Aspects of Transient Classification

Ashish Mahabal, Caltech

C-BAS, UCI, 15 Feb 2011

Collaborators

- Caltech
  - **George Djorgovski**
  - Ciro Donalek
  - Andrew Drake
  - Matthew Graham
  - Roy Williams
- JPL
  - Baback Moghaddam
  - Mike Turmon

Plus at various other institutes all over, but especially in US, India and Italy

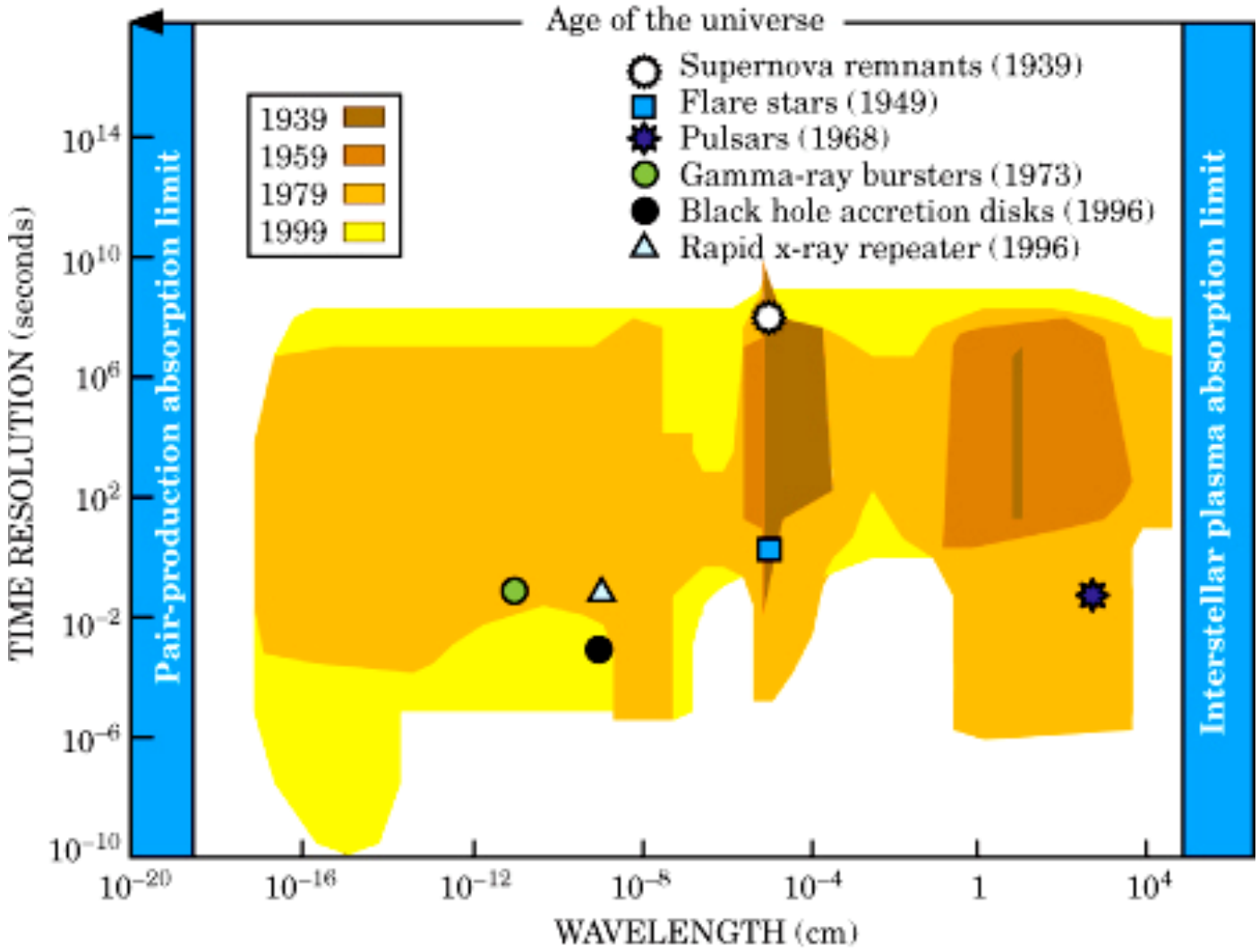
Contact:
mahabal.ashish@gmail.com
or aam@astro.caltech.edu

Expanding the Observable Parameter Space

Technology advances ➔ Expanded domain of measurements

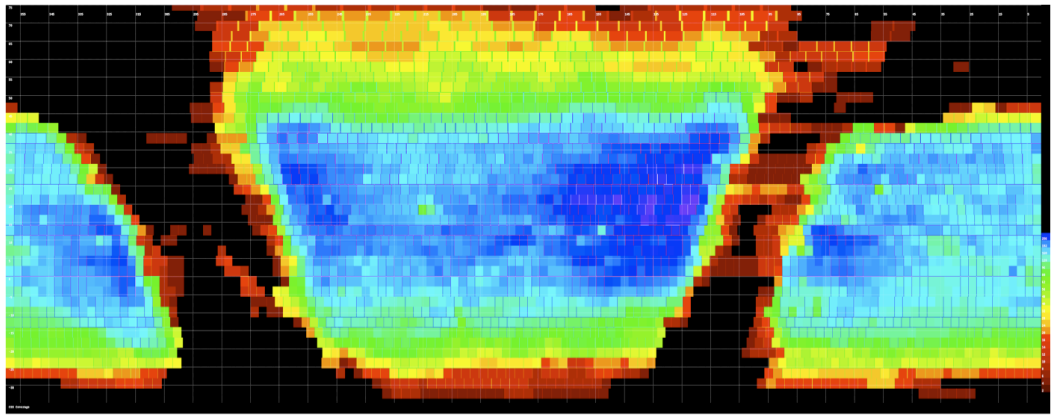
➔ Discovery of new types of phenomena

(M. Harwit)

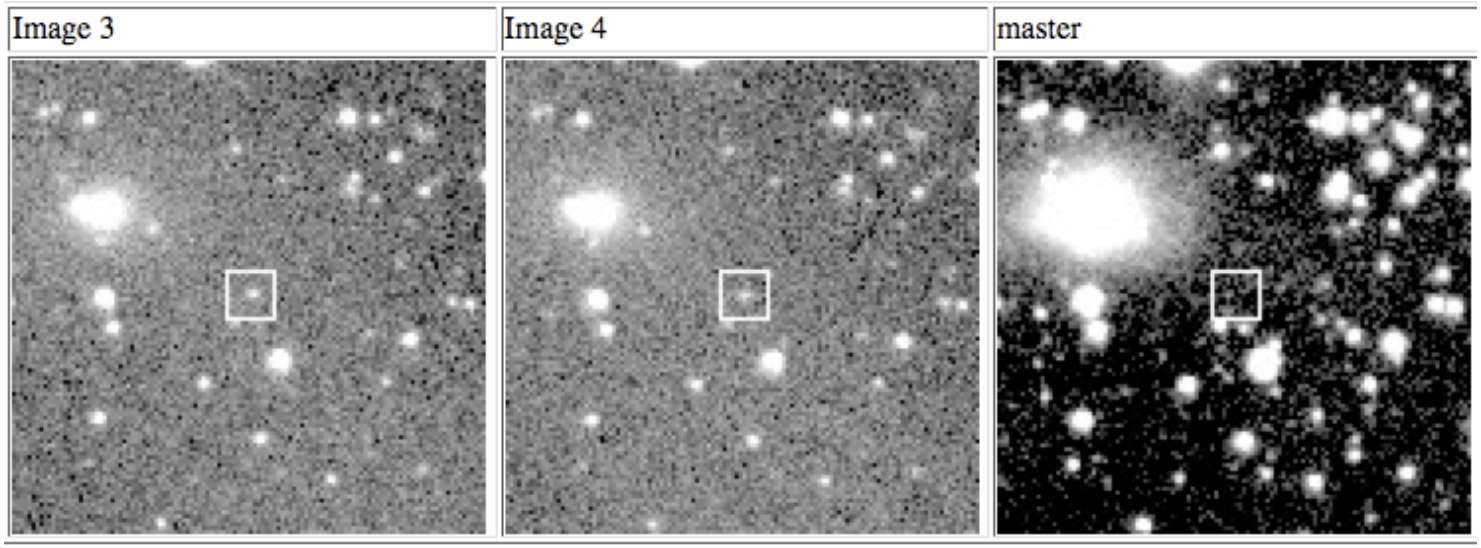


As we open up the time domain, we are bound to discover some new things!

Synoptic surveys



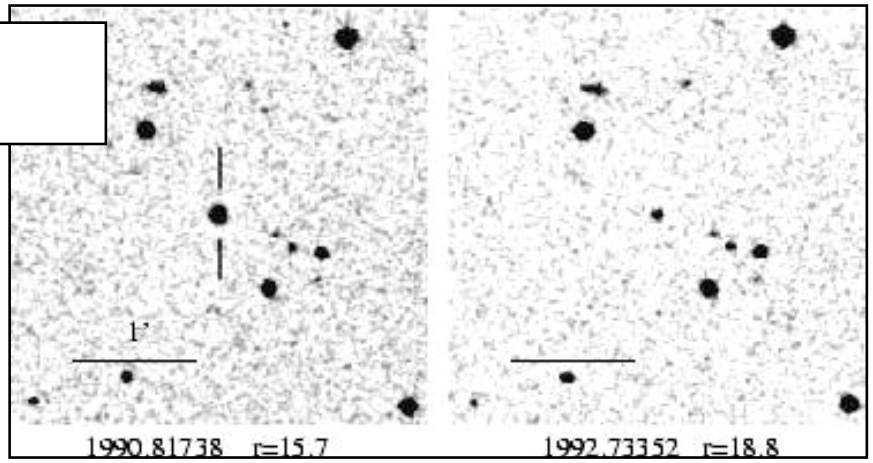
Sporadic to repeated

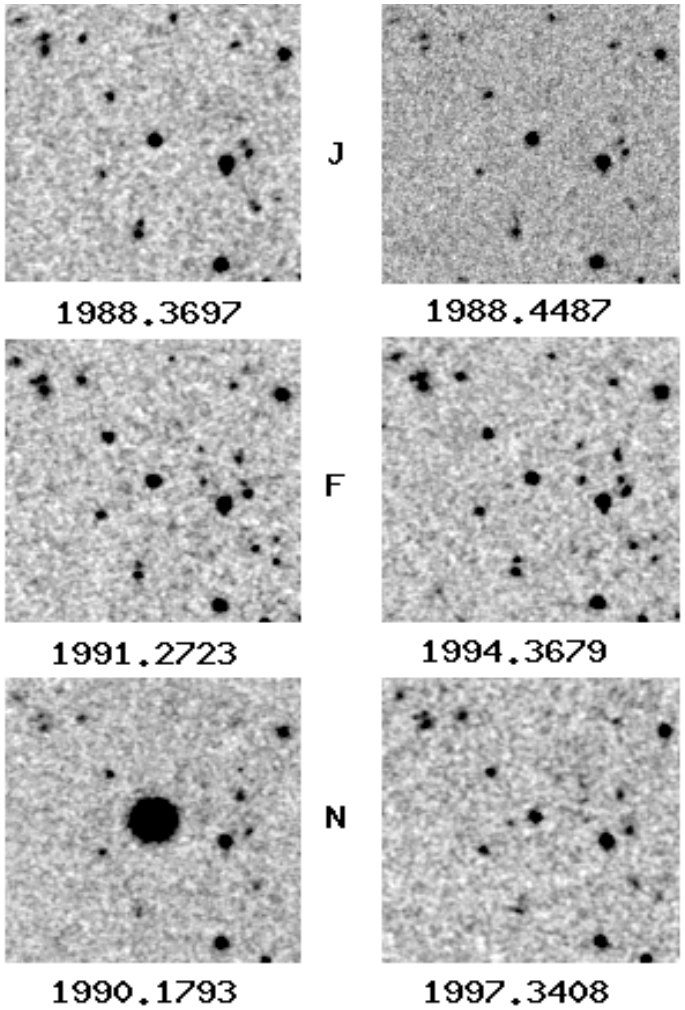


SN z=0.05
CSS 20090711

DPOSS Plate Overlap Survey

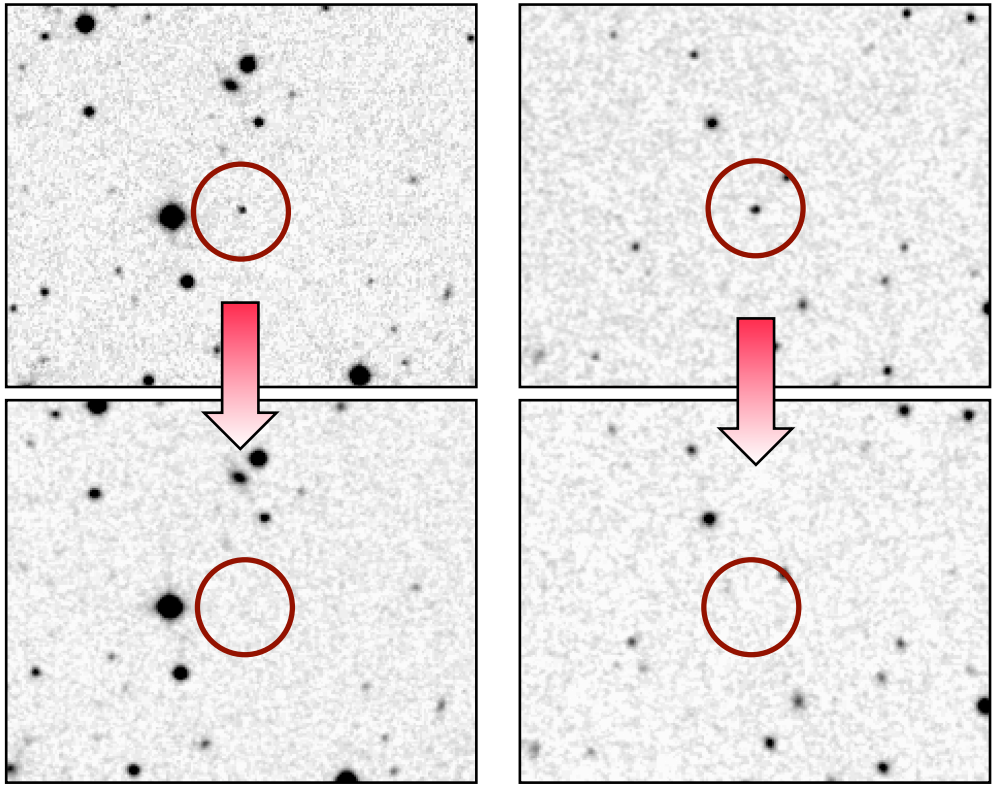
High-amplitude (non-OT) variables, mainly CVs and AGN, over the time baselines ~ a few years





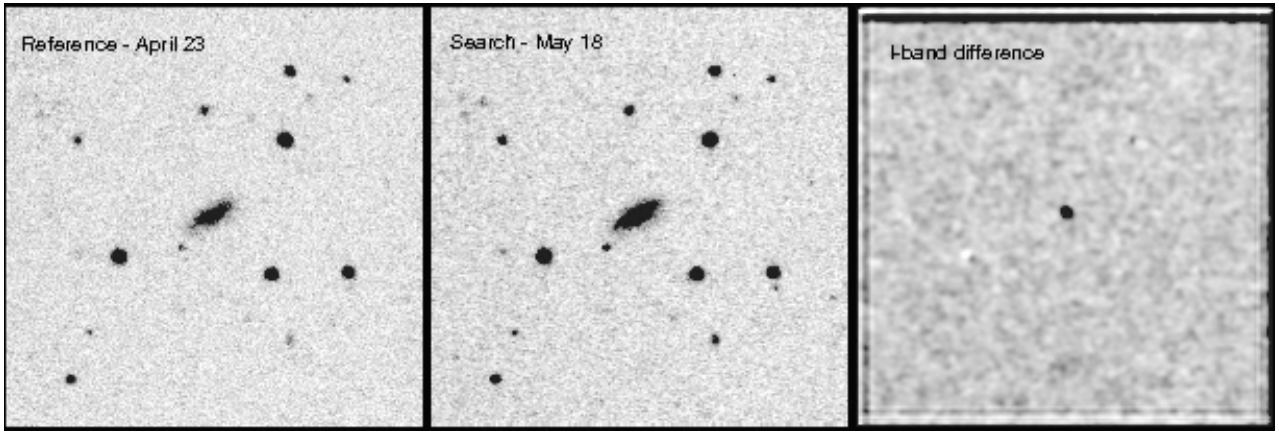
(Mahabal, Djorgovski, Granett 2001, 2003)

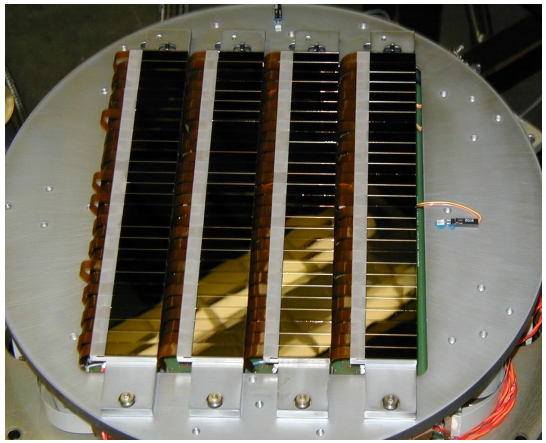
DPOSS Transients



The Palomar-Quest (PQ)
Digital Synoptic Sky Survey

- Palomar 48-in. + 112-CCD, 161 Mpix camera
- A Caltech-Yale collab. Co-PIs: C. Baltay & SGD; plus other groups worldwide (LBL, etc.)
- Many passes with up to 4 filters (*UBRI/griz*), time baselines from minutes to years
- Collected > 50 TB of data
- Operated from Aug. 2003 through Sept. 2008
- ***Key goal: Exploration of the time domain***



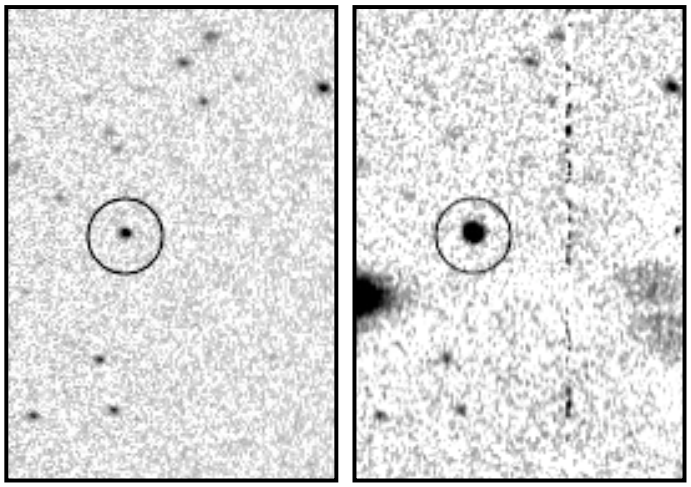


⇦ LBL SNF search (Nugent et al.)

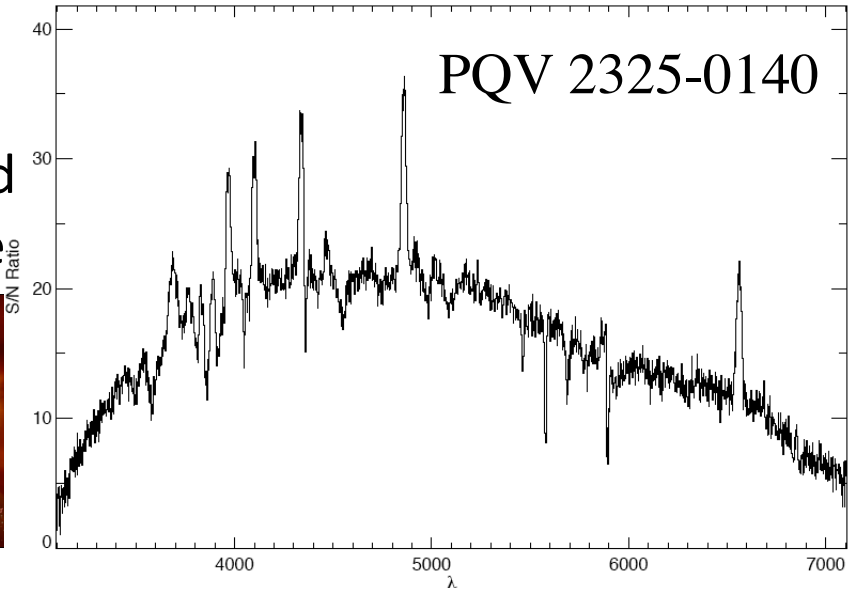
> 700 SNe discovered

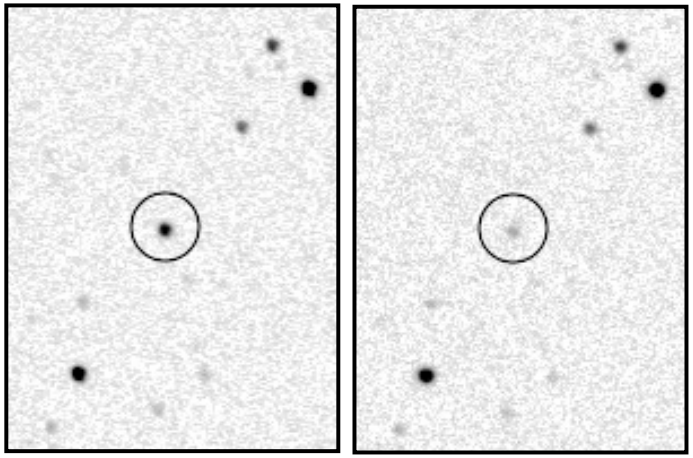
The Most Variable Sources on the Sky:

Selected in the Palomar-Quest Survey

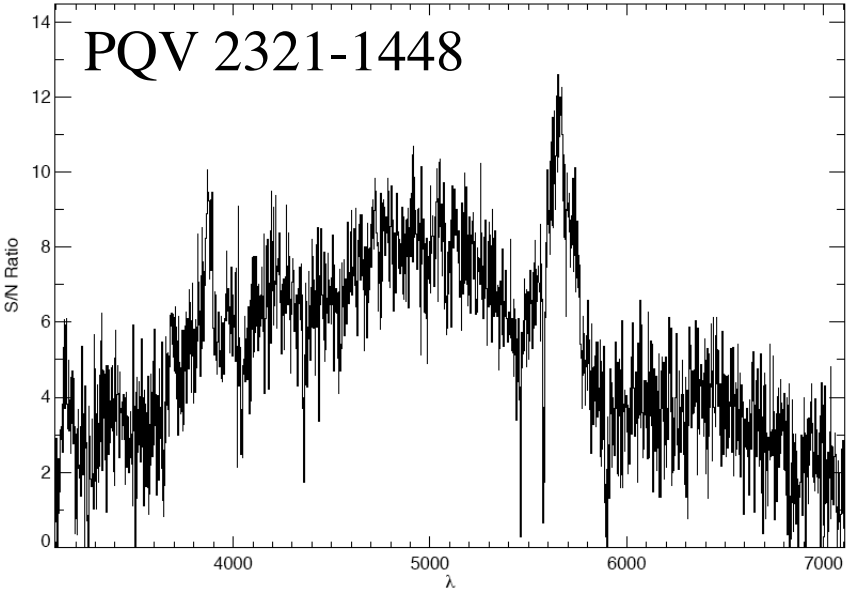


Cataclysmic Variables and Dwarf Novae





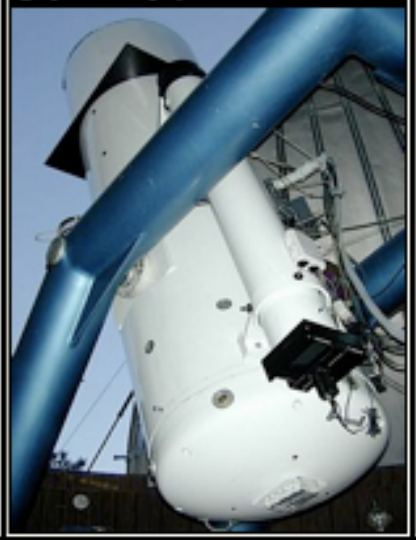
Blazars and OVV Quasars



Catalina Sky Survey(s):

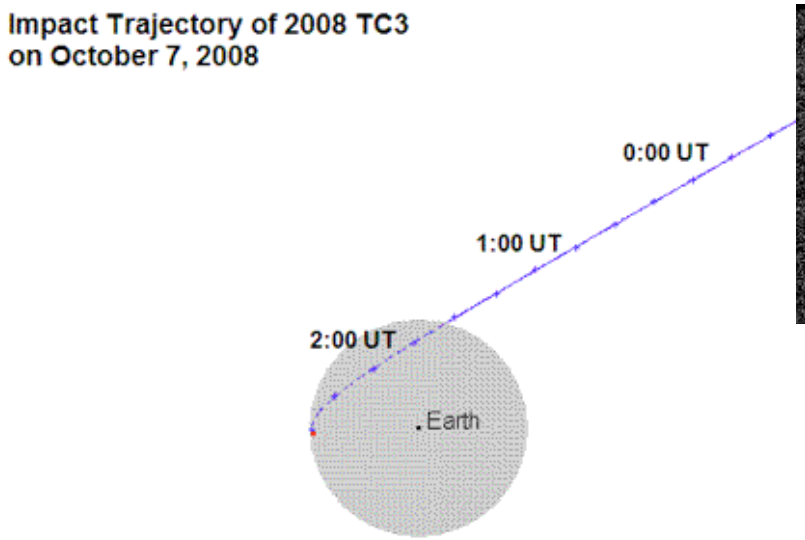
NEO survey Co-PI's:
E. Beshore & S. Larson (LPL)

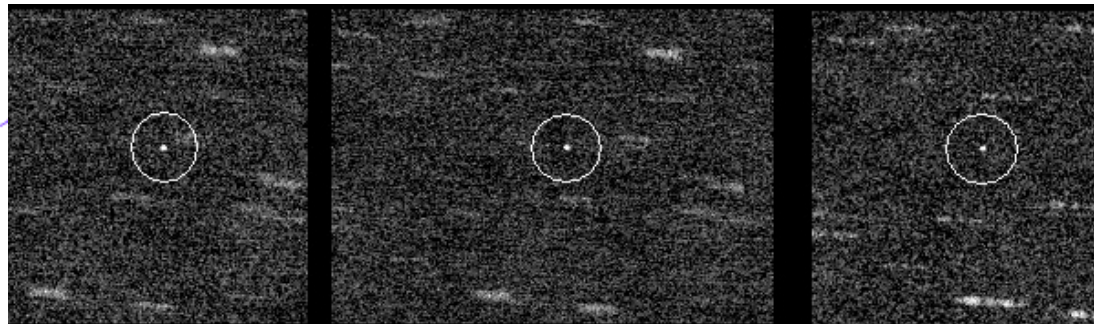
CRTS uses the data from all three Catalina NEO surveys, with a coverage of up to 2,500 deg^2 / night, and the total area coverage of ~ 30,000 deg^2

	MLS The Mt. Lemmon Survey 1.5m Cass	CSS Catalina Sky Survey 0.7m Schmidt	SSS Siding Springs Survey 0.5m Schmidt
			
Survey region (deg)	+/- 5 deg ecliptic	-25 < Dec < +70	-80 < Dec < -25
Field of View (square deg)	1.2	8.1	4.2
Mag limit (V)	21.5	19.5	19.0

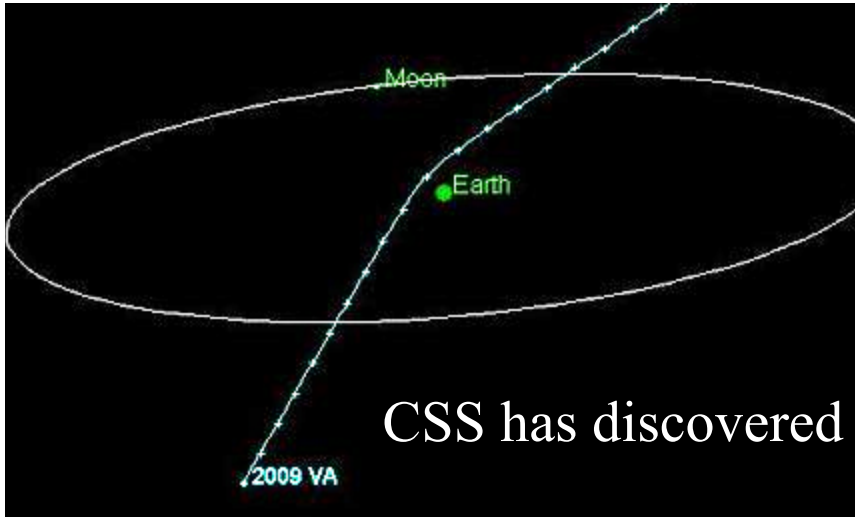
We are processing the Catalina data streams in real time to look for astrophysical transients

CSS Discoveries of Earth-Grazing Asteroids

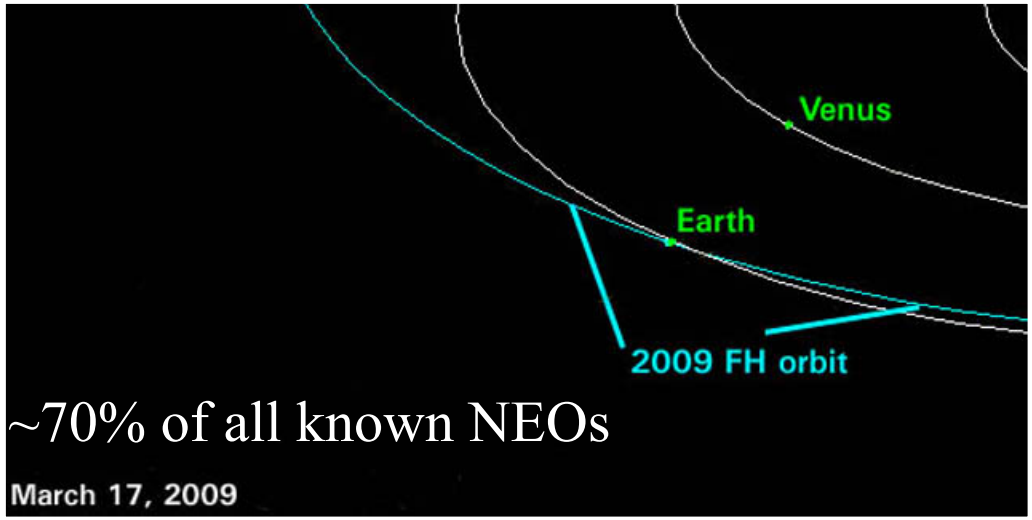




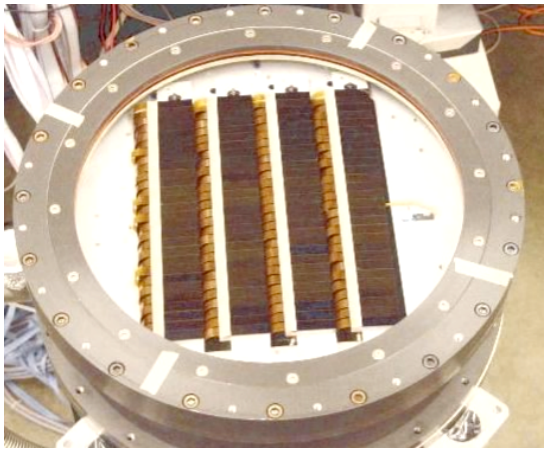
An extremely low cost
"sample return mission"



CSS has discovered

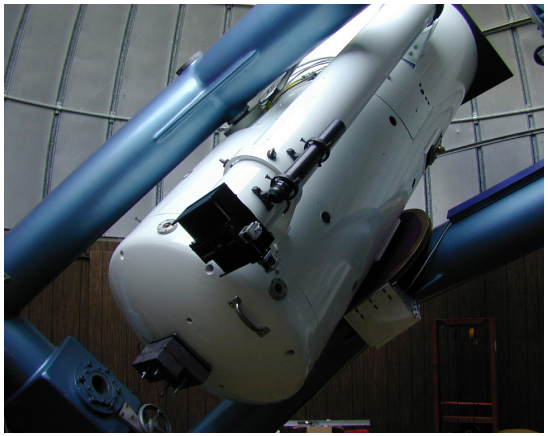


~70% of all known NEOs



PQ

CSS



GALEX, Spitzer, FIRST, …

Recent, current and future multiepoch surveys

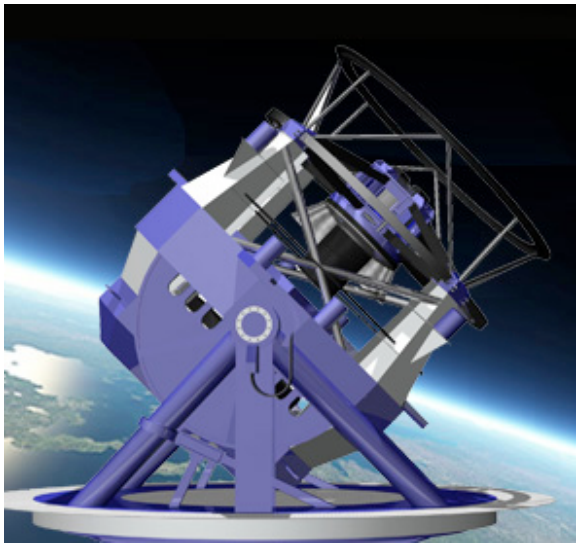
PTF

Skymapper

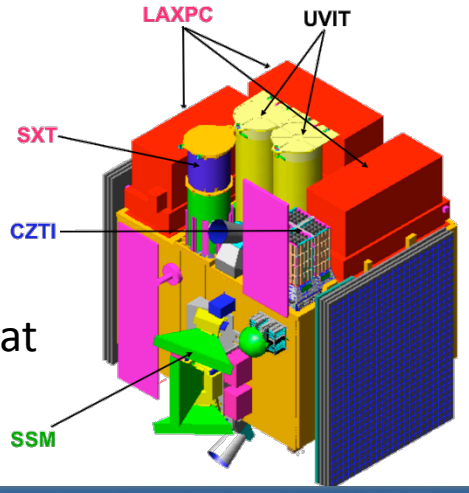
Pan-STARRS

Orders of magnitudes different.

Move towards digital movies!



LSST



Astrosat

The Palomar-Quest Event Factory

Sept. 2006

Detect ~ 1 - 2 $\times 10^6$ sources per half-night scan

Compare with the baseline sky

Find ~ 10^3 apparent transients (in the data)

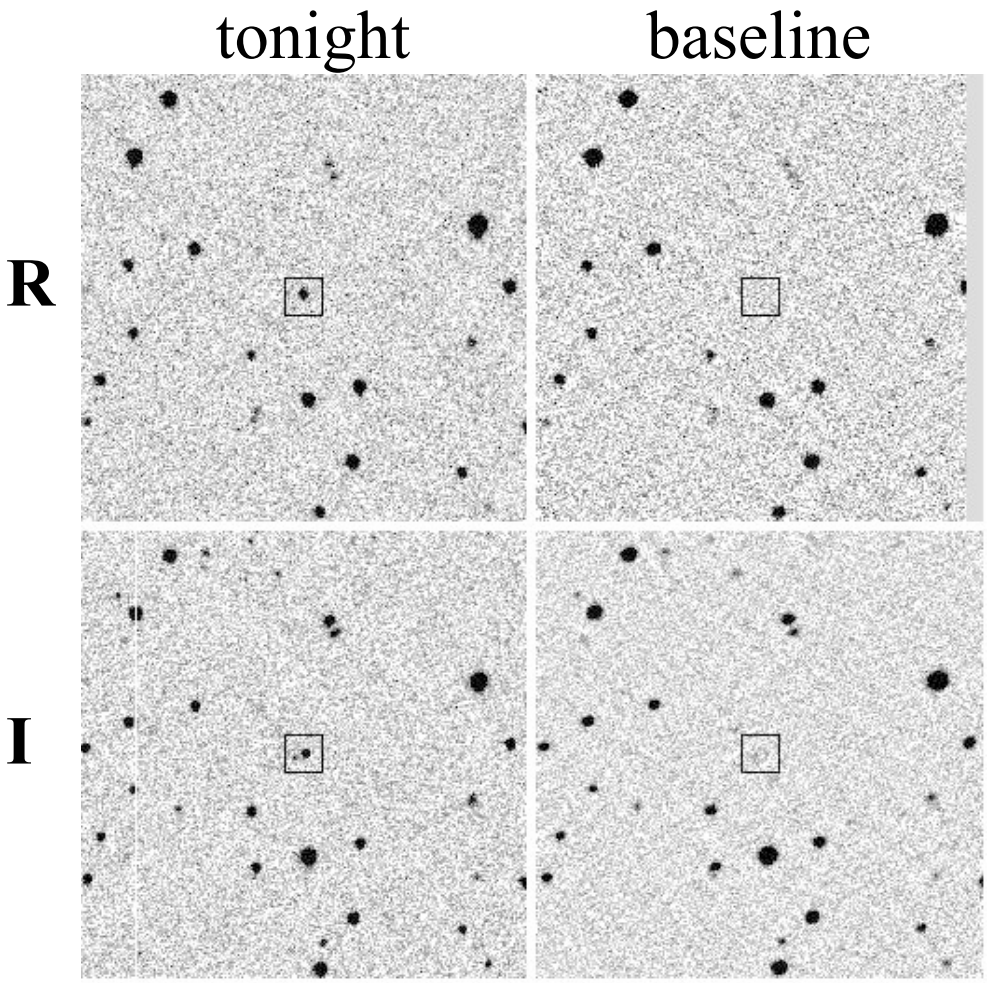
Remove instrum. artifacts

Identify ~ 2 - 4 $\times 10^2$ real transients (on the sky)

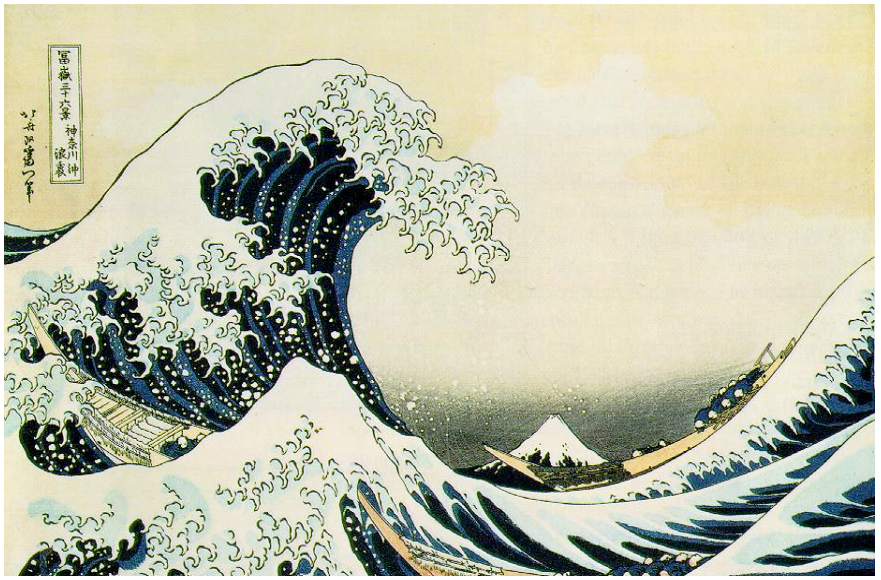
Remove asteroids

Identify ~ 1 - 10 possible Astrophysical transients

→ Classification and follow-up



The Tsunami Wave of the Future



- Now: data streams of ~ **0.1 TB / night, ~ 10 - 10^2 transients / night** (CRTS, PQ, PTF, various SN surveys, asteroid surveys)
- Forthcoming on a time scale ~ 1 - 5 years: **~ 1 TB / night, ~10^4 transients / night** (PanSTARRS, Skymapper, VISTA, VST…)
- Forthcoming in ~ 8 - 10 years: LSST, **~ 30 TB / night, ~ 10^5 transients / night**

A major, qualitative change!

- Observational follow-up needs:
  - Rapid photometric/positional monitoring
  - Rapid spectroscopy
  - Information/computation infrastructure

Transient classification technologies are essential

Time Domain Astrophysics

- **Moving objects:** Solar system, Galactic structure, exoplanets
- **Variability** < Intrinsic
  Modulation along the LOS: microlensing, ISS, eclipses, variable extinction …

Physical causes of intrinsic variability:

- Evolution (structural changes etc.), generally long time scales
- Internal processes, e.g., turbulence inside stars
- Accretion / collapse, protostars to CVs to GRBs to QSOs
- Thermonuclear explosions
- Magnetic field reconnections, e.g., stellar flares
- Line of sight changes (rotation, jet wiggles…)

Variability is known on time scales from ms to 10^{10} yr

Synoptic, panoramic surveys → event discovery

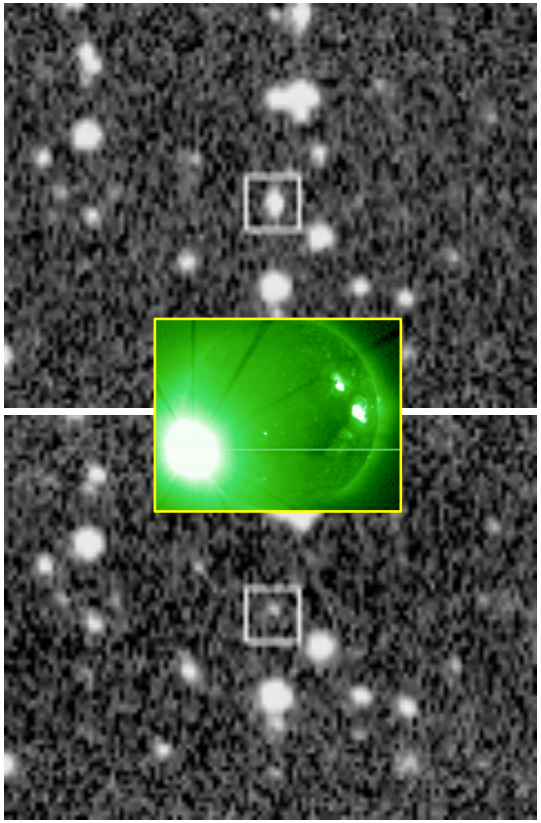
Rapid follow-up and multi-λ → keys to understanding

Intrinsically Variable Phenomena

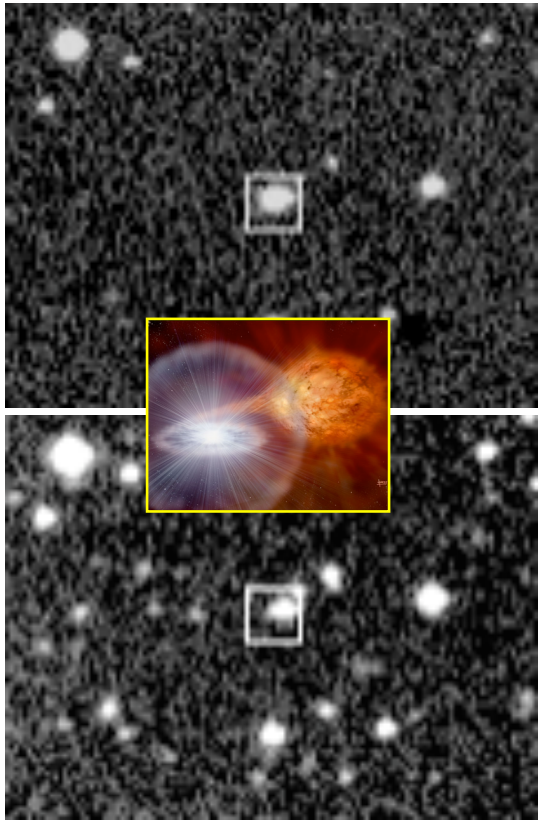
- Things we know about:
  - **Stars:** oscillations, noise, activity cycles, atmospheric phenomena (flares, etc.), eclipses, explosions (SNe, GRBs), accretion (CVs, novae), spinning beams (pulsars, SS 433, …)
  - **AGN:** accretion power spectrum, beaming phenomena
- Things we see, but don't really understand:
  - Faint fast transients
  - Archival OTs
  - Megaflares on normal stars
- Things we expect to see, and maybe we do:
  - Breakout shocks of Type II SNe
  - SMBH loss cone accretion events
  - BH mergers (LIGO, LISA?), QSO formation…?

Examples of CRTS Transients

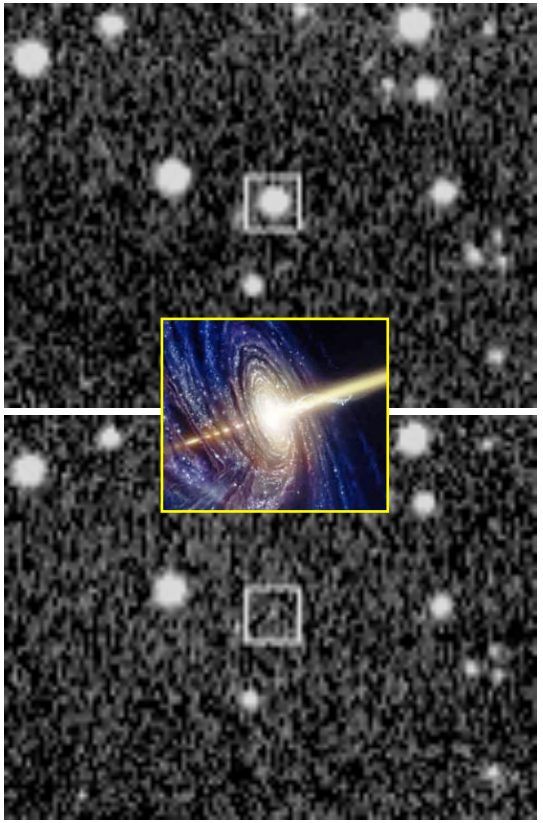
CSS090429:135125-075714
Flare star



CSS090429:101546+033311
Dwarf Nova



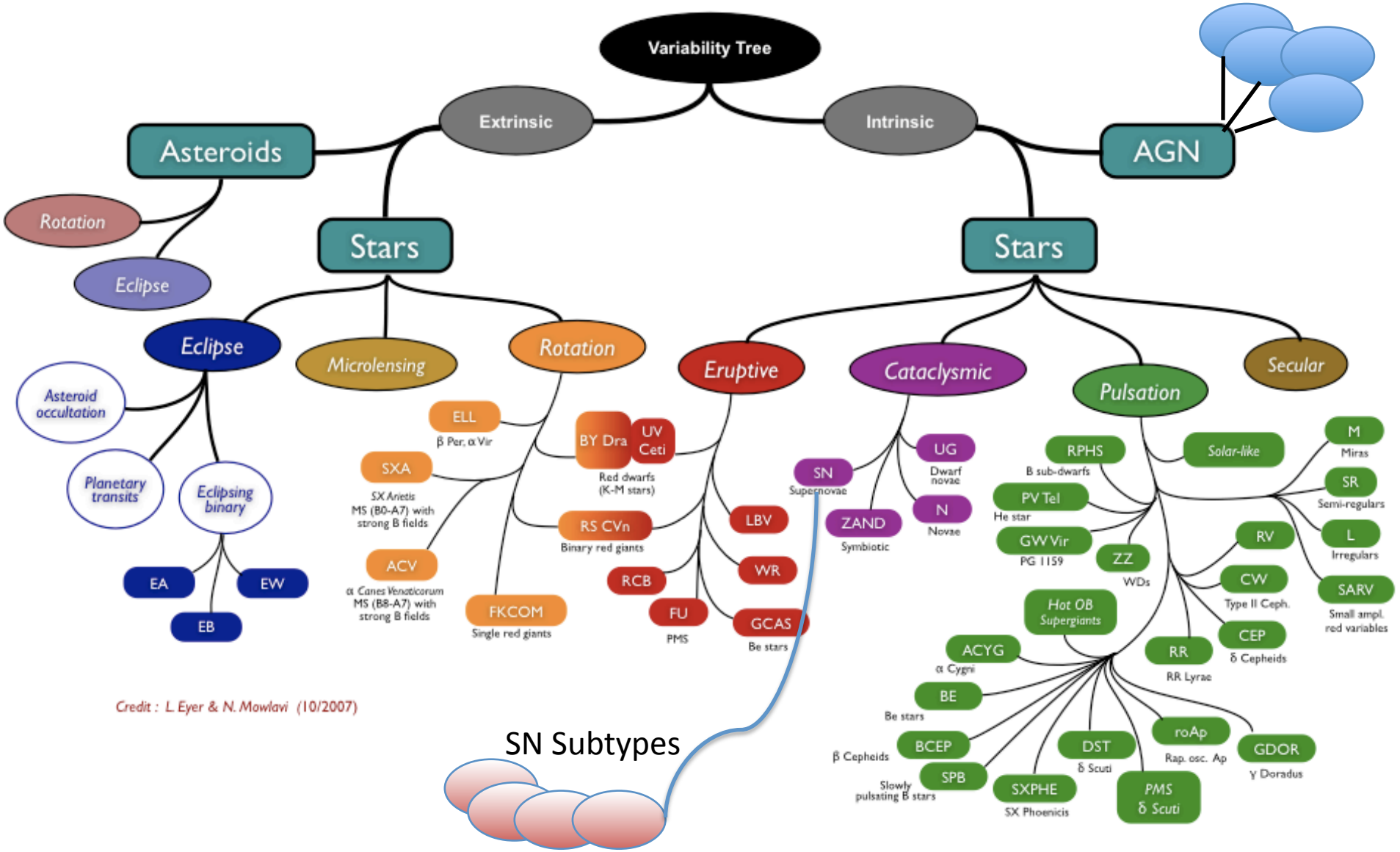
CSS090426:074240+544425
Blazar, 2EG J0744+5438



Vastly different physical phenomena, and yet they look the same!
Which ones are the most interesting and worthy of follow-up?

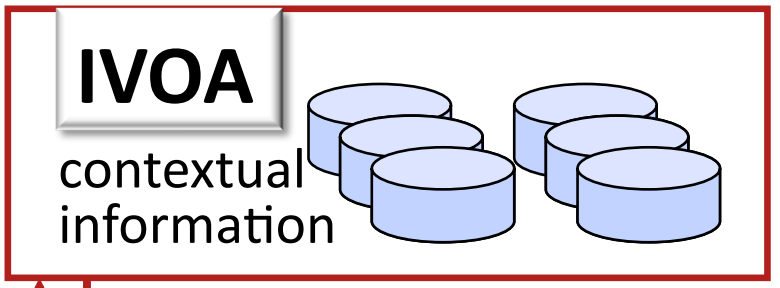
Rapid, automated transient classification is a critical need!

Semantic Tree of Astronomical Variables and Transients



Towards Automated Event Classification

A **necessity** for large synoptic surveys



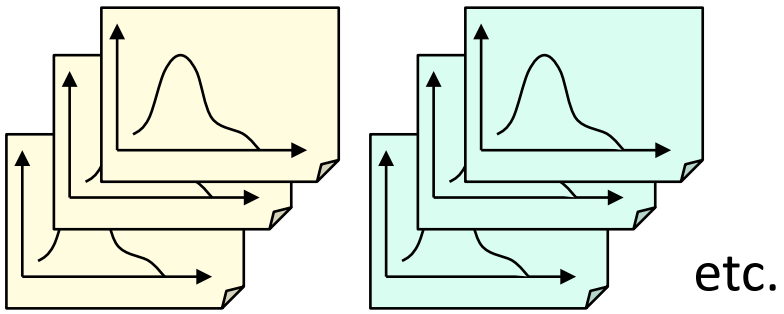
Event parameters:
$m_1(t)$, $m_2(t)$,…
α, δ, μ, …
image shape…

Event Classification Engine

P(SN Ia) = …
P(SN II) = …
P(AGN) = …
P(CV) = …
P(dM) = …
….

Classification probabilities (evolving, iterated)

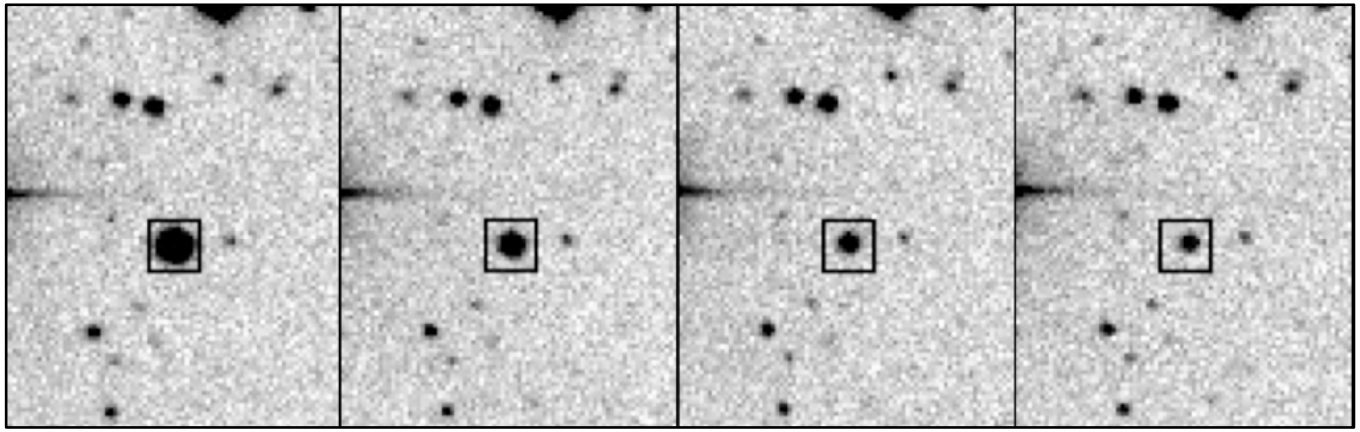
Expert and ML generated priors



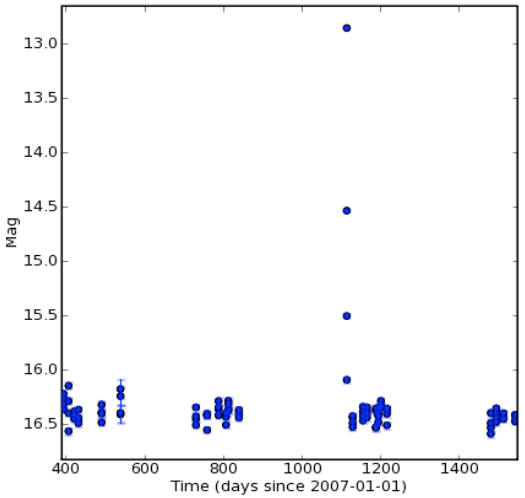
With M Turmon and B Moghaddam, JPL

What is a transient?

4 individual exposures, separated by 10 min



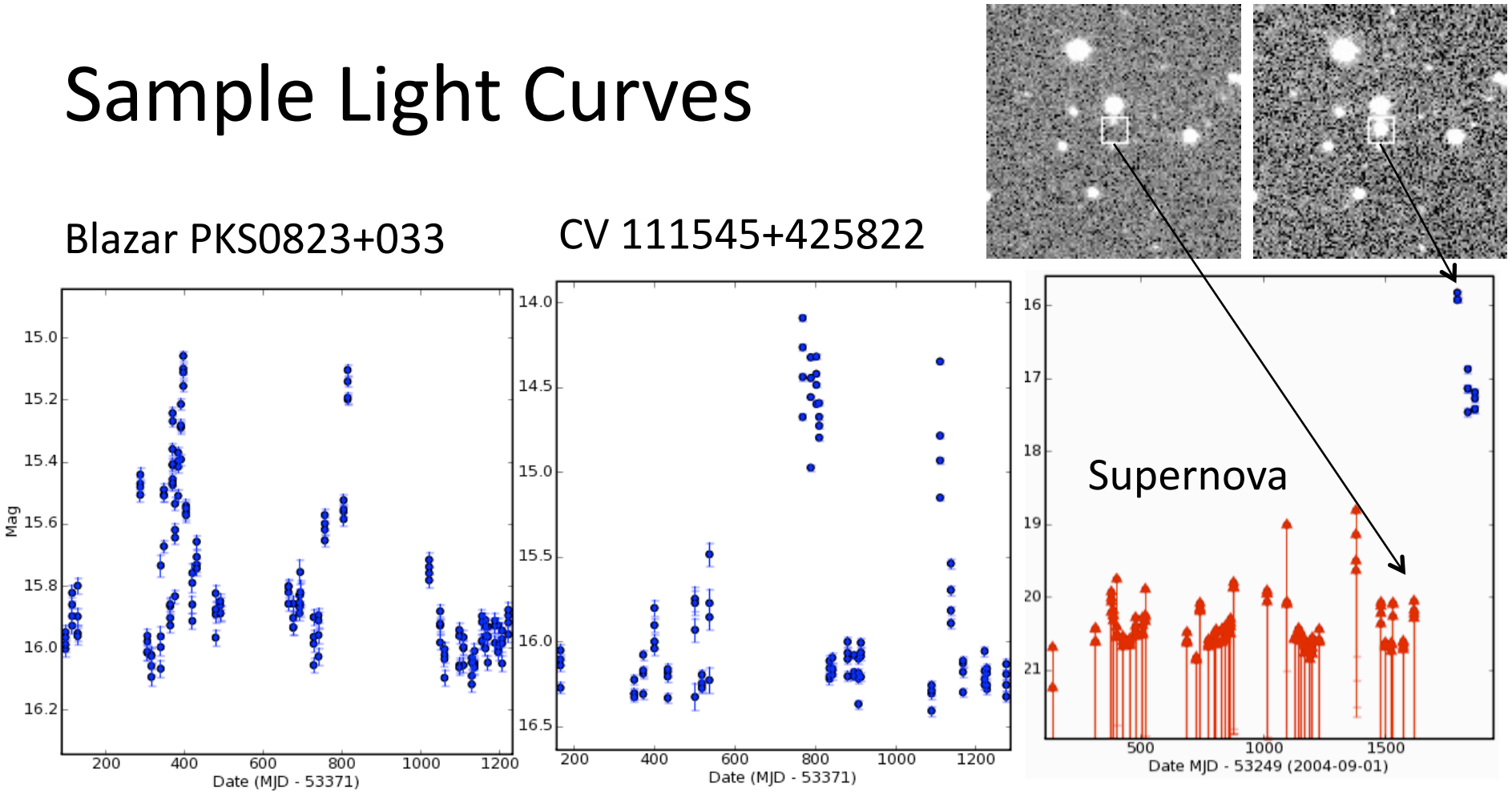
Light curve



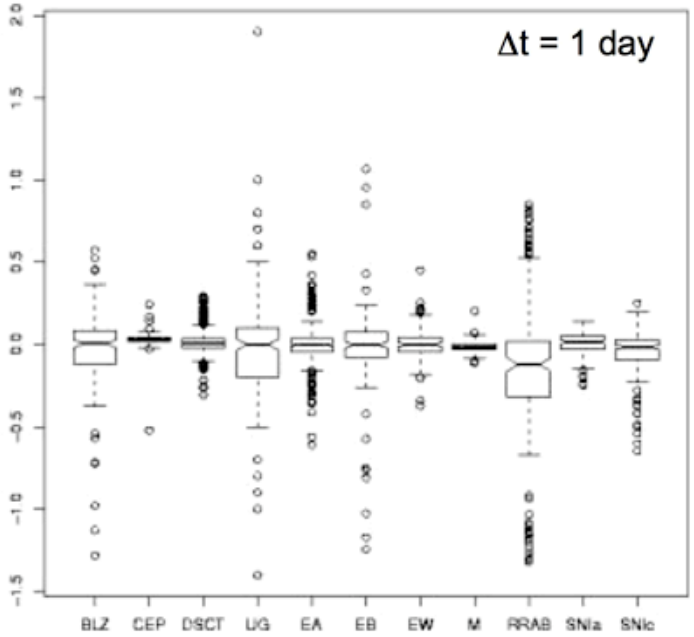
Fast transient (flaring dM), CSS080118:112149–131310

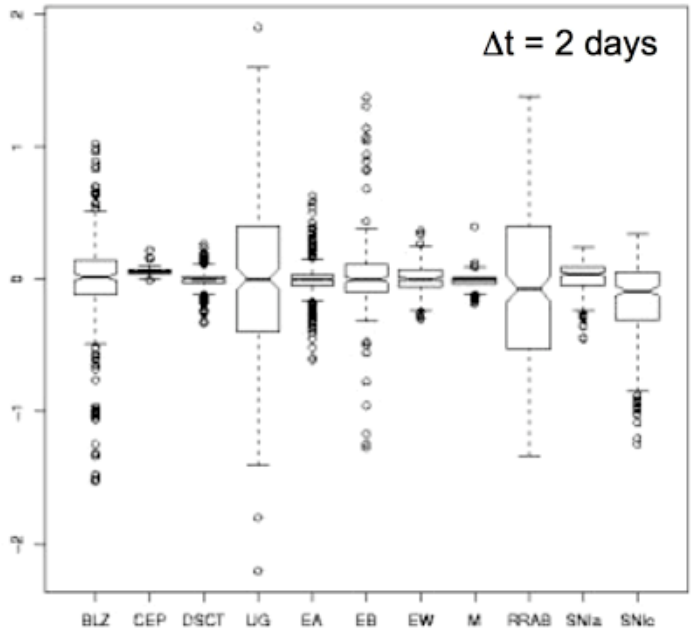
Something that has a large delta-magnitude for a small delta-time

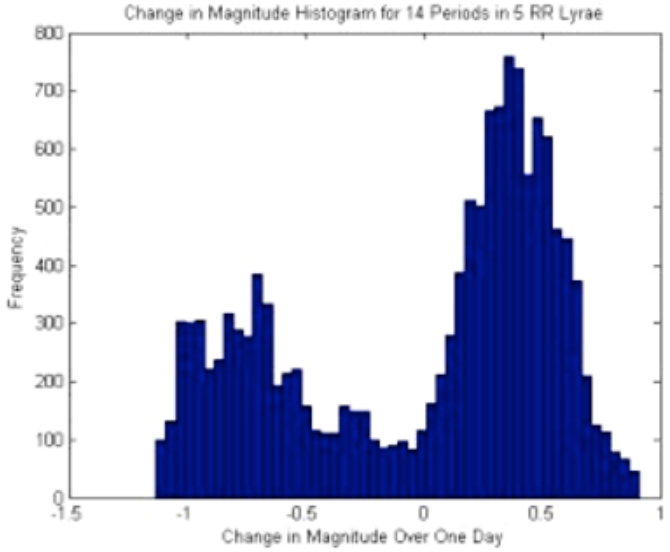
Sample Light Curves

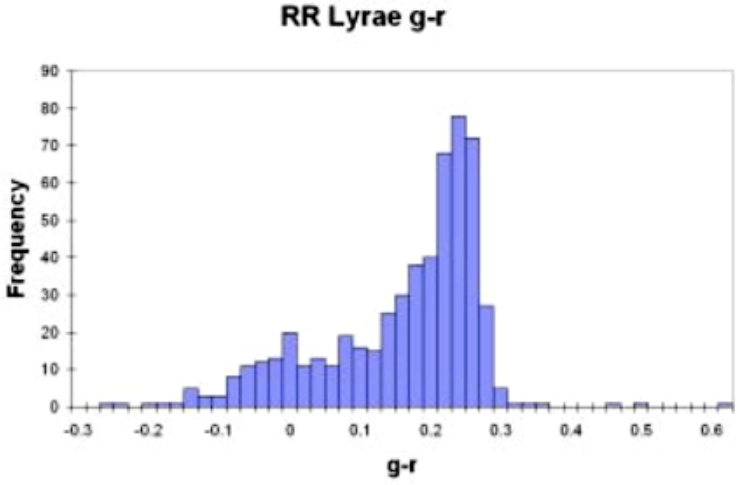


Variables and transients – the distinction is one of perception, and your aims

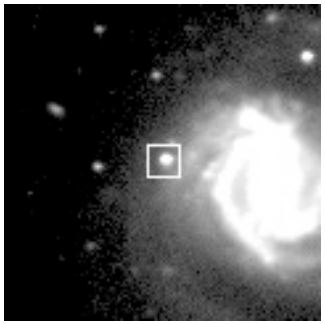

Δt = 1 day
BLZ
CEP
DSCT
UG
EA
EB
EW
M
RRAB
SNIa
SNIc

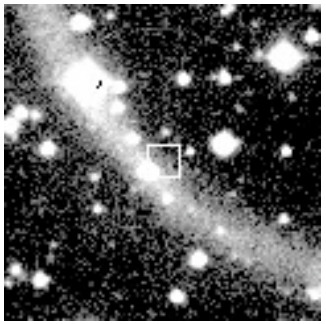

Δt = 2 days
BLZ
CEP
DSCT
UG
EA
EB
EW
M
RRAB
SNIa
SNIc

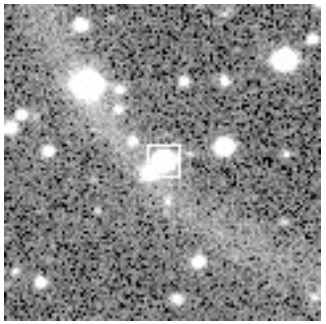

Change in Magnitude Histogram for 14 Periods in 5 RR Lyrae
Frequency
Change in Magnitude Over One Day

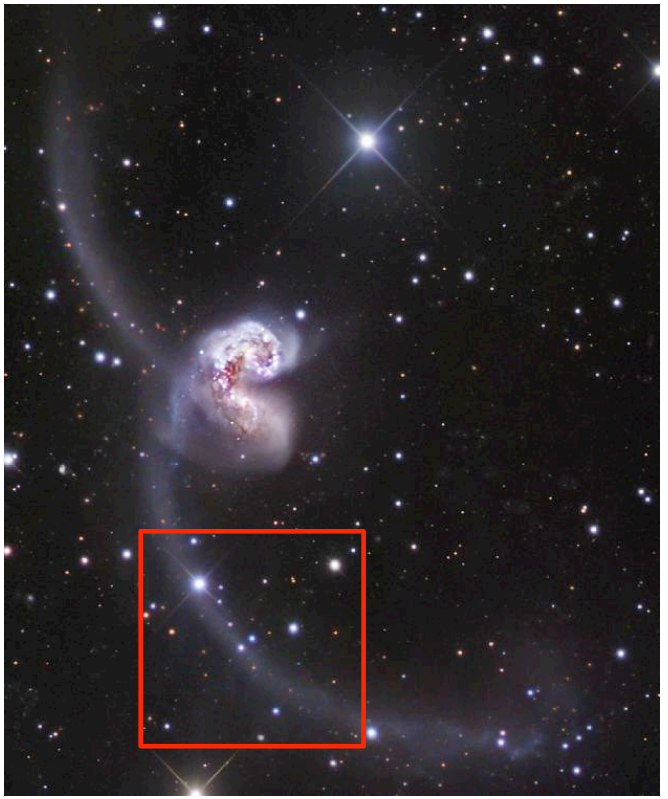

RR Lyrae g-r
Frequency
g-r

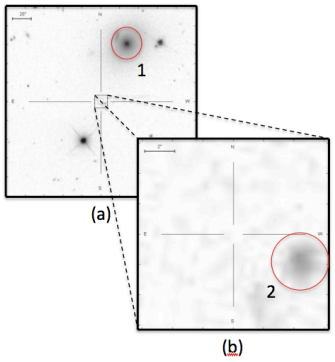
The importance of context

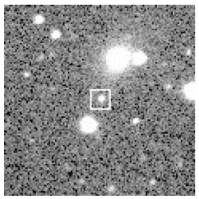


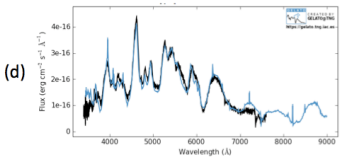




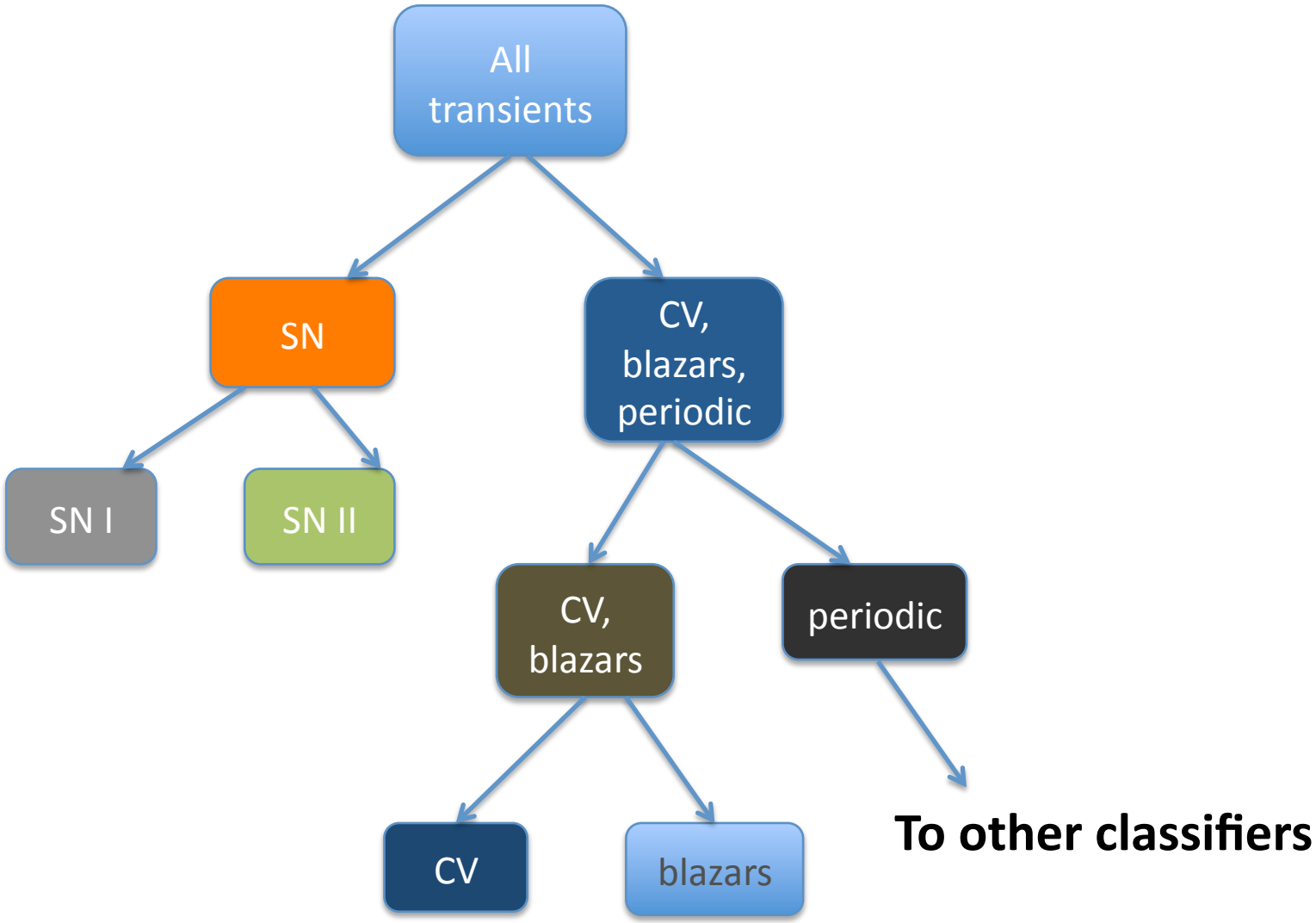








Broad, incomplete hierarchy



- More context information will help
- How to maintain uniformity?
- What when filters change?
- Uniformity of priors?
    - Number of objects
    - Their magnitude range
    - Spread over time
- Ground truth?

Questions raised by the Data paucity regime

- How many classes?
- Too few: probabilities incorrect (where do objects belonging to unrepresented classes go?)
- Too many: overlaps increase (e.g. SN of different types; variables of different types) and probability splits into smaller fractions
- What kind of winner?

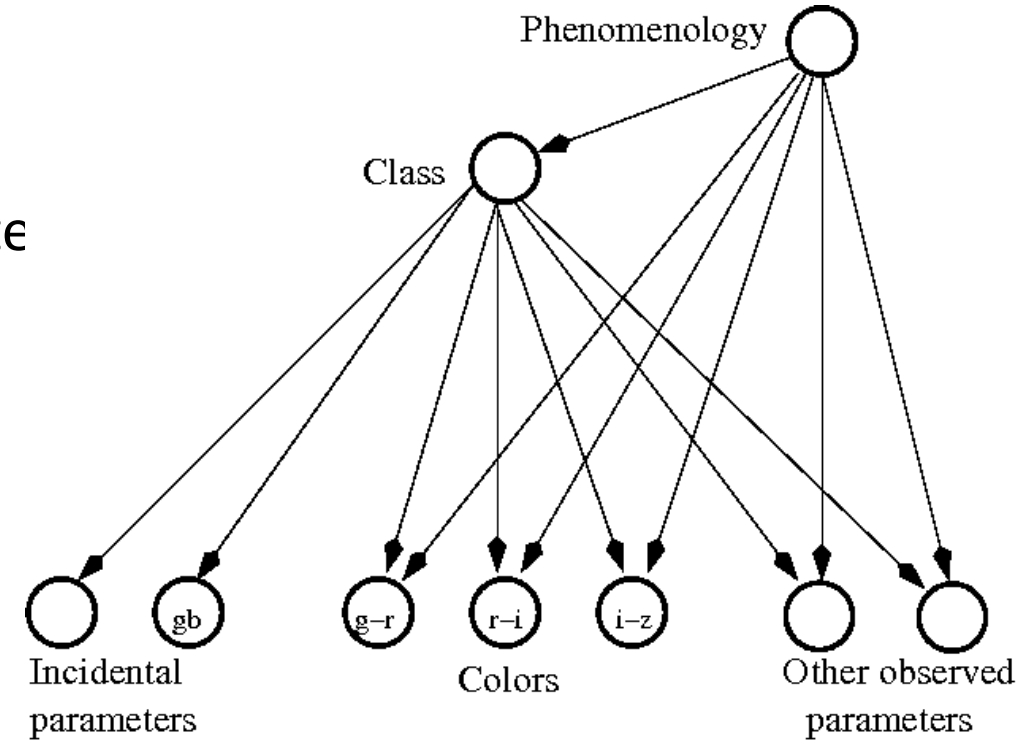
Naïve Bayes

$$P(y = k \mid x) = P(x \mid y = k)P(k)/P(x) \propto P(k)P(x \mid y = k) \approx P(k)\prod\prod_{b=1}^{B} P(x_b \mid y = k)$$

- x: feature vector of event parameters
- y: object class that gives rise to x (1<y<k)
- Certain features of x known: (position, flux)
- Others will be unknown: (color, delta-mag)
- Assumption: based on y, x is decomposable into B distinct independent classes (labeled x_b)
- This helps with the curse of dimensionality
- Also allows us to deal with missing values

Building Bayesian Networks

- Local dependencies, irrelevancies are evaluate using modeling
- Priors, likelihoods are obtained
- Data define network



Radio, for instance

Priors based on CRTS data (dm > 2)

3 colors + gb (WTA)	CV (0.65)	SN (0.71)	BL (0.33)	REST (0.23)
CV	0.72	0.08	0.08	0.13
SN	0.23	0.46	0.12	0.19
BL	0.24	0.03	0.49	0.24
REST	0.34	0.18	0.21	0.26

8% CV classified as SN, 65% of objects classified as CV are actually CV

- Winner-take-all
- At least 50%
- 40%+ and 10% diff
- allowing missing info

Based on a single set of observations

- Adding peripheral parameters like gb and distance to nearest galaxy helps
- Having additional colors is good
- More context info helps (flux in radio, x-ray etc.)
- Need to inculcate temporal information

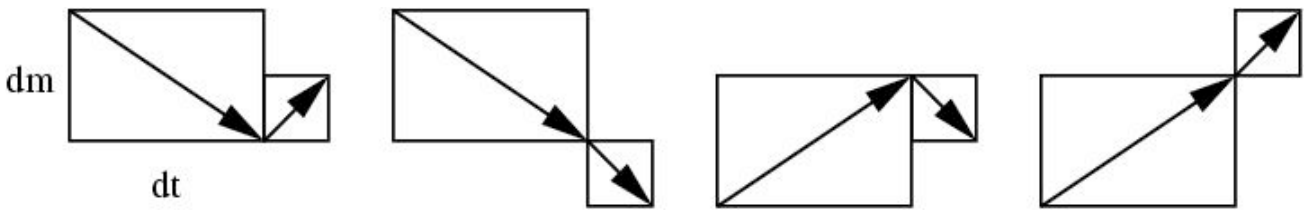
3 colors + gb + galaxy prox. (WTA)	CV (0.74)	SN (0.84)	BL (0.31)	(1-contam.)
CV	0.74	0.08	0.16	
SN	0.21	0.50	0.27	
BL	0.19	0.00	0.80	
				completeness

Transient classification mantra

- Obtain a couple of epochs in one or more filters
- Assigns probabilities for different classes
- Choose observations (filters, wavelengths) for best discrimination
- Feed the new observations back in
- Revise probabilities, choose observations, …
- Based on confirmed class (how?) revise priors

Characterization Vs. Classification

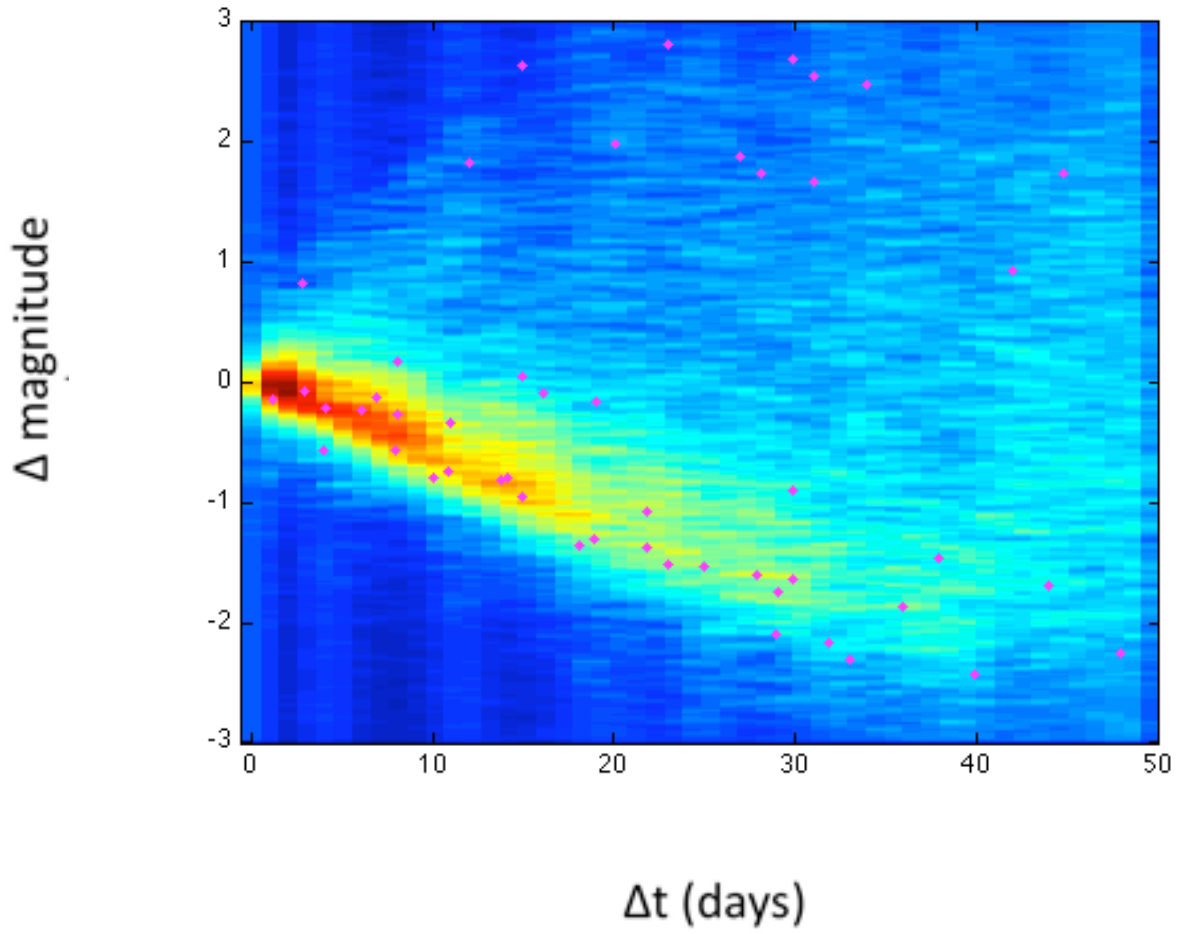
- Early focus on the extraction and dissemination of time series
- Characterizations is important
  - dm/dt
  - change of direction per unit time
  - change in periodicities (e.g,. wavelet or fourier decomposition);
  - variation in dm/dt
  - acceleration in dm/dt



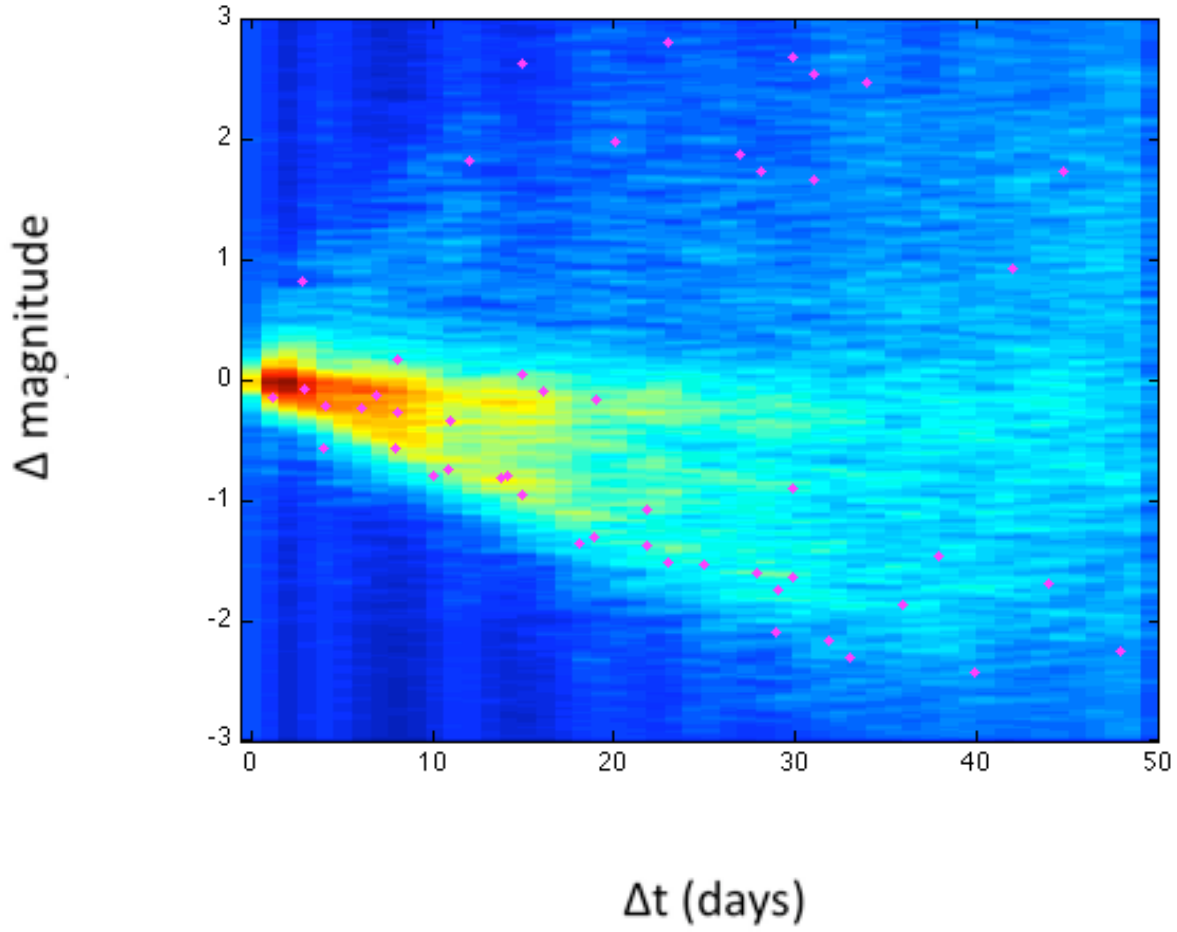
Most SNe will not become fainter and then brighten up

- Non-sparse time series (Many methods; relatively easy)
- Sparse time series (Non-trivial)
  - Non-gridded
  - Error-bars
  - Upper limits

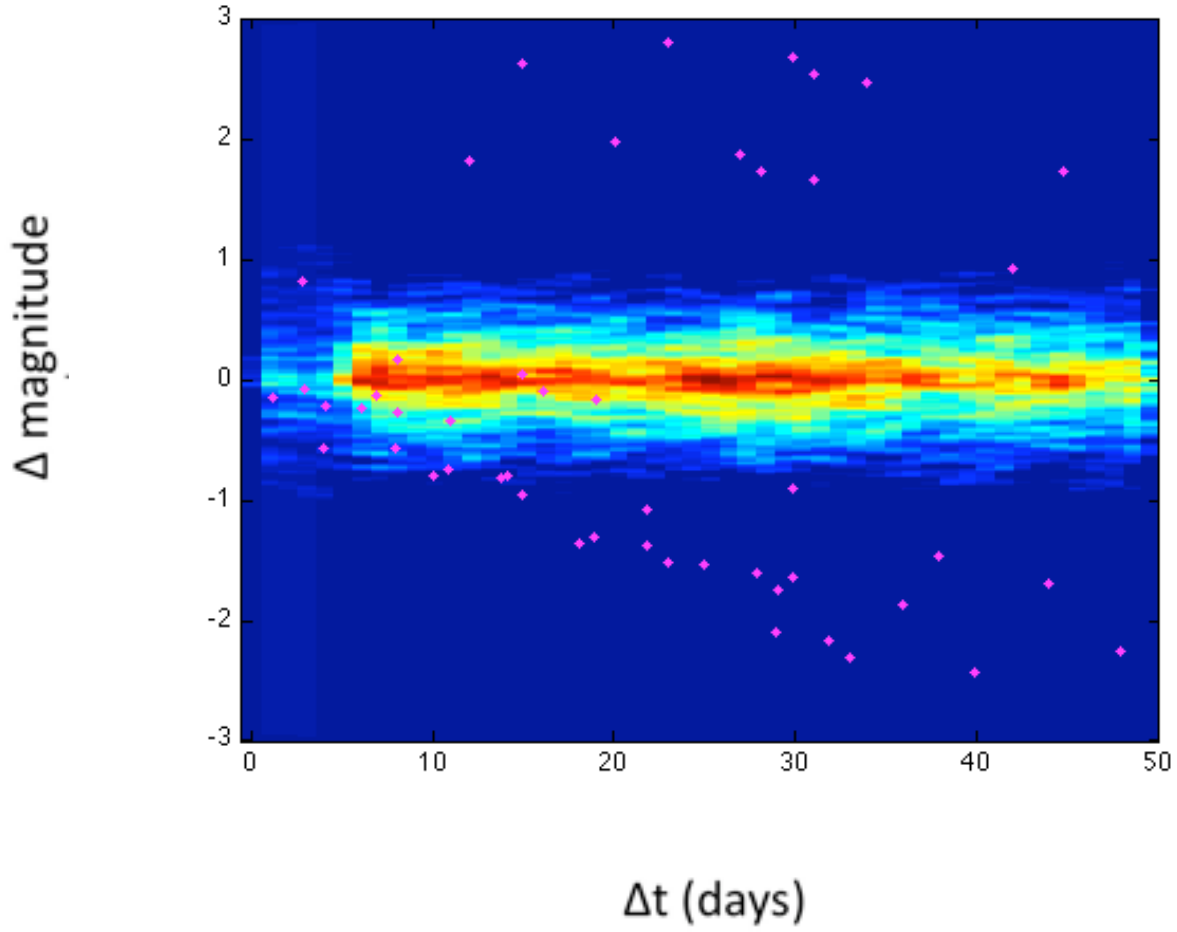
SN Ia



SN IIP



RR Lyrae



Automated Generation of Finding Charts for the Follow-Up Observing

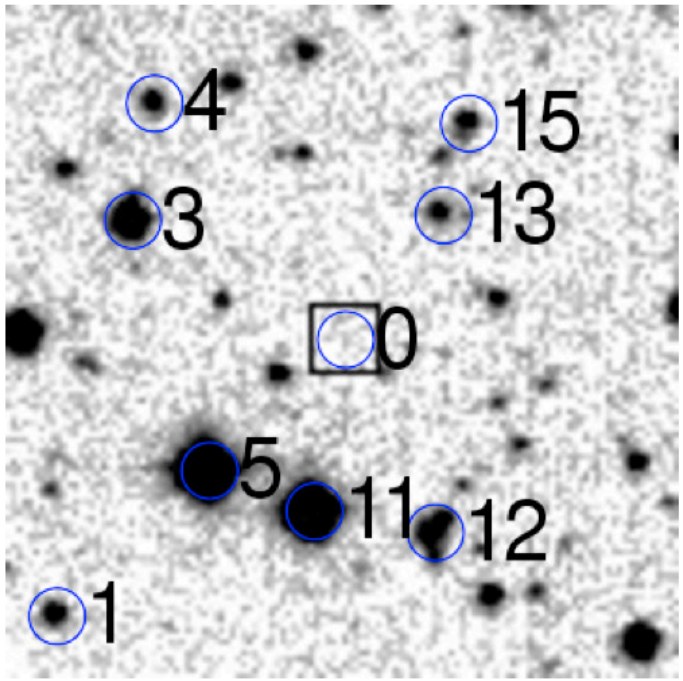
Transient CSS100320:135108+133407

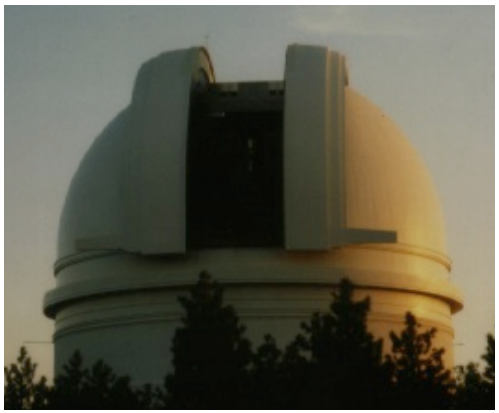
RA Dec (2000)
207.78253 13.56852

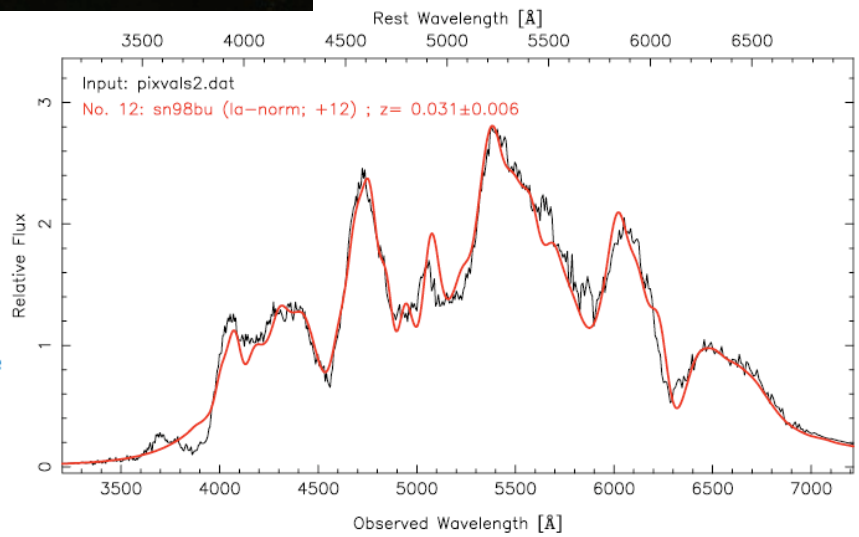
Rough Mag:
19.4

Discovery data
Current lightcurve
Pre and post-discovery CSS images
SDSS data
Images from other surveys
P60 Follow-up

Pre-discovery 5' Catalina Sky Survey coadd image (transient location marked with 0)
N is towards the top and E is to the left.







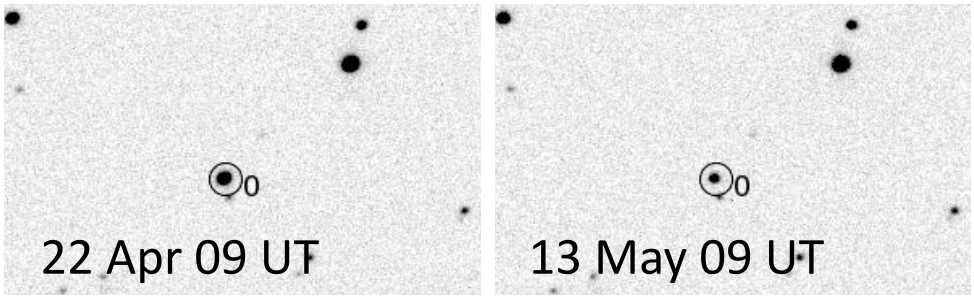
ID	RA	Dec (2000)	mag	delmag	delra (")	deldec
0	207.78253	13.56852	19.4	0.0	0.0	0.0
3	207.81002	13.58293	15.5	-3.9	96.2	51.9
4	207.80724	13.59740	18.4	-0.9	86.5	104.0
5	207.80025	13.55195	12.8	-6.6	62.0	-59.7
8	207.79109	13.56361	18.3	-1.0	30.0	-17.7

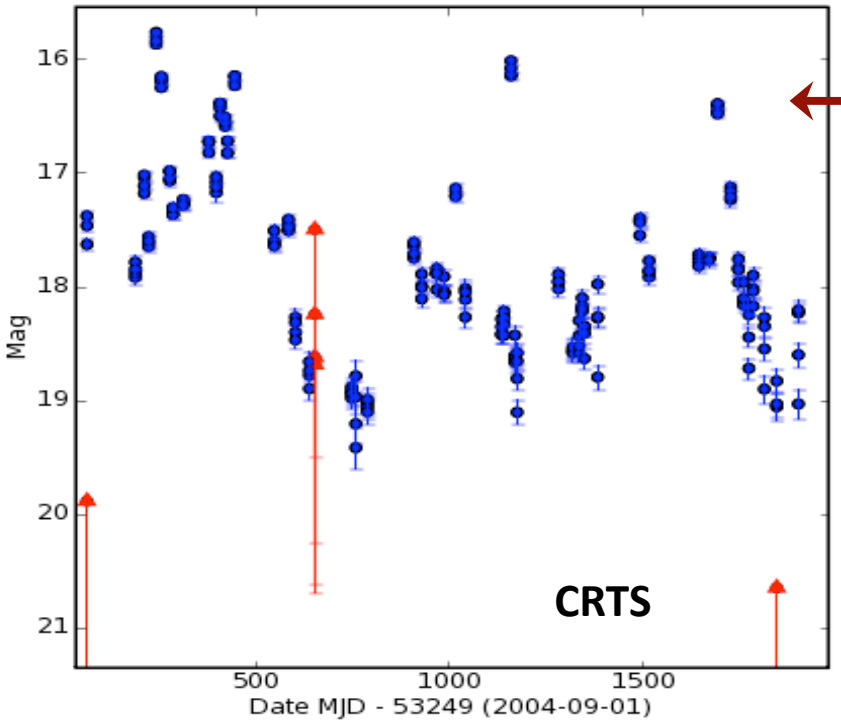
Follow-Up Observations:

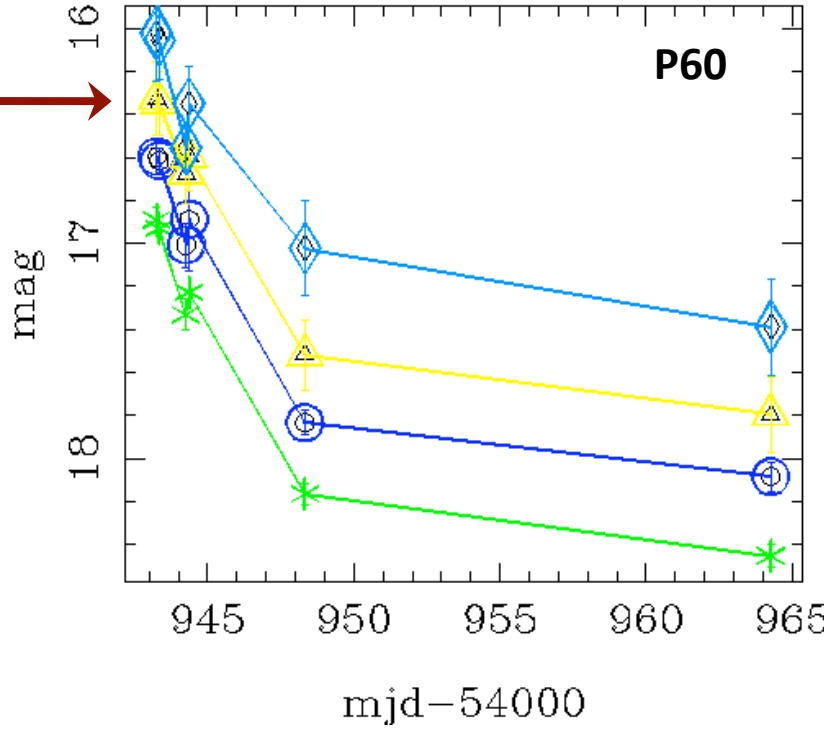
Lead: A. Mahabal

- Photometry (P60, NMSU, DAO, HTN, India, Mexico, etc.)
- Spectroscopy (Gemini N+S, Keck, P200, SMARTS, IGO, MDM)

CSS090421:174806+340401 A blazar, also monitored at OVRO in radio



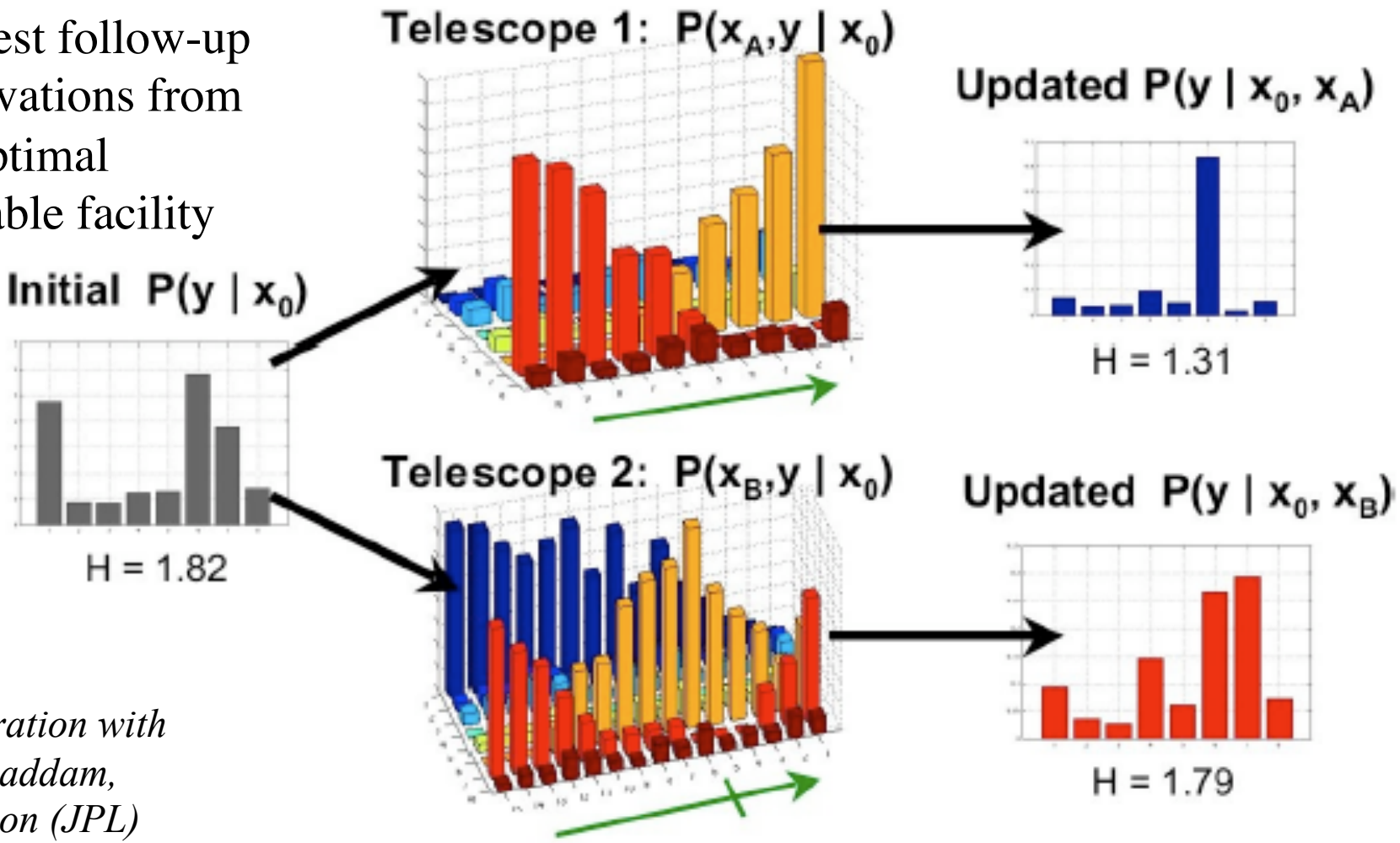




Automating the Optimal Follow-Up

What type of follow-up data has the greatest potential to discriminate among the competing models (event classes)?

Request follow-up observations from the optimal available facility



Collaboration with
B. Moghaddam,
M. Turmon (JPL)

CRTS Event Detections

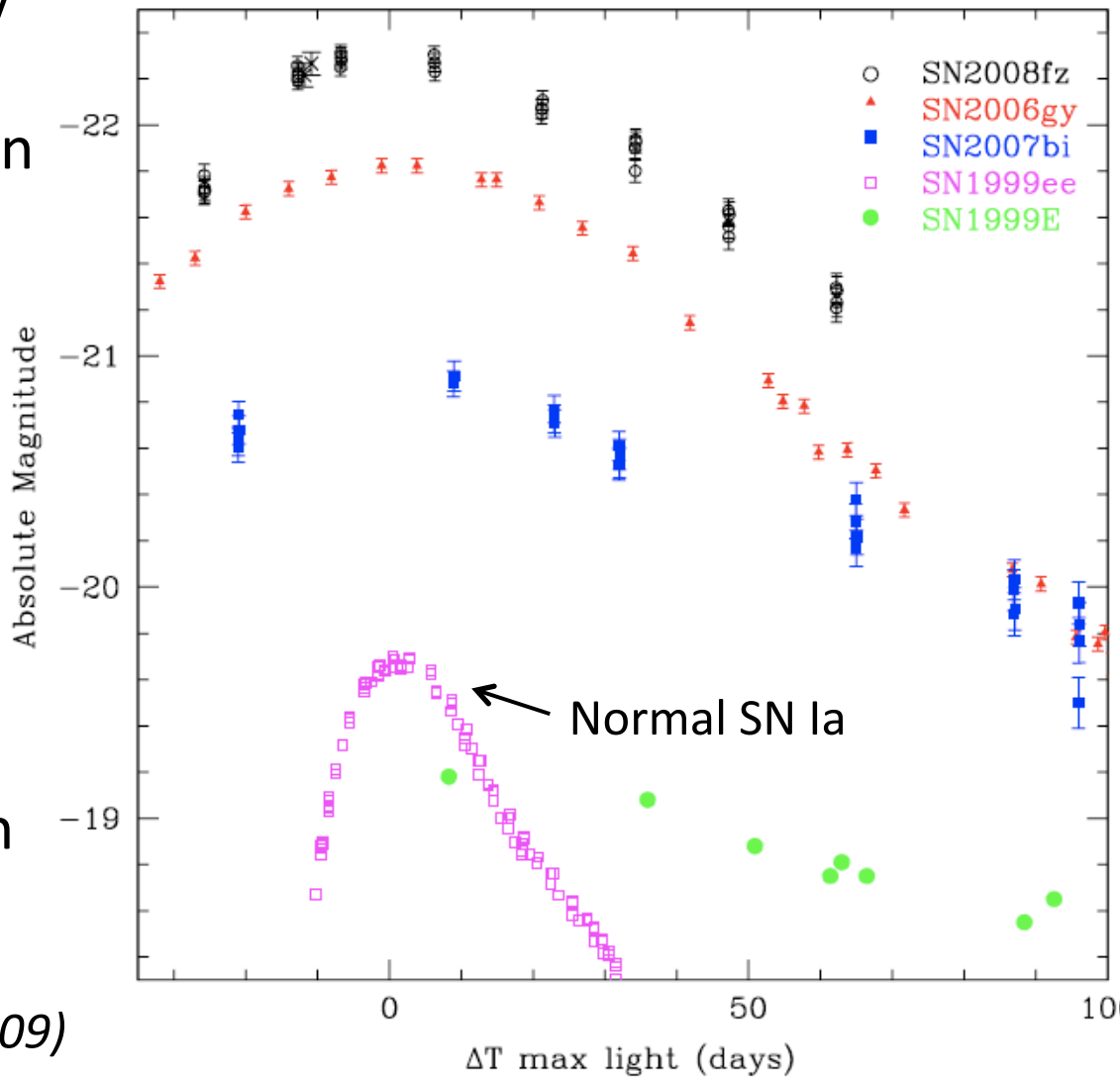
Distinct Events Detection Statistics as of 30 Nov 2010 UT:

Telescope	All OTs	SNe	CV	Blazars	Ast/Flr	CV or SN	Other
CSS	1623	432	419	97	182	240	281
MLS	670	81	17	3	60	211	316
SSS	98	13	38	6	2	16	23
Total	2391	526	474	106	244	467	620

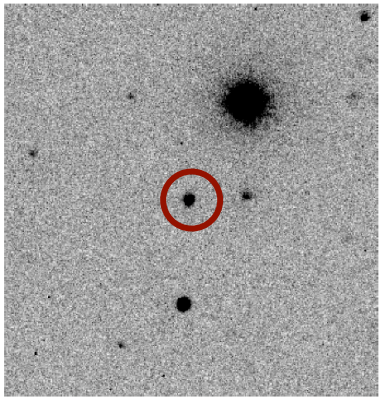
- Threshold set deliberately very high – only the most dramatic transients are pulled out in the real time
- About 1 strong transient per 10^6 source detections
- The rate of significant transients/variables is at least an order of magnitude higher
- Many events are re-detected repeatedly (not counted above)

2008fz: The Most Luminous Supernova

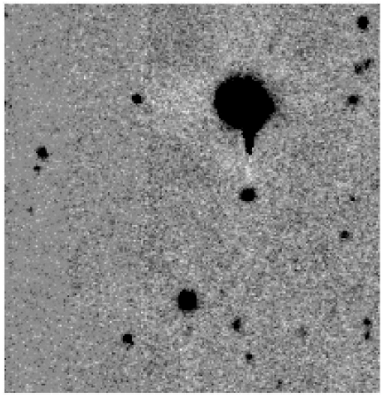
- Brightest type IIn known (5 times brighter than the Milky Way)
- Host galaxy > 50 times fainter than Milky Way
- A possible example of a pair-production SN?



Discovery



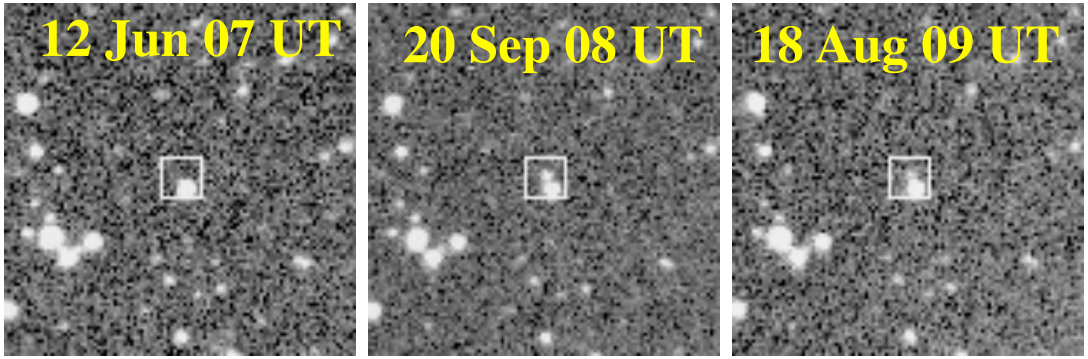
Comparison

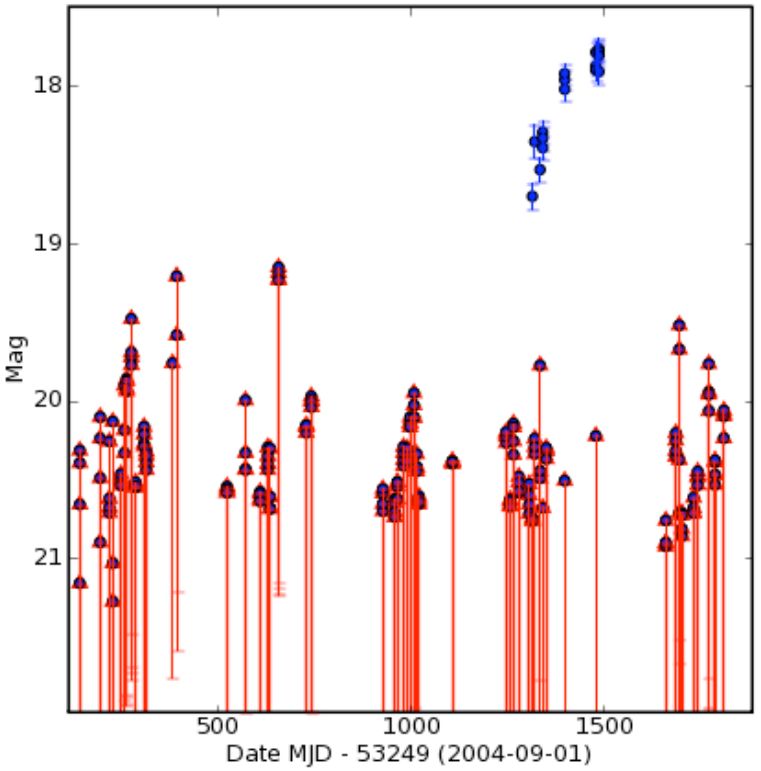


(Drake et al. 2009)

The Slow SN 2008iy

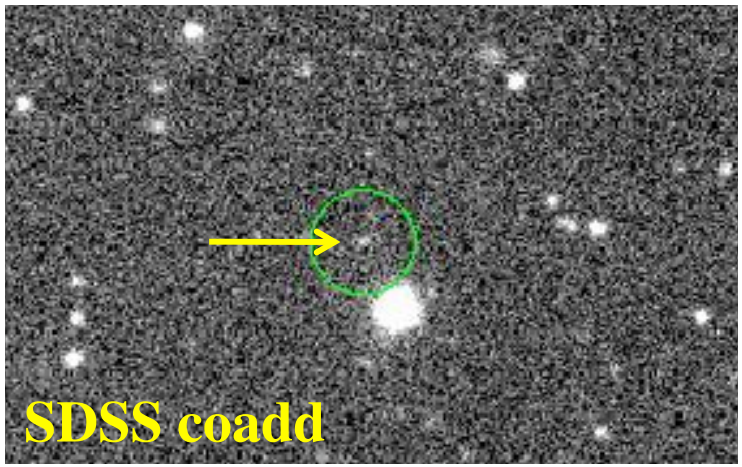
= CSS080928:160837+041627



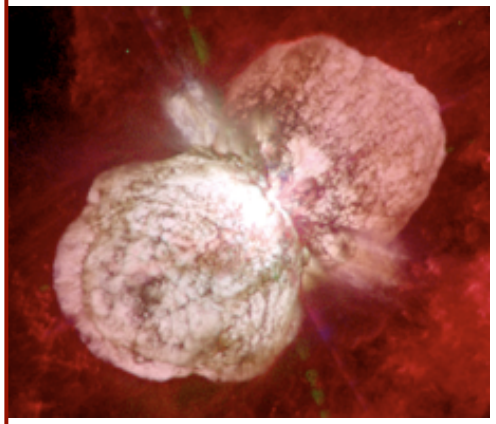


Longest-lasting type IIn at z = 0.041
it took **> 400 days** to reach the peak!

Host galaxy ***> 500 times fainter***
than the Milky Way (M ≈ −13)



Possibly from an
~ η Carinae type progenitor:
expanding SN interacts with
the material from past
outbursts

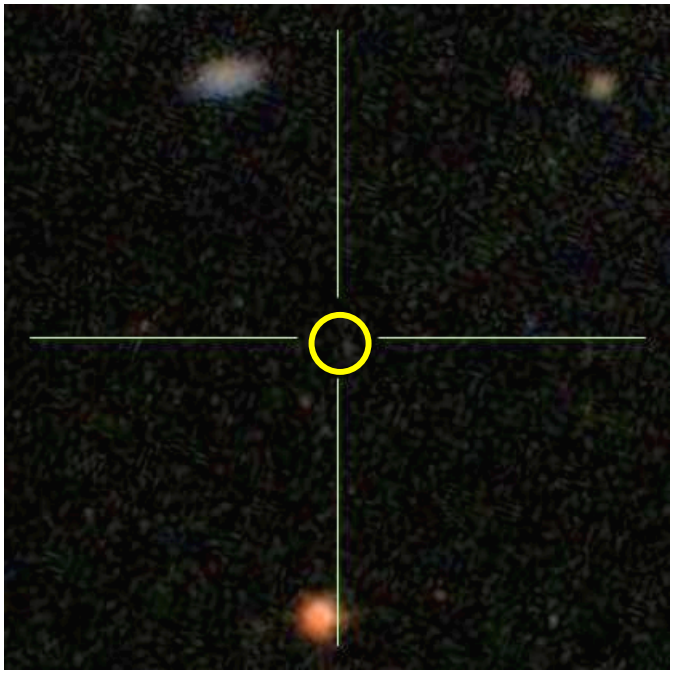


Luminous SNe in Underluminous Hosts

- A number of SNe discovered in extremely faint dwarf galaxy hosts (M ≈ –12 or –13), e.g., 2008fz, 2008iy, 2008hp, 2009aq, etc.

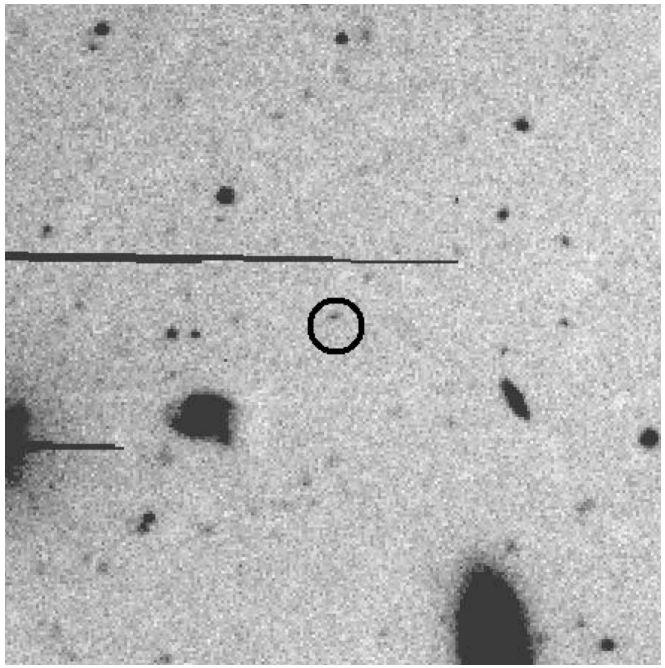
  ⇨ Huge specific SN rates (per unit stellar mass)

- Many are hyperluminous SNe ⇨ massive star progenitors?
- Low mass host ⇨ Low metallicity ⇨ Top-heavy IMF ??
- Possible connection with GRB hosts? Local Pop. III analogs?

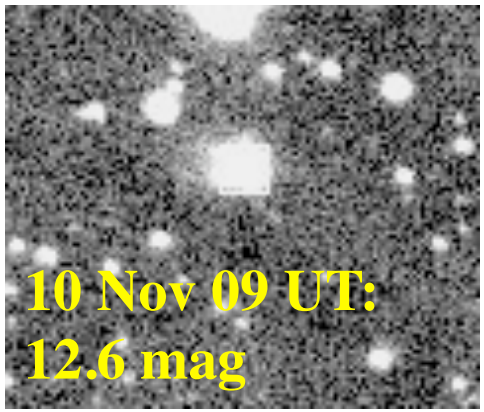


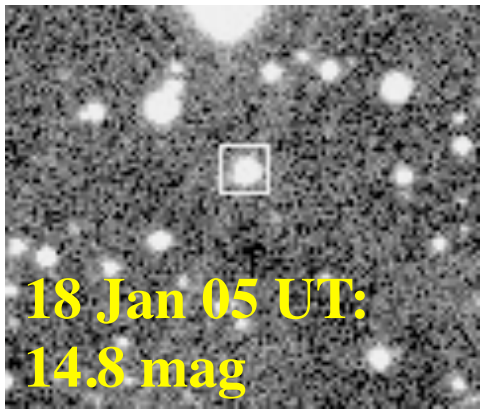
⇦ SN 2008hp
Host M_r ≈ –12.4

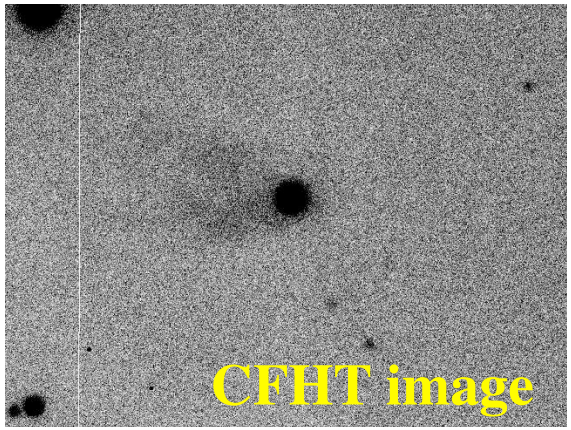
SN 2009aq ⇨
Host M_r ≈ –13

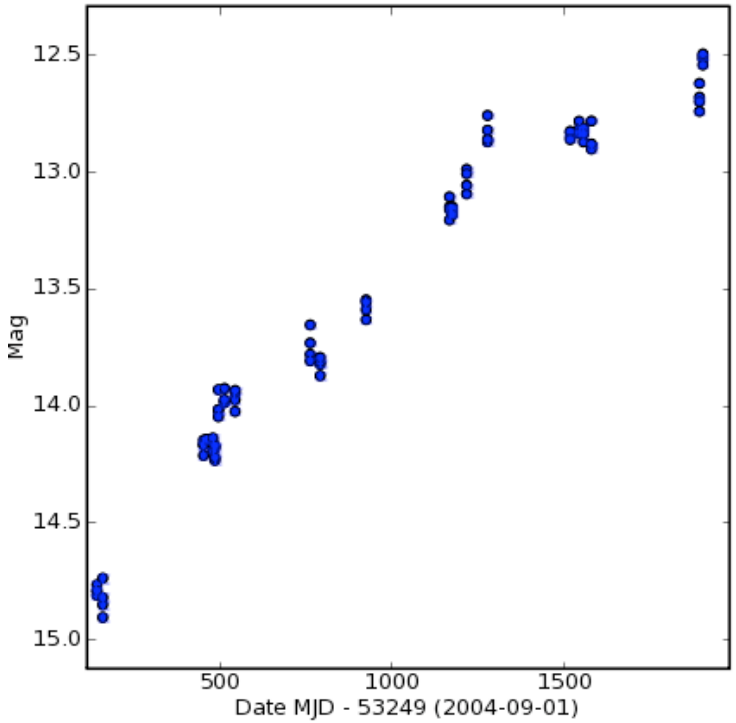


Discovery of a New FU Ori Object



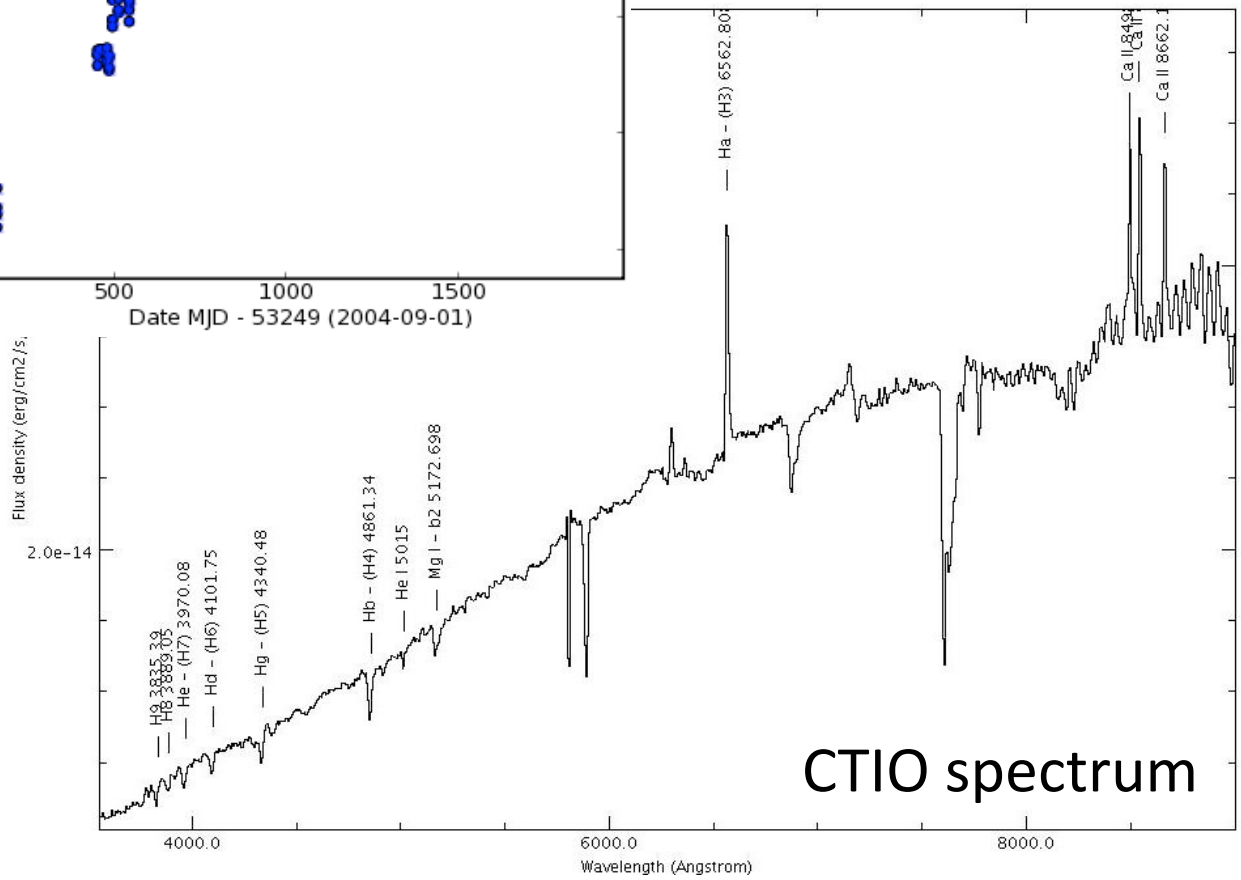






IRAS 06068–0641 =
CSS091110:060919–064155

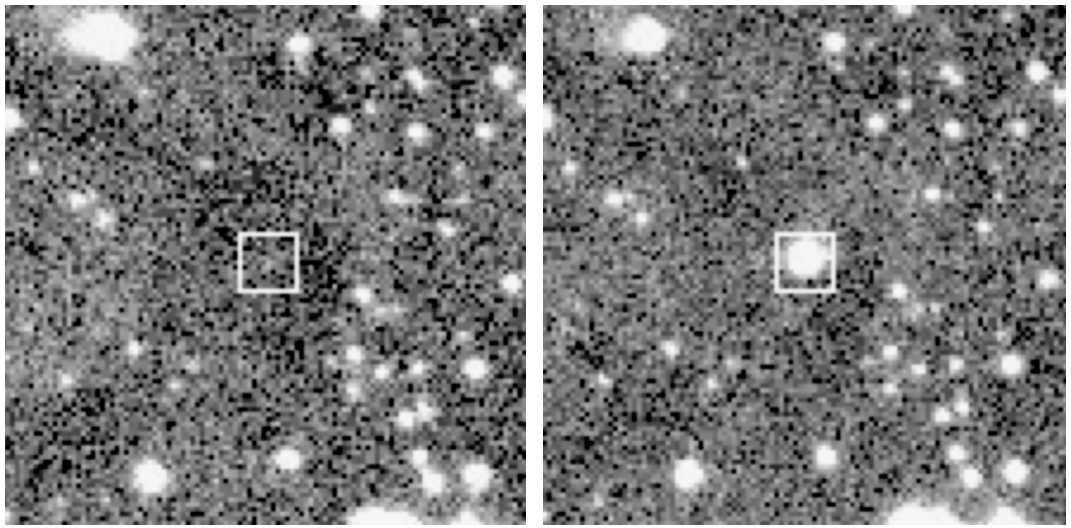
Wils et al. ATel 2307

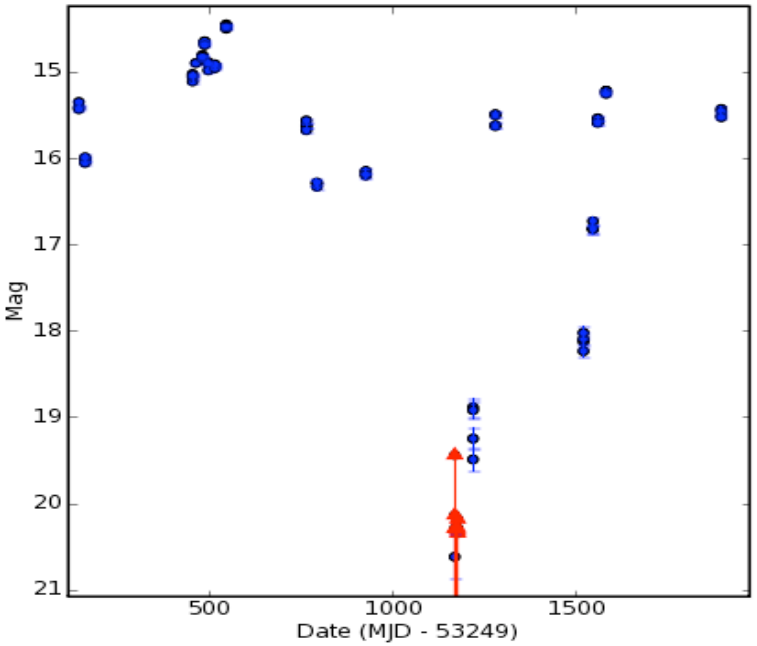


CTIO spectrum

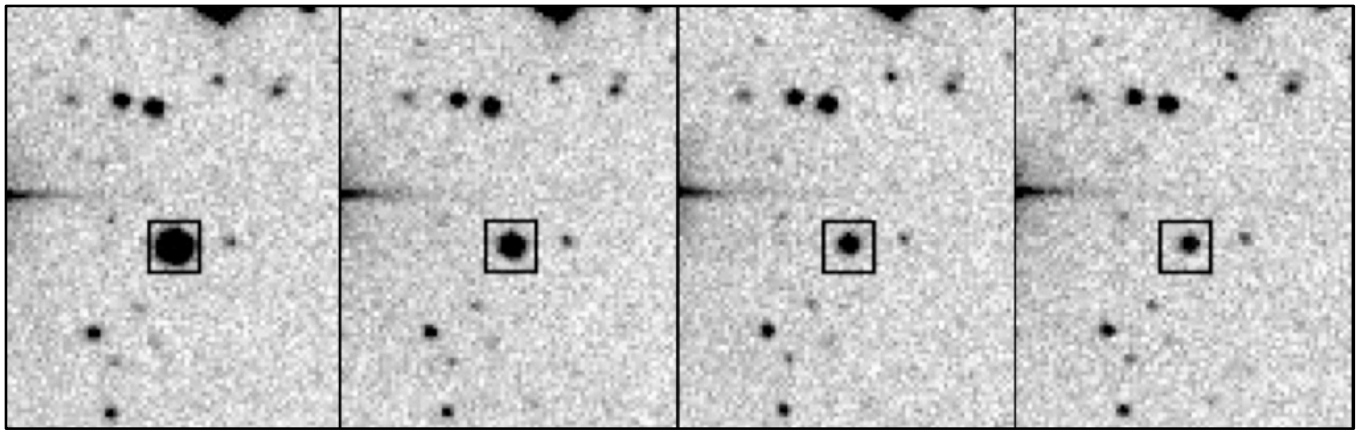
Unsettled Stars

IRAS 06068–0643 (UX Ori type) young star

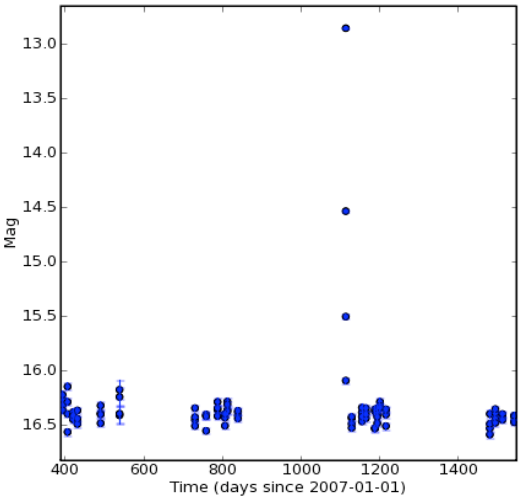




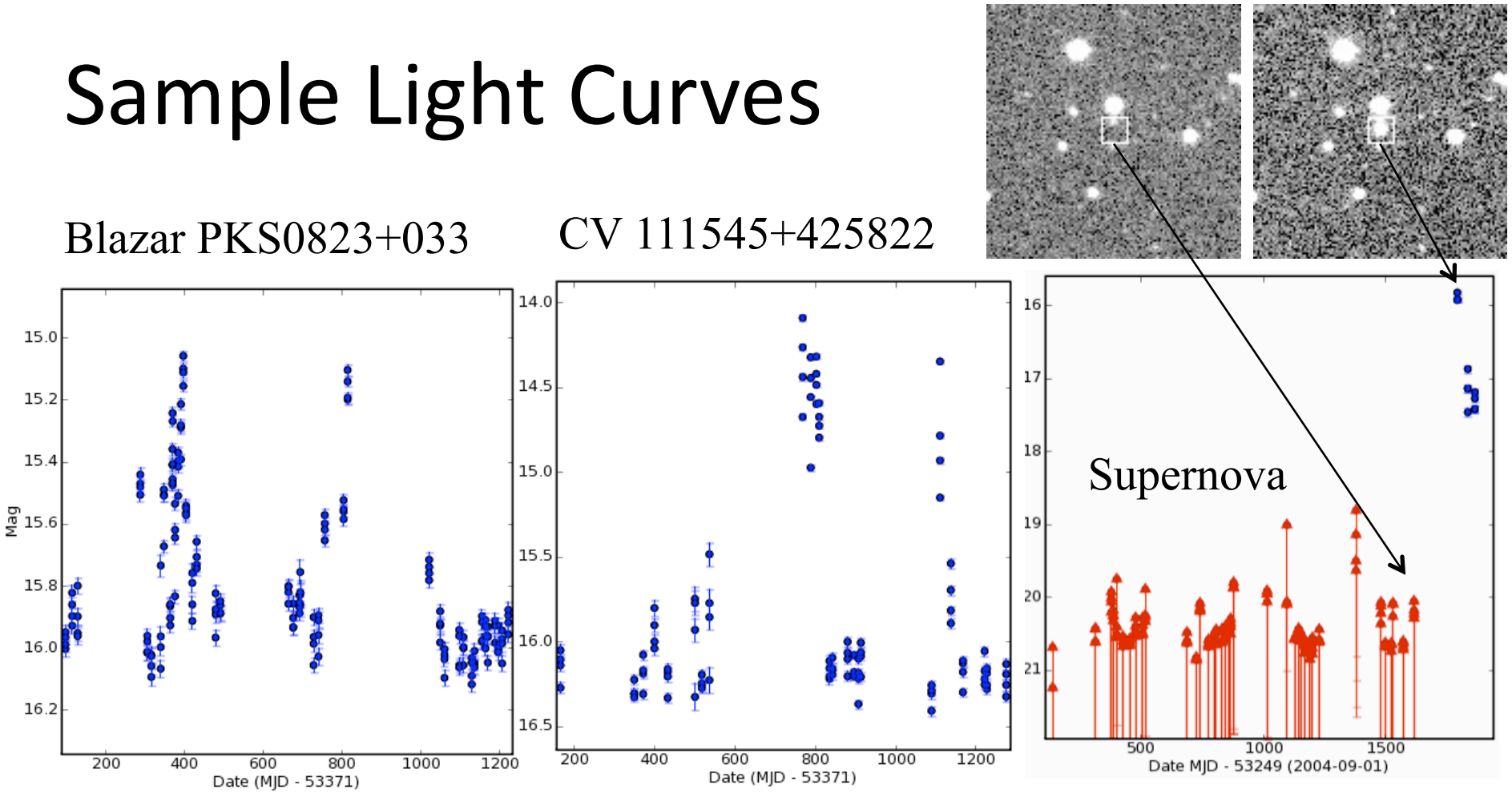
Fast transient (flaring dM), CSS080118:112149–131310
4 individual exposures, separated by 10 min



Light curve



Sample Light Curves



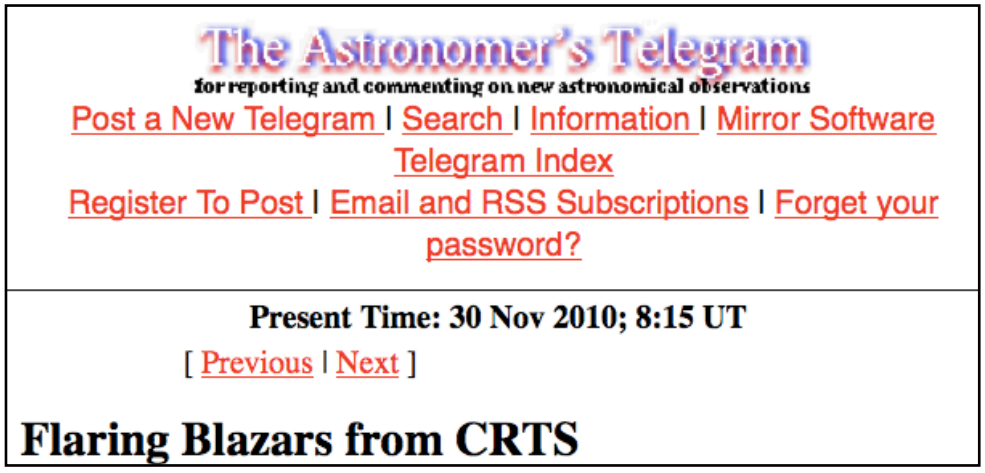
The plan is to produce light curves for every detected source in the survey (> 10^8 sources), make them publicly available, and mine that data set. Light curves are currently generated on demand for transient sources, blazars, etc.

Event Publishing / Dissemination

- Real time: VOEvents, Twitter, iApp (thousands of events)
  - Also on SkyAlert.org, feeds to the WWT, GoogleSky
- Next day: annotated tables on the CRTS website

CSS ID	RA (J2000)	Dec (J2000)	Date	Mag	CSS images	SDSS	Others	Followed	Last	LC	Classification
CSS091121:221159+263906	332.99697	26.65153	20091121	18.33	911211261084134848	no	34848	no	2009-11-21	34848	SN/Blazar mag 21
CSS091121:013728+253450	24.36768	25.58061	20091121	17.78	911211260084103595	no	03595	no	2009-11-21	03595	SN/CV
CSS091121:032627+070744	51.61364	7.12902	20091121	16.68	911211070194124436	no	24436	no	2009-11-21	24436	CV mag 21
CSS091121:033232+020439	53.13295	2.07747	20091121	16.93	911211010194134434	no	34434	no	2009-11-21	34434	CV mag 20
CSS091121:085600-051945	133.99922	-5.32906	20091121	18.17	911210040484107252	no	07252	no	2009-11-21	07252	SN CFHT mag 22 gal
CSS091120:100525+511639	151.35223	51.27742	20091120	18.80	911201520354108835	yes	08835	no	2009-11-20	08835	SN SDSS mag 21,9 gal
CSS091120:082908+482639	127.28503	48.44423	20091120	15.69	911201490314109371	yes	09371	no	2009-11-20	09371	CV/SN SDSS mag 21,6 gal?
CSS091120:004417+411854	11.07004	41.31494	20091120	17.00	911201400044145995	yes	45995	no	2009-11-20	45995	Nova M31 2009-11d
CSS091120:001019+410455	2.58044	41.08191	20091120	16.69	911201400014137919	no	37919	no	2009-11-20	37919	CV mag 20,0

- Days/weeks: ATel, CBET for selected transients (~ 200 so far)



The Astronomer's Telegram

for reporting and commenting on new astronomical observations

Post a New Telegram | Search | Information | Mirror Software

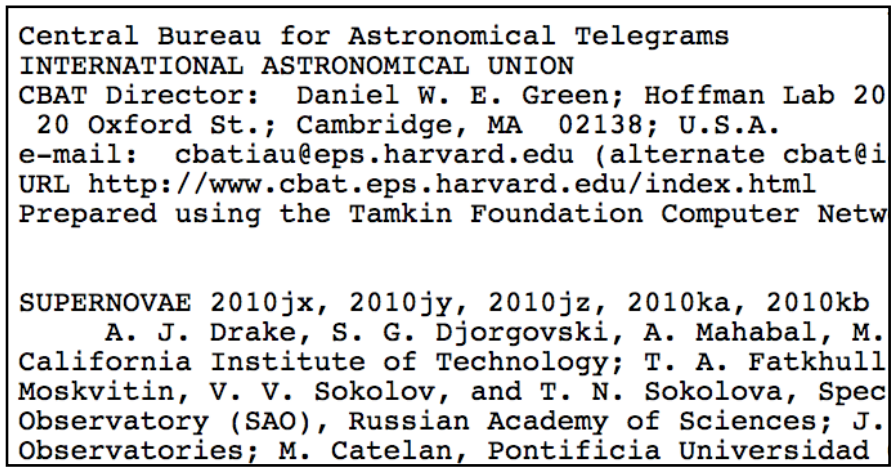
Telegram Index

Register To Post | Email and RSS Subscriptions | Forget your password?

Present Time: 30 Nov 2010; 8:15 UT

[Previous | Next]

Flaring Blazars from CRTS



Central Bureau for Astronomical Telegrams
INTERNATIONAL ASTRONOMICAL UNION
CBAT Director: Daniel W. E. Green; Hoffman Lab 20
20 Oxford St.; Cambridge, MA 02138; U.S.A.
e-mail: cbatiau@eps.harvard.edu (alternate cbat@i
URL http://www.cbat.eps.harvard.edu/index.html
Prepared using the Tamkin Foundation Computer Netw

SUPERNOVAE 2010jx, 2010jy, 2010jz, 2010ka, 2010kb
A. J. Drake, S. G. Djorgovski, A. Mahabal, M.
California Institute of Technology; T. A. Fatkhull
Moskvitin, V. V. Sokolov, and T. N. Sokolova, Spec
Observatory (SAO), Russian Academy of Sciences; J.
Observatories; M. Catelan, Pontificia Universidad

Real Time Event Publishing via *VOEvents* and *SkyAlert*

http://skyalert.org

From the CRTS stream.
Catalina Real-time Transient Survey
Position is 115.98635,21.1753 ± 0.0012
This portfolio initiated 2009-11-11 08:35:18

See context in WorldWideTelescope

PI: R. Williams

Basic event info

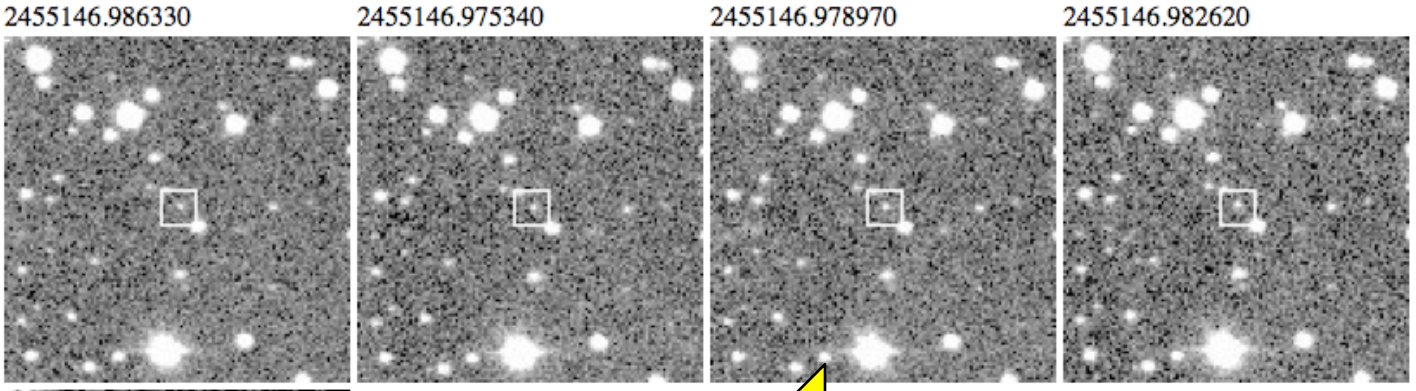
CRTS
911111210394136030
2009-11-11T11:34:58

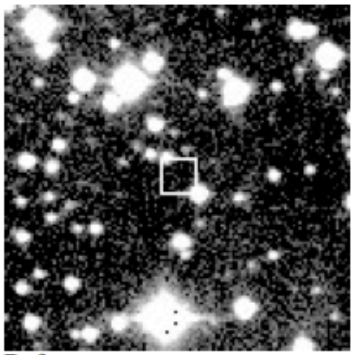
CRTSCircular
911111210394136030-2009-
2009-11-11T16:26:29

SDSS
observation
2009-11-11T16:35:19

CatalogArchives
observation
2009-11-11T16:35:26

CRTS (Catalina) Event identifier is 911111210394136030 or CSS091111:074357+211031




Reference

Finding Chart	Click here
Past CRTS images	Click here
Other images	Click here
Lightcurve	Click here
SDSS cutout	Click here
Position	(115.98635,21.1753)
Time	2009-1
Magnitude	18.559
Magnitude	18.673201

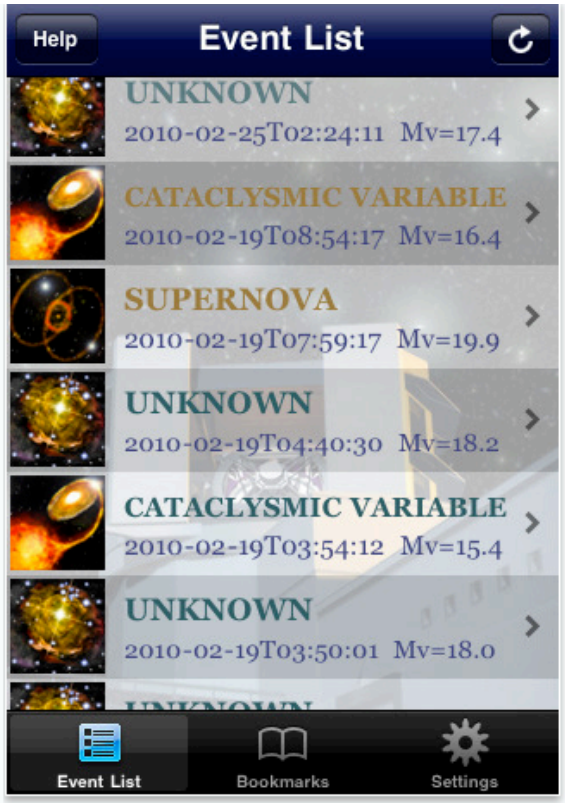
Linked VO/archival data for classif. and follow-up

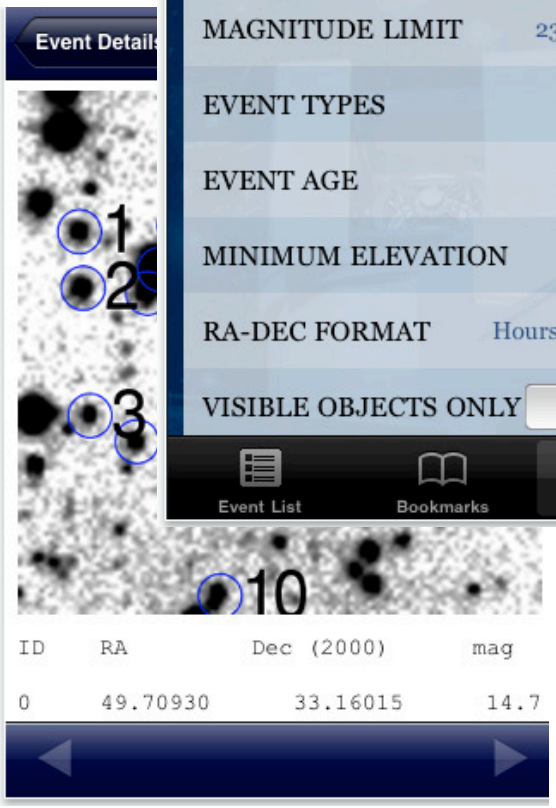
Dynamically growing portfolio

Subscribe to VOEvents via email, RSS, Atom feed, etc.

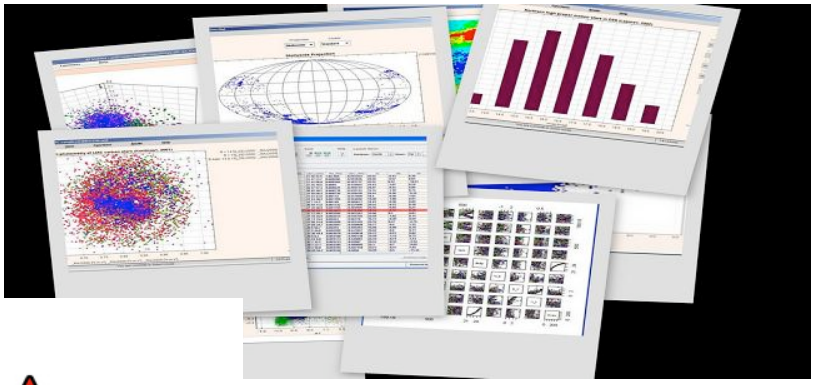
Twitter and iApp Event Distribution

A. Drake, R. Williams (CIT)
B. Truax (DLD, LLC)
J. Myers (LSST)





- Existing tools like NED, CDS/Alladin/Vizier, ADS abstracts server
- VO is larger; IVOA; VO-I etc.



What is DAME

DAME is a joint effort between University Federico II, INAF-OACN, and Caltech aimed at implementing (as web application) a scientific gateway for data analysis, exploration, mining and visualization tools, on top of virtualized distributed computing environment.

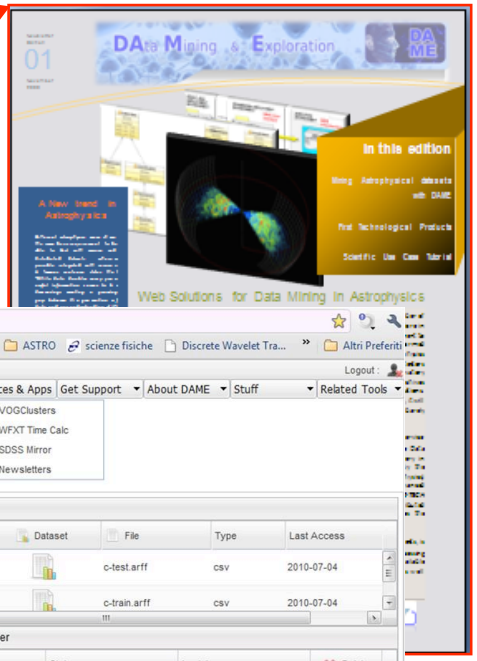
http://voneural.na.infn.it/

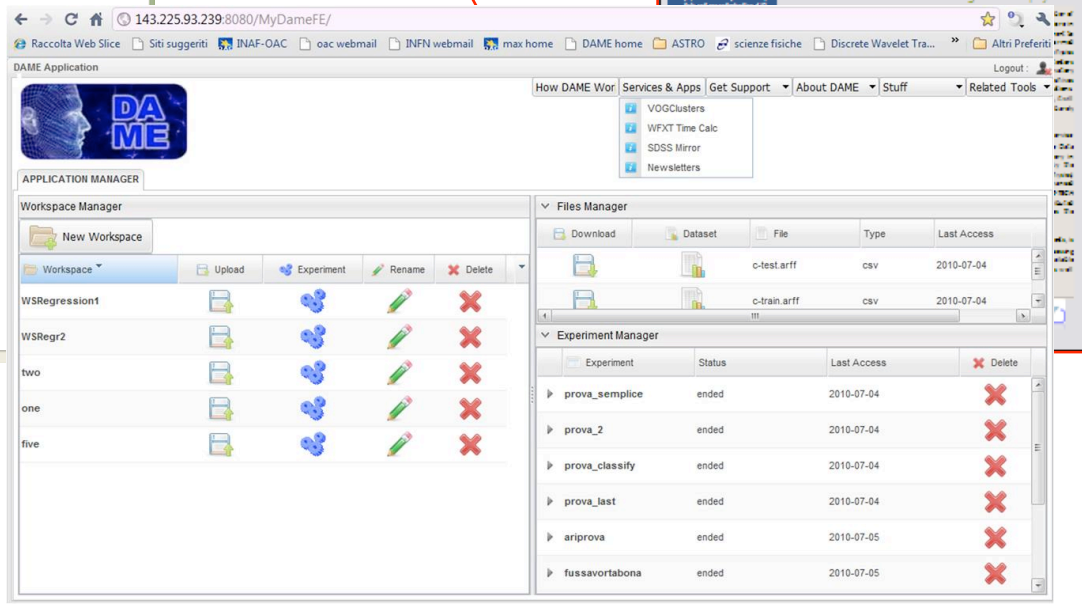
Technical and management info

Documents

Science cases

Newsletter





http://143.225.93.239:8080/MyDameFE/

Web application PROTOTYPE

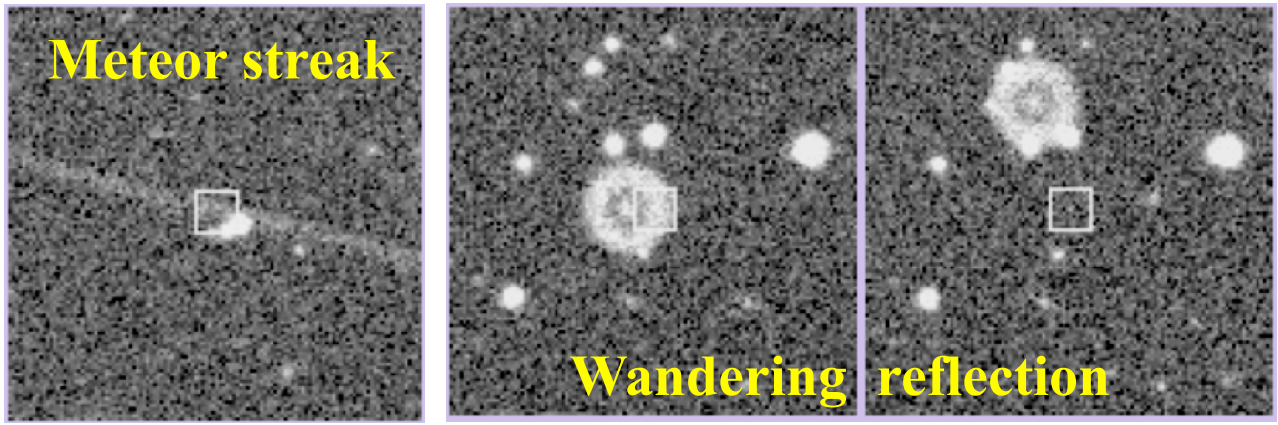
(ALPHA release)

brescia@na.astro.it

Harvesting the Human Pattern Recognition

Recognizing the artifacts (false transients)

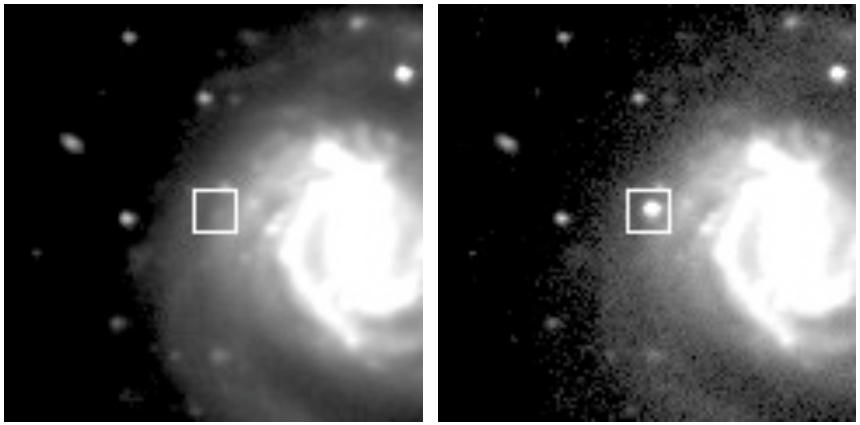
Contextual information is essential



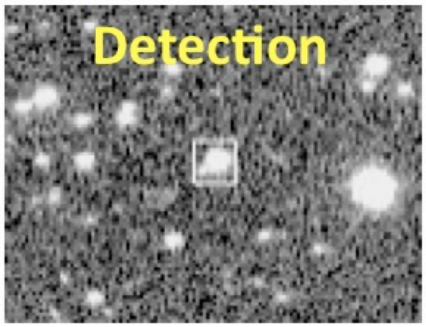
A more sophisticated case uses a **prior (expert) knowledge:**

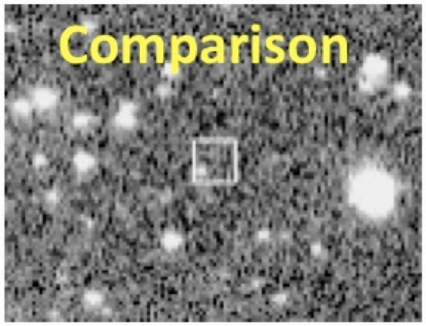
Star-like transient apparently associated with a non-coincident galaxy a likely Supernova

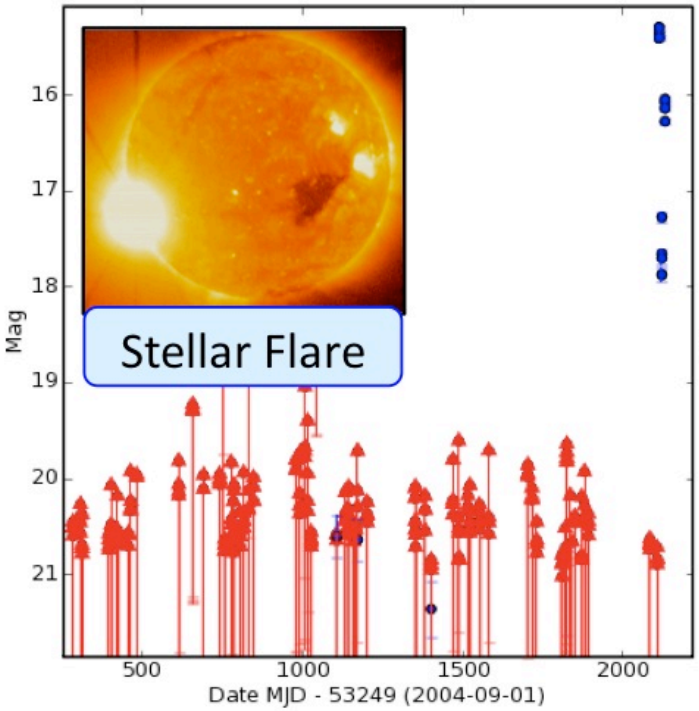
Spiral host galaxy
a possible Type II

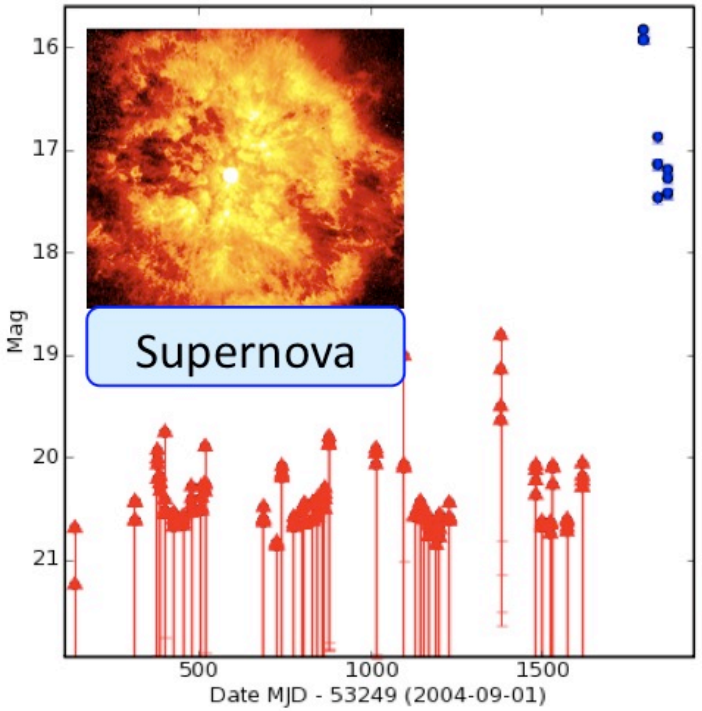


How to capture this and teach a machine to do the same thing?



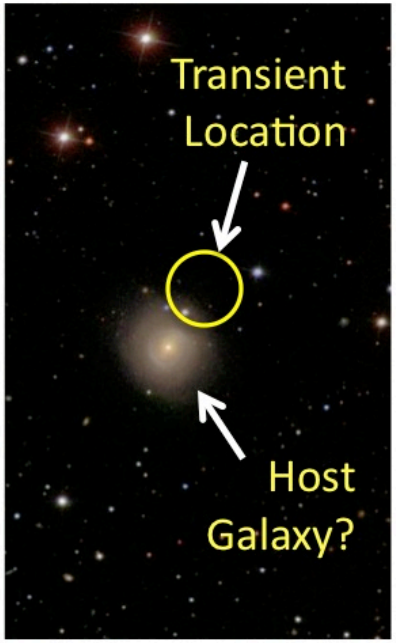






Use Case Scenario:

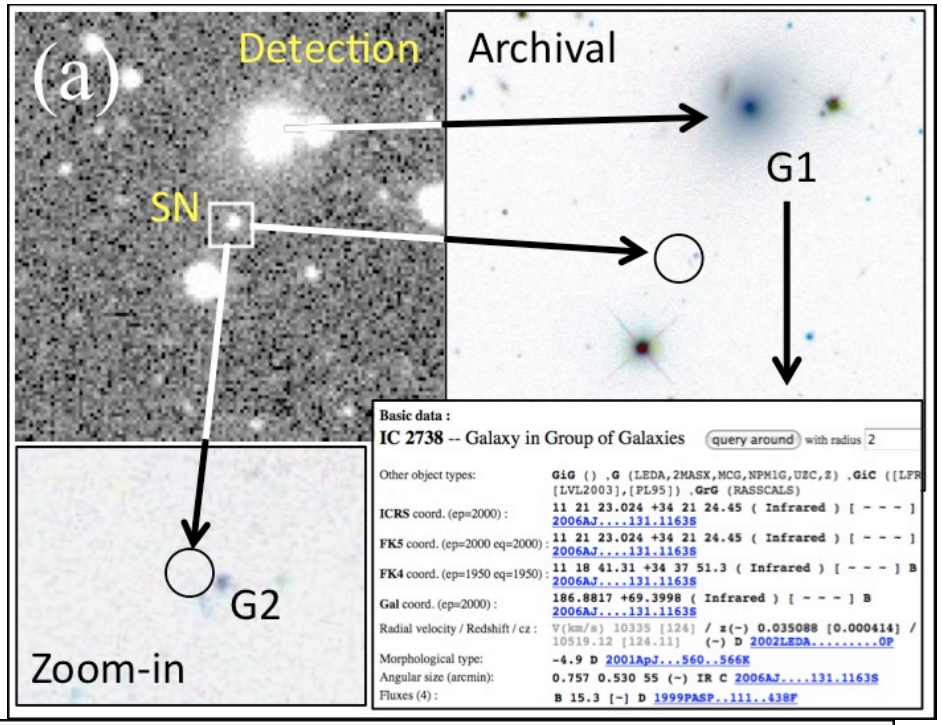
Light curves are ambiguous, but the presence of a possible host galaxy suggest that it is a Supernova

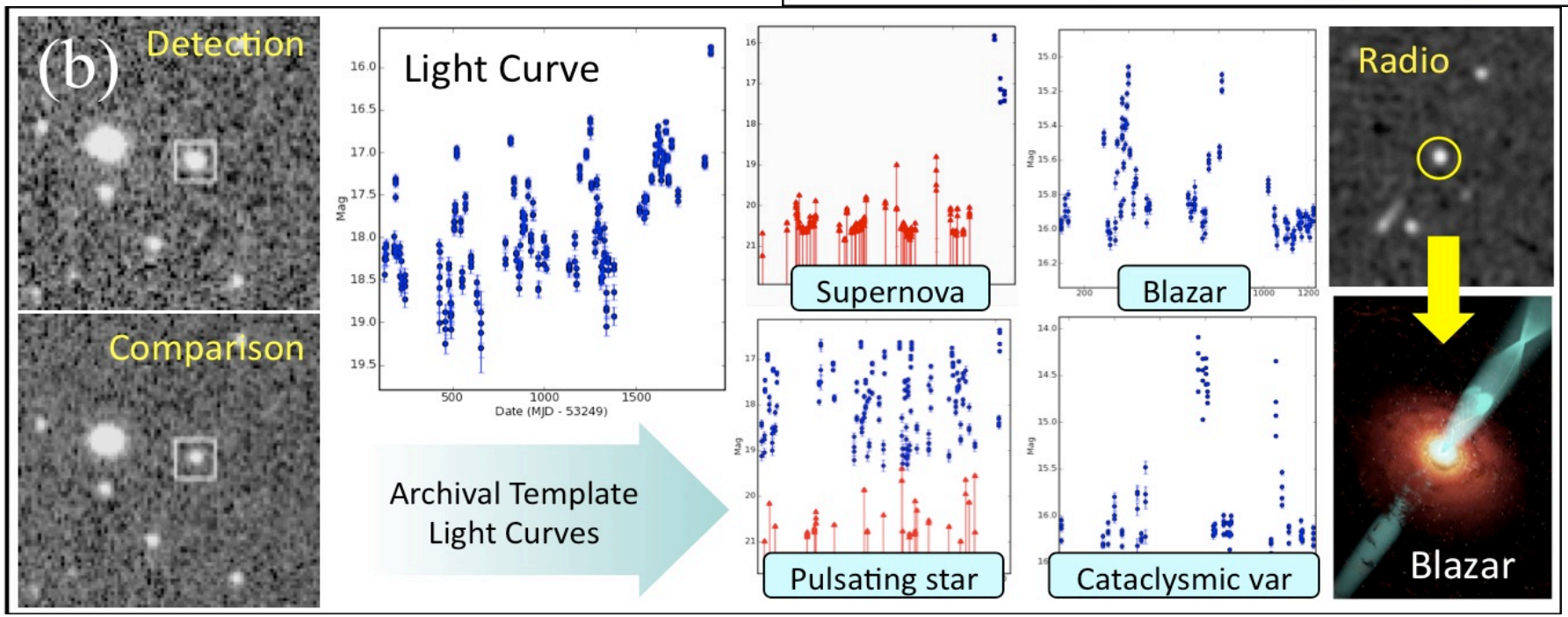


Use Case Scenarios:

(a) Archival data on potential host galaxies provides the more likely choice,

(b) Presence of a radio source discriminates between a CV and a blazar





SkyDiscovery is a Caltech Virtual Astronomy initiative

You are logged in as djorgovski. Log out.

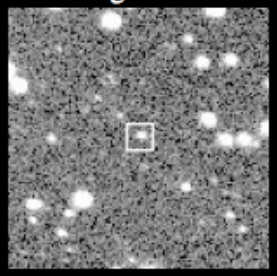
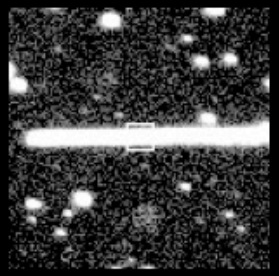
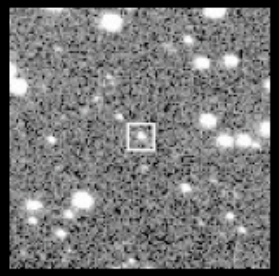
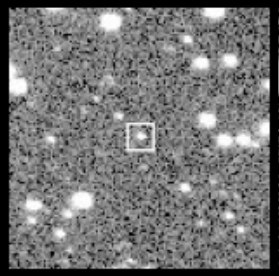
SkyDiscovery.org

Humans and Machines Working Together

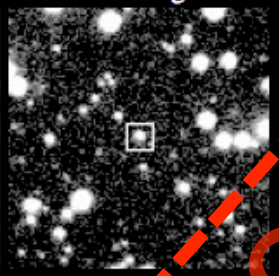
Citizen Scientists Making Discoveries

Event 9387

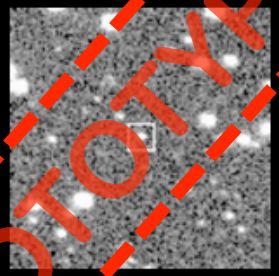
New Images

Reference Image



GIF

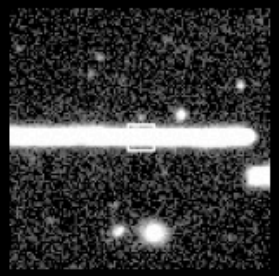
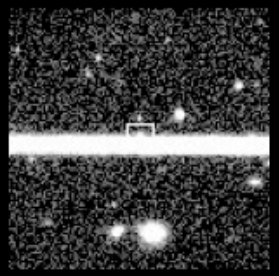
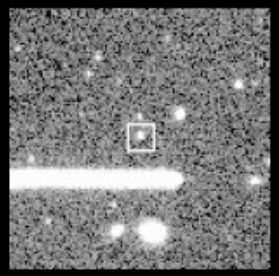
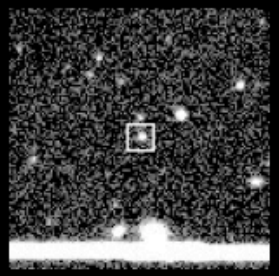
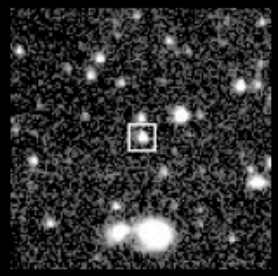
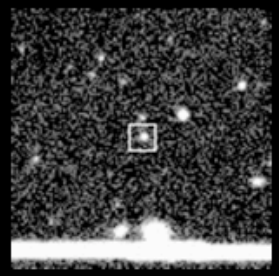


PROTOTYPE

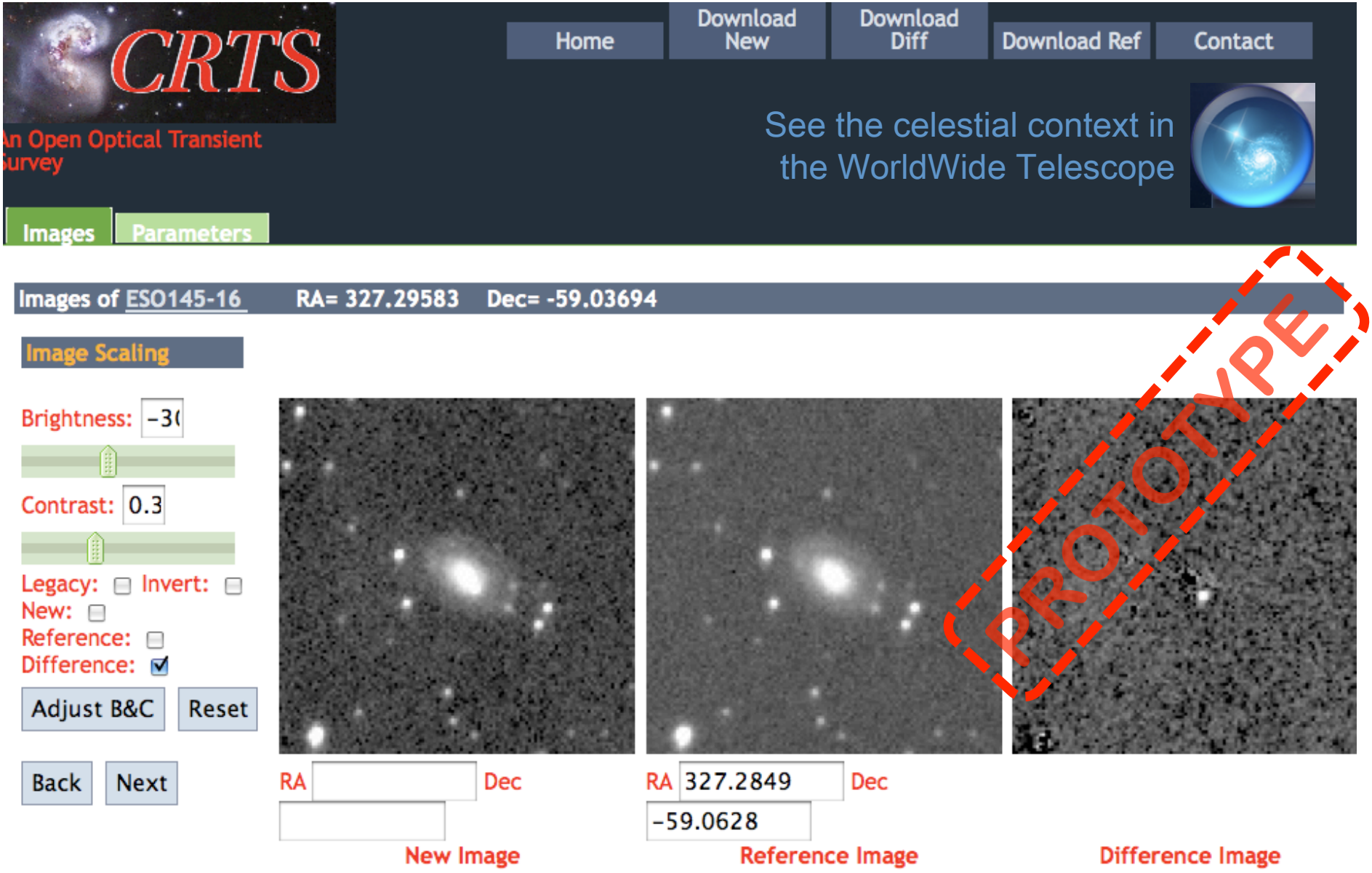
Is there a satellite trail?

Yes | No | Unsure | Help

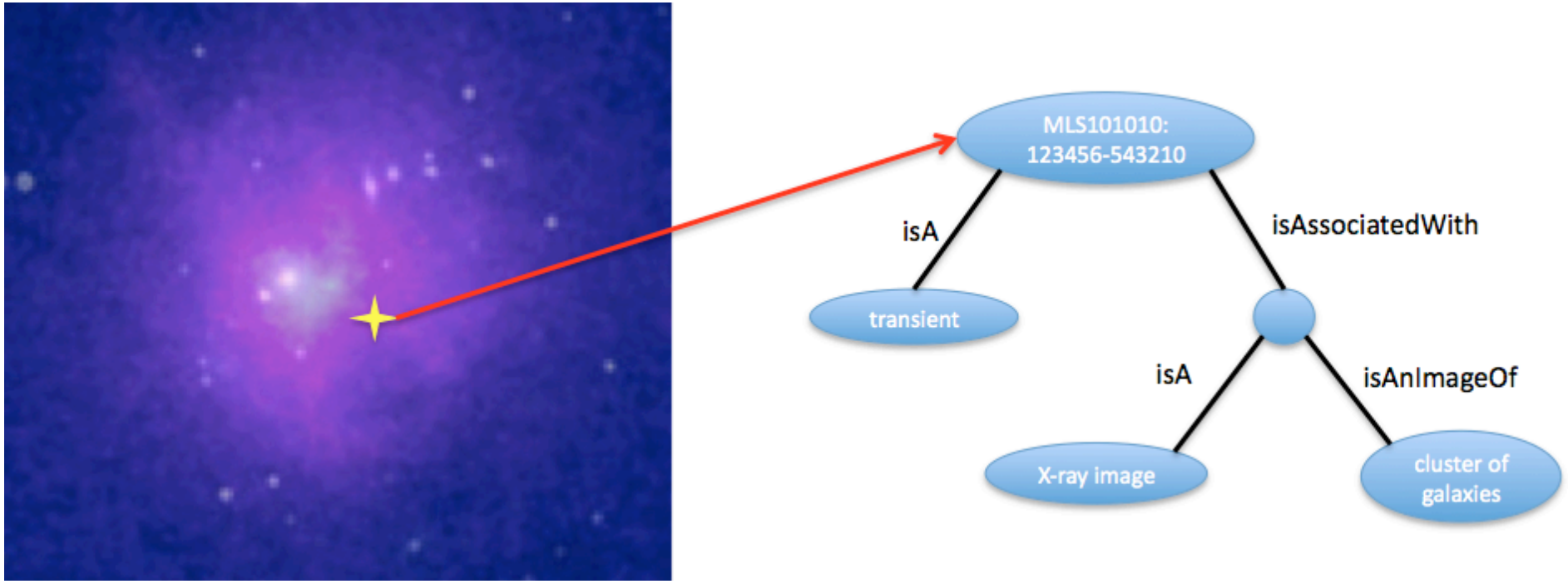
Bold lines, such as those shown below, are caused by satellites in orbit and can confuse the detection software. Is there a satellite trail in any of the images?

Citizen Science Supernova Hunt



Developing the Interface Between Carbon-Based and Silicon-Based Minds



Human-annotated images (via *SkyDiscovery.org*)

⇨ Semantic descriptors

⇨ Machine processing

⇨ Novel algorithms

Video pictures ala sixth sense: Portfolios of transients (or any object for that matter) – automagically updating lightcurves, SEDs etc.

This is the sense in which we are moving

Summary

Minority Report like interfaces in open sims.