Rubin Camera (October 31, 2022, SLAC)



LSSTCam, 21 “rafts”, 189 CCDs

LSSTCam – 21 rafts (189 CCDs, 3.2 Gpix)



Interactive viewer: https://dirac.us/kw3

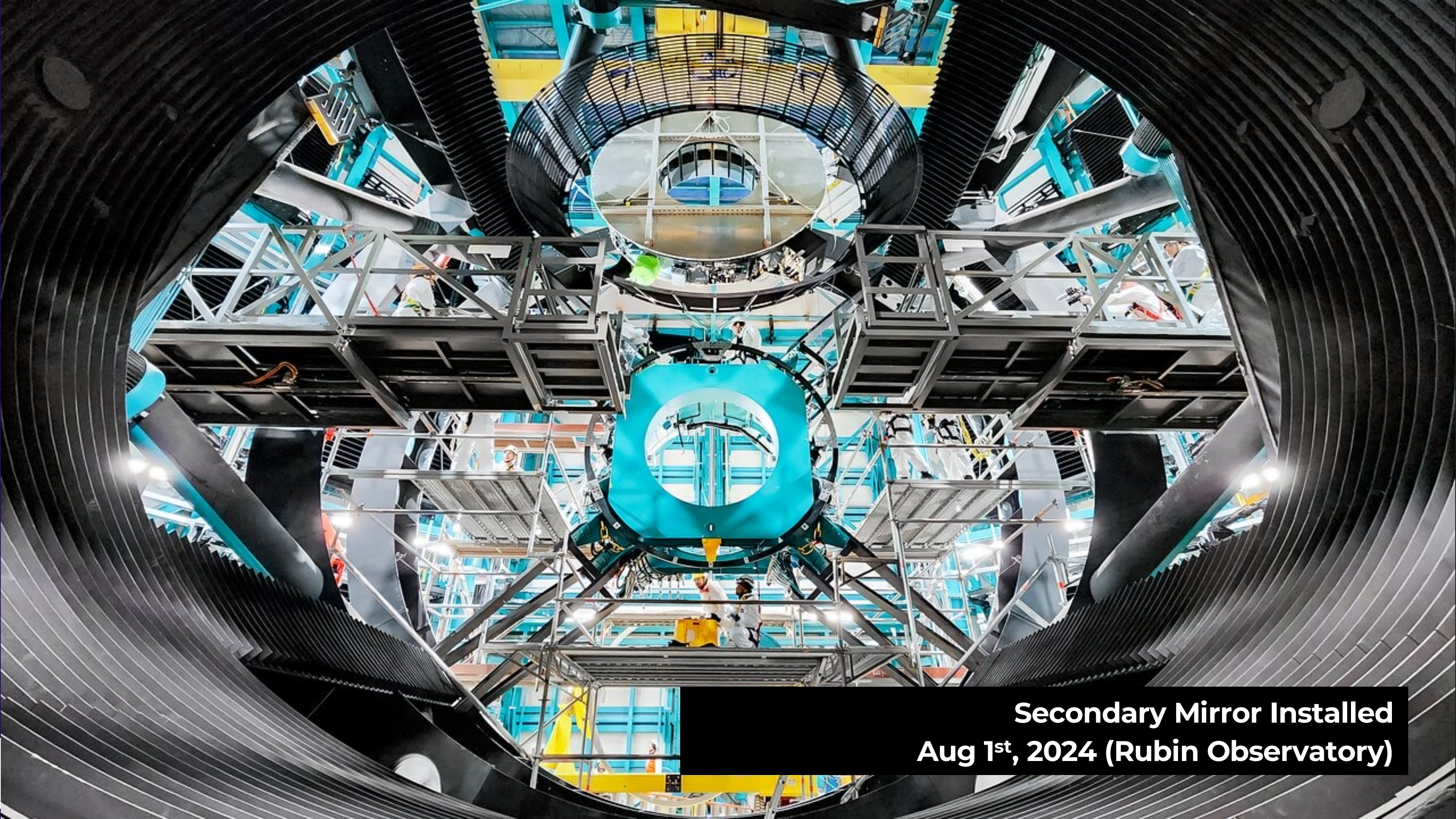

Packed for the move!
May 6th, 2024 (SLAC)

Arrival to Chile
May 15th, 2024 (Santiago)

Arrival to Chile
May 15th, 2024 (Cerro Pachon)

Arrival to Chile

May 15th, 2024 (Rubin Observatory)



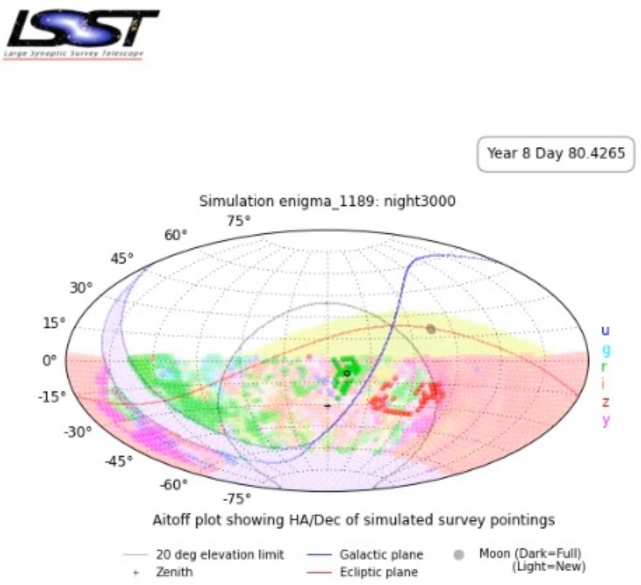
Secondary Mirror Installed
Aug 1st, 2024 (Rubin Observatory)

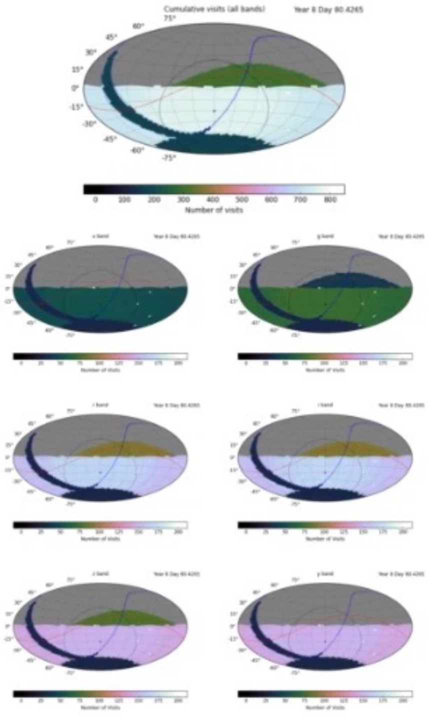
ComCam Reinstalled
Aug 23rd, 2024 (Rubin Observatory)

Primary Mirror Installed

Oct 3rd, 2024 (Rubin Observatory)

LSST: The Legacy Survey of Space and Time





Rubin will execute a single* survey designed to support all four science themes.

How to think about LSST:

- 500 pointings per night
- 2 visits to each pointing (~20 min apart)
- 10 deg^2 per visit, to r~24^{th} mag

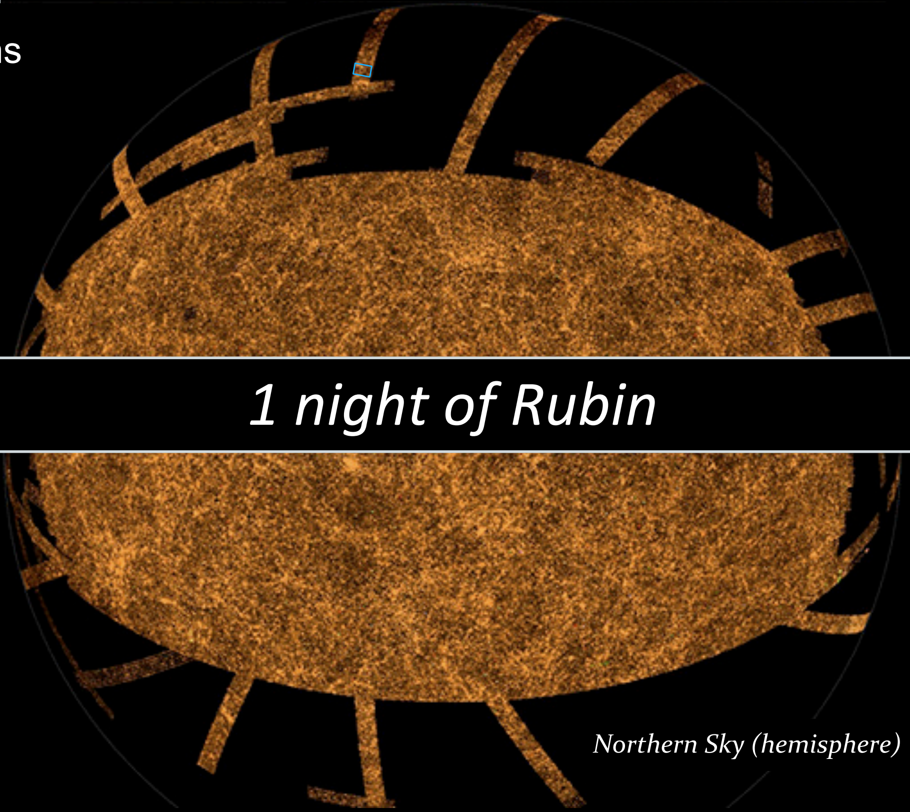
- ~4000 unique deg^2 surveyed per night
- Repeat for ~3300 nights.

(*) There's also smaller (<10% of time) set of "special survey programs" designed to explore extreme corners of discovery space.

10 years of SDSS

Over 1.2 billion observations
of stars and galaxies

Approximately 20 TB of
raw imaging data



1 night of Rubin

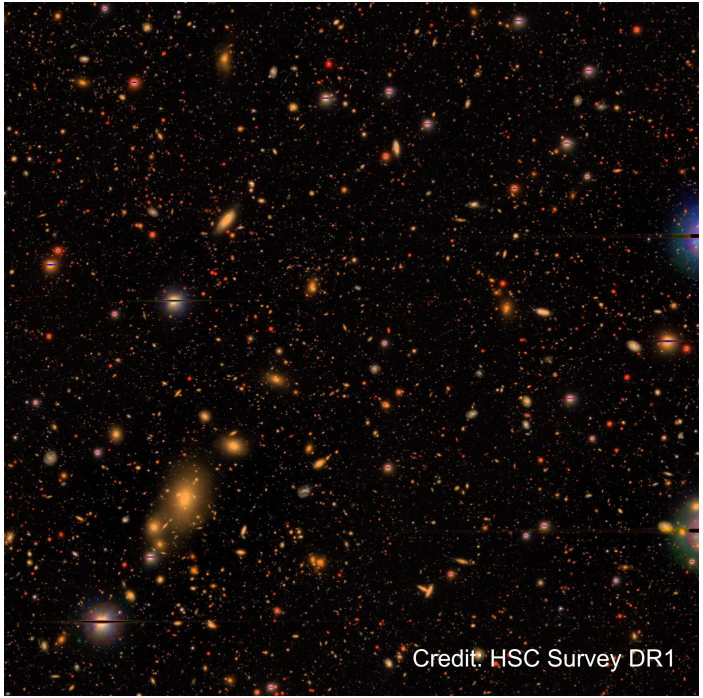
Northern Sky (hemisphere)

SDSS III

VERA C. RUBIN OBSERVATORY

Data Products: Images and Catalogs

Images (instrument-signature corrected)

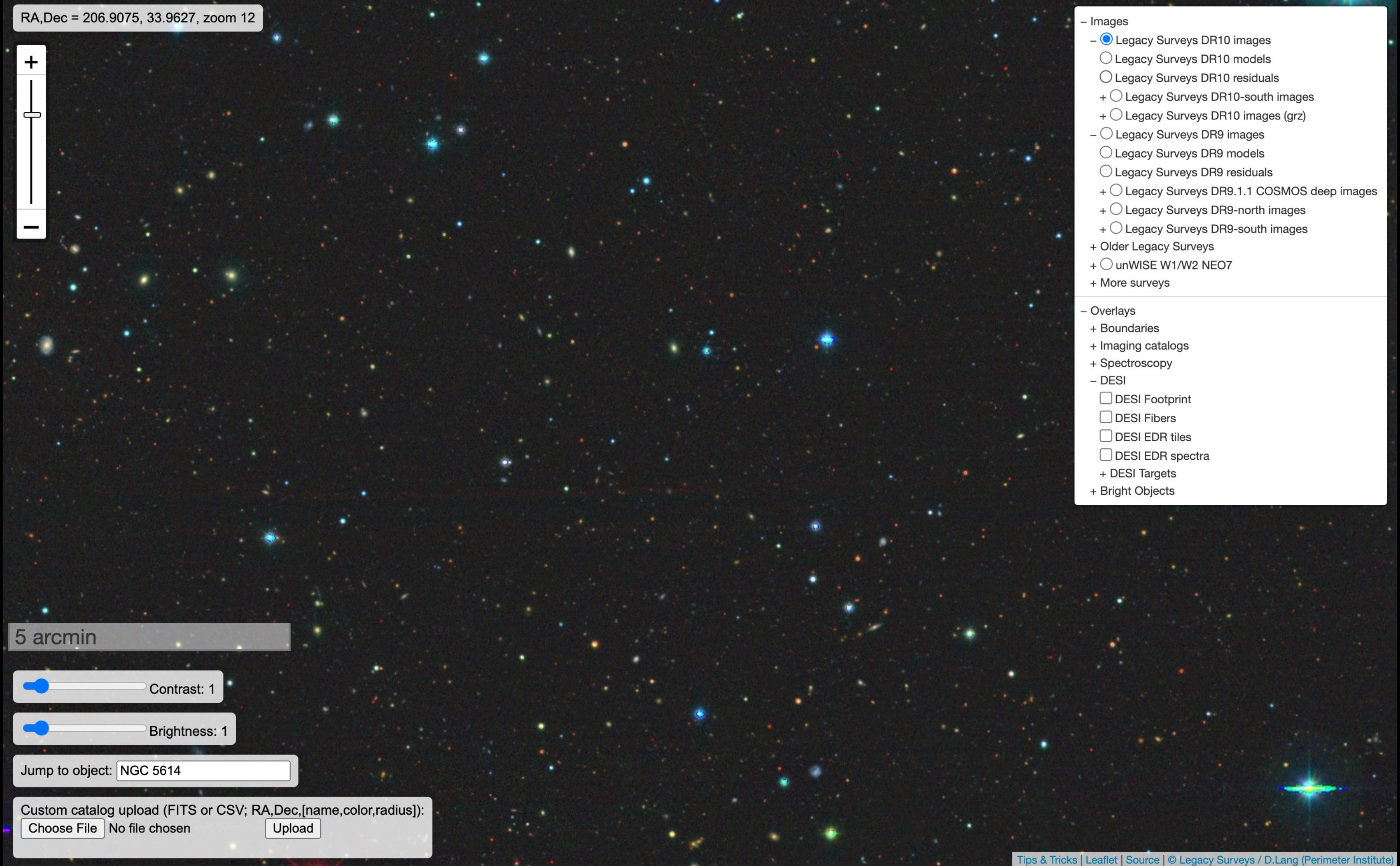

Credit: HSC Survey DR1

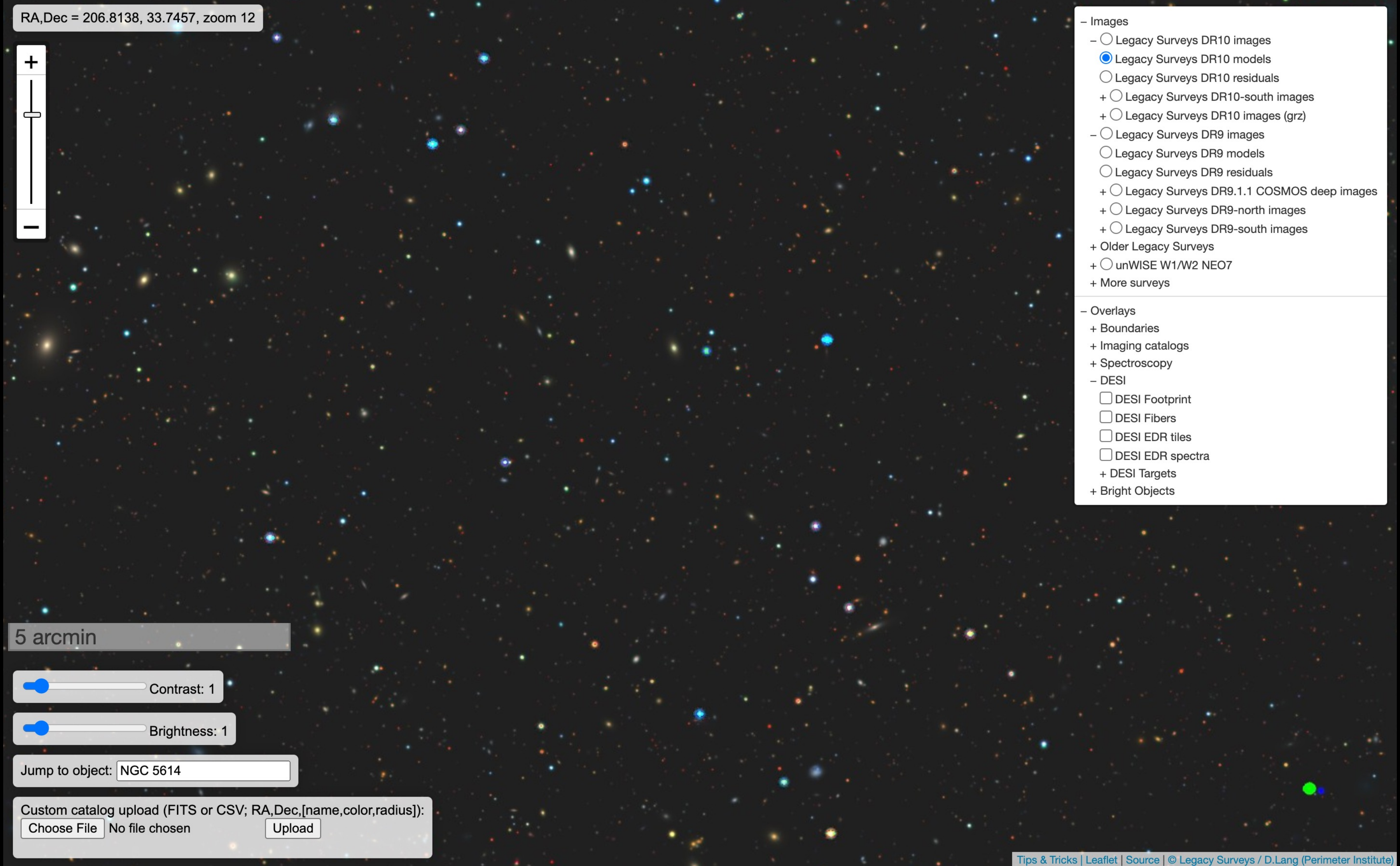
~6 PB/yr, raw (~30 PB/yr processed)

Catalogs (positions & shapes of sources in images; compression)

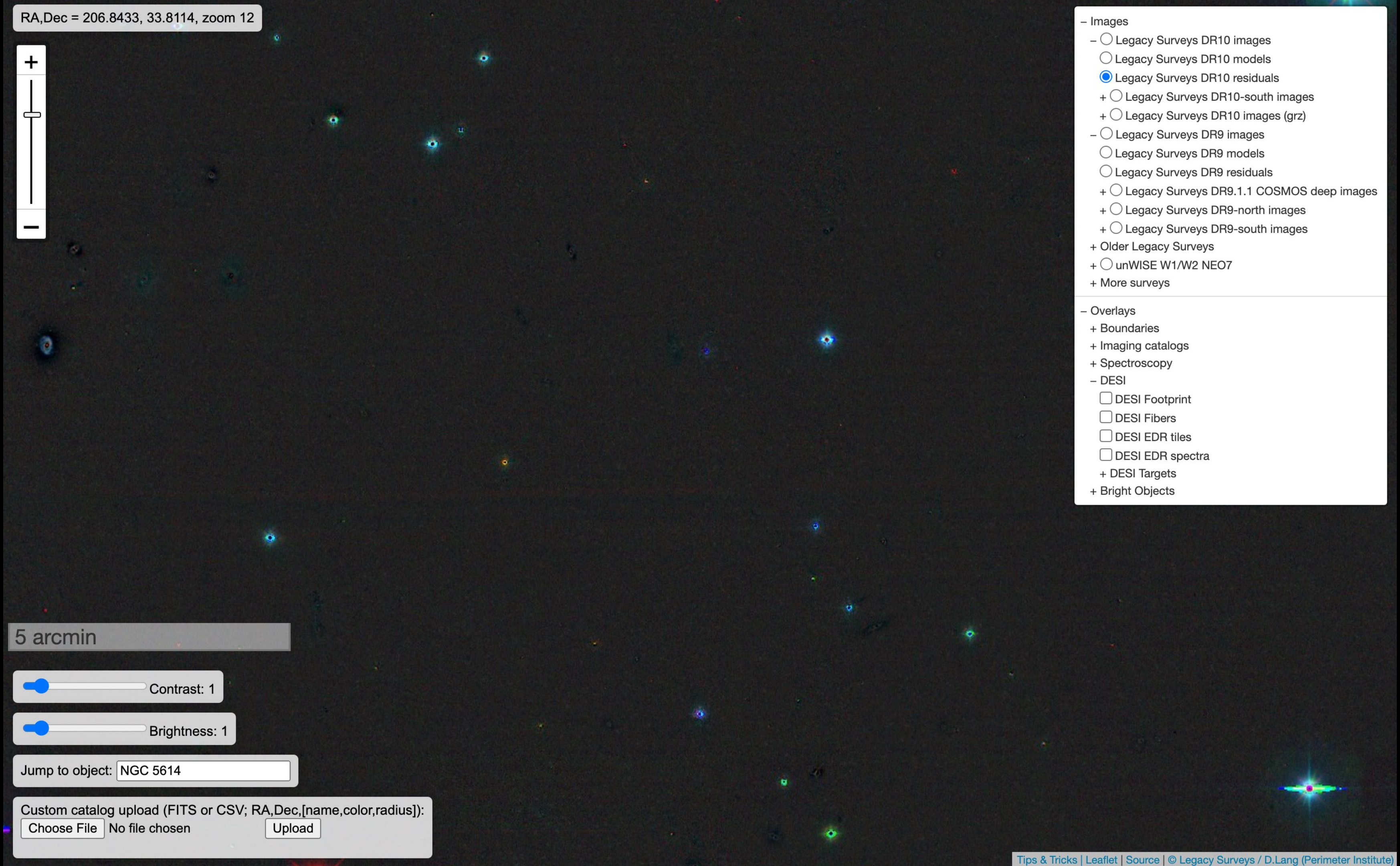
	run	ra	dec	mjdstart	mjdend	node	inclination	mu0	nu0
	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter
1	94	336.432779182593	-1.04429400326262	51075.2332107	51075.45501377	286.855205	0.009477	336.432666780969	-1.05150869113324
2	109	396.241808760683	-1.25055686854694	51078.3907829	51078.47494369	283.391746999999	0.008279	36.2418791557514	-1.25818616551591
3	125	350.469742676909	-1.25274979437412	51081.2557589	51081.49528898	287.818732	0.007781	350.469664290045	-1.25966106871904
4	211	402.581109220535	-1.26517002274148	51115.307	51115.46205485	283.219780000001	0.007975	42.5811958160864	-1.27212059068193
5	240	375.189677877875	-1.26440348484946	51132.185032	51132.24885143	290.578187	0.010103	15.1896568538969	-1.27446183768339
6	241	403.029547848757	-1.26513669884245	51132.2621497	51132.30359089	266.715505000001	0.005148	43.0296301780866	-1.2686924465153
7	250	15.3571787184549	-1.03608421024635	51133.183	51133.36699888	62.095899	0.024055	15.3568830986486	-1.01856644905285
8	251	85.8800045707678	-1.00945333791977	51133.3780874	51133.40792506	11.252511	0.037496	85.8798262897452	-1.04560781770085
9	256	-8.28409345414711	-1.05720709740967	51134.11449765	51134.1335707401	58.141704	0.024019	351.715731123619	-1.03519263888099
10	259	368.375160843834	-1.04718589849226	51134.16041463	51134.39053563	299.408811	0.007597	8.37511083423279	-1.05427670762748
11	273	371.502721507982	-1.25773504769026	51136.164	51136.38085276	286.541530000001	0.008068	11.5027059004616	-1.26577186829035
12	287	396.486846917758	-1.15429721929214	51138.2276088	51138.40424222	295.298232	0.007857	36.4868777389229	-1.16200488663283
13	297	61.1510214928285	-1.15372111751772	51139.293	51139.37260943	92.038416	0.040845	61.1503219956783	-1.13275302170583
14	307	437.534478187481	-0.881981301739572	51140.336	51140.40856068	318.930778	0.012093	77.5345678518198	-0.892598371473539
15	308	107.889971514459	-0.886601061804192	51140.42	51140.50499965	271.301292	0.009619	107.890113955159	-0.883854835297584
16	727	235.005814476688	-1.2718775415526	51251.4677052	51251.49712062	289.317131	0.006134	235.005735183148	-1.26689551184036
17	745	160.35314661036	-1.06151355809997	51257.2457361	51257.49442104	289.613188	0.005337	160.35320906976	-1.05738121457293
18	752	145.143535674411	-1.26788775985279	51258.2008694	51258.4960876901	269.084259	0.009016	145.14364675899	-1.26040794022503
19	756	117.406345241493	-1.05628803448366	51259.1213462	51259.4534920099	251.702115	0.008736	117.406457395396	-1.05003528632892
20	994	438.724151650507	2.98684990306946	51457.45150621	51457.50209137	275.044012000001	15.000775	78.5083218724765	-1.27790383425565
21	1000	359.688420776475	49.1147184578696	51458.2045791901	51458.27672523	181.180816000001	-92.815716	50.30232857747	-1.15241058132403
22	1006	424.134866081084	-1.15955512228338	51458.46843552	51458.4912403	250.771684999999	0.001138	64.1348889620757	-1.15968664695317
23	1009	309.422714173226	-1.05509044188473	51459.18267622	51459.27223683	292.702508	0.004268	309.422638849396	-1.05631833550749
24	1010	13.9590135587482	-3.52523511485643	51459.28625595	51459.31776801	94.929341	2.488812	13.9433618984923	-1.06720362610654
25	1011	356.84619283143	13.7989646493031	51459.36074413	51459.42832922	275.004574000001	15.001812	357.08183302725	-1.05705500327294
26	1013	418.117811163019	-1.15991607429602	51459.44527024	51459.48631885	276.389441	0.00342	58.1178655758971	-1.16203438809252
27	1022	364.056259733763	13.9623875876541	51463.30057066	51463.38100935	274.974967000001	15.005361	4.08325855790555	-1.04112249443299
28	1024	37.8150553881238	-9.46406356801865	51463.39785123	51463.45797293	94.927513	9.991926	37.3182347436184	-1.04373948097949
29	1033	-41.3792270544886	-1.05326958451555	51464.17174245	51464.26918107	37.794552	0.016415	318.620716693897	-1.03714674711003
30	1035	366.987859907795	13.7302914415656	51464.30925102	51464.3818117	274.965525	14.999218	6.93052148978 41	-1.25995233165474

~60 Bn rows/year (~3 trillion/yr for the “forced source” table)


RA,Dec = 206.9075, 33.9627, zoom 12
+
−
5 arcmin
Contrast: 1
Brightness: 1
Jump to object: NGC 5614
Custom catalog upload (FITS or CSV; RA,Dec,[name,color,radius]):
Choose File No file chosen
Upload
– Images
– Legacy Surveys DR10 images
Legacy Surveys DR10 models
Legacy Surveys DR10 residuals
+ Legacy Surveys DR10-south images
+ Legacy Surveys DR10 images (grz)
– Legacy Surveys DR9 images
Legacy Surveys DR9 models
Legacy Surveys DR9 residuals
+ Legacy Surveys DR9.1.1 COSMOS deep images
+ Legacy Surveys DR9-north images
+ Legacy Surveys DR9-south images
+ Older Legacy Surveys
+ unWISE W1/W2 NEO7
+ More surveys
– Overlays
+ Boundaries
+ Imaging catalogs
+ Spectroscopy
– DESI
DESI Footprint
DESI Fibers
DESI EDR tiles
DESI EDR spectra
+ DESI Targets
+ Bright Objects
Tips & Tricks | Leaflet | Source | © Legacy Surveys / D.Lang (Perimeter Institute)



RA,Dec = 206.8138, 33.7457, zoom 12
5 arcmin
Contrast: 1
Brightness: 1
Jump to object: NGC 5614
Custom catalog upload (FITS or CSV; RA,Dec,[name,color,radius]):
Choose File No file chosen
Upload
– Images
– Legacy Surveys DR10 images
Legacy Surveys DR10 models
Legacy Surveys DR10 residuals
+ Legacy Surveys DR10-south images
+ Legacy Surveys DR10 images (grz)
– Legacy Surveys DR9 images
Legacy Surveys DR9 models
Legacy Surveys DR9 residuals
+ Legacy Surveys DR9.1.1 COSMOS deep images
+ Legacy Surveys DR9-north images
+ Legacy Surveys DR9-south images
+ Older Legacy Surveys
+ unWISE W1/W2 NEO7
+ More surveys
– Overlays
+ Boundaries
+ Imaging catalogs
+ Spectroscopy
– DESI
DESI Footprint
DESI Fibers
DESI EDR tiles
DESI EDR spectra
+ DESI Targets
+ Bright Objects
Tips & Tricks | Leaflet | Source | © Legacy Surveys / D.Lang (Perimeter Institute)


RA,Dec = 206.8433, 33.8114, zoom 12
+
−
5 arcmin
Contrast: 1
Brightness: 1
Jump to object: NGC 5614
Custom catalog upload (FITS or CSV; RA,Dec,[name,color,radius]):
Choose File No file chosen
Upload
− Images
− Legacy Surveys DR10 images
Legacy Surveys DR10 models
Legacy Surveys DR10 residuals
+ Legacy Surveys DR10-south images
+ Legacy Surveys DR10 images (grz)
− Legacy Surveys DR9 images
Legacy Surveys DR9 models
Legacy Surveys DR9 residuals
+ Legacy Surveys DR9.1.1 COSMOS deep images
+ Legacy Surveys DR9-north images
+ Legacy Surveys DR9-south images
+ Older Legacy Surveys
+ unWISE W1/W2 NEO7
+ More surveys
− Overlays
+ Boundaries
+ Imaging catalogs
+ Spectroscopy
− DESI
DESI Footprint
DESI Fibers
DESI EDR tiles
DESI EDR spectra
+ DESI Targets
+ Bright Objects
Tips & Tricks | Leaflet | Source | © Legacy Surveys / D.Lang (Perimeter Institute)

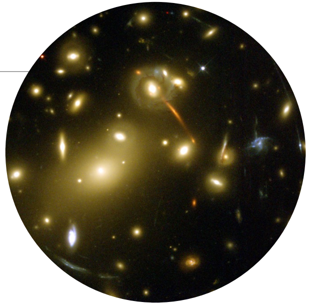
Rubin, Science, and AI

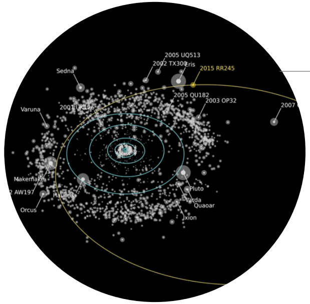
LSST Science Themes

LSST Overview: [Ivezic et al. (2019)]
LSST Data Products Definition: [Juric et al. (2013)]

Probing Dark Matter & Dark Energy

- Strong & Weak Lensing
- Large Scale Structure
- Galaxy Clusters, Supernovae



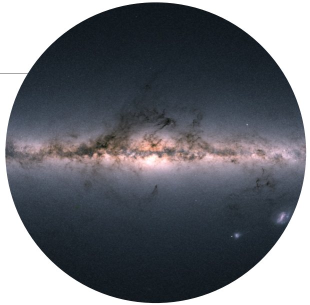


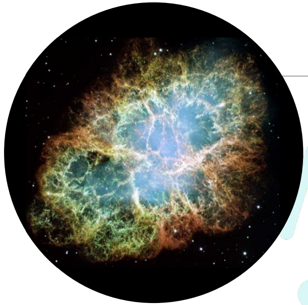
Inventory of the Solar System

- Comprehensive small body census
- Comets and ISOs
- Planetary defence

Mapping the Milky Way

- Structure and evolutionary history
- Spatial maps of stellar characteristics
- Reach well into the halo

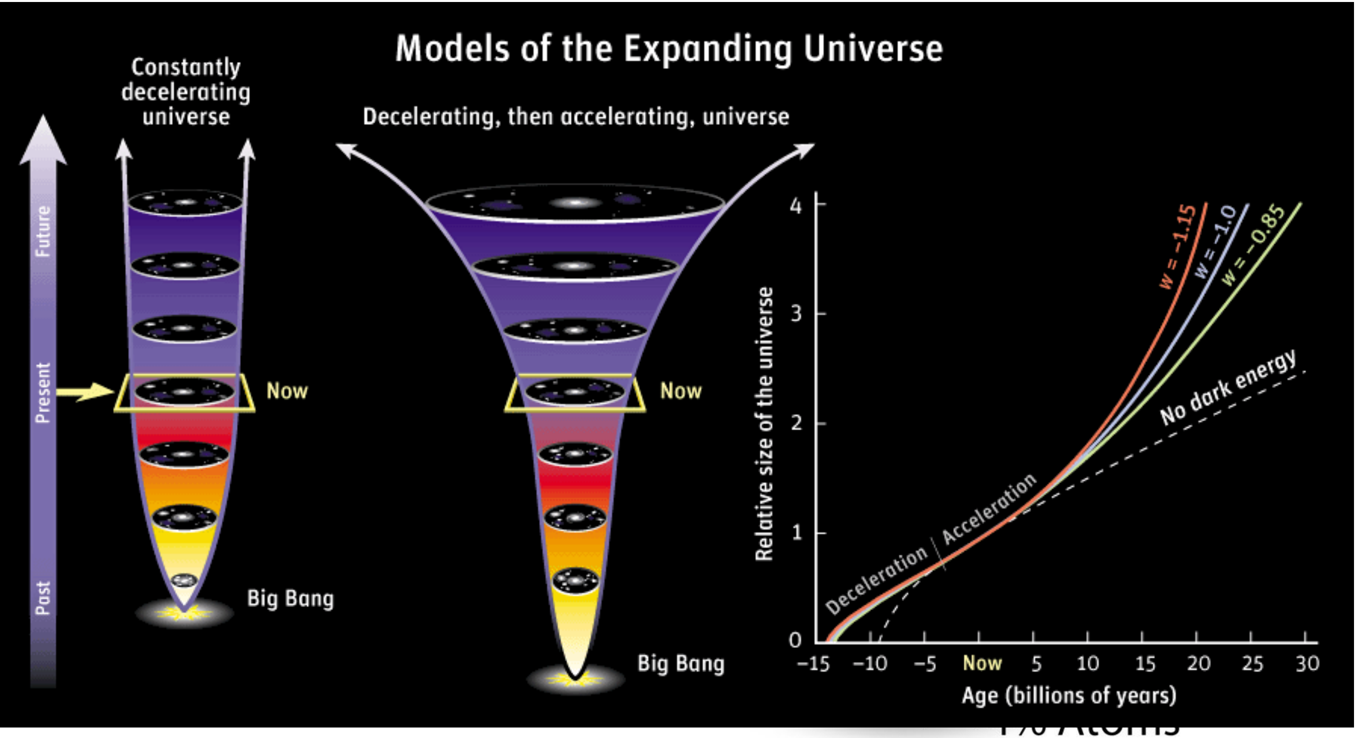


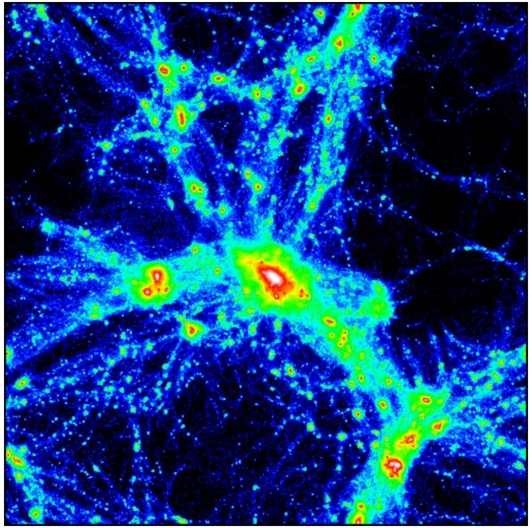


Exploring the Transient Optical Sky

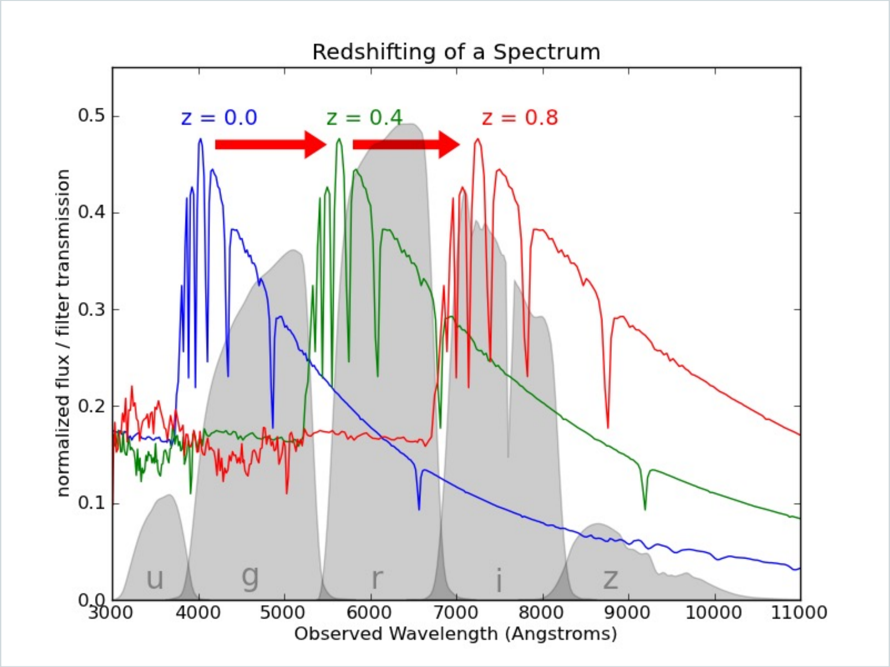
- Variable stars, Supernovae
- Fill in the variability phase-space
- Discovery of new classes of transients

Probing Dark Matter and Dark Energy

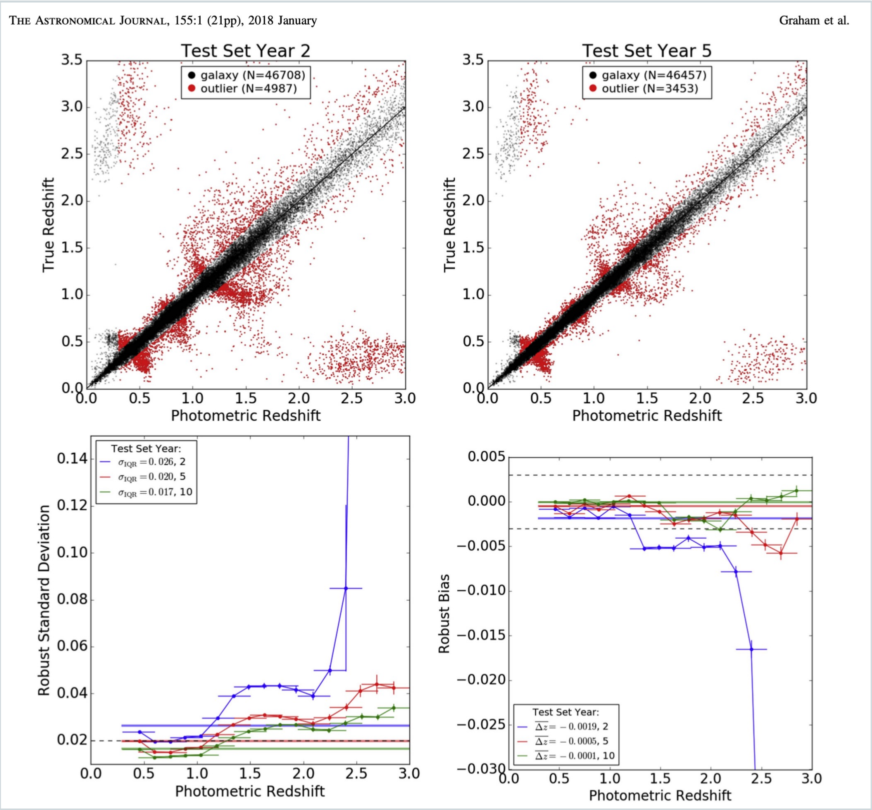




Probing Dark Matter and Dark Energy



☝ https://ogrisel.github.io/scikit-learn.org/sklearn-tutorial/tutorial/astronomy/regression.html

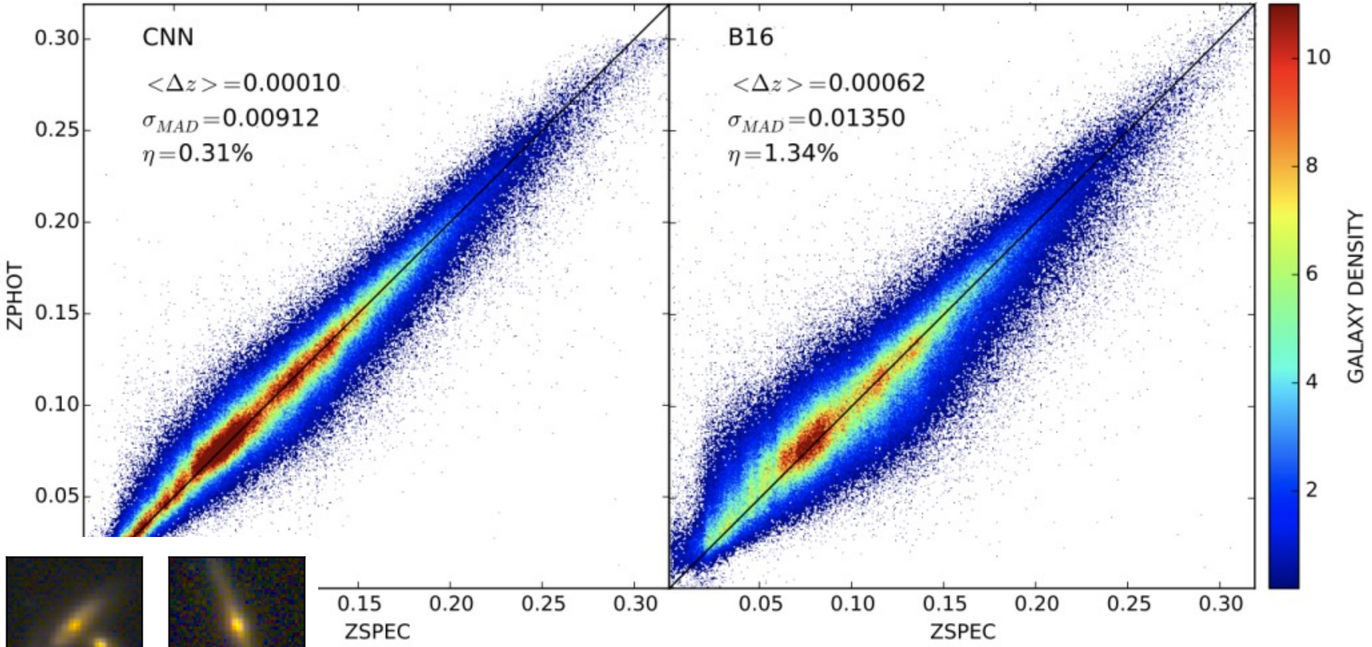


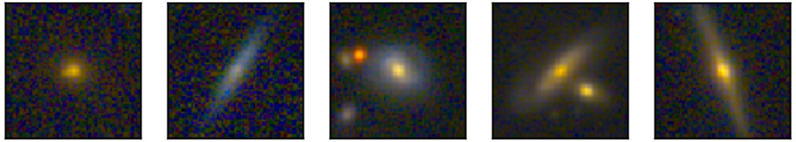
Graham et al. (2018)

Photo-Z Estimation with AI

Photo Z model trained on 400k **images** of galaxies from SDSS, for which spectroscopic redshifts are available.

Outperforms best known photo-Z estimators.





z=0.258 z=0.039 z=0.094 z=0.163 z=0.078

Pasquet et al. 2018

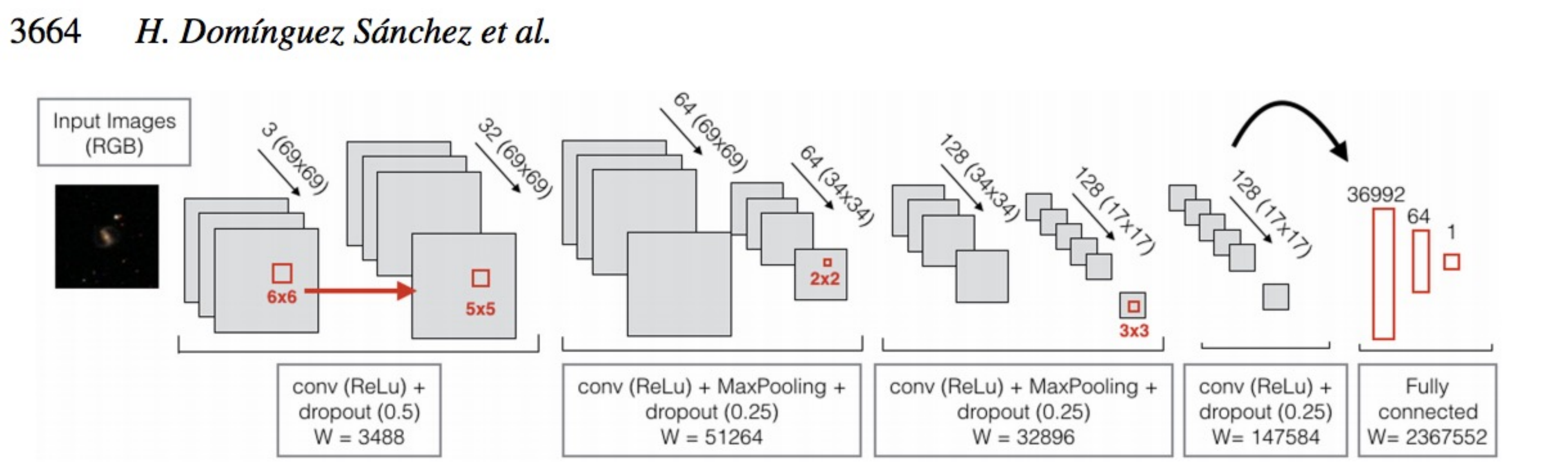
Galaxy classification

The bright spiral galaxy M51 and its fainter companion

(https://www.sdss4.org/science/)

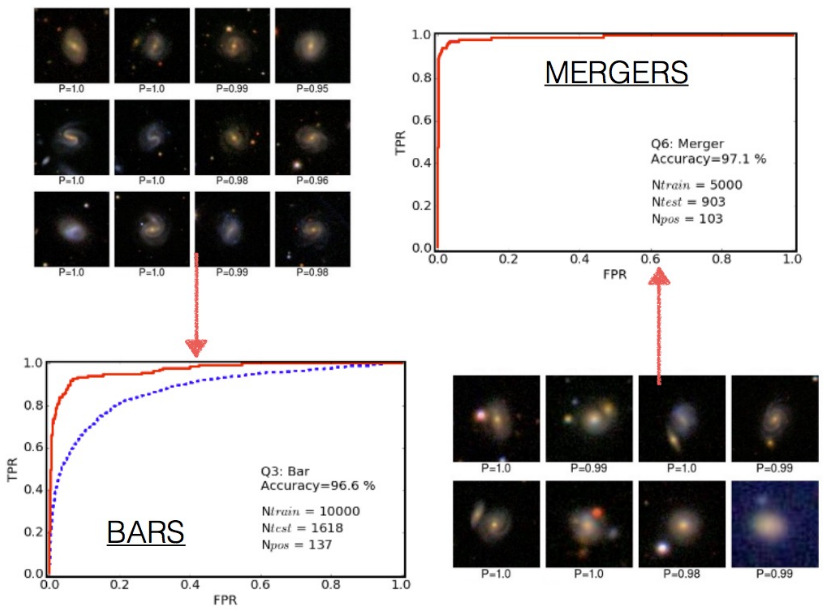
Can an AI find merging or barred galaxies?

A (simple) CNN trained on ~15k galaxies, used to classify 670k galaxies in SDSS.
Four convolutional layers, and a fully connected layer (2 million parameters).



Dominguez Sanchez et al. (2018)

How well does it do?



Outperforms SVM-based models.

Large accuracy (> 97\%) for distinguishing between disk features/bars/edge or face on galaxies/etc..

Note: this is using a fairly simple CNN - improvements are likely.

Most impressively, it's conceptually simple!

```
#========= Model definition=======

#Convolutional Layers

model = Sequential()
model.add(Convolution2D(32, 6,6, border_mode='same',
                    input_shape=(img_channels, img_rows, img_cols)))
model.add(Activation('relu'))
model.add(Dropout(0.5))

model.add(Convolution2D(64, 5, 5, border_mode='same'))
model.add(Activation('relu'))

model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Dropout(0.25))

model.add(Convolution2D(128, 2, 2, border_mode='same'))
model.add(Activation('relu'))

model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Dropout(0.25))

model.add(Convolution2D(128, 3, 3, border_mode='same'))
model.add(Activation('relu'))

model.add(Dropout(0.25))

#Fully Connected start here
#------------------------------#

model.add(Flatten())
model.add(Dense(64, activation='relu'))
model.add(Dropout(.5))
model.add(Dense(1, init='uniform', activation='sigmoid'))

print("Compilation...")

model.compile(loss='binary_crossentropy',optimizer='adam',metrics=['accuracy'])
```

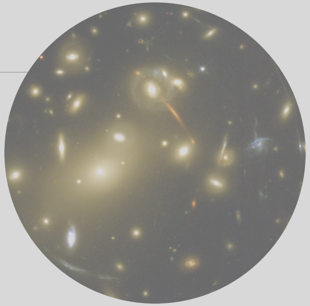
Implementation in Keras
(courtesy of M. Huertas-Company)

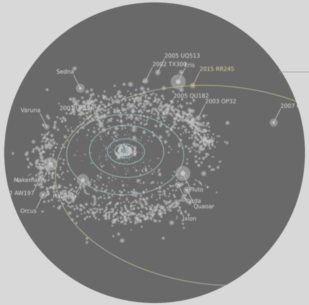
LSST Science Themes

LSST Overview: [Ivezic et al. (2019)]
LSST Data Products Definition: [Juric et al. (2013)]

Probing Dark Matter & Dark Energy

- Strong & Weak Lensing
- Large Scale Structure
- Galaxy Clusters, Supernovae



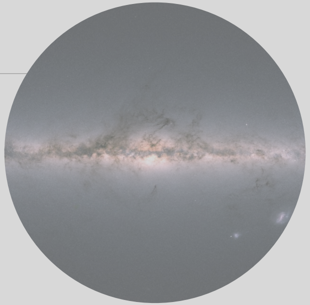


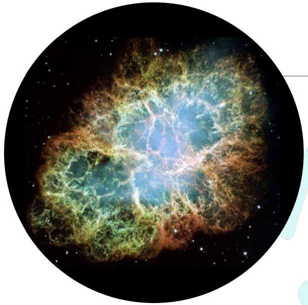
Inventory of the Solar System

- Comprehensive small body census
- Comets and ISOs
- Planetary defence

Mapping the Milky Way

- Structure and evolutionary history
- Spatial maps of stellar characteristics
- Reach well into the halo



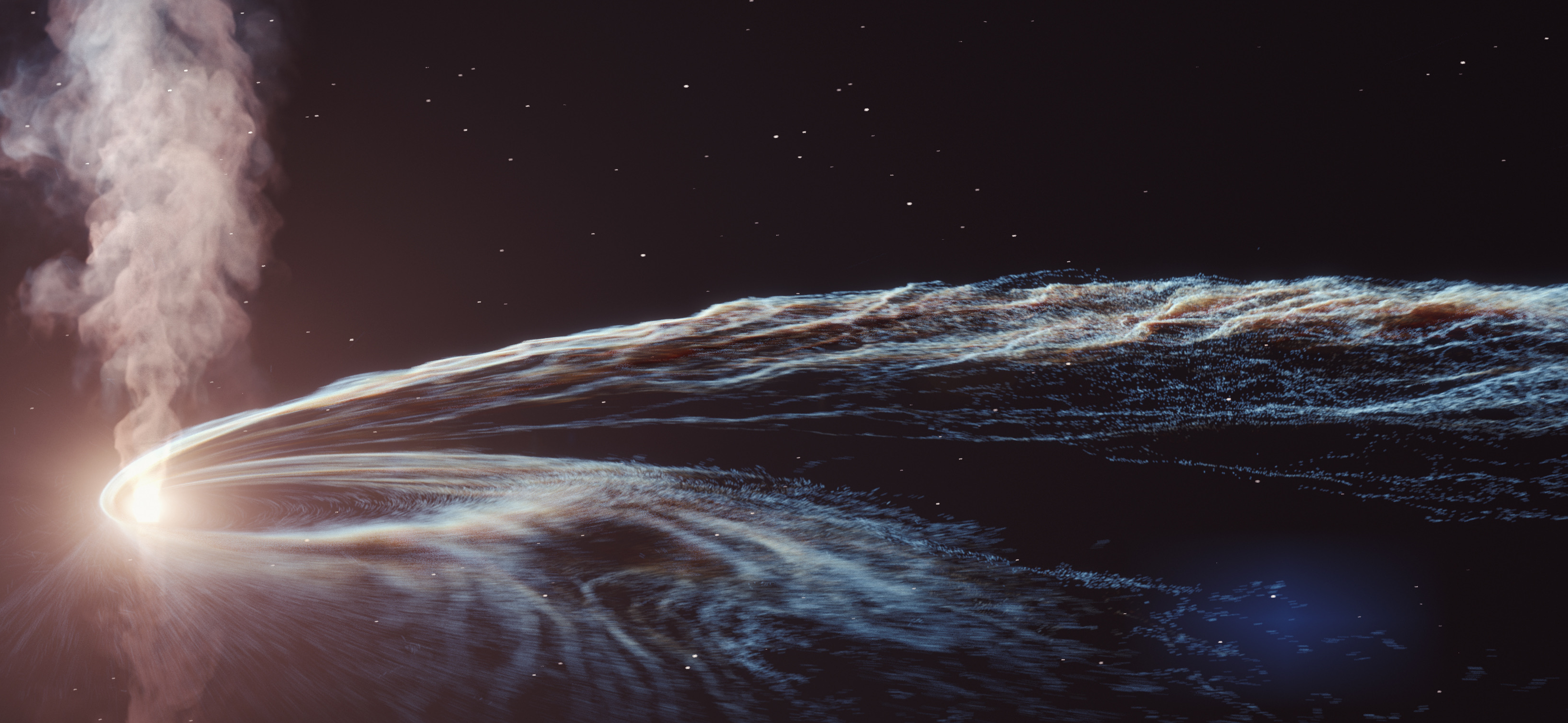


Exploring the Transient Optical Sky

- Variable stars, Supernovae
- Fill in the variability phase-space
- Discovery of new classes of transients

The Transient Universe and Time-Domain Astronomy

ZTF Alert System
Maria Patterson et al.



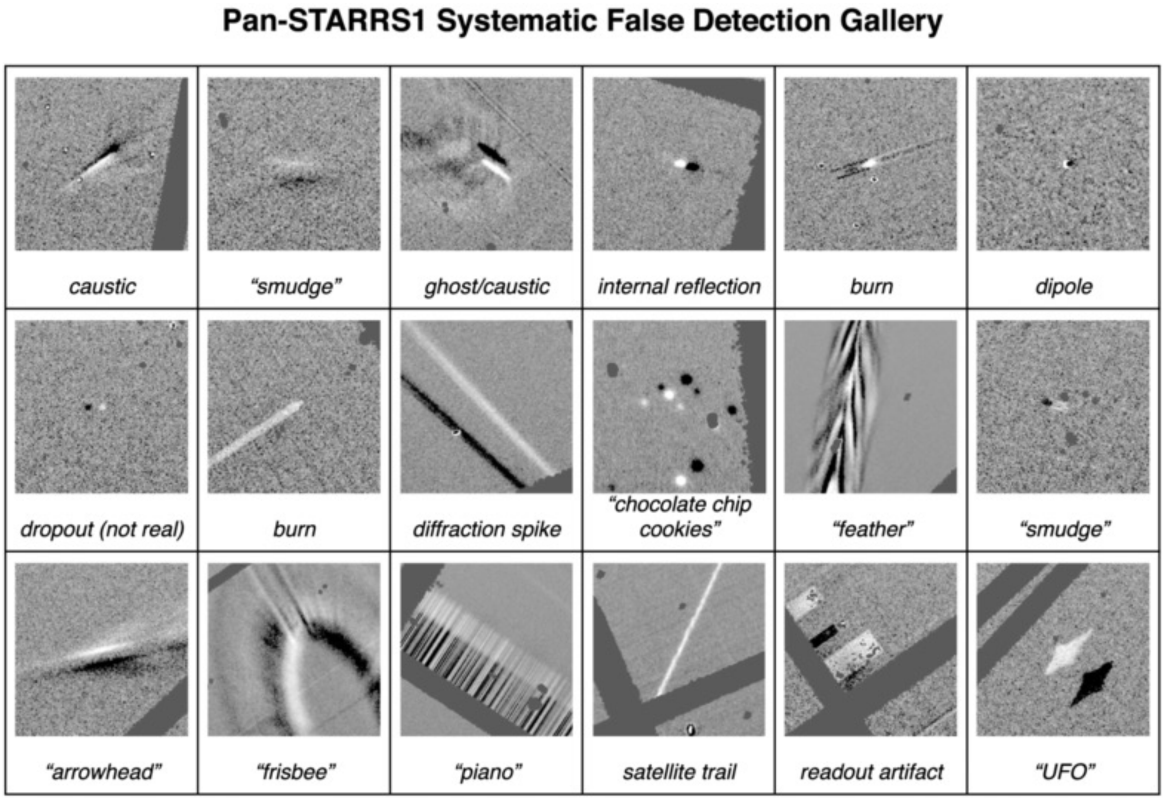
When a star gets too close to a supermassive black hole, the gravitational forces tear it apart with some of the star's material thrown back in space to form a disk around the black hole. One such event, called a Tidal Disruption Event (TDE), was discovered by ZTF in April 2019 and is believed to be the source of a high-energy neutrino caught by the IceCube Neutrino Observatory.

Above: Tidal Disruption Events: Stars Shredded by Black Holes

Studying time domain: Image Differencing

Deep Lens Survey (Witman et al. 2002)

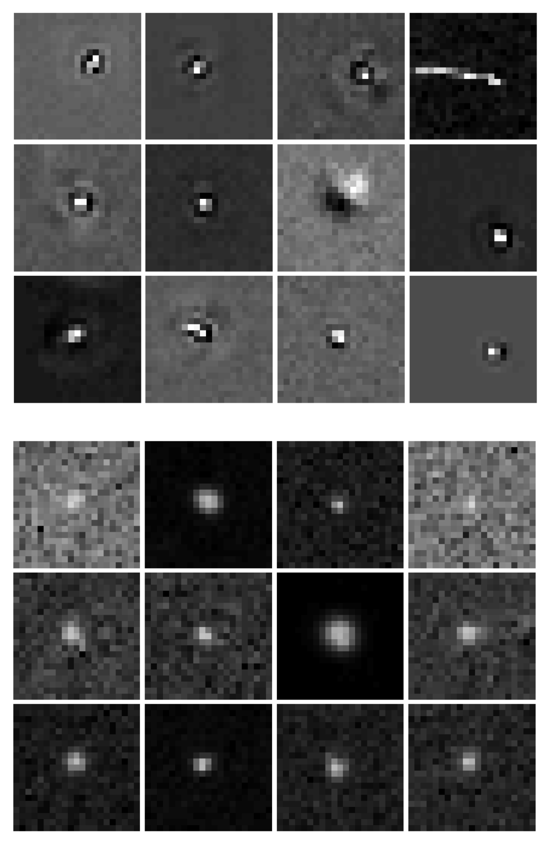
Differencing introduces (many) artifacts



Imperfect image differencing can lead to spurious detections in image differences, that usually outnumber (by ratios up to 100:1) the real objects.

PanSTARRS Survey, Denneau et al. (2013)

The cleanup: “Real-Bogus Classifiers”



6 *H. Brink et al.*

Set	Selected	Feature	Description
RB1		mag	USNO-B1.0 derived magnitude of the candidate on the difference image
		mag_err	estimated uncertainty on mag
		a_image	semi-major axis of the candidate
	✓	b_image	semi-minor axis of the candidate
		fwhm	full-width at half maximum (FWHM) of the candidate
	✓	flag	numerical representation of the SExtractor extraction flags
	✓	mag_ref	magnitude of the nearest object in the reference image if less than 5 arcsec from the candidate
	✓	mag_ref_err	estimated uncertainty on mag_ref
	✓	a_ref	semi-major axis of the reference source
	✓	b_ref	semi-minor axis of the reference source
		n2sig3	number of at least negative 2 σ pixels in a 5×5 box centered on the candidate
		n3sig3	number of at least negative 3 σ pixels in a 5×5 box centered on the candidate
		n2sig5	number of at least negative 2 σ pixels in a 7×7 box centered on the candidate
	✓	n3sig5	number of at least negative 3 σ pixels in a 7×7 box centered on the candidate
	✓	flux_ratio	ratio of the aperture flux of the candidate relative to the aperture flux of the reference source
		ellipticity	ellipticity of the candidate using a_image and b_image
	✓	ellipticity_ref	ellipticity of the reference source using a_ref and b_ref
	✓	nn_dist_renorm	distance in arcseconds from the candidate to reference source
		magdiff	when a reference source is found nearby, the difference between the candidate magnitude and the reference source. Else, the difference between the candidate magnitude and the limiting magnitude of the image
	✓	maglim	True if there is no nearby reference source, False otherwise.
		sigflux	significance of the detection, the PSF flux divided by the estimated uncertainty in the PSF flux
		seeing_ratio	ratio of the FWHM of the seeing on the new image to the FWHM of the seeing on the reference image
	✓	mag_from_limit	limiting magnitude minus the candidate magnitude
		normalized_fwhm	ratio of the FWHM of the candidate to the seeing in the new image
	✓	normalized_fwhm_ref	ratio of the FWHM of the reference source to the seeing in the reference image
	✓	good_cand_density	ratio of the number of candidates in that subtraction to the total usable area on that array
	✓	min_distance_to_edge_in_new	distance in pixels to the nearest edge of the array on the new image
New	✓	ccdid	numerical ID of the specific camera detector (1 – 12)
		sym	Measure of symmetry, based on dividing the object into quadrants
	✓	seeingnew	FWHM of the seeing on the new image
	✓	extracted	number of candidates on that exposure found by Sextractor
	✓	obsaved	number of candidates on that exposure saved to the database (a subset of extracted)
		pos	True for a positive (i.e., brighter) residual, False for a negative (fading) one
	✓	gauss	gaussian best fit sqaured difference value
		corr	gaussian best fit correlation value
		scale	gaussian scale value
	✓	amp	gaussian amplitude value
	✓	l1	sum of absolute pixel values
		smooth1	filter 1 output
		smooth2	filter 2 output
		pca1	1st principal component
		pca2	2nd principal component
Test		empty	zero for all candidates (i.e., no information)
		random	a random number generated for every candidate (i.e., pure noise)

Table 1. List of all of the features used in our analysis. The first set of features, labeled ‘RB1’, were first introduced by Bloom et al. (2011) and we repeat here their Table 1. The second, labeled ‘New’ is introduced here. The last set of features, called ‘Test’ serves as a benchmark for feature selection in §3.1, where we expect good features to perform better than these. The check-marked as ‘selected’ represent the optimal subset found by our incremental feature selection algorithm in §3.1.

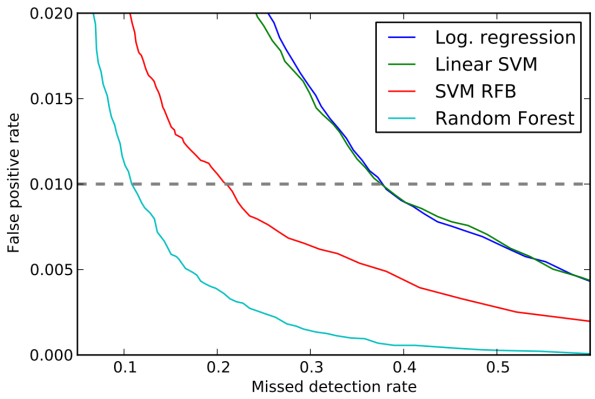


Figure 3. Comparison of a few well known classification algorithms applied to the full dataset. ROC curves enable a trade-off between false positives and missed detections, but the best classifier pushes closer towards the origin. Linear models (Logistic Regression or Linear SVMs) perform poorly as expected, while non-linear models (SVMs with radial basis function kernels or random forests) are much more suited for this problem. Random forests perform well with minimal tuning and efficient training, so we will use those in the remainder of this paper.

Problem: In transient searches, image differencing generates many artefacts (”false positive detections”). These overwhelm real candidates (by ~100:1).

Solution: RF-based classifiers.

Pioneering work by Bloom et al (2011)
Figures from Brink et al (2012)

Real Bogus, Dark Energy Survey (~2015)

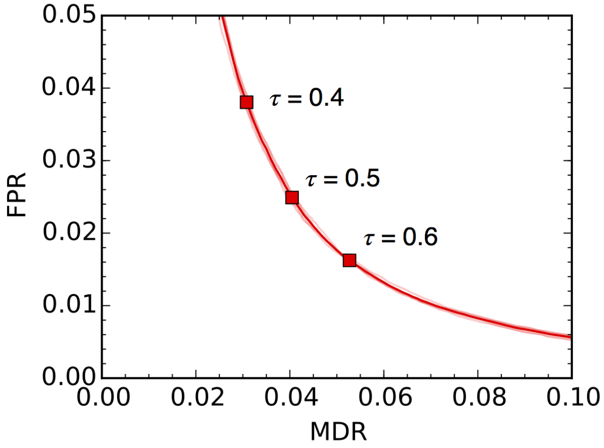


Fig. 7.— 5-fold cross-validated receiver operating characteristics of the best-performing classifier from §3.5. Six visually indistinguishable curves are plotted: one translucent curve for each round of cross-validation, and one opaque curve representing the mean. Points on the mean ROC corresponding to different class discrimination boundaries τ are labeled. $\tau = 0.5$ was adopted in DES-SN.

TABLE 4

CAN ON REPROCESSED DES Y1 TRANSIENT CANDIDATE

	No ML	ML ($\tau = 0.5$)	ML / No ML
N_c [a]	100,450	7,489	0.075
$\langle N_A/N_{NA} \rangle$ [b]	13	0.34	0.027
ϵ_F [c]	1.0	0.990	0.990

[a] Total number of science candidates discovered.

[b] Average ratio of artifact to non-artifact detections in human scanning pool.

[c] `autoScan` candidate-level efficiency for fake SNe Ia.

- **Raw false detection rates of 13:1**
- **Post-filtering rates of 1:3**

State of the art: ZTF braa

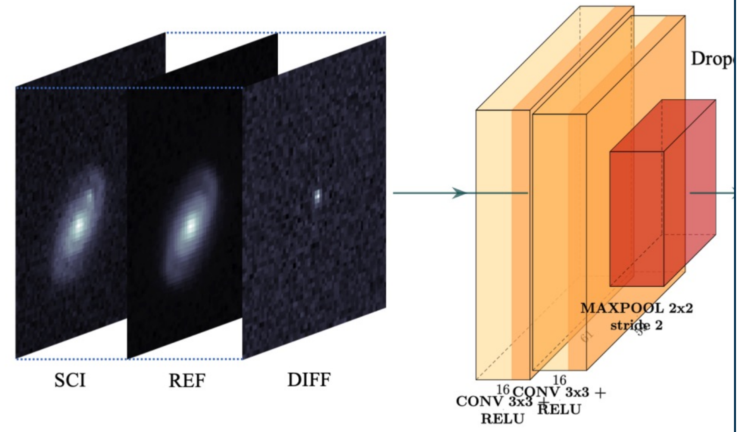


Figure 7. Architecture of the custom VGG6 model. The L^2-norm
stacked to form 63x63x3 triplets that are input into the model. ReL
a sigmoid activation function is used for the output layer that prod
Section 2.2 for the details.

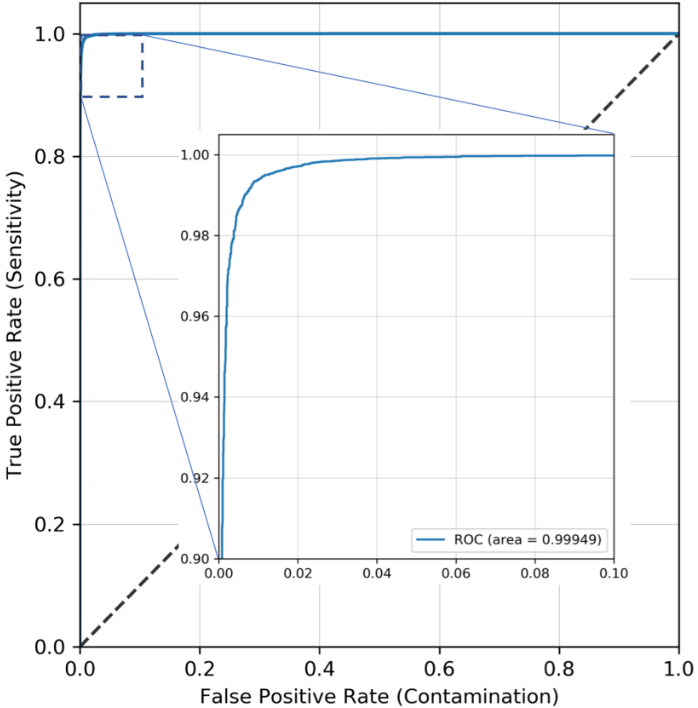
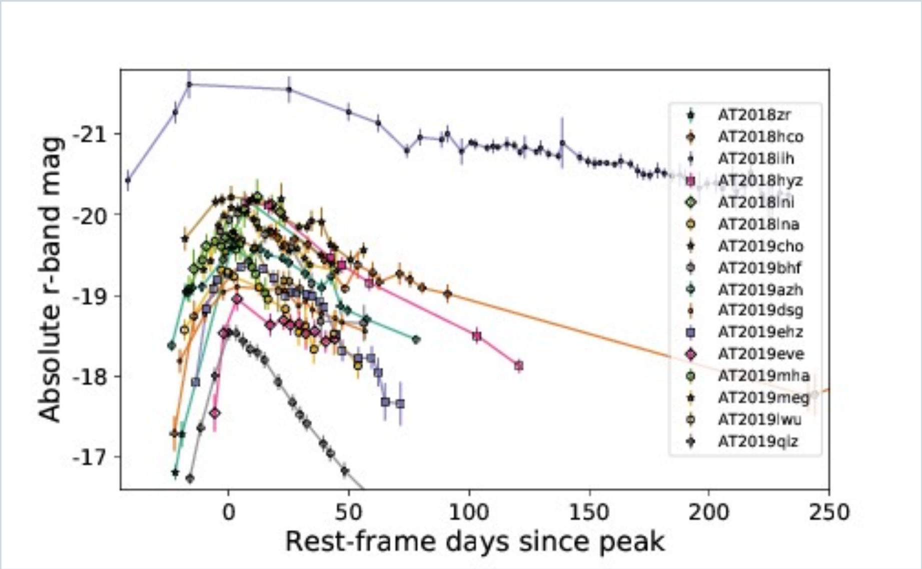


Figure 10. ROC curve of **braai** version *d6_m7* that is deployed in production as of June 2019.

Duev et al. (2019): “Bogus-Real Adversarial AI”

The Transient Universe and Time-Domain Astronomy

ZTF Alert System
Maria Patterson et al.



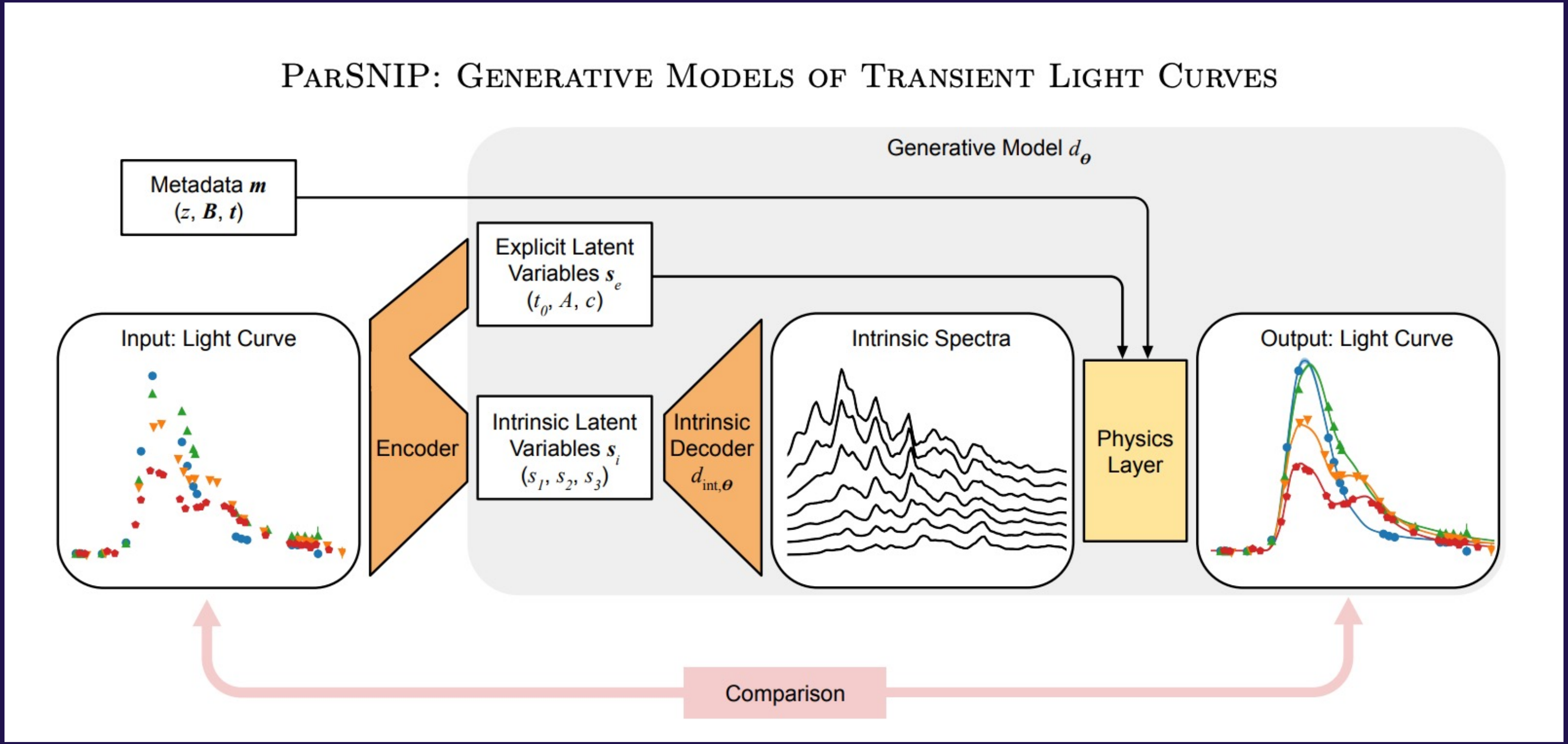
When a star gets too close to a supermassive black hole, the gravitational forces tear it apart with some of the star's material thrown back in space to form a disk around the black hole. One such event, called a Tidal Disruption Event (TDE), was discovered by ZTF in April 2019 and is believed to be the source of a high-energy neutrino caught by the IceCube Neutrino Observatory.

Above: Tidal Disruption Events: Stars Shredded by Black Holes

What am I looking at? Real-time transient classification.

ParSNIP: Generative Models of Transient Light Curves with Physics-Enabled Deep Learning

Boone (2021)

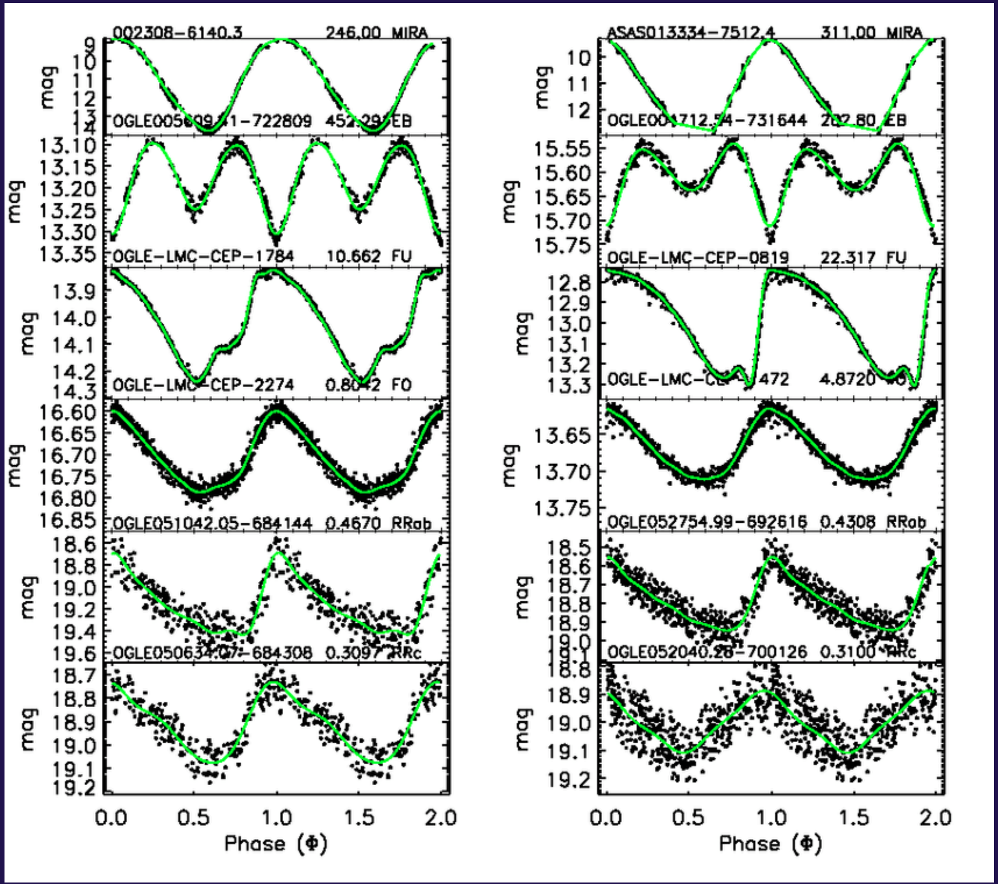


Hybrid model: uses a neural network to model the unknown intrinsic diversity of different transients and an explicit physics-based model of how light from the transient propagates through the universe and is observed.

2x better (contamination) at Type-Ia SNe identification relative to PLAsTiCC SOTA model

Periodic Variability Classification

Classifying variable sources in (noisy, sparse) survey datasets.



Variable examples: Deb & Singh (2009)

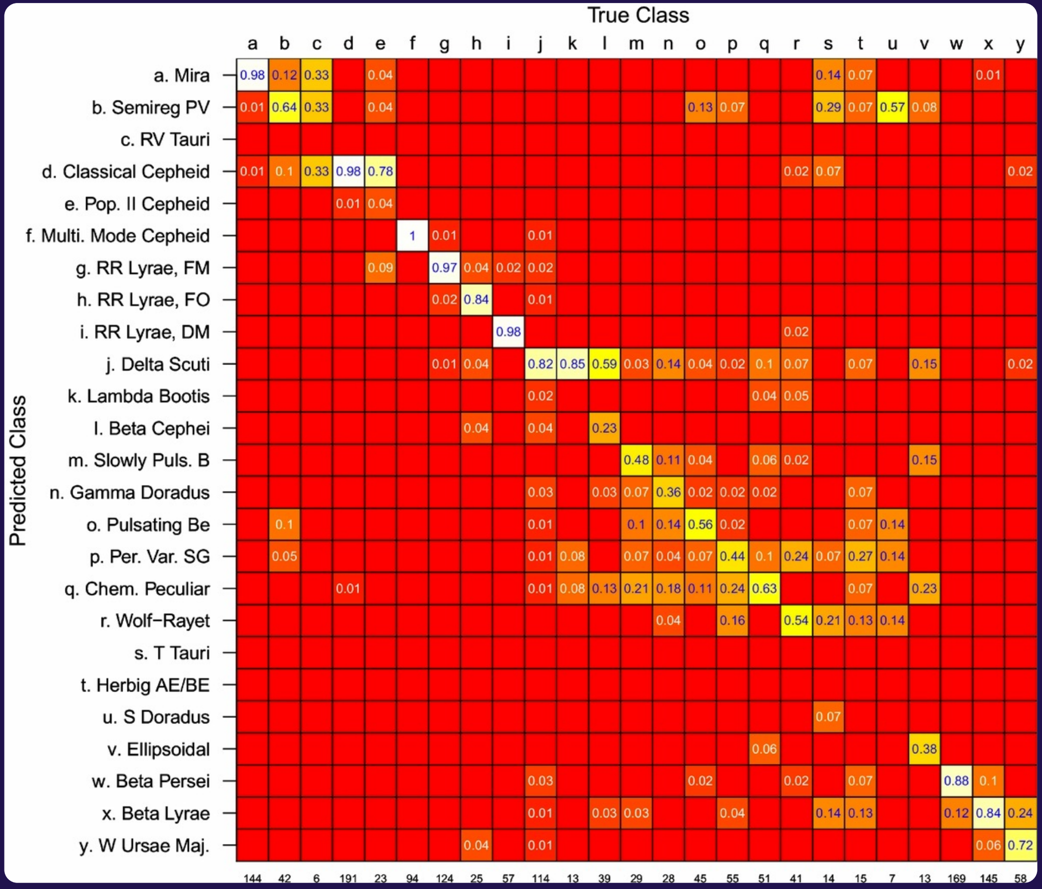
Periodic Variability Classification

Classifying variable sources in (noisy, sparse) survey datasets.

RF-based classifiers trained on features computed from time-series of well-known variables.

Outperformed all other classifiers (by ~25%).

Extremely efficient discovery tool (e.g., >95% for pulsational variables).

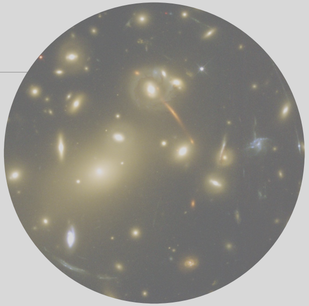


Classifier: Richards et al. (2011)

LSST Science Themes

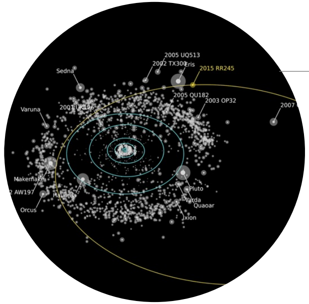
LSST Overview: Ivezic et al. (2019)
LSST Data Products Definition: Juric et al. (2013)

Probing Dark Matter & Dark Energy



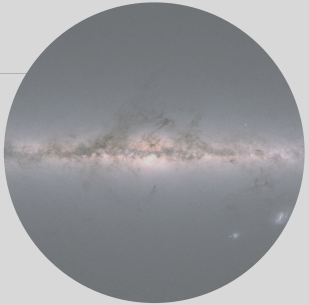
- Strong & Weak Lensing
- Large Scale Structure
- Galaxy Clusters, Supernovae

Inventory of the Solar System



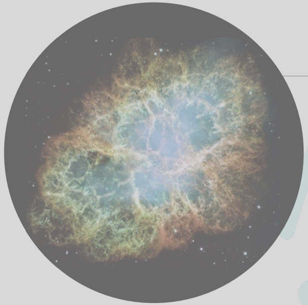
- Comprehensive small body census
- Comets and ISOs
- Planetary defence

Mapping the Milky Way

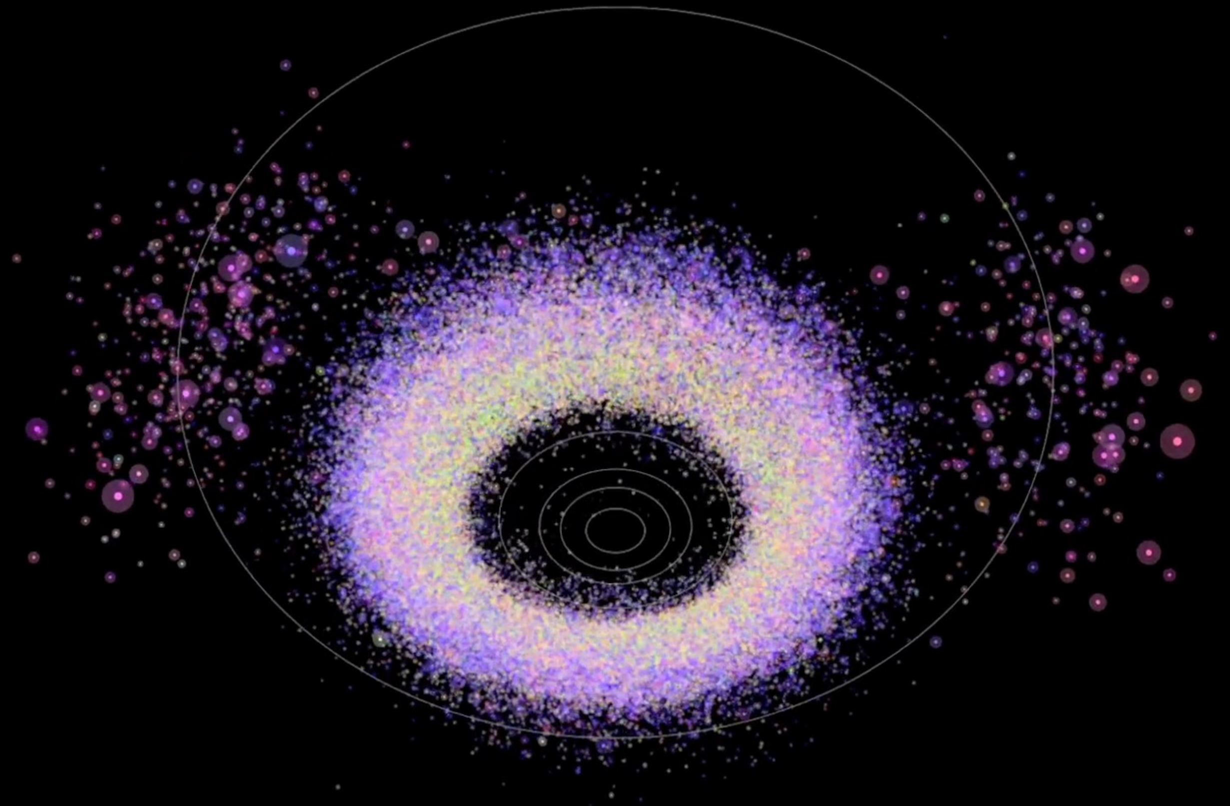


- Structure and evolutionary history
- Spatial maps of stellar characteristics
- Reach well into the halo

Exploring the Transient Optical Sky



- Variable stars, Supernovae
- Fill in the variability phase-space
- Discovery of new classes of transients



Animation: SDSS Asteroids
Alex Parker, SwRI

Exploring the Solar System

An unprecedented census of the Solar System

Animation: SDSS Asteroids (Alex Parker, SwRI)

Estimates: Lynne Jones et al.

The LSST data should increase the number of known objects between 5x-30x, depending on the population.

	Currently Known*	LSST Discoveries**	Typical number of observations+
Near Earth Objects (NEOs)	~25,500	100,000	(D>250m) 60
Main Belt Asteroids (MBAs)	~1,000,000	5,000,000	(D>500m) 200
Jupiter Trojans	~10,000	280,000	(D>2km) 300
TransNeptunian Objects (TNOs) + Scattered Disk Objects (SDOs)	~4000	40,000	(D>200km) 450
Comets	~4000	10,000	?
Interstellar Objects (ISOs)	2	>10	?

These objects will be well-characterized (orbits, light curves, absmag estimates), and discovered with an exceptionally well understood selection function.

Impacts and Planetary Defense

At present, we've discovered only ~40% of asteroids capable of causing continent-wide destruction.

The LSST will bring that number up to ~70%.

With advanced software (THOR) and cadence changes, we could go as high as ~85%.

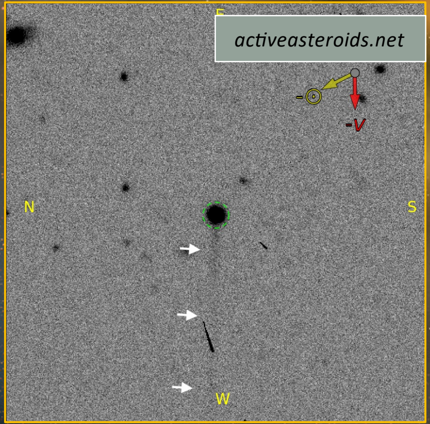
Chelyabinsk, Russia February 15, 2013.

Jones et al. (2019)

Solar System Volatiles

Colin Chandler
LINCC Postdoc & PS
Active Asteroids

comets →

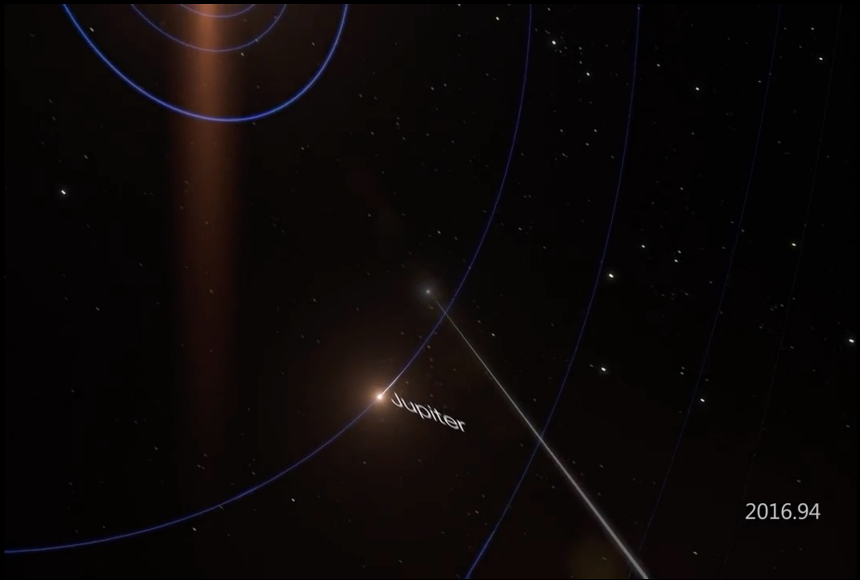


← accretion

Active asteroids? ~30 known
Active Centaurs? ~20 known
Active Quasi-Hildas? ~10 known

Interstellar Objects: Messengers from Exoplanetary Systems

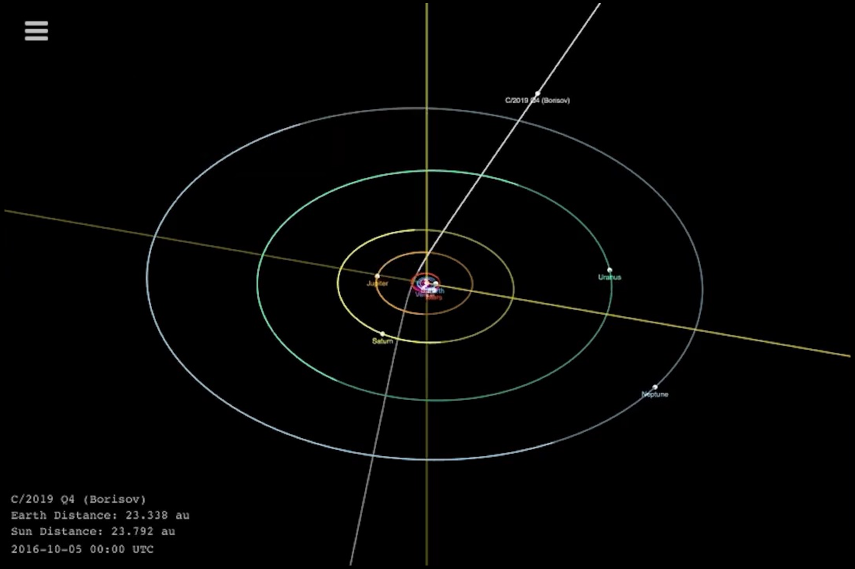
1I/2017 U1 ('Oumuamua)



https://www.youtube.com/watch?v=iwv1RxtsmA0

Credit: ESO, M. Kornmesser, L.Calcada.

C/2019 Q4 (Borisov)

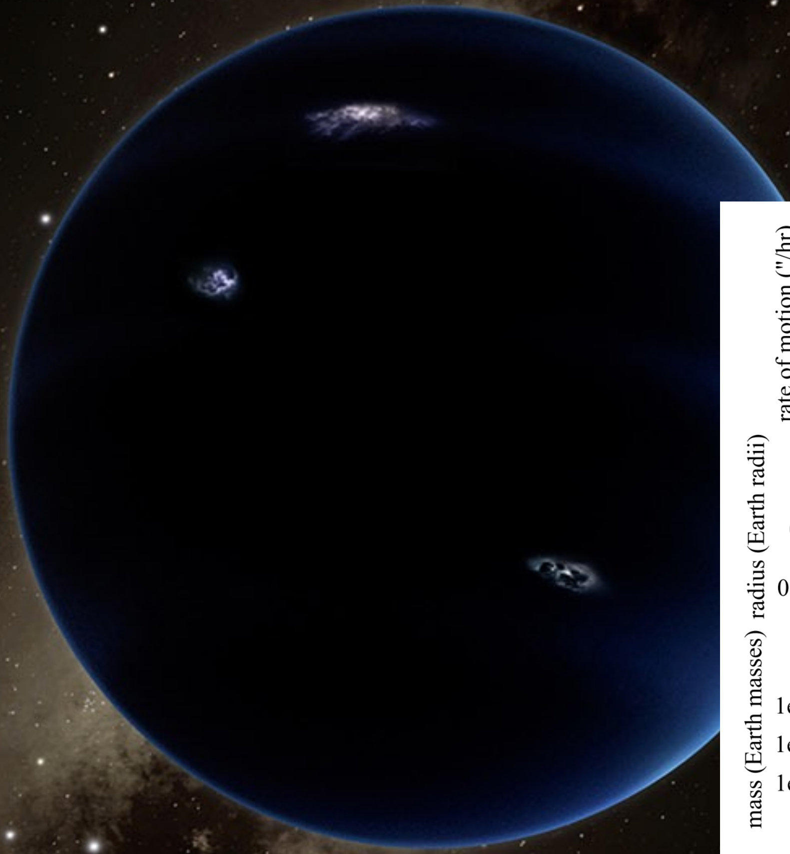


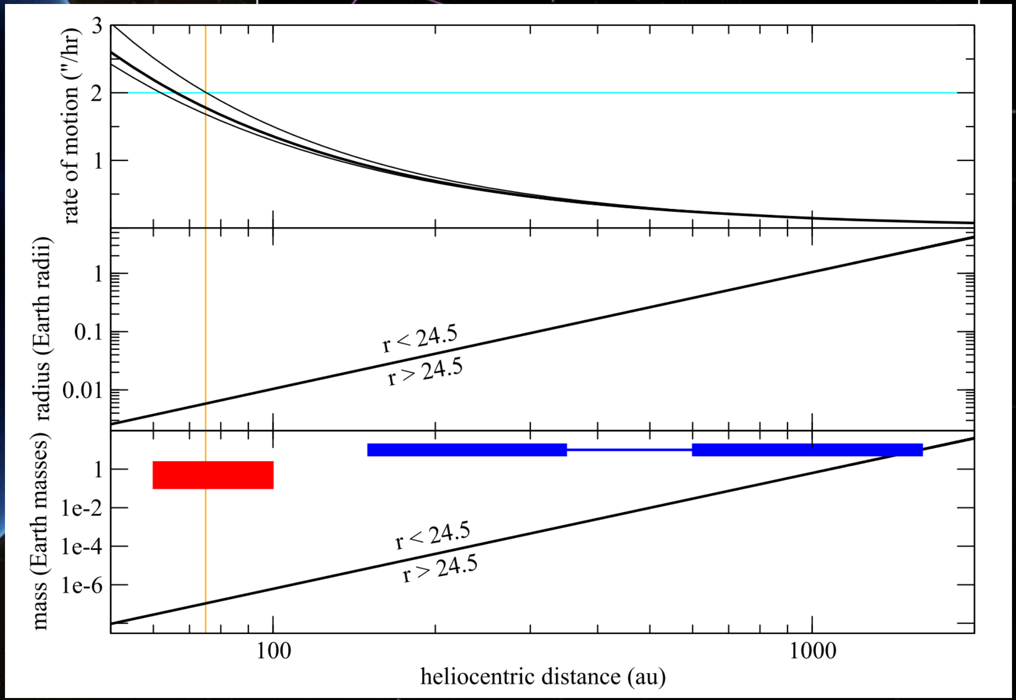
https://www.youtube.com/watch?v=vqMJo3DHOfg

Credit: NASA/JPL-Caltech and Steve Spaleta (Space.com)

LSST will discover anywhere from 1 ISO /year to 1/month.
Up-close exploration of material from other planetary systems.

Planet Nine





A planet that may exist somewhere on the outskirts of the Solar System.

Rubin will directly and exhaustively survey ~60% of the sky for the proposed unknow planets, and indirectly set strong constraints on their (non)existence by characterizing the KBO population.

But… we first have to identify them in the data!

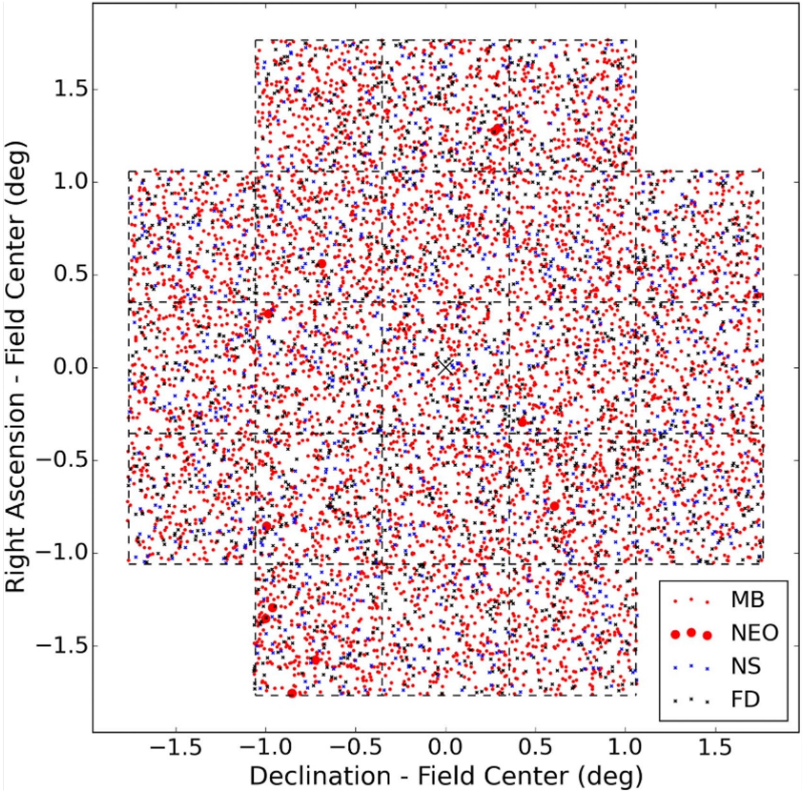
Problems in Solar System Discovery

1. Asteroid Linking
2. Optimal detection of fast-moving asteroids (planetary defense)
3. Deep discovery (dwarf planets, Planet 9)

Finding Asteroids

Credit: a sequence of images of asteroid (1078) Mentha by the UK Spaceguard Centre (https://spaceguardcentre.com)

1. Asteroid Linking Problem



Above: Simulation by Veres & Chesley (2017)

Deep sky images have high object densities. Unclear which object is which.

Problem statement

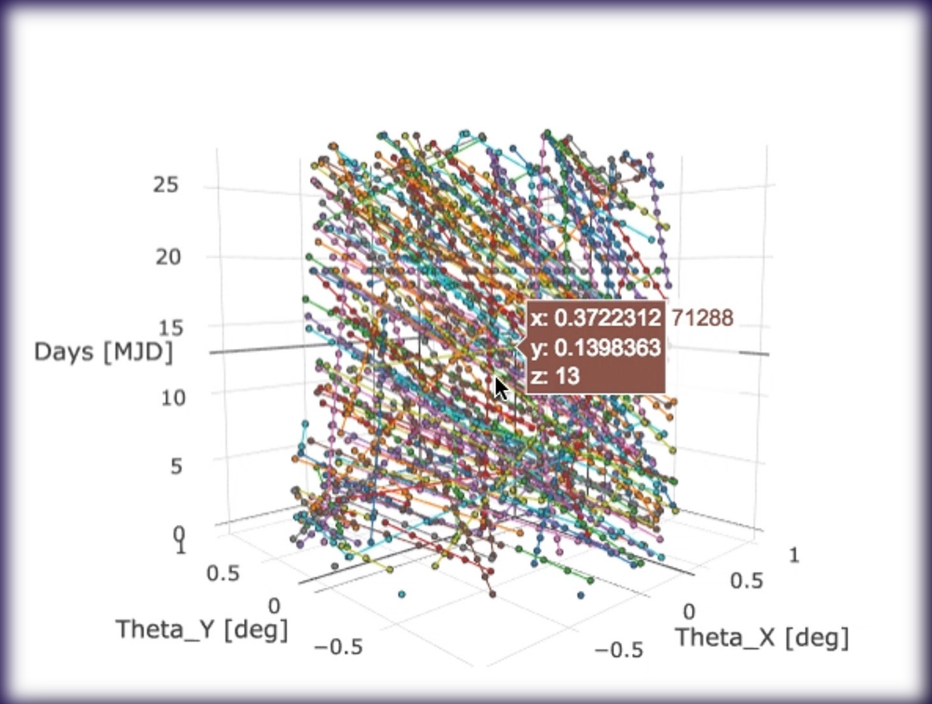
Given O(150M) 2D data points, find O(1M) groups of O(10M) objects that satisfy gravitational equations of motions.

Non-AI solutions exist for:

- Simplified problem where pairs of images are taken closely space in time (O(100) cores)
- Arbitrary distribution in time, but assumption the objects are not near the Earth (O(10k) cores)

Tracklet-less Asteroid Discovery Algorithm

Tracklet-less
Heliocentric
Orbit
Recovery



Joachim Moeyens,
THOR algorithm author

Transforming observations into a space where correct associations are approximately straight lines in 3D space (detectable with Hough transform)

Joachim Moeyens et al. (2021); http://github.com/moeyensj/thor

Trawling the NOIRLab Source Catalog (NSC)

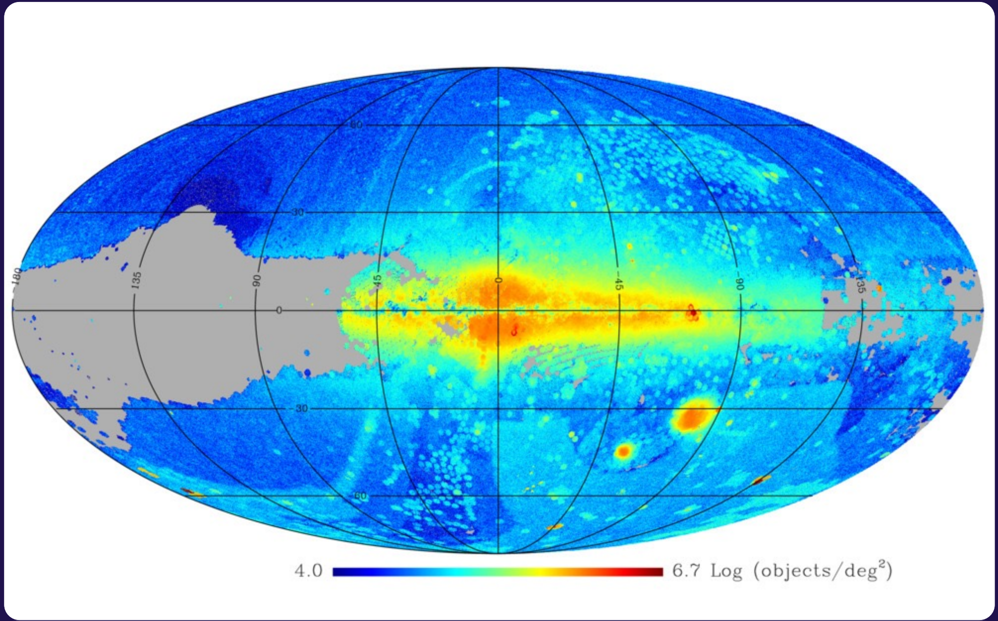
ASTEROID INSTITUTE
A PROGRAM OF B612

DiRAC

- The NOIRLab Source Catalog (NSC) is a catalog of nearly all of the public imaging data in [NOIRLab's Astro Data Archive]

- 68 billion individual source measurements
- Dominated by DECam + Blanco 4m measurements (3/4 of all exposures)
- Deep (~23rd magnitude in most filters)

- ~1.7 billion don't appear to be static (i.e., could be asteroids).



NSC DR2 Object Density Map

Initial run on 0.2% of data (15% of September 2013)

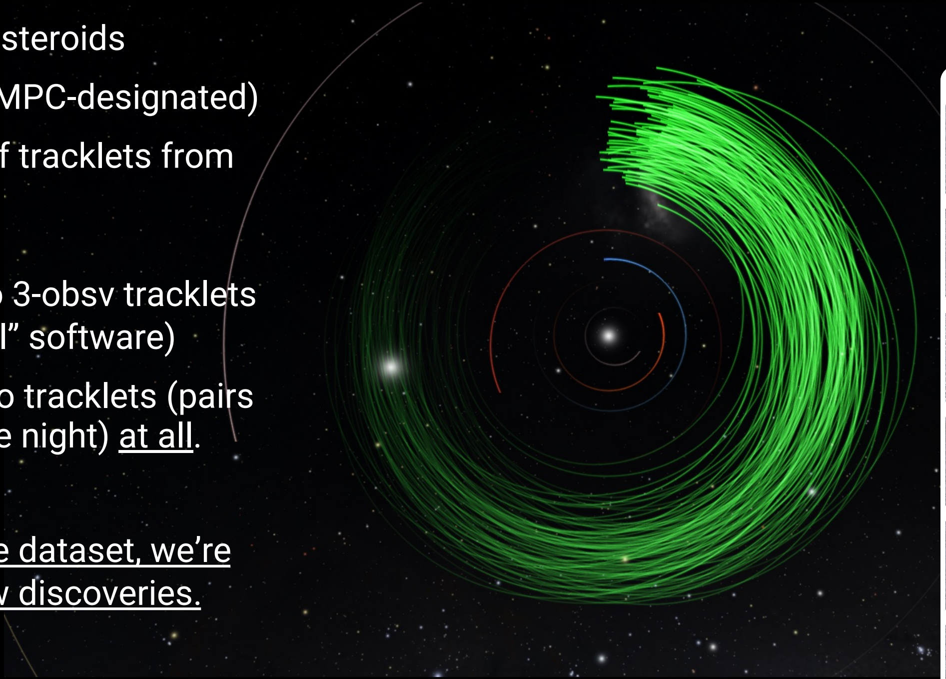
ASTEROID INSTITUTE
A PROGRAM OF B612

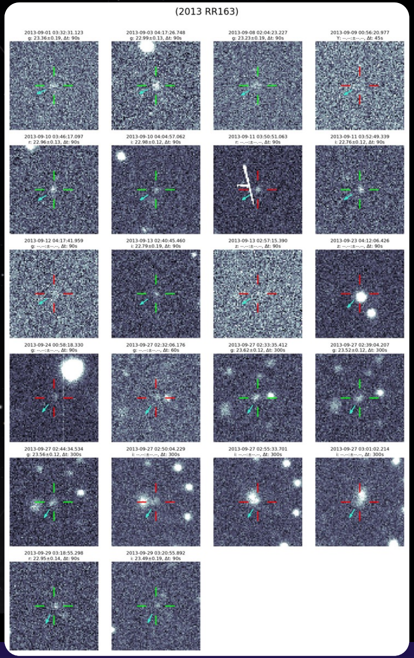
DiRAC

- Identified ~1200 known asteroids
- Linked 104 new objects (MPC-designated)
- Also pulled in a number of tracklets from the ITF.

- The vast majority have no 3-obsv tracklets (unidentifiable with "usual" software)
- A number of them have no tracklets (pairs of observations in a single night) <u>at all</u>.

- <u>Extrapolating to the whole dataset, we're looking at ~10-40k of new discoveries.</u>

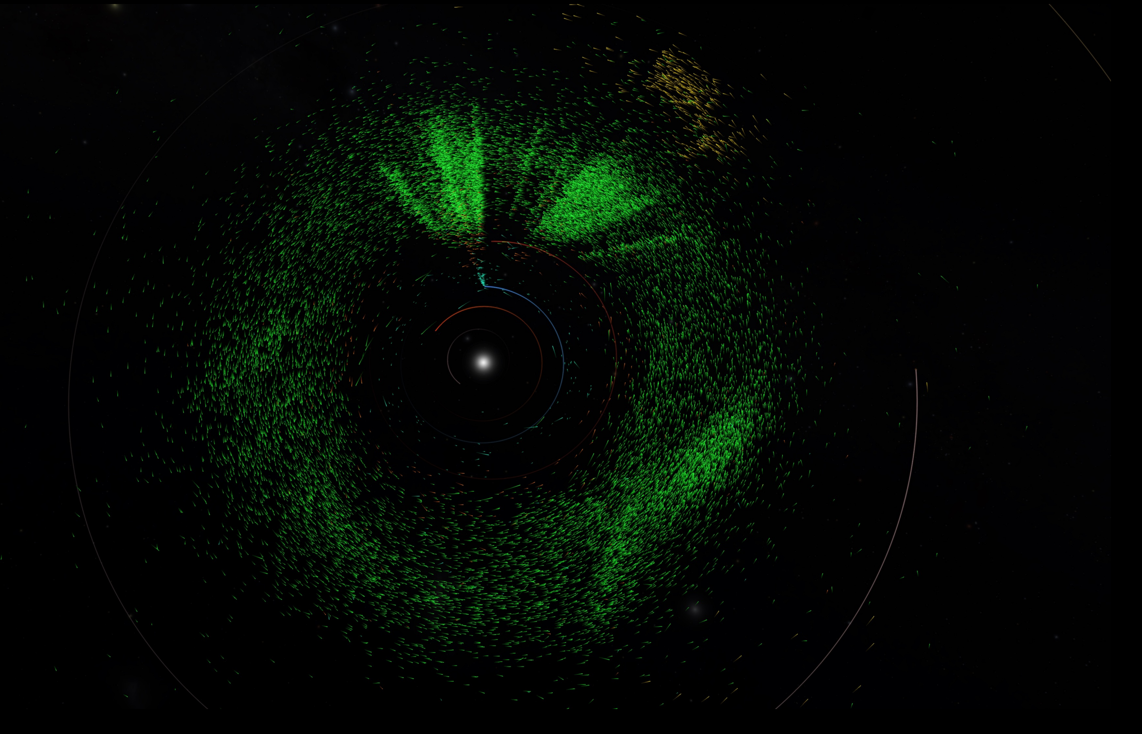




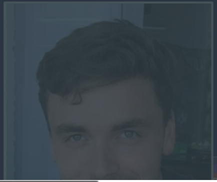
Moeyens et al. (2021)

27,000 new candidates from the NSC

- Used a scalable version of THOR implemented and run on Google Cloud
- For scale: typical world-wide discovery rate today is ~25k/yr
- Most candidates are main belt asteroids
- ~100 NEOs
- A few objects on cometary orbits



Moeyens et al. (in prep)



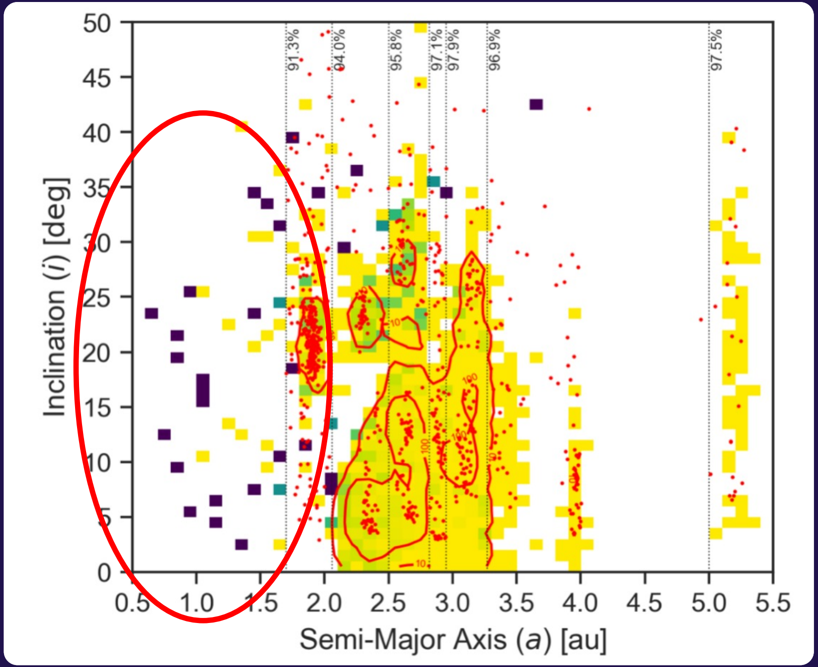
Joachim Moeyens,
THOR algorithm author

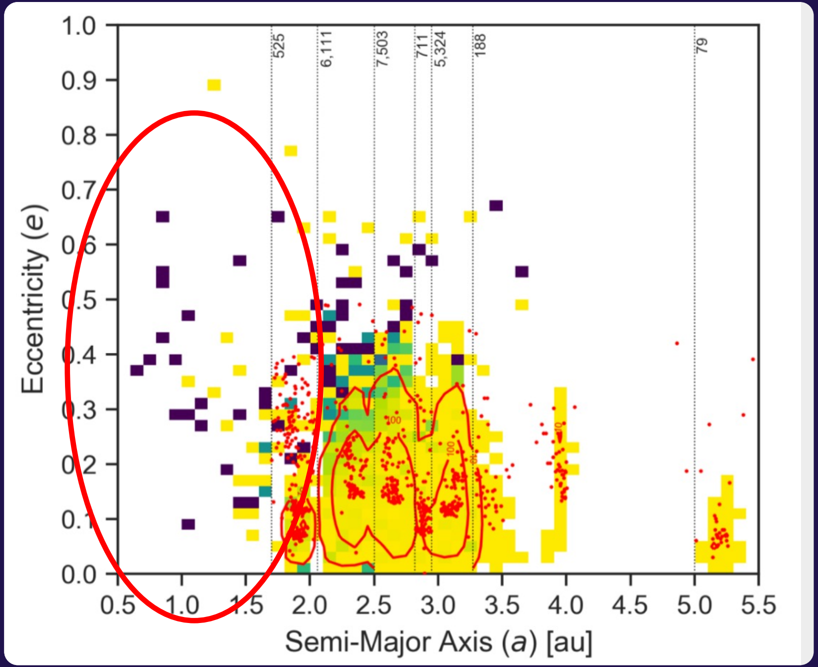
26 May 2023

2022 IAU PhD Prize Winners Announced

- Division B Facilities, Technologies and Data Science: **Joachim Moeyens**, USA, "The Characterization and Discovery of Solar System Small Bodies in Modern Astronomical Surveys"

But… we miss too many to dangerous objects.





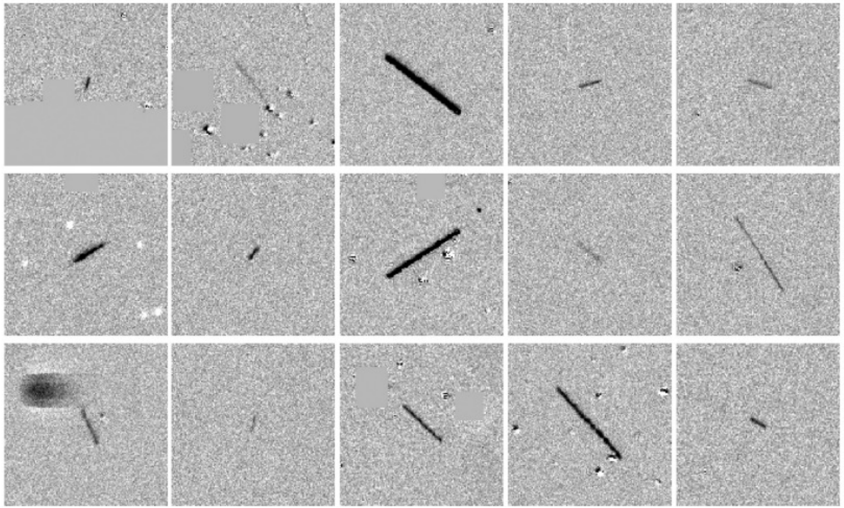
Too computationally intensive for the current algorithms; AI emulators may be the way to make this work.
Opportunity for transfers of knowledge from HEP.

Moeyens et al. (2021)

2. Optimal Streak Detection

Fast moving asteroids leave streaks in images (think motion blur)…

Fast-moving == close. Sometimes, they impact the Earth. 9 found so far, hours before impact (all small)



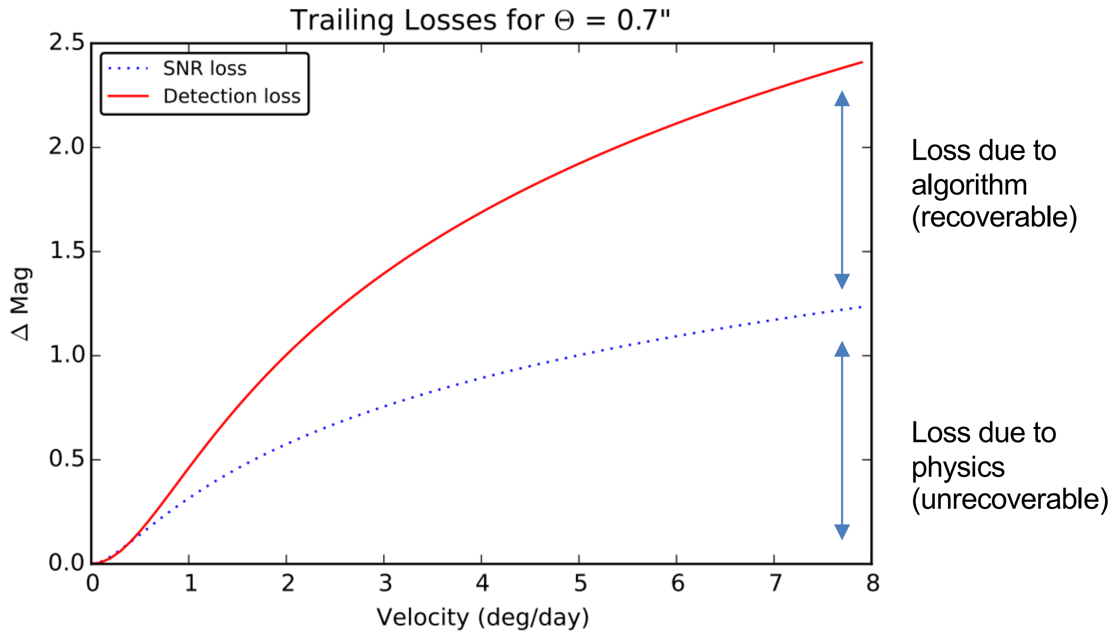
2022 WJ1 (#C8FF042) over London, ON, Canada
(Photo by Rob Weryk)

The time from first detection to ~100%-impact likelihood determination was ~1hr.

Above: Duev et al. (2019)

2. Optimal Streak Detection

Our current image processing algorithms are optimal for point source detections, not streaks.



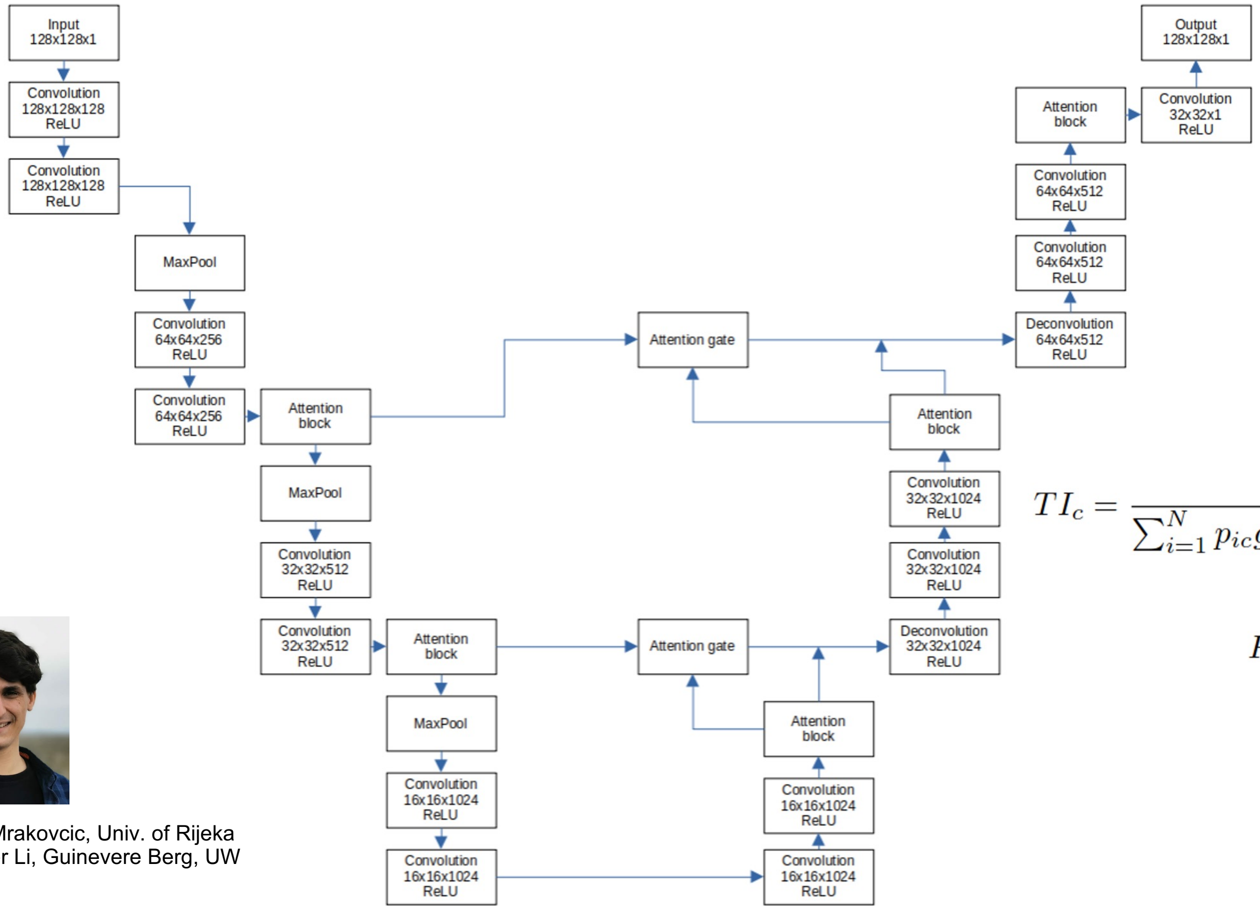
Radon-transform type algorithms exist for streak detection, but the performance is not ideal and they suffer from being confused by artefacts.

<u>Seems ideal for a CNN to solve.</u>

Nearly perfect, unlimited, easy-to-generate training sets are possible.

Above: Jones et al. (2018)

Convolutional Neural Network for Streak Detection



Focal Twersky loss

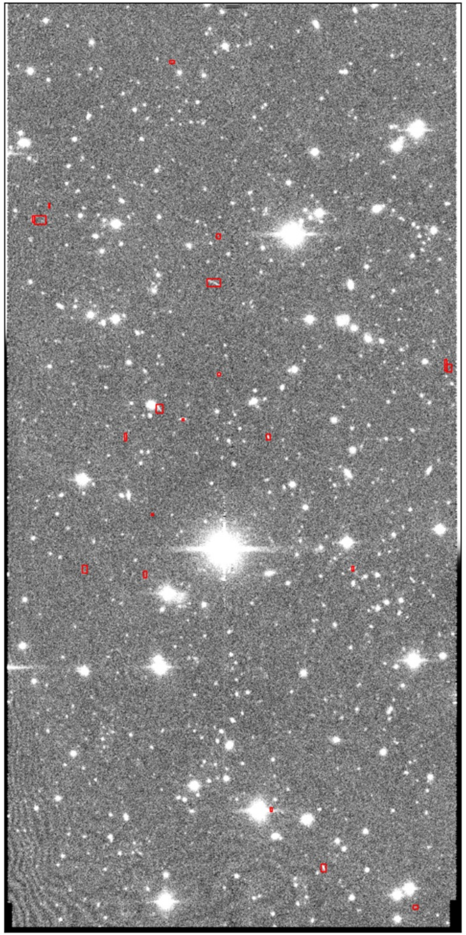
$$TI_c = \frac{\sum_{i=1}^{N} p_{ic} g_{ic} + \epsilon}{\sum_{i=1}^{N} p_{ic} g_{ic} + \alpha \sum_{i=1}^{N} p_{i\bar{c}} g_{ic} + \beta \sum_{i=1}^{N} p_{ic} g_{i\bar{c}} + \epsilon}$$

$$FTL_c = \sum_{c} (1 - TI_c)^{1/\gamma}$$

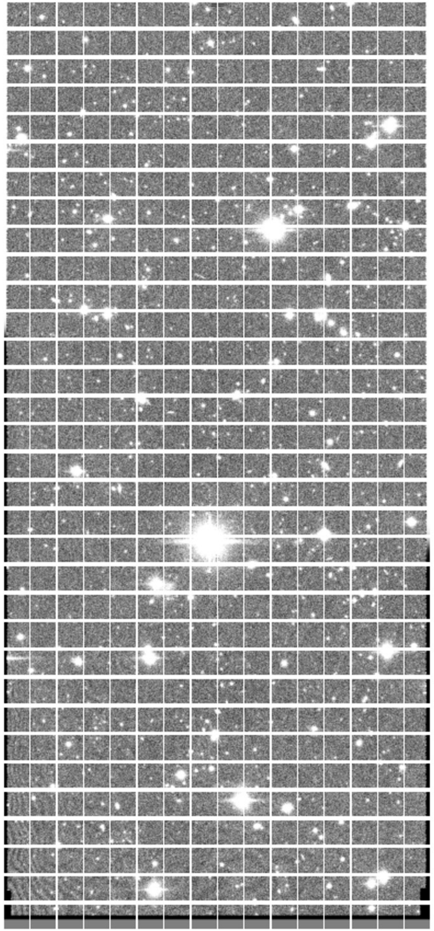
Abraham et. al. (2018)

Karlo Mrakovcic, Univ. of Rijeka
Chester Li, Guinevere Berg, UW

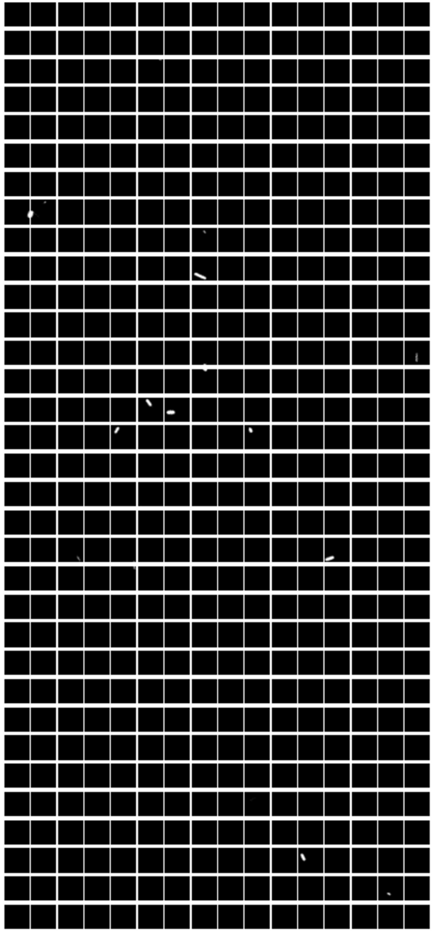
Injected calexp

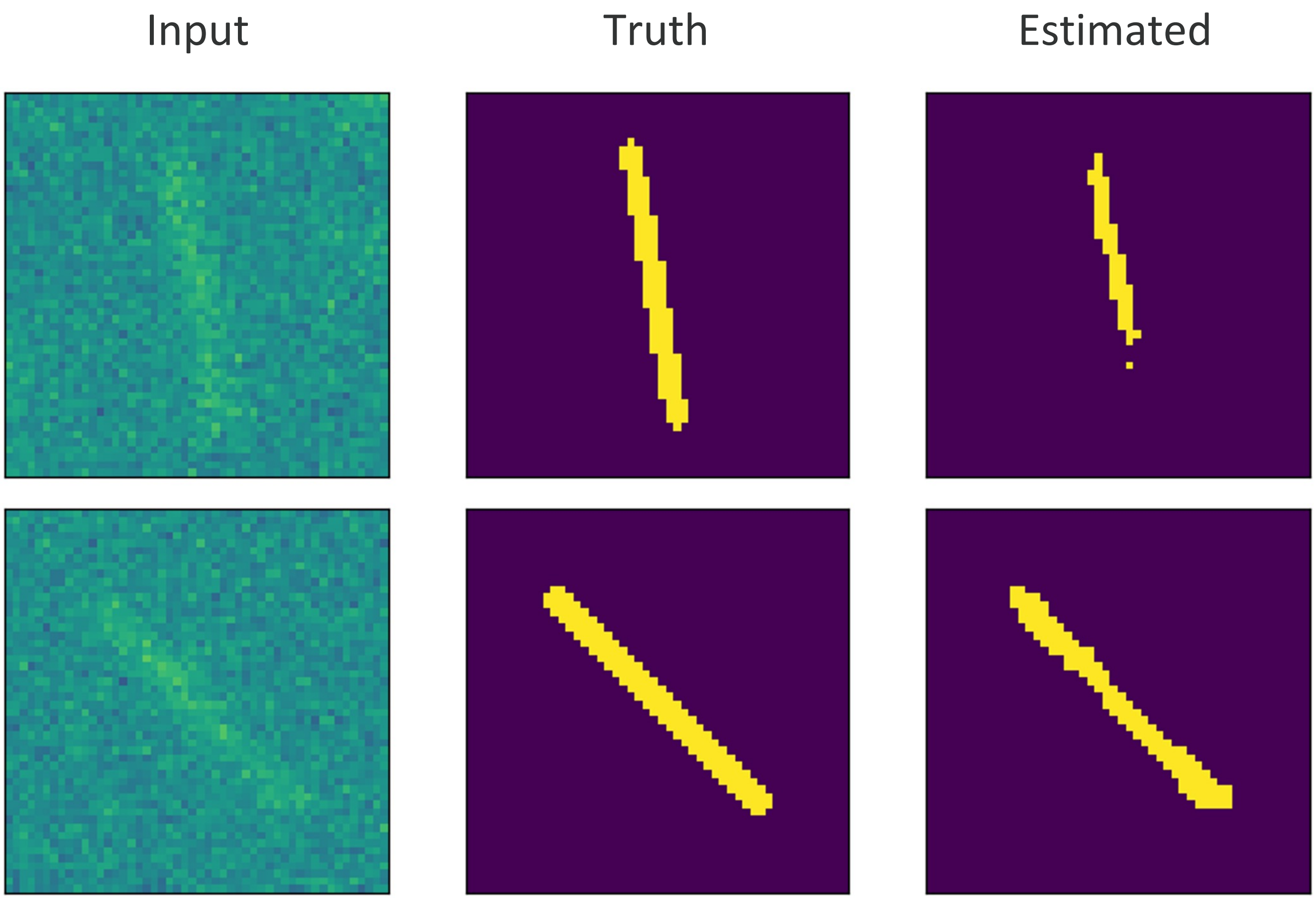


Split injected calexp



Detection mask




Input
Truth
Estimated

Training

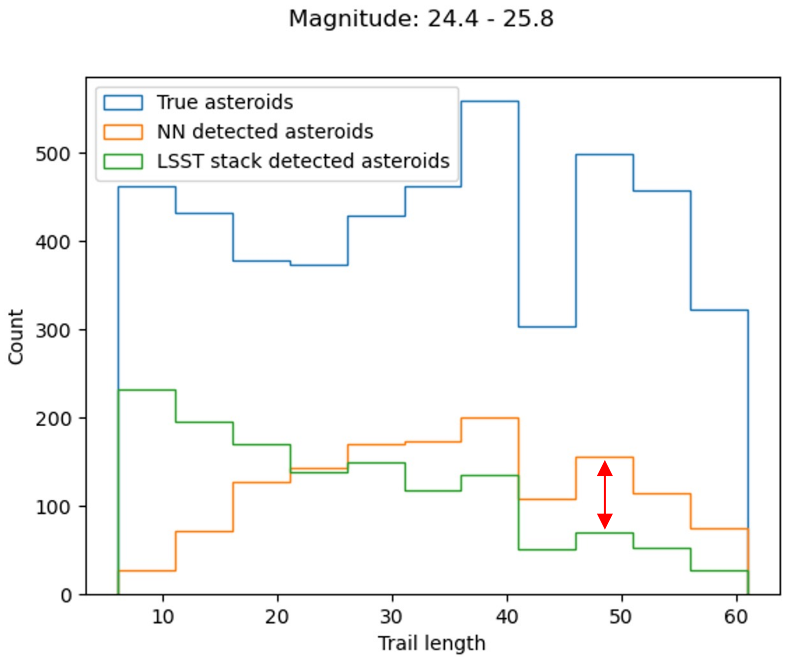
Karlo Mrakovcic, Univ. of Rijeka
Chester Li, Guinevere Berg, UW

- Model trained on S3DF on SLAC
- ~85 hours on 8 GPU-s
- Training set: 10 000 HSC images with synthetic injections of asteroids
- Test set: 100 HSC images with synthetic injections of asteroids
- Magnitude range: 20 mag - 26 mag
- Speed range: 1 deg/day - 10 deg/day
- Inference time: ~1 minute on 8 GPUs per 200 CCDs (one Rubin camera image)

Preliminary results: 2x better

- We can detect more asteroids than LSST pipeline
- CNN successfully detects faint asteroids down to magnitude 27.5.
- Slower drop off due to trailing losses

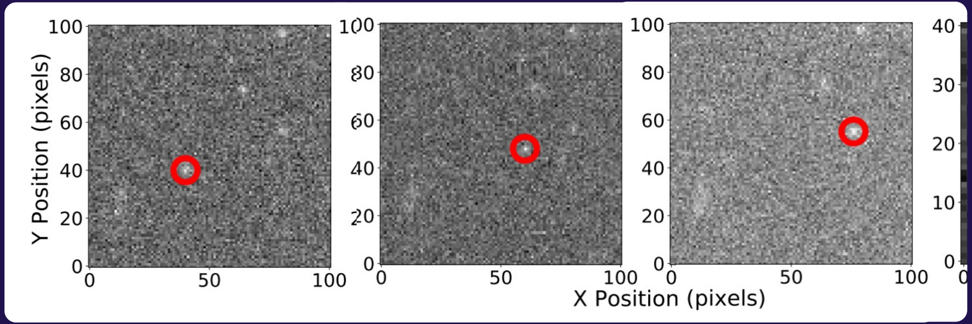
2x better for very fast moving objects (>5 deg/day) than PSF-detection (the current Rubin algorithm,)

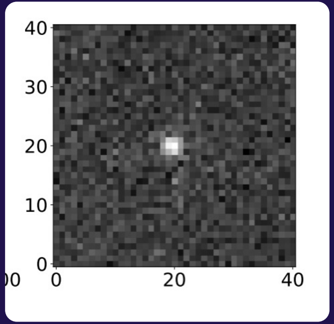


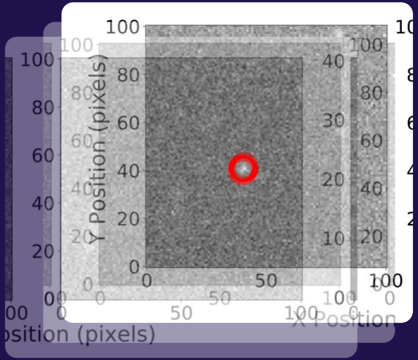
AND, AGAIN, AI COULD HELP *SAVE* THE WORLD.

3. Deep Exploration of the Outer Solar System

How do we detect very faint objects?







Whidden, Kalmbach, Connolly et al. (2019)

How do we go deeper?

- We take multiple, short-exposure, images, and produce a co-add digitally compensating for the motion of the object. This increases the signal to noise and makes an object detectable.

- Q: But what if you don't know where the object is, and it's undetectable on individual exposures frames?

- A: try all possible motion compensation vectors.

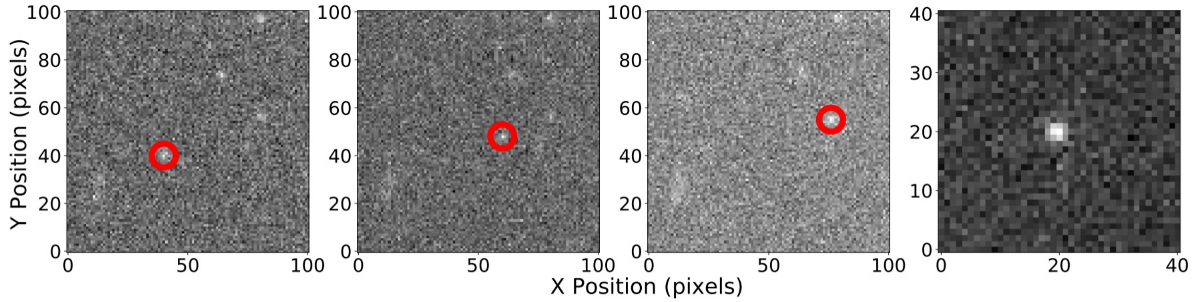
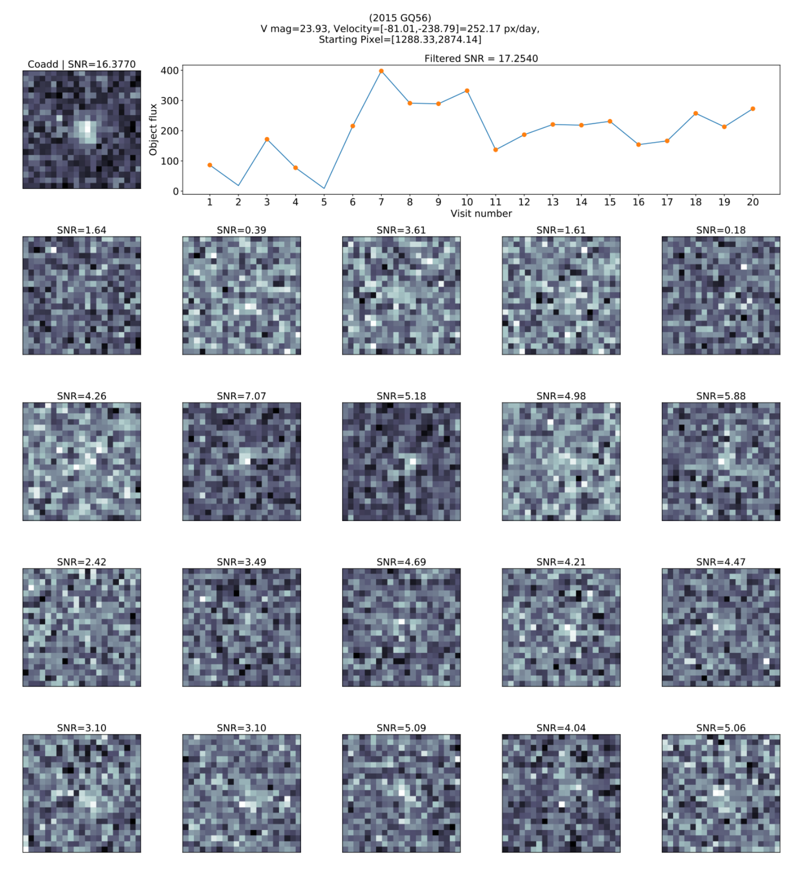


Figure 1. Shifting and stacking of individual images along the asteroid's trajectory creates a single point source in the stacked image.

Whidden, Kalmbach, Connolly et al. (2019)

Shift-and-stack (KBMOD)

Smotherman et al. (2021)



KBMOD (Kernel Based Moving-Object Detection)
https://github.com/dirac-institute/kbmod

Uses GPUs to try out billions of plausible trajectories (motion vectors).

Incredibly computationally expensive; to do all of Rubin may need 5000+ GPUs operating 24/7.

Likely an excellent AI problem: detect objects in "movies" that are below the SNR threshold in individual frames but with the right motion compensation trajectories result in an object with a signficant SNR.

Left: An example object, coadded SNR=16, per-epoch snr <5

Impact if solved

- Nobel Prize for Planet X discovery (in Physics, of course)

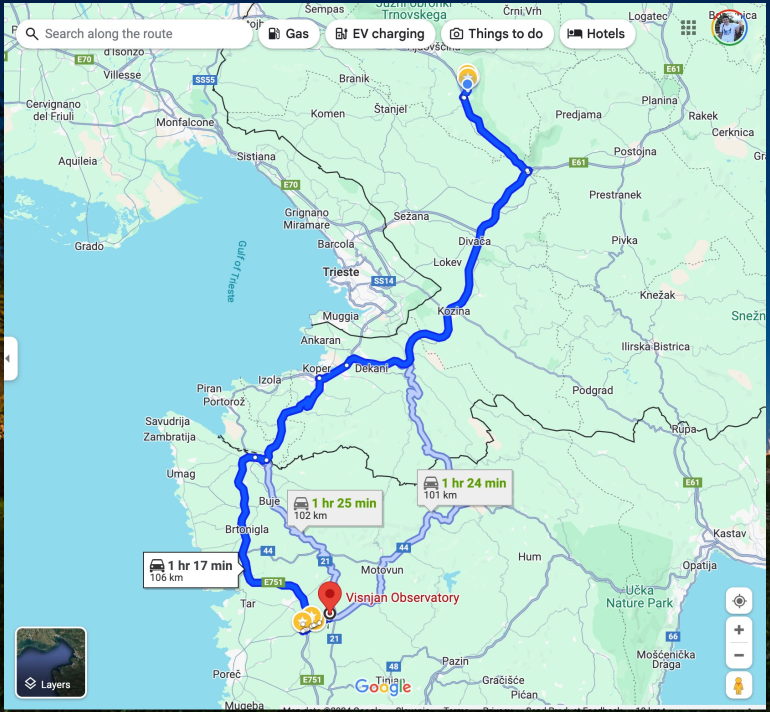
- Find numerous (>100) dwarf planets (objects of same or large size than Pluto)
- Increase the number of known outer solar system objects by 5-10x (beyond what LSST will already find)
- Enable the understanding of Solar System formation (many of the clues are in the outer solar system)

- If made to run in real-time, enable similar advantages for Planetary Defense use cases (finding dangerous asteroids); enable the use of cheap CMOS chips taking movies to repeatedly scan the skies for dangerous/interesting objects.

A final note: LLM-accelerated research, for everyone

Visnjan Observatory, Croatia

https://astro.hr



1m f/2.9 25 ton Telescope

Višnjan Observatory
L01

SCIENCE MISSION

For more than 25 years, our main science mission is primarily dedicated to operational asteroid observation, and to serve the community in educating students and future experts in STEM with the emphasis on astronomy.

As a member of the International Asteroid Warning Network (IAWN), we are amongst the top five observatories in the world in collecting more near-Earth object(NEO) measurements to determine if they pose a threat to Earth.

EDUCATIONAL EXCELLENCE

Our main mission is to motivate, educate and support children, students, teachers and parents to express their talents and to grow into experts, educators, and leaders capable of creating and managing the world that is about to come.

Višnjan School of Astronomy

Višnjan School of Astronomy is a international program aimed for high school students (grades 8-12). Because of often foreign participants, the activities at VSA take place in English language. The selection of students is made by reviewing of their motivation letters, results achieved in competitions and letters of recommendation of their teachers. The school usually takes 8-10 days, and for the difference of similar Višnjan's educational programs, it includes also night work and outdoor data collection (astronomical observations). To maintain high level of individual work between mentors and team members, the program is planned for 15-20 high school students in average, divided into 4-5 groups, working on interesting projects in the fields of astronomy and astrophysics. For more information, please visit the [Visnjan School of Astronomy](https://en.sci.hr/educational-programs/visnjan-school-of-astronomy/).

My VSA 2024 Project Pitch Slides

The Project:

Build the fastest, meanest, online asteroid position lookup service, ever ☺.

The Plan

- By Tuesday: learning the basics and setting things up
- By Wednesday: set up our own cloud server
- By Thursday: set up the automated MPSky service
- By Friday: set up the basic web interface
- By Saturday: have VONEOF-ers try it out!
- By Sunday: add graphical display
- By Monday: have a working website ready for users worldwide

This is a rough, ambitious, sketch of a plan – it *will* change!

My team of five excellent high-school students (operating at 35 degC temperatures!)

We recruited a 6th team member: ChatGPT 4o

Now add some code to measure how long it takes for the fetchEphemerides function to run, and display the runtime in small font at the end of the page, in milliseconds. We want to give the user a sense for how fast the server was in returning the data.

Here's the updated code that measures the runtime of the `fetchEphemerides` function and displays the runtime at the end of the page in milliseconds:

html Copy code

```html
<!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <title>Ephemerides Data</title>
    <style>
```

Result: mpsky.org

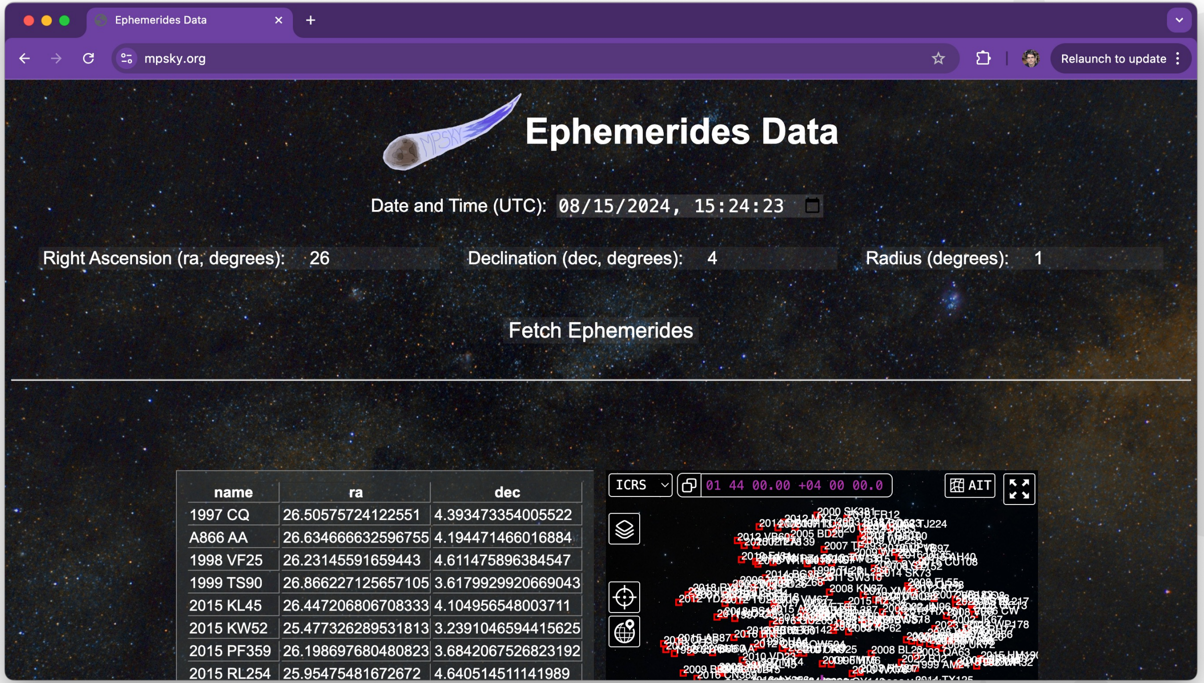
https://mpsky.org

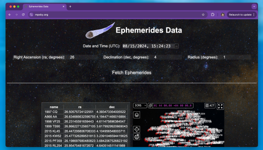
Note: Only one observatory location (L01), and only one night (August 15, 2024)

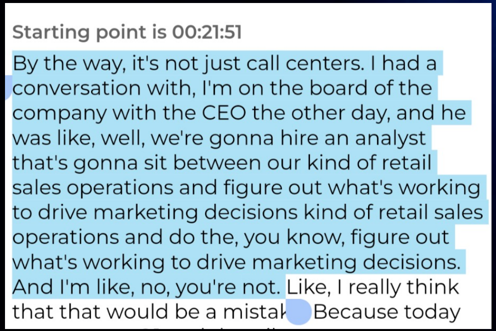
Written and deployed by a team of five high school students with no prior experience with JavaScript, web programming, server maintenance, etc.

Delivered AHEAD OF SCHEDULE (Saturday).

ChatGPT was an INCREDIBLE accelerant!







Starting point is 00:21:51

By the way, it's not just call centers. I had a conversation with, I'm on the board of the company with the CEO the other day, and he was like, well, we're gonna hire an analyst that's gonna sit between our kind of retail sales operations and figure out what's working to drive marketing decisions kind of retail sales operations and do the, you know, figure out what's working to drive marketing decisions. And I'm like, no, you're not. Like, I really think that that would be a mistak Because today

All-In Podcast
680K subscribers

Construct SQL queries via ChatGPT?

Complex visualizations?

Data exploration?

Complete research tasks (with o1-preview)?

We're all be getting a team of AI RSEs & analysts to work with!

https://youtu.be/43Rd-y2xe84?si=B8ZC1OD13Oo6sst2&t=994

Rubin is nearly here & with AI we can maximize its discovery potential

> After nearly 10 years of construction, Rubin Observatory will enter commissioning in ~a month, with early data previews expected next year. The largest astronomical sky survey in human history is about to begin.

> The next challenge is extracting knowledge from the data: AI will play a major role in doing so optimally.

> The applications of AI are still novel in astronomy: possibilities for collaboration are abundant!

DiRAC

Thank You! Questions?

https://dirac.astro.washington.edu

A Universe Understood Through Data-Intensive Discovery

UNIVERSITY of WASHINGTON