# Palomar Transient Factory
# Data Processing & Archive

Frank Masci (IPAC / Caltech)

iPTF Summer School, July 2016

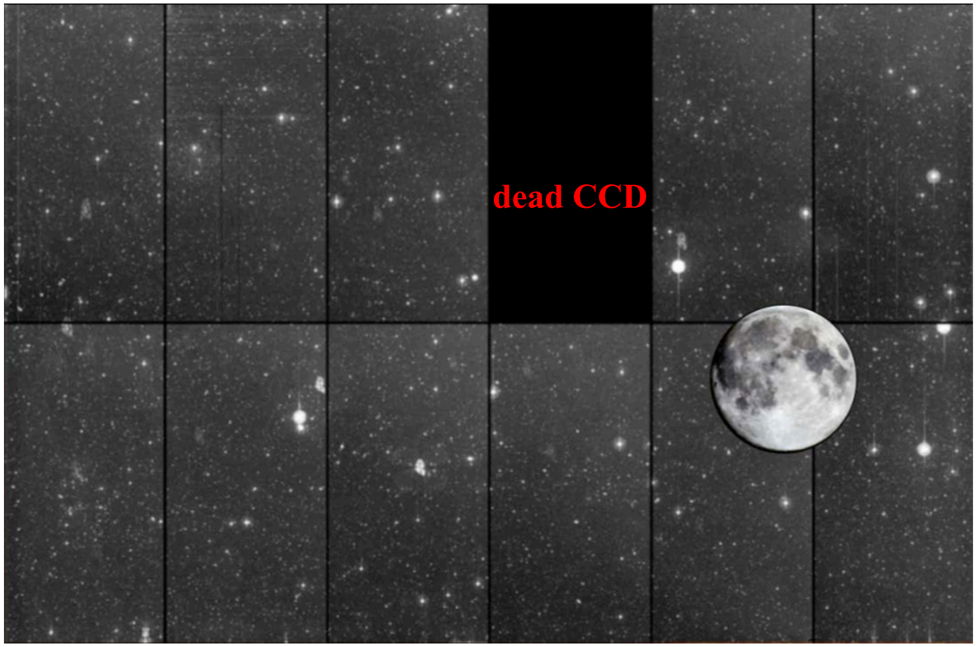
# The Palomar Transient Factory (PTF)

- PTF is a robotic wide-field time-domain survey in the optical
- Carried out on 48″ Schmidt telescope on Mt Palomar
- Can survey up to 3000 deg$^2$ per night to $R \sim 20.5$ ($5\sigma$)
- Median seeing ~ 2 arcsec.
- Cadence (repeatability): 1 min to 5 days: can discover SNe, variable stars, exoplanets, asteroids…
- PTF science operations ran from Mar 2009 – Dec 2012
- Became the *intermediate* PTF (iPTF) after this

Edwin Hubble, 1949

# The PTF Survey Camera

- Based on the original CFHT 12K camera
- 7.26 deg$^2$ FOV, 11 working CCDs (2048 × 4096 pixels each); ~ 92 megapixels in total
- Pixel size on sky at FOV center is ~ 1 arcsec
- Typically 60 second exposures with 30 second readouts
- On average ~ 20 to 30 exposures per survey field per year
- Generates >100 TB of image data products and $10^{12}$ extracted sources **per year**



# Data processing overview

- Data flows through multiple pipelines, creating a variety of science products tailored for different scientific purposes. These pipelines run on different timescales.

- **Photometric (or frame processing) pipeline:** daily (end-of-night) processing to produce high quality instrumentally-calibrated images and source catalogs

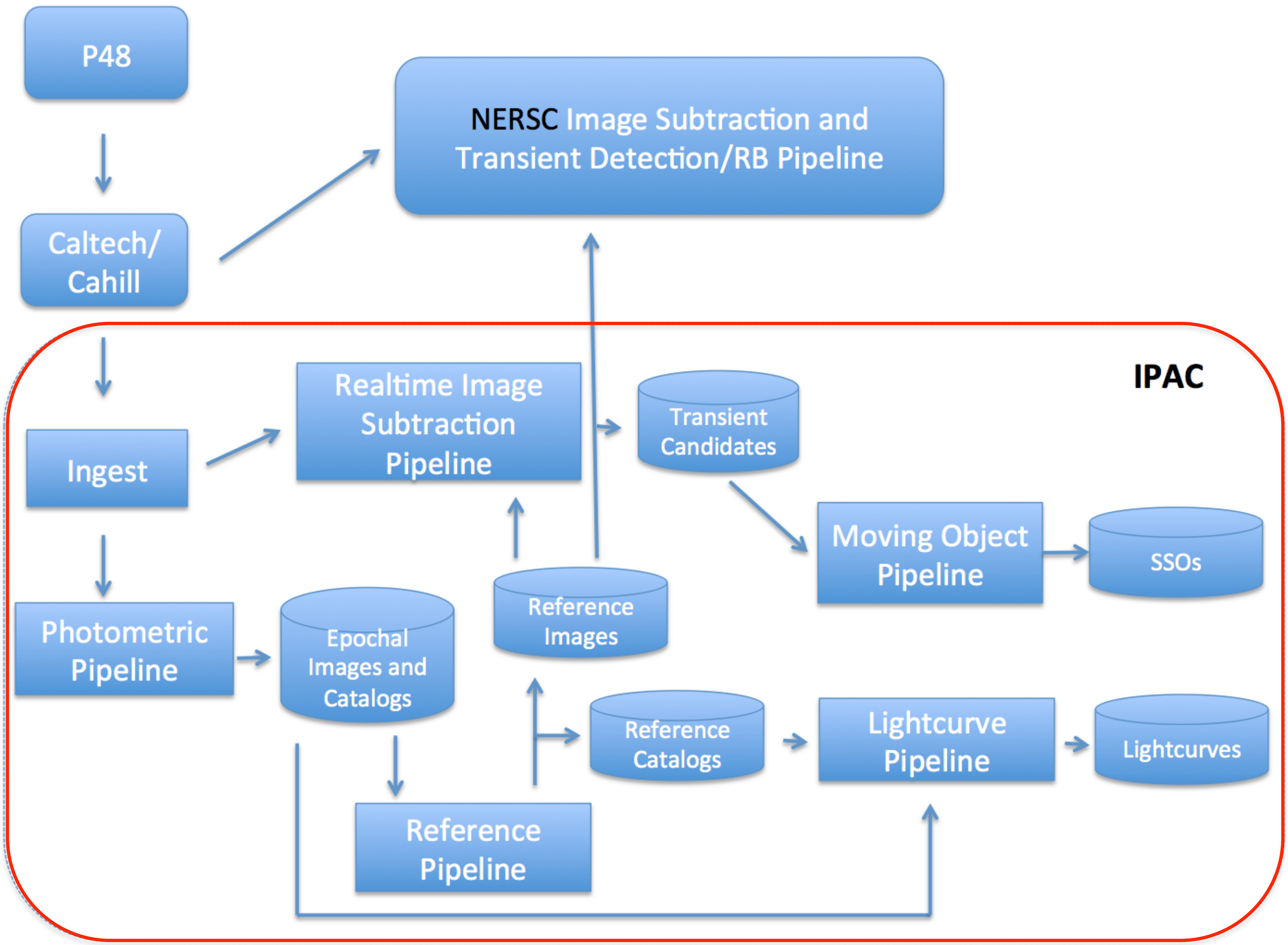
- **Reference image pipeline:** combines high quality frames into deeper images (coadds) – products are used in the real-time and lightcurve pipelines below. Reference images are periodically made, depending on availability of good data for a given field/chip (more later).

- **Lightcurve (or relative-photometry) pipeline:** uses source catalogs from the photometric pipeline to create high precision photometric lightcurves. Also periodically made.

- **Real-time pipeline:** runs throughout a night to support transient-discovery via image-differencing. Outputs feed into various science marshals: extragalactic; galactic; solar system

- Interfacing with the above: an advanced data archive with exploratory tools to support long-term data curation and public distribution – storage of raw data, processed images, and source catalogs

# Data Flow



P48
Caltech/ Cahill
NERSC Image Subtraction and Transient Detection/RB Pipeline
IPAC
Ingest
Realtime Image Subtraction Pipeline
Transient Candidates
Moving Object Pipeline
SSOs
Photometric Pipeline
Epochal Images and Catalogs
Reference Images
Reference Catalogs
Lightcurve Pipeline
Lightcurves
Reference Pipeline

# Data Transfer

- Raw CCD camera-image data from individual exposures are packaged into Multi-Extension FITS formatted files at the telescope.
- First transferred via a microwave link to the San Diego Supercomputing Center; then to Cahill at Caltech.
- At Cahill the data forks to two places: **NERSC** at LBNL, and **IPAC** at Caltech:
- **NERSC:** developed first version of the realtime pipeline: image-differencing and transient-discovery with machine learned vetting; designed around SN discovery; still operating today.
- **IPAC:** all flavors of processing described on slide 4, including a reimplementation of the realtime/transient-discovery pipeline for future use (became live ~ mid 2014);
  - also maintains the central archive for all PTF/iPTF products; served through IRSA

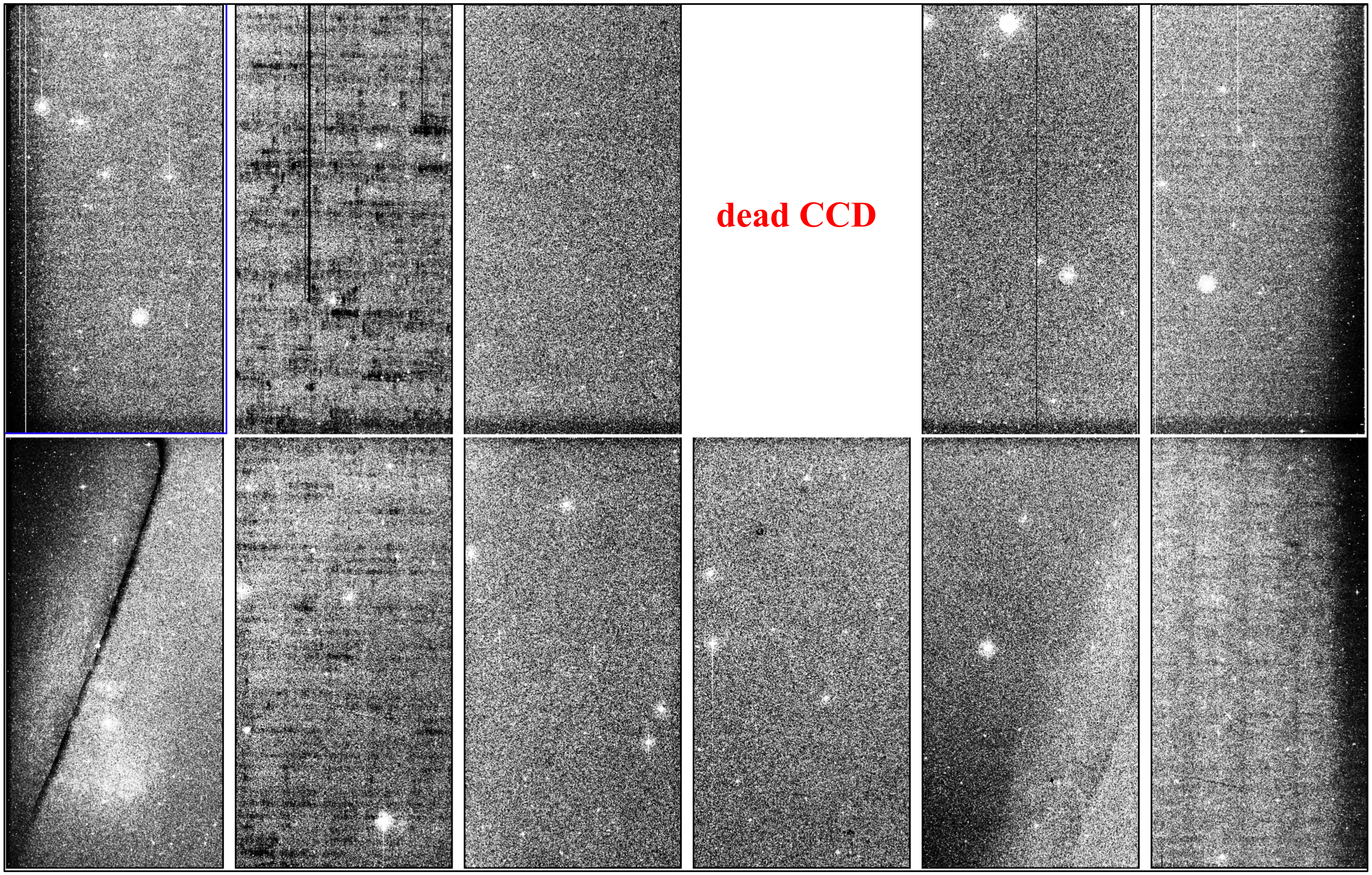
# Infrared Processing and Analysis Center

- IPAC is a Multi-mission Science Center (IRAS, ISO, Spitzer WISE, Herschel, Planck, 2MASS…)
- iPTF generates ~1TB of data every 4-5 days.
- iPTF cluster has 24 machines with 240 cores total
- Roughly 0.7 PB of spinning disk
- Databases, archive and file servers; tape backup
- Maintains archive data retrieval interfaces/services

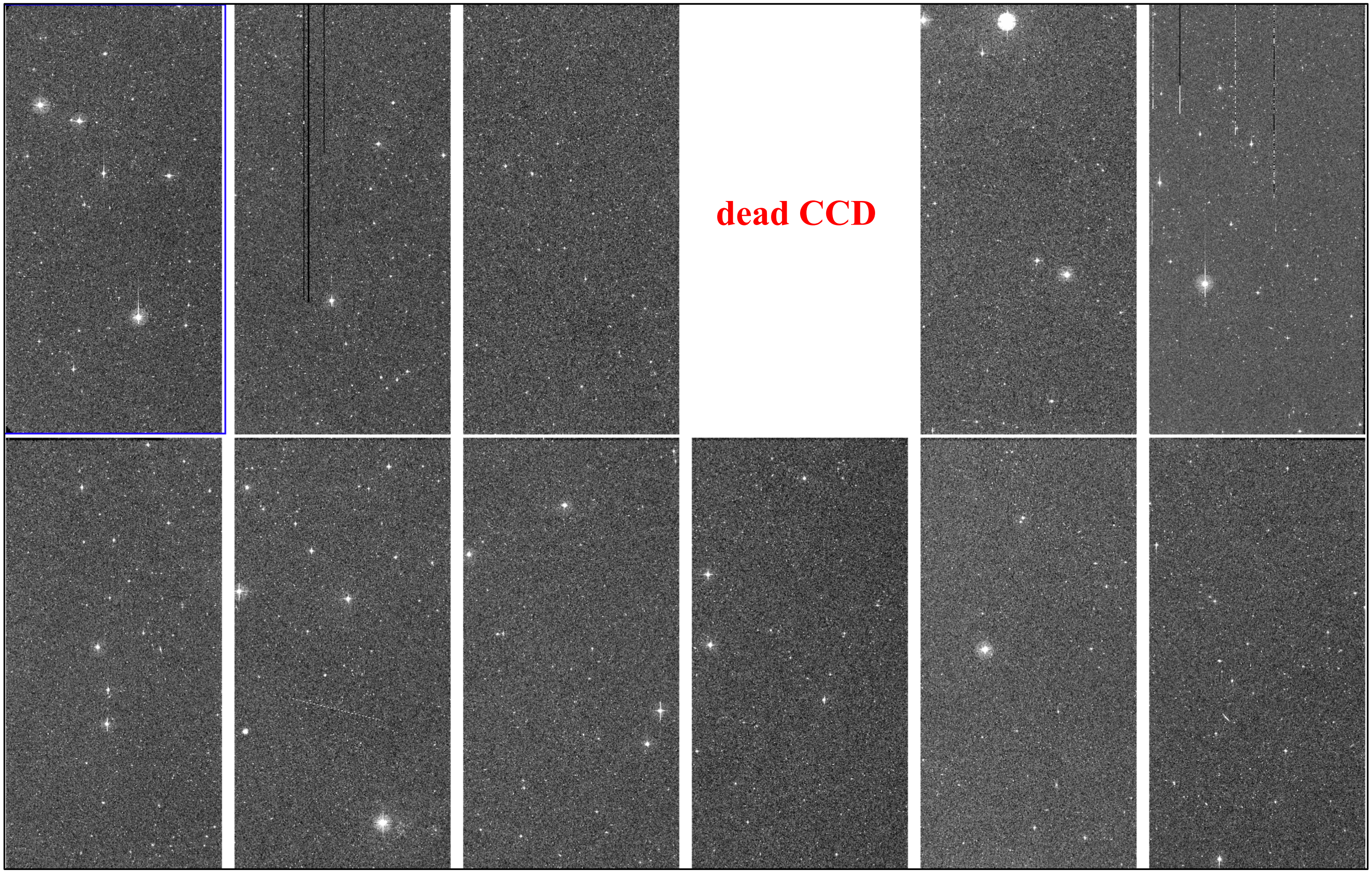
You are here


PTF data lives here (IPAC)

# Raw CCD images from one exposure



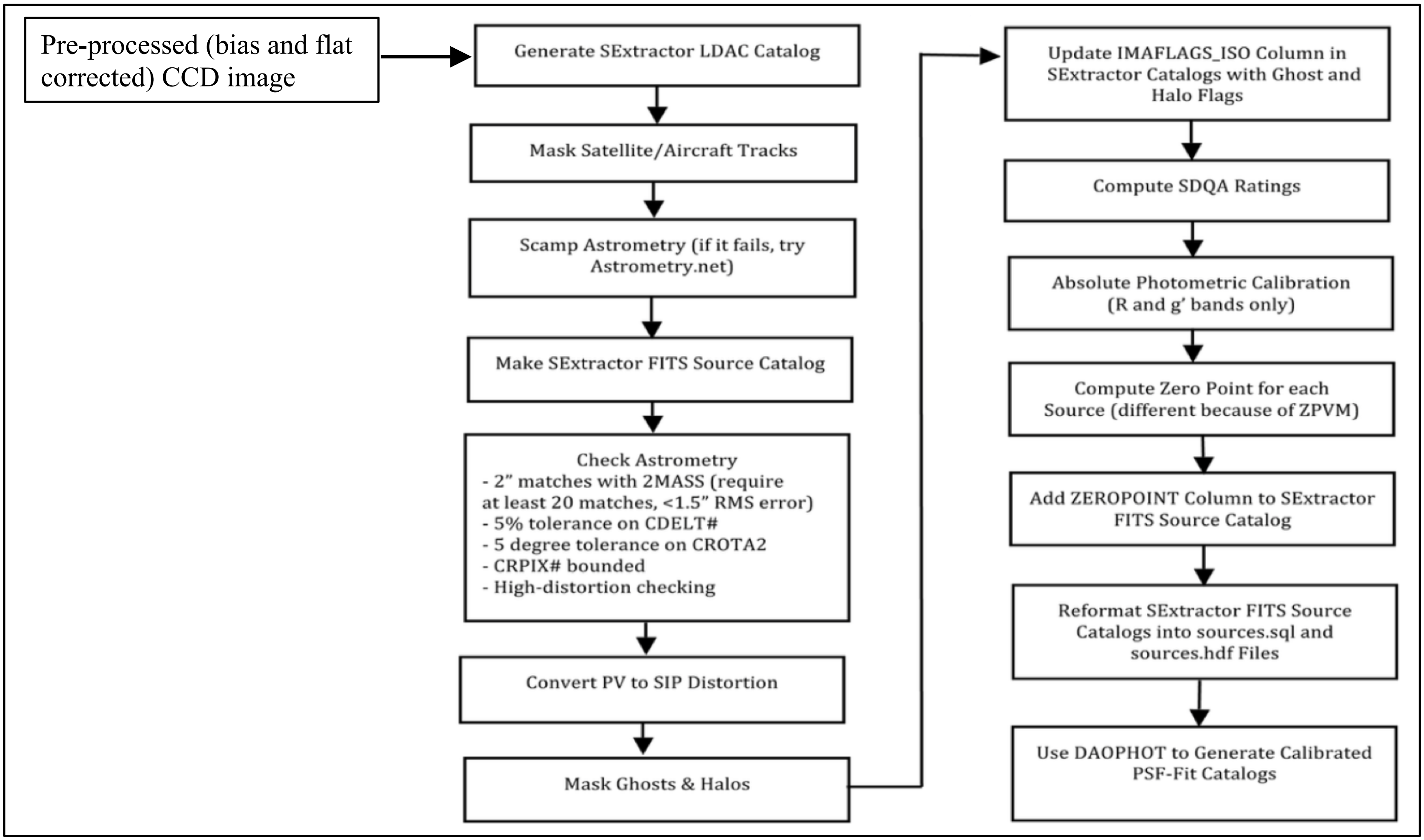
# Processed CCD images from same exposure



# Photometric (frame-processing) pipeline

- Triggered at the end of the night, after all the data has been received.
- Instrumental calibrations are derived from an entire night's worth of data. More specifically, flat-field maps are derived from the on-sky data.
- Photometric calibration is from a nightly model-fit using the SDSS overlap region: fits for color and airmass terms, spatial and time-dependent throughput variations. Accuracy is ~ few percent.
- Astrometric (and distortion) calibration is done at the individual CCD-image level against a combined SDSS and UCAC4 catalog. Typically good to 0.15 arcsec in unconfused regions.
- Outputs are calibrated single-CCD FITS images with bit-masks and accompanying source catalogs in FITS binary table format – both aperture and psf-fit photometry is provided.
    - These products are archived at IPAC and available 1 – 3 days after observation.
    - Subsets are made publically available periodically (more later).

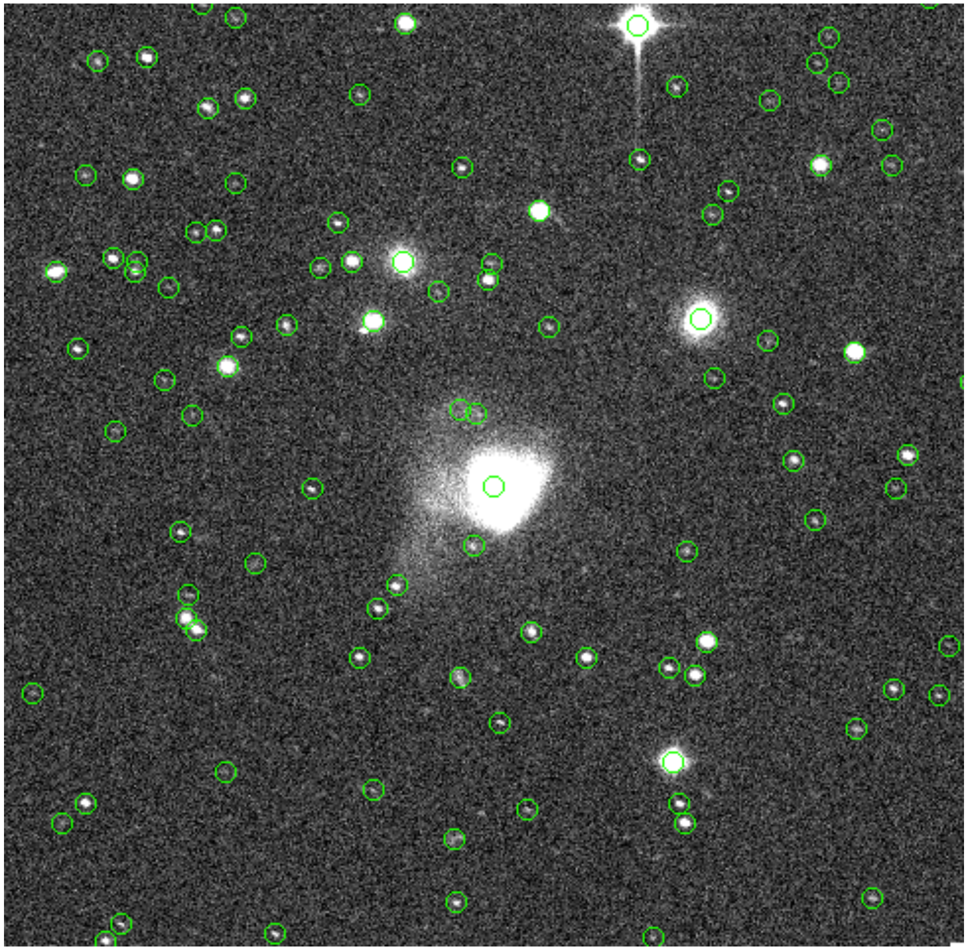
# Photometric (frame-processing) pipeline steps




*from Laher et al., 2014, PASP, 126, 674*

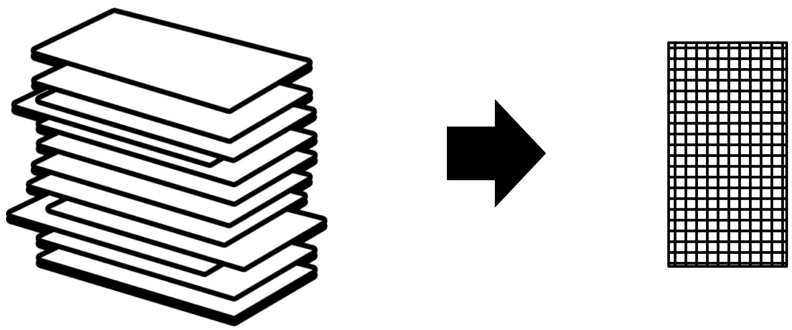
# Photometric (frame-processing) pipeline output

- Zoom-in on *R*-band image of Arp 220 (galaxy merger); 8 arcminutes across.
- Extractions from aperture photometry catalog are overlaid.
- Single-epoch (CCD-) based products archived:
  -- processed CCD image with metadata
  -- bit-mask image identifying bad pixels
  -- calibrated aperture photometry catalog
  -- PSF-fit photometry catalog
  -- raw image data also available
  -- all products above are in FITS format
  -- nightly processing log



# Reference image (co-addition) pipeline

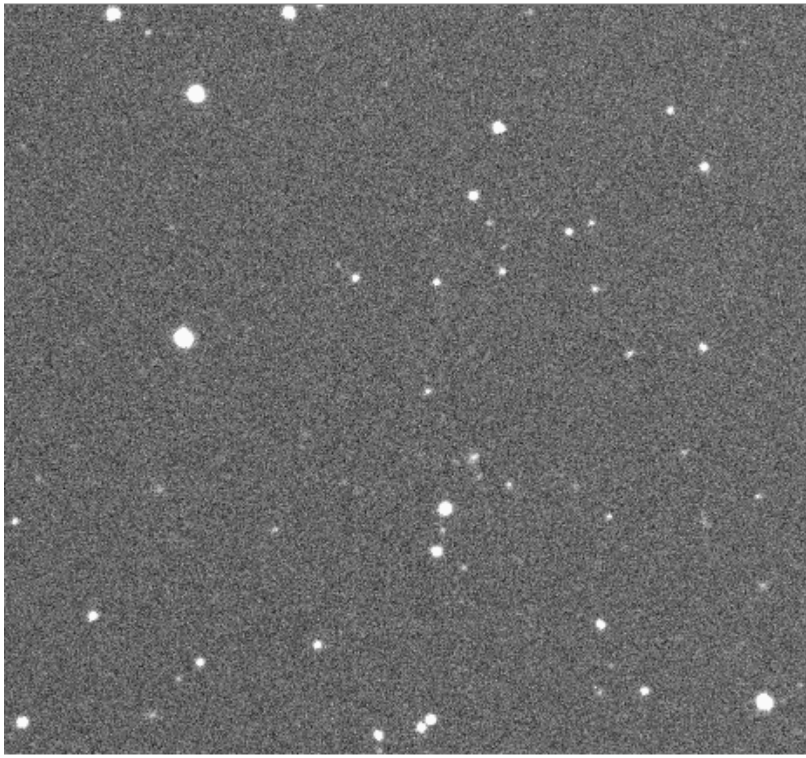
- When enough individual CCD exposures accumulate, the "reference image" pipeline is triggered
- This pipeline coadds (combines) the "best" image data for a given CCD, field, and filter: i.e., with best seeing, photometric conditions, and astrometry
- Images are first reprojected, then combined using pixel-stack averaging with outlier trimming.



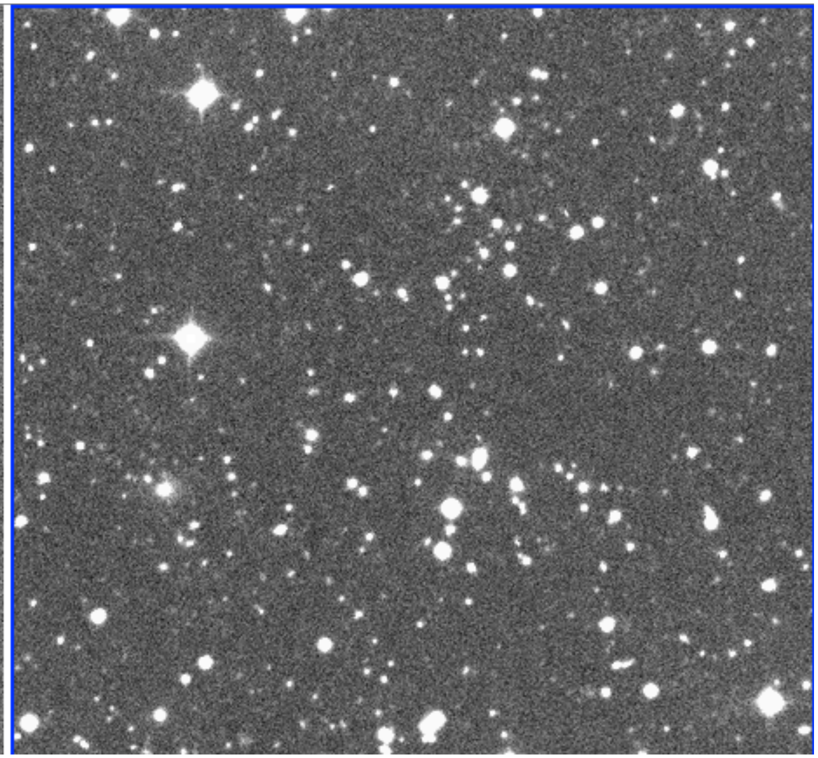
- The coadds are images of the "static" sky as represented by the state of the input CCDs used
  - ➢ deeper than the individual exposures: currently, stacks are ~ 5 to 50 images deep.
- Source catalogs are also generated from these images: both PSF-fitting and aperture (SExtractor)
- Reference image products: images, coverage maps, and catalogs are publically available.
- Products support the real-time (image subtraction) and light-curve pipelines

# Reference image example

Single image, 60 sec exposure in R filter

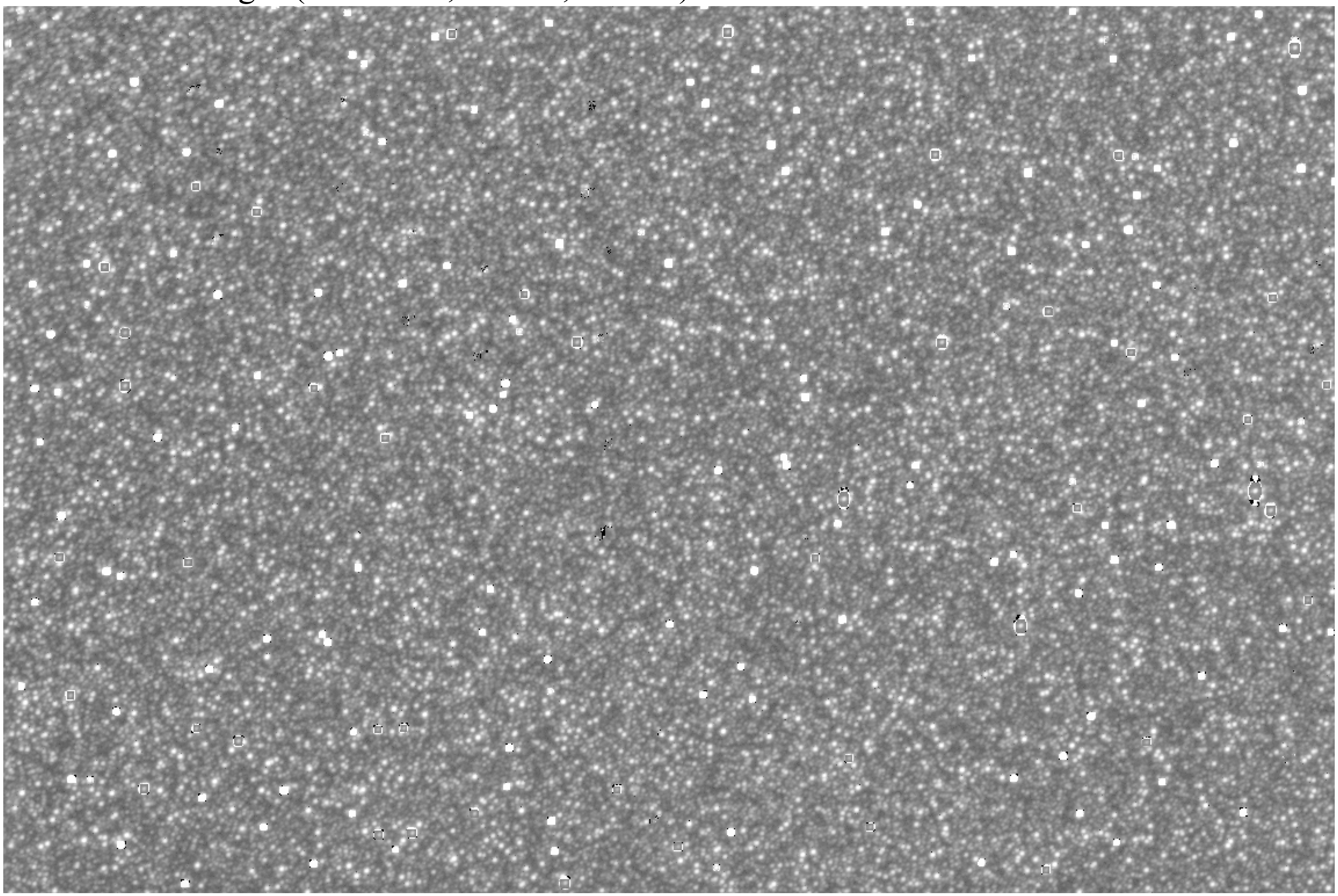


Stack of 34 images (field 5257, CCD 7)



# Reference image example (near galactic center)

Stack of 10 images (field 1549, CCD 5, R filter)



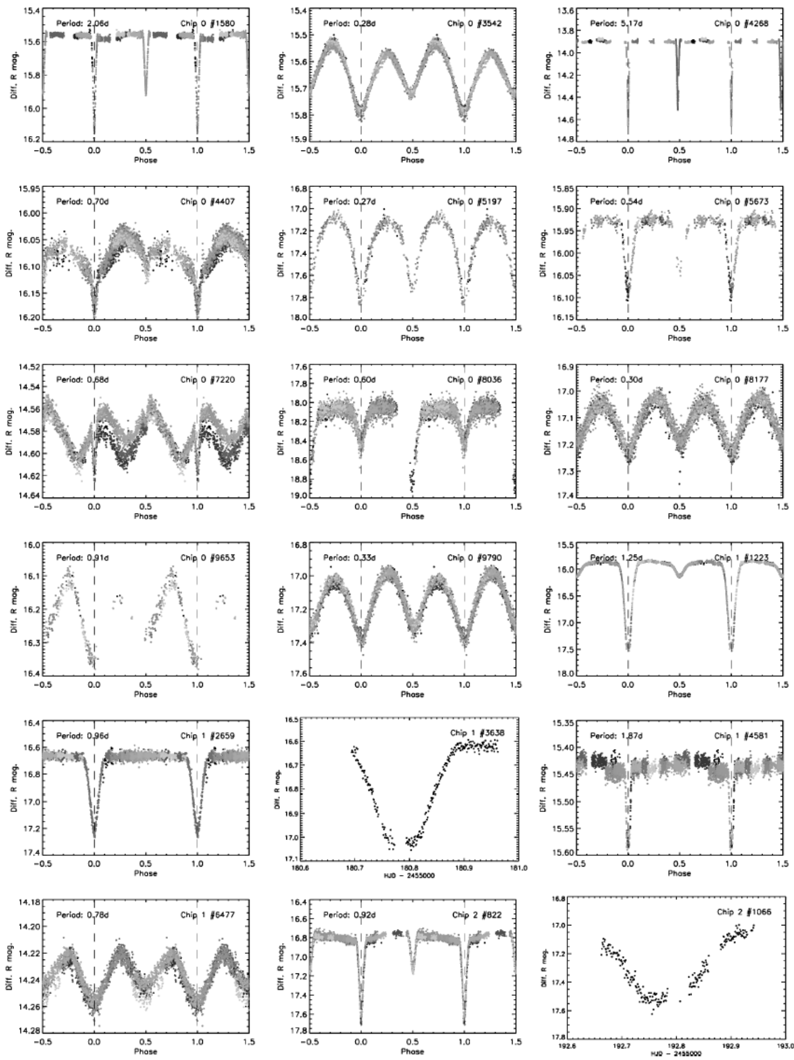
# Deeper image co-adds
# (courtesy: Jason Surace)

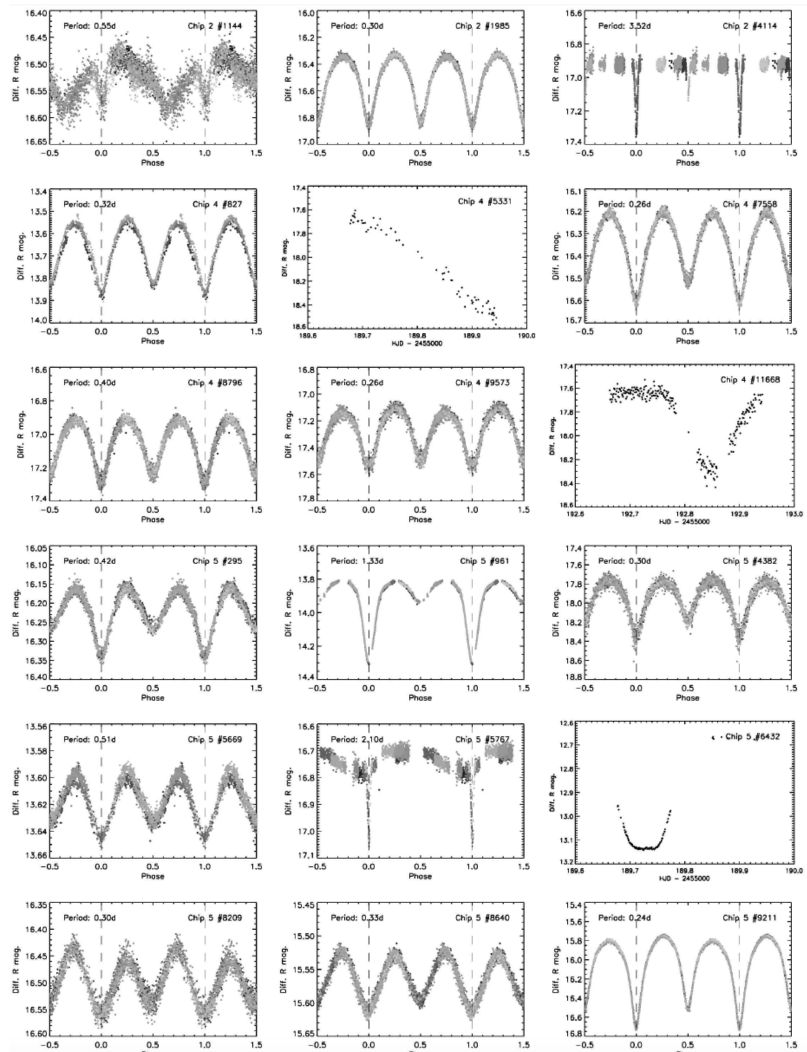
# Lightcurve (relative photometry) pipeline

- At the end of each night, all detected sources from the photometric pipeline are matched against the reference-image source catalog for a given field, CCD, and filter
- The “cleanest” least variable sources are used as anchors for the relative photometric calibration
- Individual image gain-correction factors are computed using a least-squares fitting method
- Application of these refined gain-correction factors improves the overall relative calibration to a few millimag for bright sources
- This pipeline is triggered on timescales of typically 1 to 2 weeks
- A lightcurve database for a subset of fields and epochs will be publically available in Dec 2016

# Example PTF lightcurves from the Orion project

Binary star lightcurves; from Van Eyken et al. (2011)



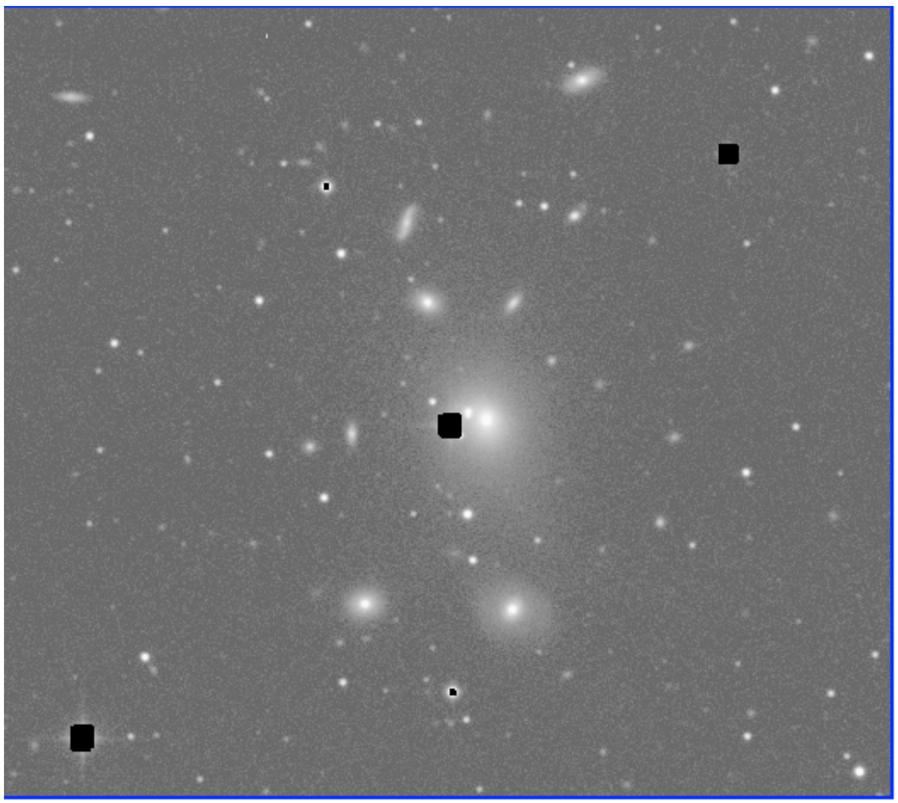


# Real-time pipeline overview

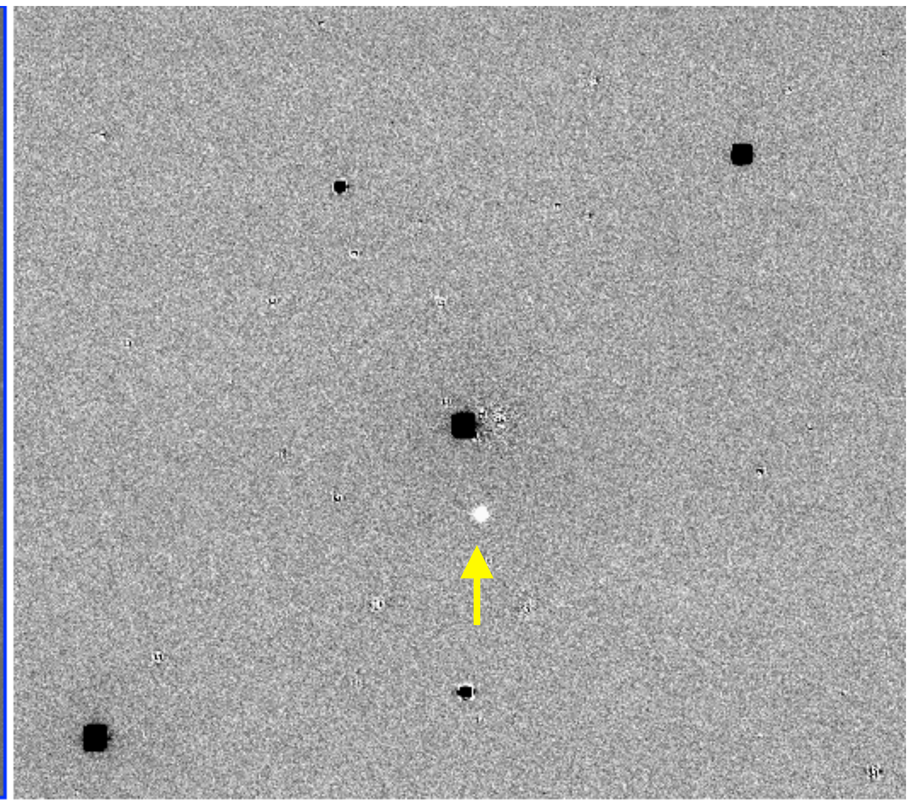
- Uses image-differencing against the reference-image library to extract transient candidates.
- Candidates are then automatically "scored" using machine learning.
- Data is processed in near real-time as it's received; turnaround is 10-25 minutes from telescope to vetted transient candidates
- Outputs are used for same-night follow-up of "interesting" candidates
  - ➢ pushed to an external gateway for pickup by the science marshals: galactic, extragalactic, solar-system, and generic ToO alerts
- Difference images and transient-source catalogs are astrometrically and photometrically calibrated
  - ➢ both aperture and PSF-fit photometry is performed
- Transient sources from this pipeline also feed a "streak detection" module to find fast-moving objects and a moving-object pipeline to construct moving-object tracklets
- Products from this pipeline are not publically available at this time.

# Example difference image

*science image exposure (zoom)*

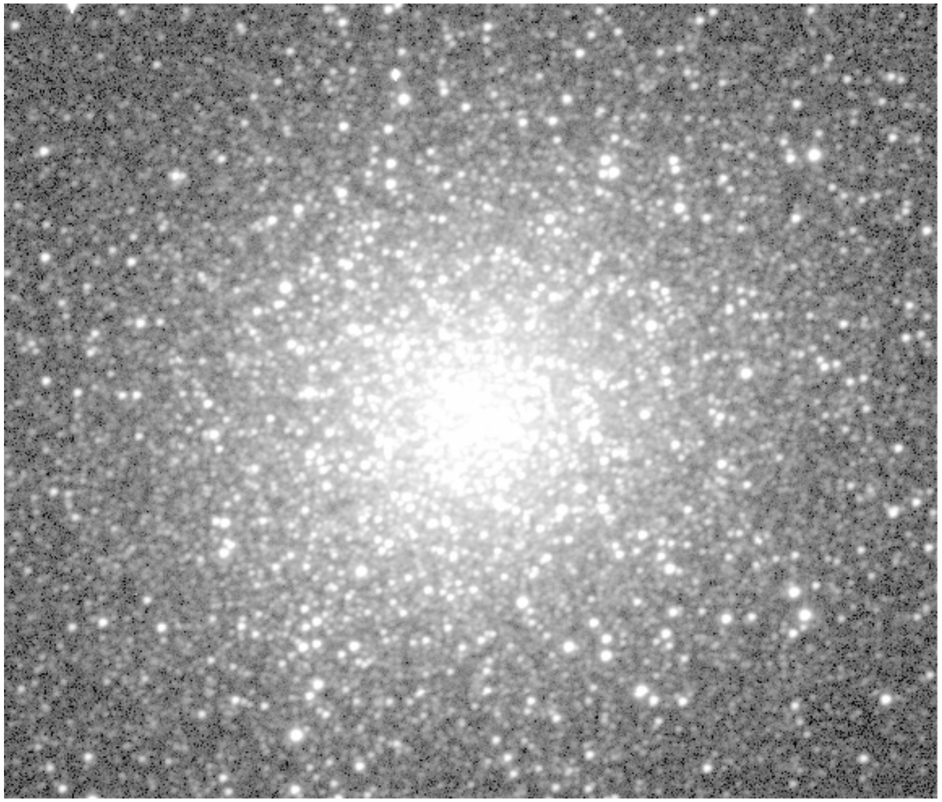


*"exposure minus reference"*

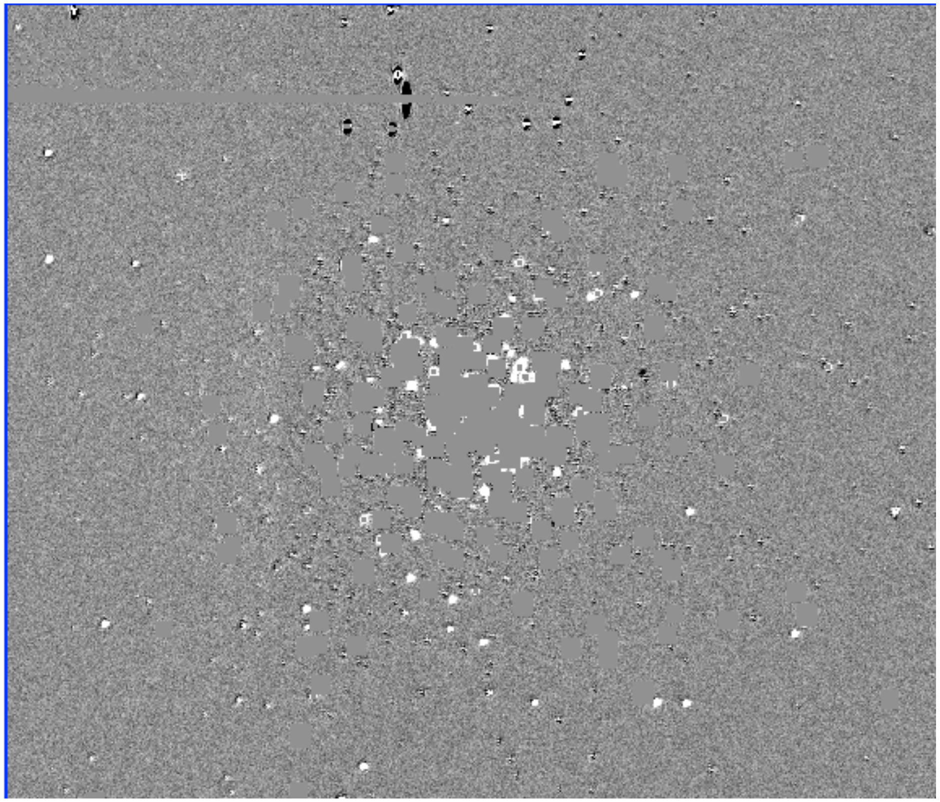


# Example difference image:
# zoom on M13 globular cluster

*science image exposure* (~ 9' x 9' *zoom*)



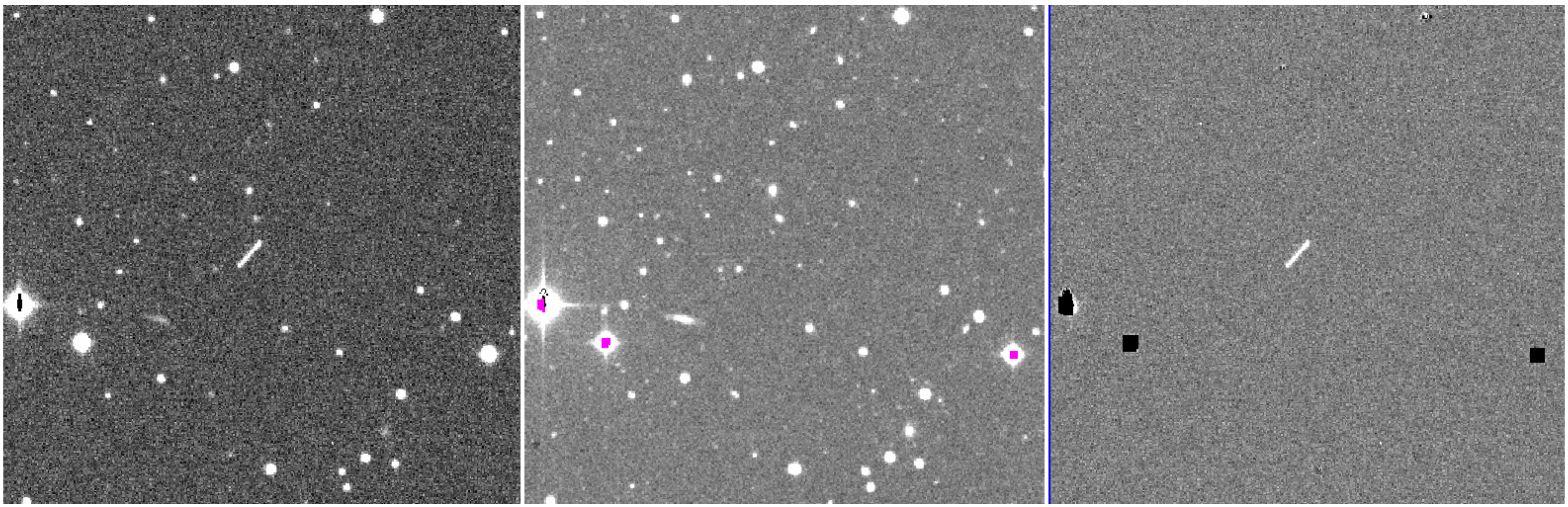
"*exposure minus reference*"



Lots of RR-Lyrae variables!

# Example difference image with a streaking asteroid

*science image exposure* *reference image* *exposure minus reference*



asteroid 2009 HK73

# What products are (will be) available

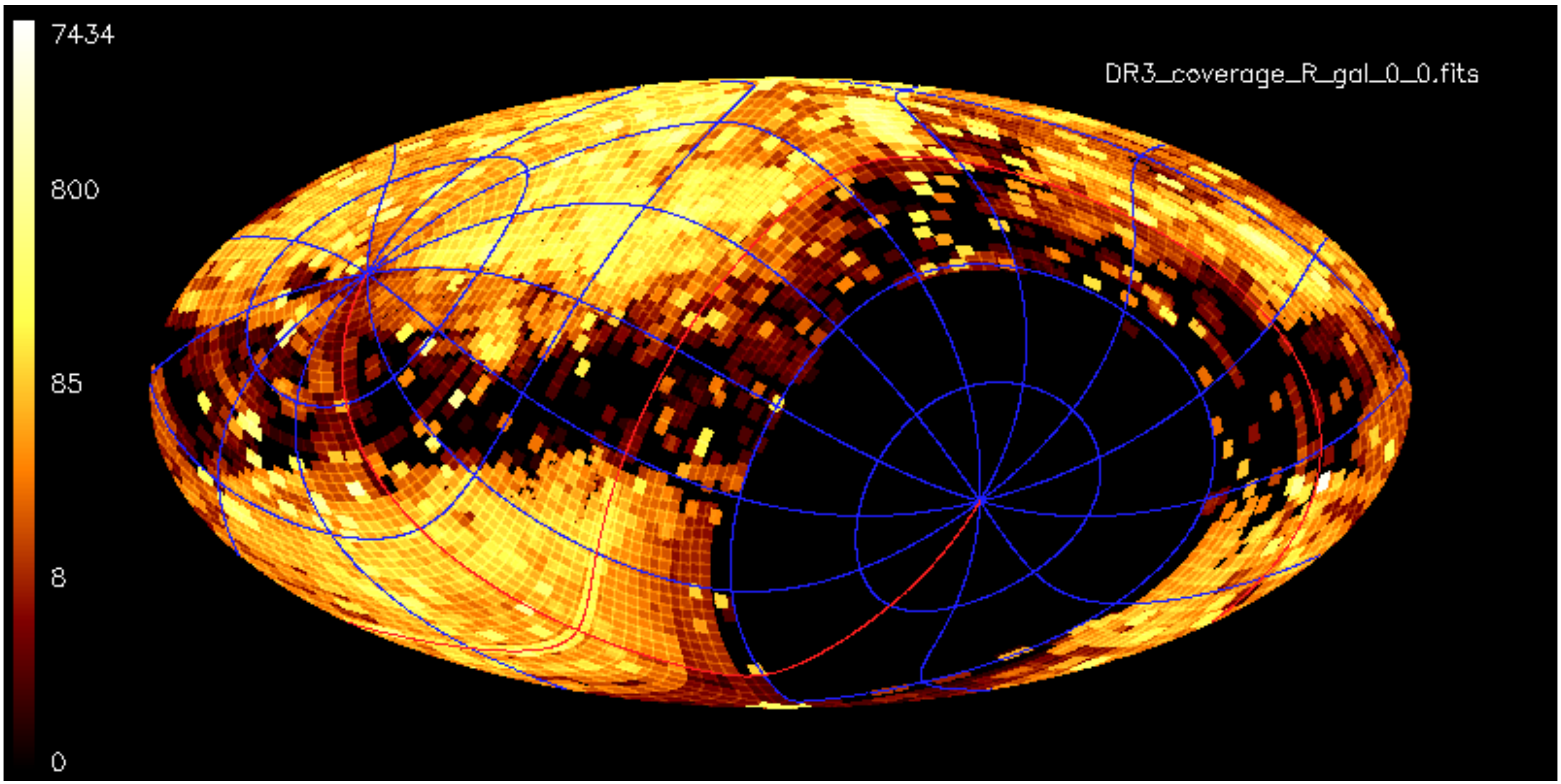
- Science-quality products available now to the public (**Data Release 2**):
  - Epochal images in *R* and *g* bands with QA metadata acquired **Mar 1, 2009 – Dec 31, 2012**
  - Accompanying source-catalog table files for these epochal images
  - Reference images (co-adds) for fields with enough epochs observed during the same period
  - Accompanying source-catalog table files for these reference images

- In early September, 2016, **Data Release 3** will add to the above:
  - ~ 650,000 more epochal images and catalogs from data acquired **Jan 1, 2013 – Jan 28, 2015**
  - ~ 9,600 more reference images (co-adds) with accompanying source catalogs.

- In early December 2016:
  - A lightcurve database, accessible via a user-interface; ~ 600 million LCs (demo later)
  - A source database based on epochal-image extractions, also accessible via a user-interface
  - A source database based on reference-image (co-add) extractions
  - These databases will contain a subset of the epochal data from **Mar 1, 2009 – Jan 28, 2015**

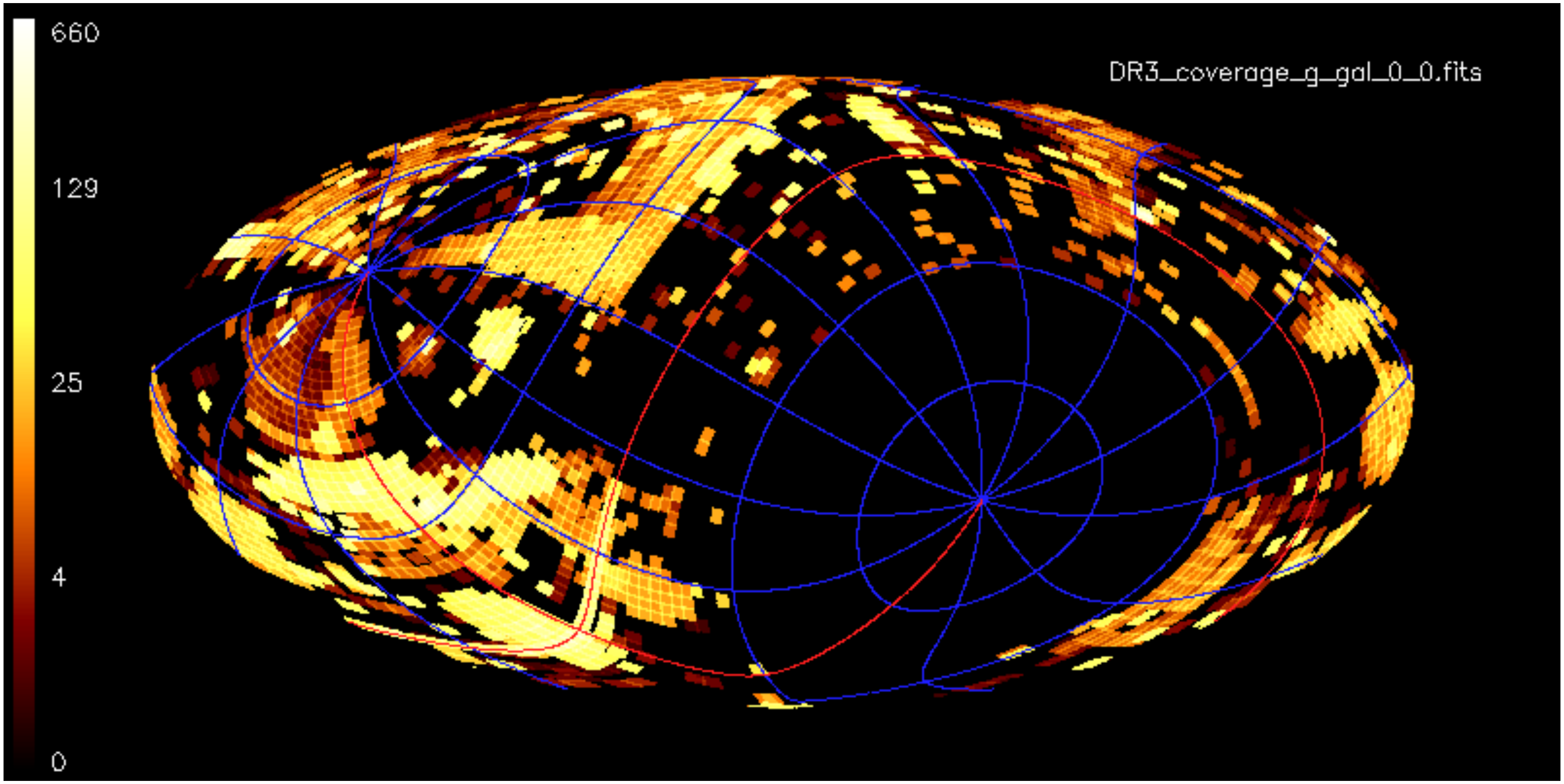
# Sky depth-of-coverage for epochal image products (R-band)

All epochs from the archive that will be publically available following the Fall 2016 release:
**~ 3.47 million R-band images.** Projection is galactic, centered at *l, b* = 0, 0.



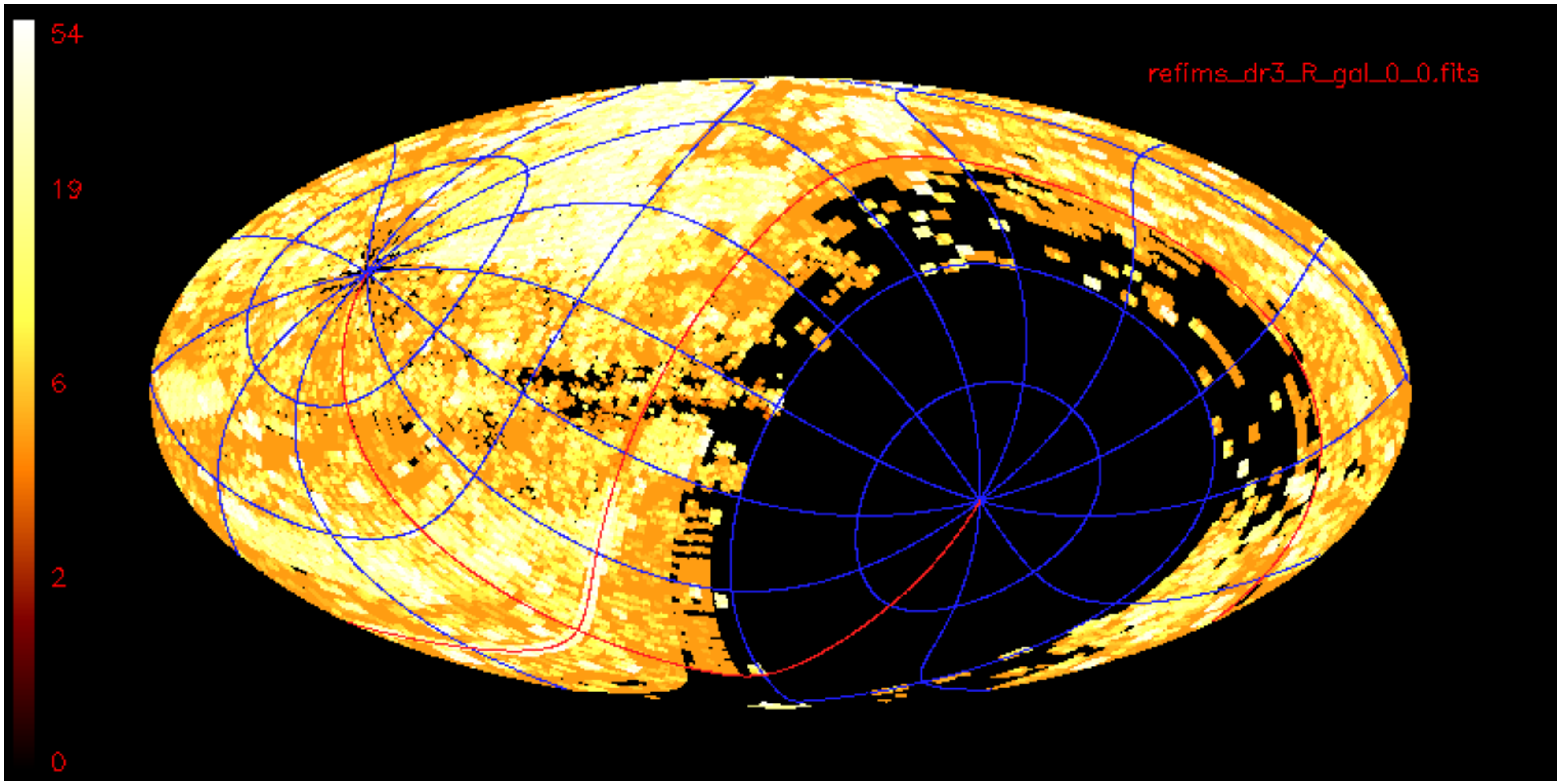
# Sky depth-of-coverage for epochal image products (g-band)

All epochs from the archive that will be publically available following the Fall 2016 release:
**~ 0.72 million g-band images.** Projection is galactic, centered at *l, b* = 0, 0.



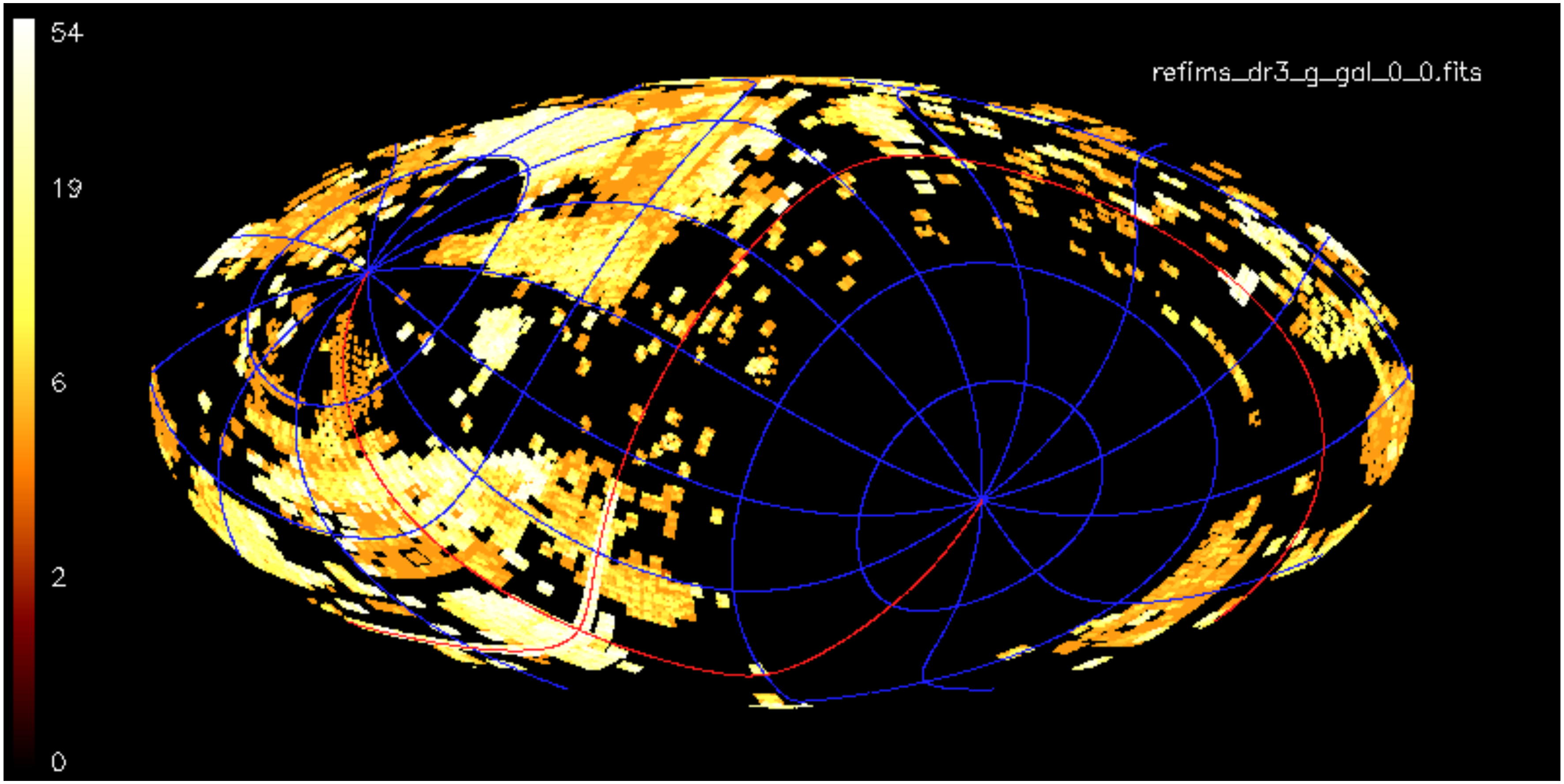
# Sky depth-of-coverage for co-add products (R-band)

All co-add images from the archive that will be publically available following the Fall 2016 release: **41,200 R-band co-add images.** Projection is galactic, centered at *l, b* = 0, 0.



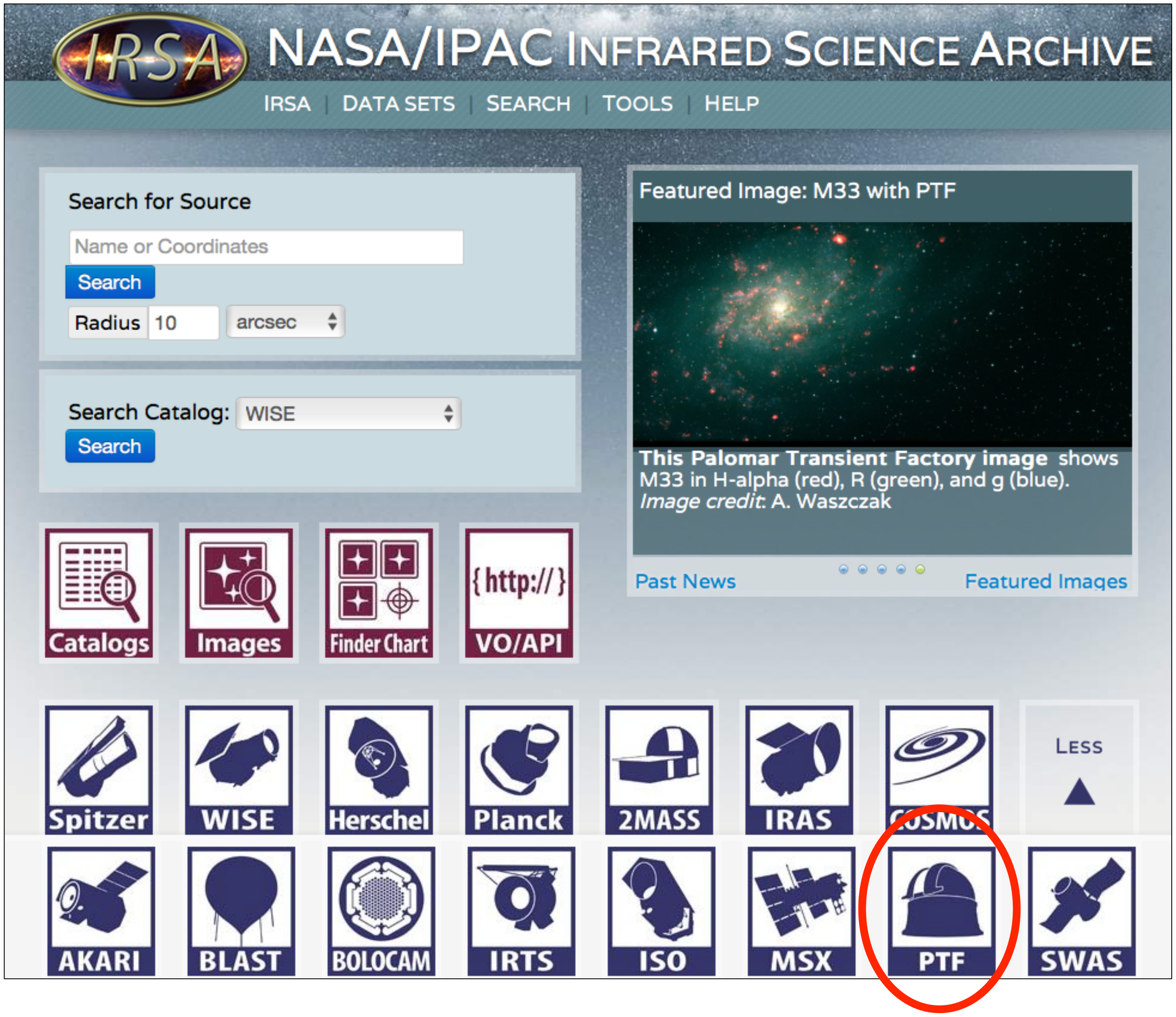
# Sky depth-of-coverage for co-add products (g-band)

All co-add images from the archive that will be publically available following the Fall 2016 release: **15,240 g-band co-add images.** Projection is galactic, centered at *l, b* = 0, 0.



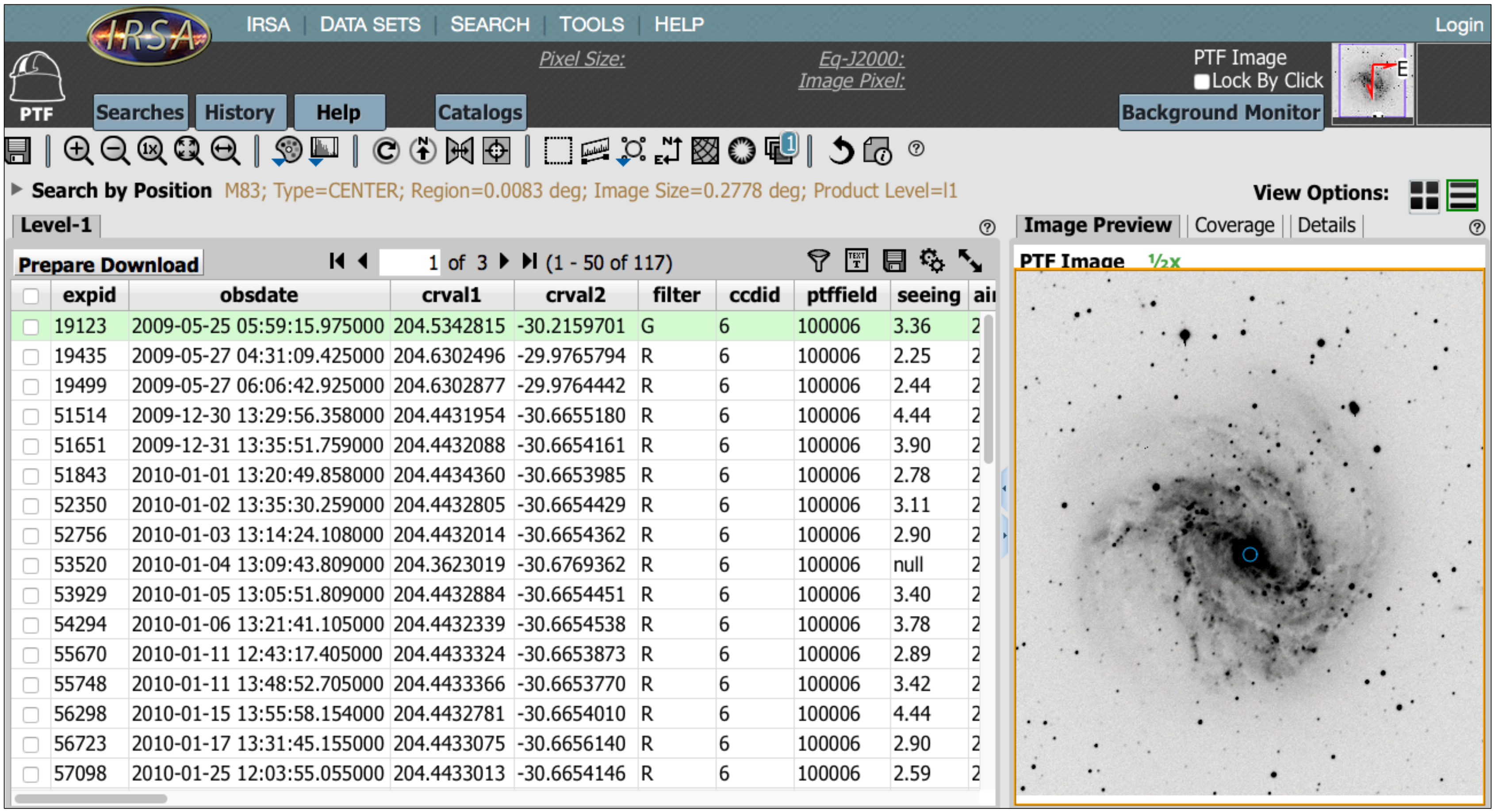
# Where to access PTF data products

*http://irsa.ipac.caltech.edu/frontpage/*



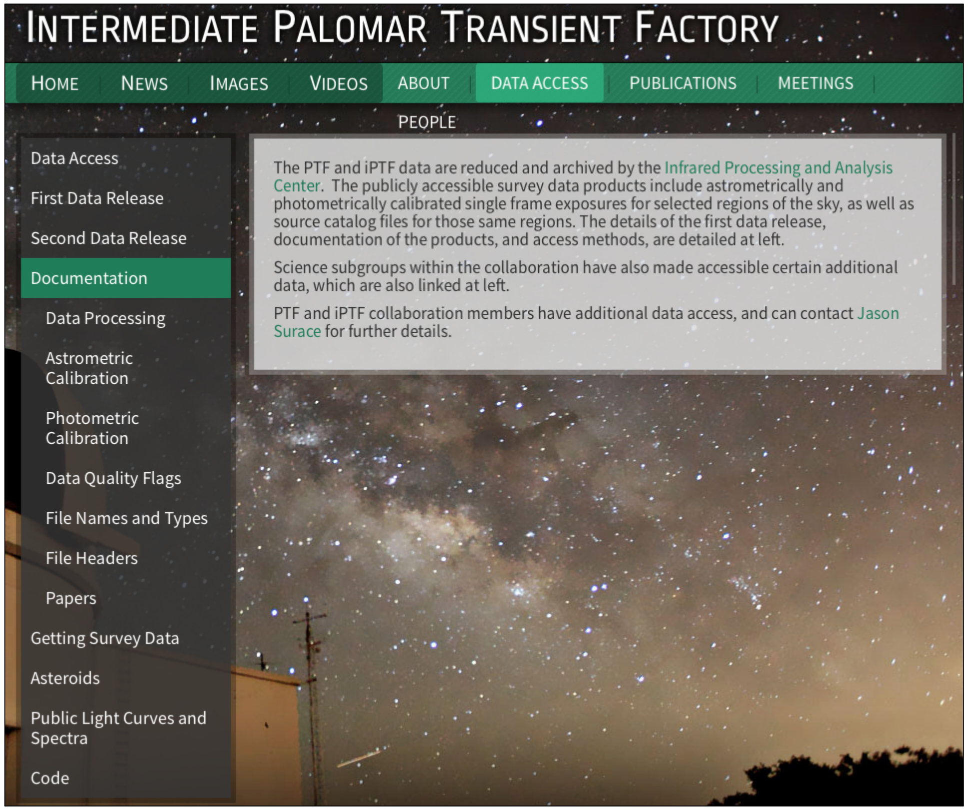
# PTF Image Service at IRSA

- CCD-based image and catalog files can be searched by position or object name falling therein via a GUI
- Known Solar-System objects also recognized
- Can also search and retrieve data via a command-line application program interface



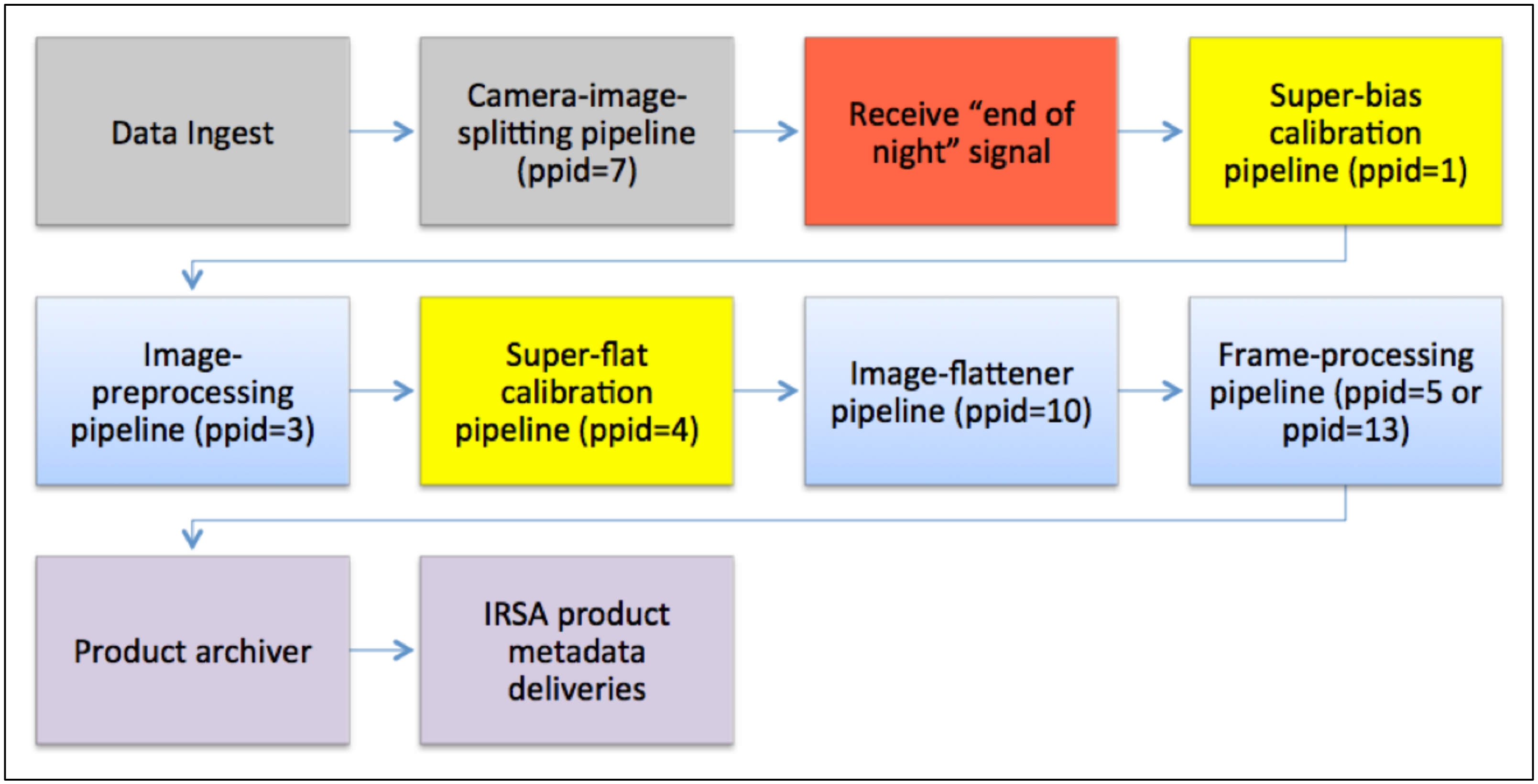
# More information…

- More details on PTF image processing and data archiving at IPAC:
  Laher et al., 2014, PASP, 126, 674
  *http://iopscience.iop.org/article/10.1086/677351/pdf*
- Documentation on how to access and use the PTF public data products:
  *http://www.ptf.caltech.edu/page/data_access*



# Back up slides

# PTF daily processing steps



# Photometric calibration of PTF using SDSS

- Described in Ofek et al., 2012, PASP, 124, 62
- Uses frames that overlap with SDSS footprint to fit a global linear model for nightly data
- Enables calibration of all CCD images observed during a night
- Absolute precision (with respect to SDSS) is ~ 0.02 – 0.04 mag.
- Primary outputs: a global ZP value per image and a spatially-binned ZP residuals map (ZPVM)
- These ZP estimates are only applicable to *mag_auto* instrumental magnitudes
- $R^{inst}$ and $g^{inst}$ below are *mag_auto* (SExtractor) instrumental magnitudes

details). For observations taken using the $R$-band[18] filter, we fit the following model:

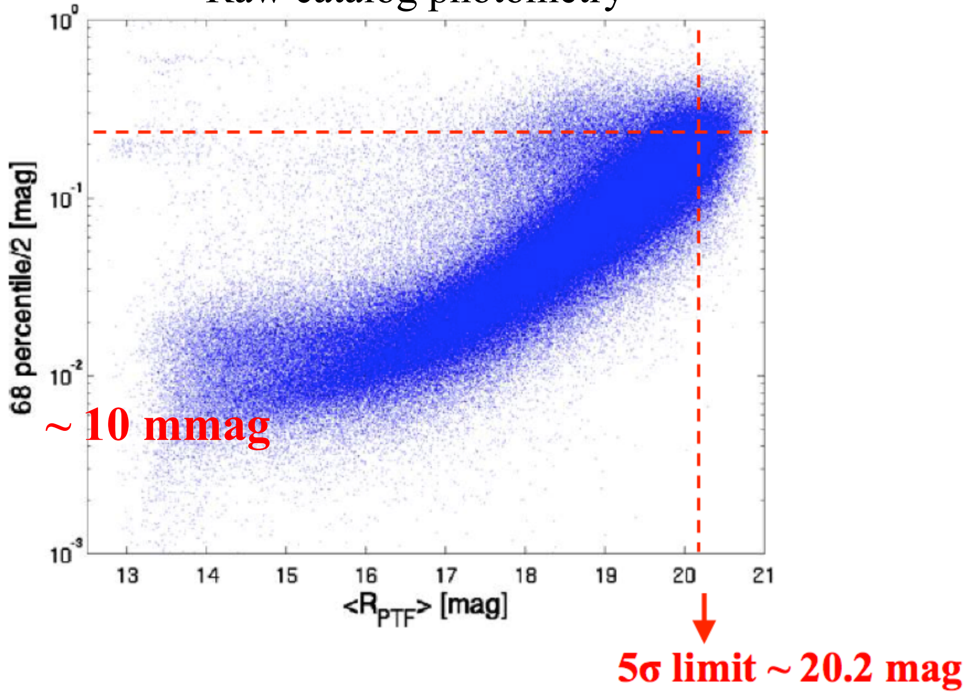
$$\begin{aligned} r_{\rm SDSS} - R^{\rm inst}_{\rm PTF} = {\rm ZP}_R + \alpha_{c,R}(r_{\rm SDSS} - i_{\rm SDSS}) + \alpha_{a,R}{\rm AM} \\ + \alpha_{ac,R}{\rm AM}(r_{\rm SDSS} - i_{\rm SDSS}) + \alpha_{t,R}(t - t_m) \\ + \alpha_{t2,R}(t - t_m)^2 - 2.5\log_{10}(\delta t), \end{aligned} \quad (1)$$

while for $g$-band observations we fit

$$\begin{aligned} g_{\rm SDSS} - g^{\rm inst}_{\rm PTF} = {\rm ZP}_g + \alpha_{c,g}(g_{\rm SDSS} - r_{\rm SDSS}) + \alpha_{a,g}{\rm AM} \\ + \alpha_{ac,g}{\rm AM}(g_{\rm SDSS} - r_{\rm SDSS}) + \alpha_{t,g}(t - t_m) \\ + \alpha_{t2,g}(t - t_m)^2 - 2.5\log_{10}(\delta t). \end{aligned} \quad (2)$$

# Photometric performance

Raw catalog photometry




Refined (relative) photometry for lightcurve DB