ZTF Operations Readiness Review

Frank Masci & the ZTF Science Data System Team

September 2017

Charter to the Board

[with slide numbers / comments for reference]

1. Are the requirements for data ingest, processing, and archiving at the beginning of survey operations agreed and understood? **[slides 4 & 11]** Are the requirements and schedule for alerts agreed and understood? **[slides 8, 13, 17]**

2. Are data access and distribution policies clear and stable enough to implement? **[slides 17 & 18]**

3. Does the ZTF Data System design allow it to meet the requirements and goals for survey operations? **[12, 13, 14]**

4. Is the ZTF Data System implementation on track so that enough of the system is available at the beginning of survey operations to meet the requirements?
   a. What components are still under development? **[slides 18 & 19]**
   b. Is there a detailed schedule showing intermediate milestones remaining to be completed prior to operations? **[20]**

5. Have the interfaces to the telescope and Observing System been tested, and has end-to-end data flow from the Observing System to the Data System been verified? **[slides 11 & 19]**

6. Will the file systems, databases, and archive access methods handle the data rates?
   [(i) pipeline to archive: yes; (ii) archive user-load, access rate and sustainability: unknown]
7. Is the calibration plan adequate for meeting requirements? Does it include a list of data sets required, and a schedule for obtaining them? **[slides 20, 21, 22]**

8. Is the plan for commissioning and science-validation defined well enough to achieve objectives and allow start of science survey? **[slides 20, 21, 22]**

9. Is the staffing currently assigned to complete the Data System, generate calibrations, perform analyses and test the system end-to-end adequate? This should account for any planned leave... **[last bullet on slide 22, then slide 23]**

Outline

- Project overview, facts, figures and key dates
- Data System objectives, requirements and Goals
- Status of data processing pipelines, sub-systems & deliverables
- Data access policies and implementation status
- Schedule and priorities
- Calibration and tuning plan
- Staffing
- Contingencies
- Typical “day in the life”: task orchestration
- Concerns

Ben Rusholme: will expand on status of data transfer and overall software infrastructure

Steve Groom: will summarize hardware procurement status, readiness of archive, user-interfaces & tools

The ZTF Science Data System (ZSDS)

- Housed at the Infrared Processing and Analysis Center (IPAC), Caltech.

- Responsibilities for ZTF:
  - data transfer from P48 telescope to IPAC
  - ingestion and archiving of *all raw* data acquired during project (both engineering & science)
  - all science-data processing pipelines
  - long-term archiving of data products, curation, user-interfaces, and APIs for data retrieval
  - near-realtime generation of flux-transient alerts and metadata for follow-up
  - near-realtime generation of products to enable NEA discovery (streaks & tracklets)
  - publishing of data quality metrics, diagnostics, analysis and reporting
  - maintenance of pipelines, operations, databases, fileservers, and archive infrastructure
  - documentation and user support

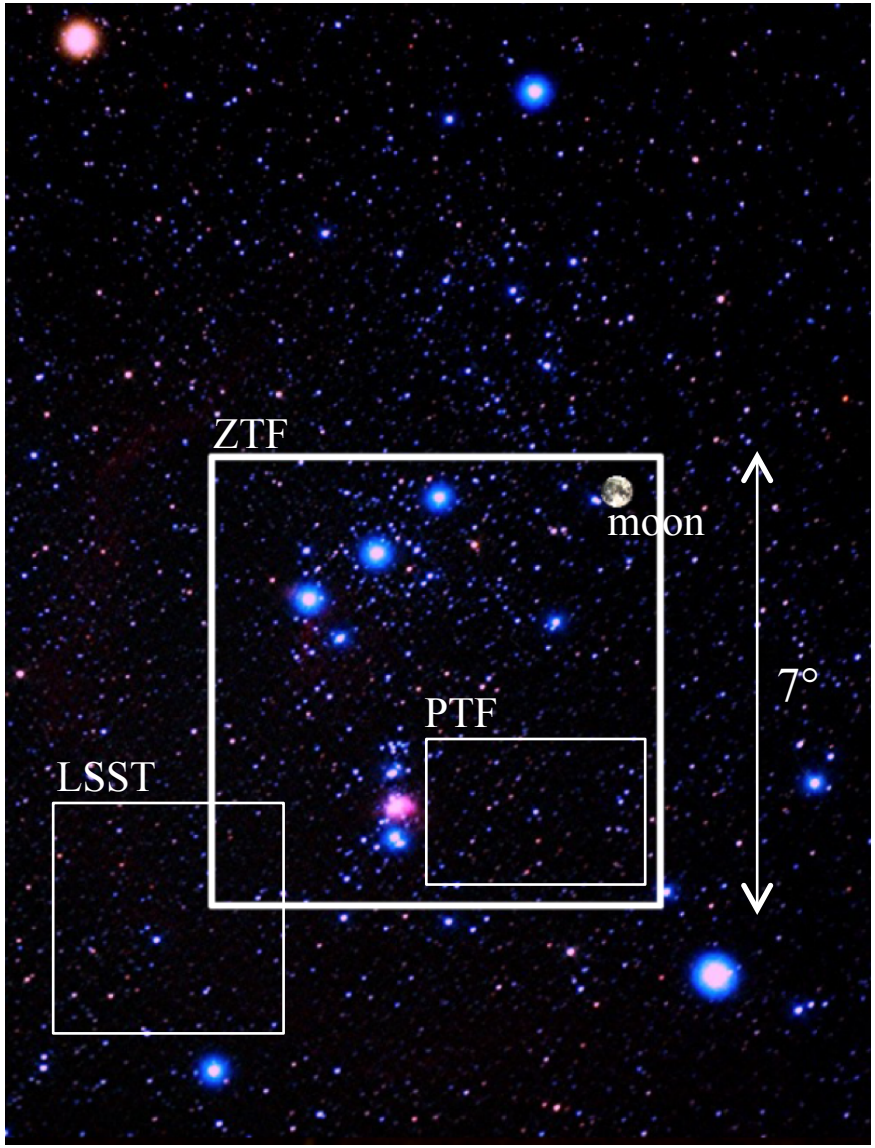
ZSDS Staff

- **Ben Rusholme:** data link from P48 to IPAC; pipeline job scheduling/executive; optimization; software/configuration management; hardware config.; alert distribution infrastructure (Kafka)
- **David Shupe:** astrometric calibration; source-matching and relative photometry pipeline
- **Russ Laher:** pipeline infrastructure; integration and testing; data ingest; pipeline executive; database schemas and stored procedures; bias- and flat-generation pipelines;
- **Steven Groom (and staff; IRSA Lead):** pipeline/archive interface design; system engineering; hardware shopping, costing and planning.
- **Frank Masci (ZSDS Lead):** instrumental and photometric calibration; reference-image generation; image-subtraction; extraction; moving-objects; algorithms; analysis; documentation…
- **David Flynn (and staff; ISG Lead):** system-engineering and hardware
- **Ed Jackson:** database administration
- **Jason Surace:** image simulation; data analysis
- **Ron Beck:** pipeline operations
- **David Imel (IPAC manager):** budgeting and personnel

Total FTEs as of Sep 1, 2017: 4.2

Total FTEs during development: 5.75

ZTF at a glance



- A fast, wide-area time-domain survey:
  - fast, young, and rare flux transients
  - counterparts to gravitational wave sources
  - low-*z* Type Ia SNe for cosmology
  - variable stars & eclipsing binaries
  - Solar System objects
- Active detector area: ~ 47 deg^2
- Areal survey rate: 3760 deg^2 / hour
- Single exposure depth (5σ): $r \sim 20.5$ mag.
- Median image quality (r): ~ 2.2” (FWHM)
- Nominal survey duration: 3 years
- Number of filters: 3 (g, r, i)
- Survey entire Northern visible sky to $\delta \sim -28°$

Key Project Dates

- Engineering commissioning: **Sep 6 – Oct 1, 2017 [delayed]**
- Science validation: **Oct 2 – Dec 31, 2017**
- Start of science survey: **Jan 2018**
- Commencement of public alerts: **Apr 1, 2018**
- First public data release: **Jan 2019**
- Future public data releases: **Jul 2019, Jan 2020, Jul 2020**
- End of science survey: (on-sky operations): **Dec 2020**
- Final public data release: **Jan 2021**

High level Data System
Objectives & Requirements

From the MSIP (NSF) proposal & ZTF Management Plan (03 / 12 / 2014):

- Sustain processing & storage for three years of operations (initially: 2017 – 2019)
- Leverage the existing PTF Data System and Archive infrastructure
- Scale data processing and storage to sustain a 15 x PTF data-acquisition rate
- Generate data products similar to those as PTF and additionally:
  - Provide a lightcurve retrieval / search tool
  - Real-time transient alerts for public consumption, beginning in survey year two
- Support public release of archived products every six months, with the first occurring at the end of survey year one
- Process data and generate alerts at all galactic latitudes in the Northern visible sky

More concrete specifications on products, requirements & goals, aligned with the science proposed to NSF and by project partners (known at the time) came together in a document by E. Bellm (12 / 16 / 2015):

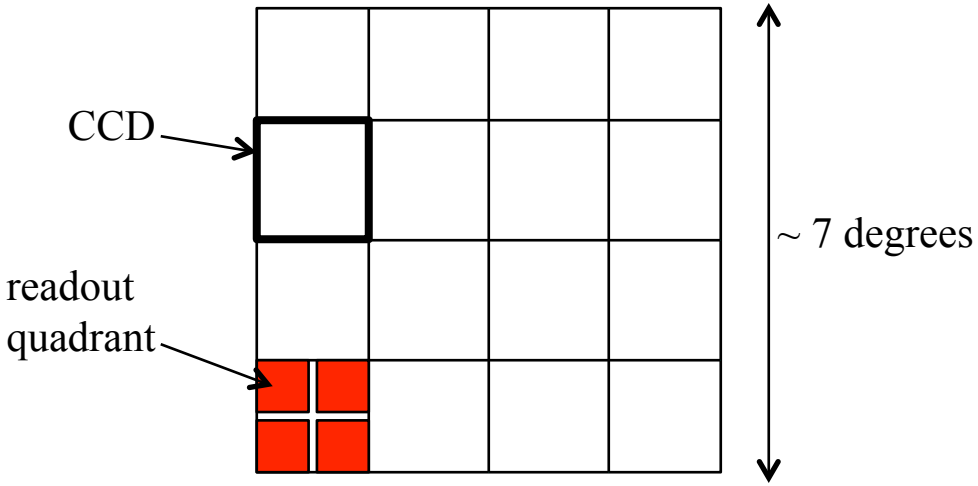
- Data System development was guided primarily by this document.
- Two additional requirements presented a huge challenge:
  - Include a database of sources detected by PSF-fit photometry from each epochal image.
  - Maintain a database of lightcurves generated by positionally matching all epochal image extractions.
  - Following analyses on source statistics, these are not feasible; arrived at a compromise for serving lightcurves
- Timing requirements:
  - > 95% of the images acquired at P48 need to arrive at IPAC within 10 min (goal: 5 min)
  - > 95% of the images received at IPAC must be processed with alerts published in < 10 min (goal: 5 min)

Decision made in early 2017:

- Commence public alerts earlier than planned: now in April 2018.

Raw Camera Image Data & terminology

- One camera exposure: 16 CCDs; each ~ 6k x 6k pixels
- Image data packet transmitted is per CCD (four readout-quadrant images & overscans)
- 16 CCD-based image files are acquired every 45 sec (30s exposures)
- Full camera exposure = one survey field on sky: ~ 1.3 GB (no compression)



Basic image-unit for pipeline processing from which all file-based products are derived is a ~3k x 3k readout quadrant:

- processed (epochal) science images
- reference image (co-adds)
- difference images
- source extraction table files
- lightcurve (source-match) files

Data product volumes, statistics & expectations

Estimates per night, based on an average on-sky duration of ~ 8hr 40min

- Number of on-sky camera exposures per night: ~ 700; calibration exposures: ~ 100
- Raw image data volume: ~ 1 TB (no compression)
- Raw incoming data rate: ~ 230 mega bits per second (no compression)
- Data product volume: ~ 3.8 TB per night (real-time products only)
- Number of **point source** transient events (flux and motion-induced): ~ 1 million (0.1x – 5x)
  - following machine-learned vetting of > 5σ events; sky-location dependent
- Number of streaks (candidates for "fast-moving" objects following ML vetting): ~ 1000
- Number of single exposure extractions: ~ 1 billion (PSF-fit based); ~ 300 million (aperture-based)
  - sky-location dependent
- Number of single readout-quadrant image products (science, difference, mask, catalogs): ~ 230,000

For nominal three-year survey (number of observing nights / year: ~ 260):

- Volume of data products: ~ 3 PB
- Number of single-exposure extractions: ~ 800 billion (PSF-fit based); ~ 230 billion (aperture)
- Number of reference images (co-adds in static library for image subtraction): ~ 282,000 (~ 55 TB)

ZTF Pipelines and run frequency

Overall, there are 9 interdependent pipelines, grouped into four categories.

- All implemented and tested on simulated camera-image data; some pipelines also tested using real camera data.
- All baseline archival products, formats, and methods for access are finalized (next slide).

Raw data ingestion and initial processing:

1. Raw data ingest, archival of raw images and storage of metadata in database [*realtime*]
2. Raw-image uncompression, splitting into readout-quadrant images, floating bias correction, QA metrics [*realtime*]

Calibration-image generation:

3. Bias-image derivation from stacking calibration images acquired in afternoon [*before/after on-sky operations*]
4. High-ν flat (pixel-to-pixel responsivity) from stacking illum. flat-screen exposures [*before/after on-sky operations*]

Real-time science-level processing:

5. Instrumental calibration of readout-quadrant images: includes astrometric and photometric calibration [*realtime*]
6. Image-subtraction with transient-event extraction (point sources & streaks), alert packets & distribution [*realtime*]

Ensemble-based (collective-image/catalog) processing:

7. Reference-image generation (co-addition of epochal images from 5) [*when sufficient good quality data available*]
8. Source-matching/lightcurves with relative photometric refinement; inputs from 5 & 7 [*every month, TBD*]
9. Moving object tracks, orbit-fitting, QA; from linking point-source events from 6 [*end of night, 3-4 day window*]

Baseline deliverables & data access portals

1. **Instrumentally calibrated, <u>readout-quadrant based</u> epochal image products:**
   - images with photometric zero-points from PSF-fit photometry; astrometric solutions; bit-mask images
   - two source catalogs per image: PSF-fitting and aperture photometry
   - difference images with accompanying PSFs
   - ➢ archive via GUI or API at IPAC; can interface with Moving-Object Search Tool (MOST)

2. **Reference images (co-adds), coverage, unc maps, and two source catalogs per image**: PSF-fitting and aperture
   - ➢ archive via GUI or API at IPAC

3. **Lightcurves and collapsed-metrics using "object-based" searches by position and/or metrics**
   - ➢ archive via GUI or API at IPAC; can interface with lightcurve viewer/analyzer

4. **Match-files: all lightcurves per readout-quadrant:** from source-matching of epochal PSF-fit extractions
   - ➢ restricted (galactic marshal)

5. **Raw image data (CCD-based files with metadata) and image calibration products used in pipelines**
   - ➢ archive via GUI or API at IPAC

Baseline deliverables & data access portals

6. **Alert (point-source event) stream** from real-time image-differencing pipeline: packetized with metadata, 30 day photometric histories, upper limits, ML-scores, cutouts on new, reference and difference images, …
   - ➢ transmitted to UW; access via specific science/filtering channel using "Kafka consumers"

7. **Products to support realtime Solar System/NEO discovery and characterization:**
   - streaks (fast moving objects) from difference images: metrics, ML-scores, and cutouts
   - moving object tracks from linking point-source events; known objects are tagged
   - ➢ ZTF-depot (restricted audience) for human vetting
   - ➢ Tracks for new (unknown) objects and new measurements on known objects will be delivered to MPC

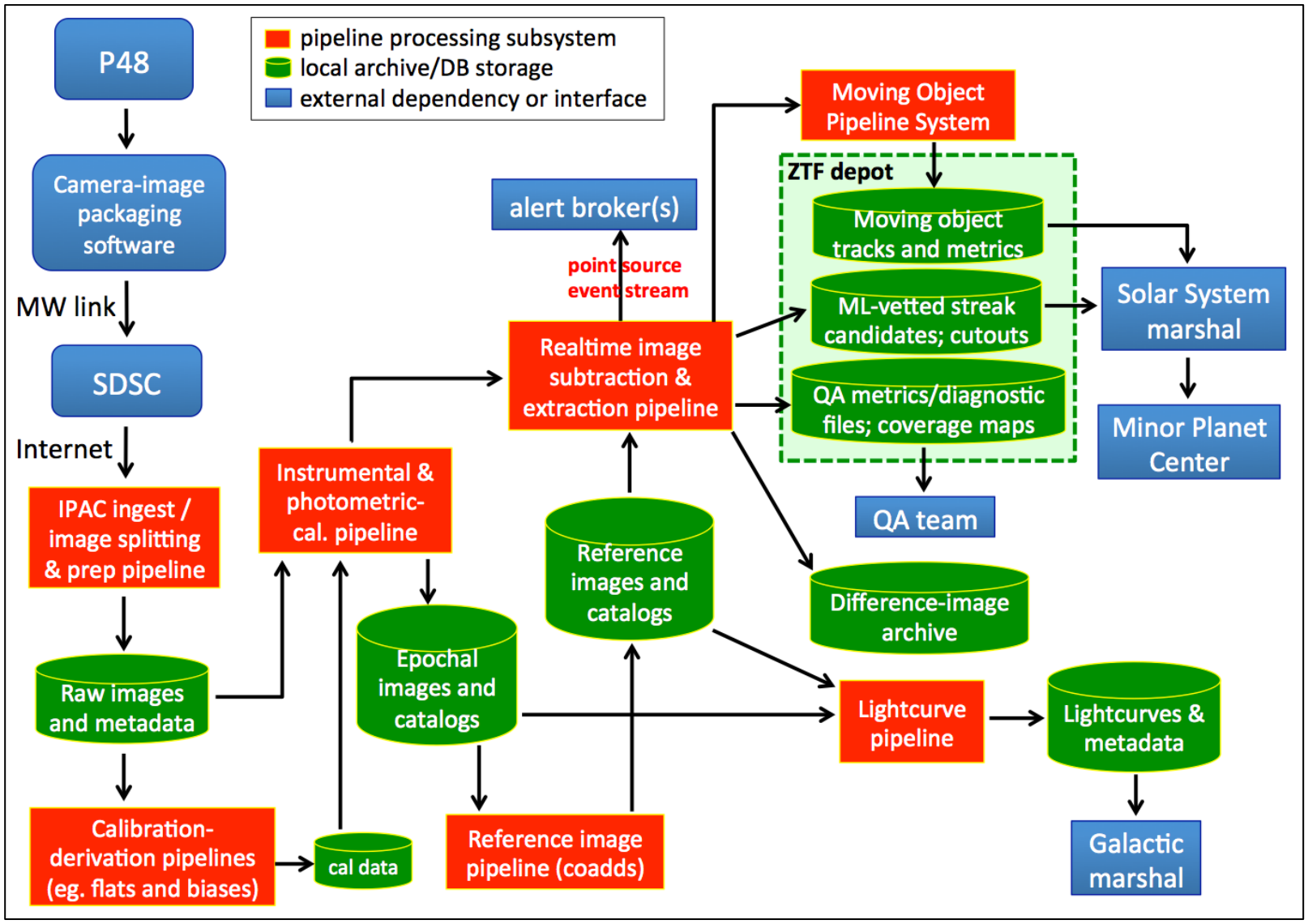
8. **Quality assurance metrics, summary statistics, and coverage maps for performance monitoring**
   - ➢ ZTF-depot (restricted audience)

9. **Documentation:** pipeline descriptions, cautionary notes, recipes, and tutorials on data-retrieval
   - ➢ *http://web.ipac.caltech.edu/staff/fmasci/home/ztf_pipelines_deliverables_V0.pdf*

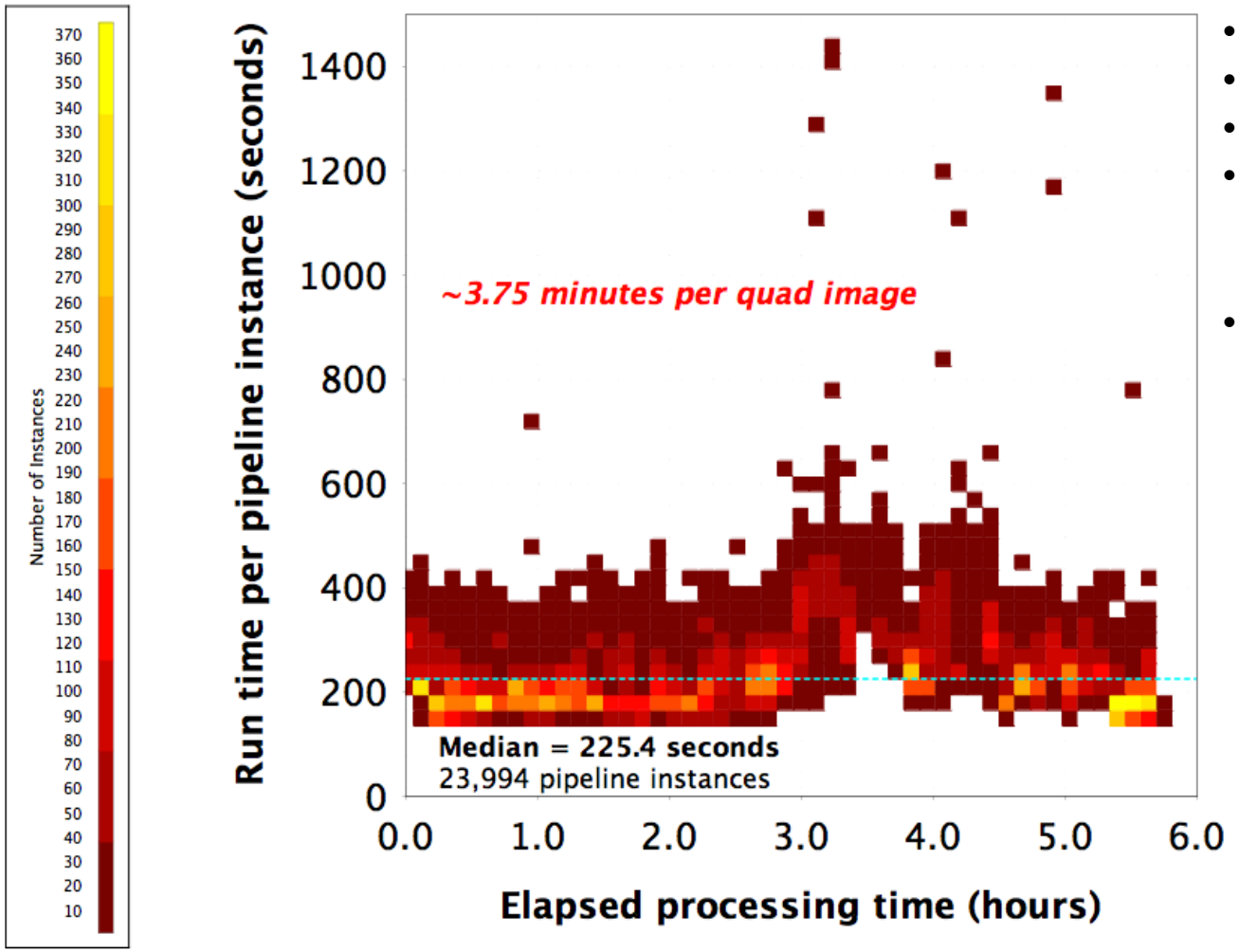
Ancillary (internal only):

All difference image events, photometry, QA metrics, and source-features are stored in a database (pipeline ops side)

Overall processing & data flow



ZTF real-time pipeline runtime (processing unit = one readout-quad image)



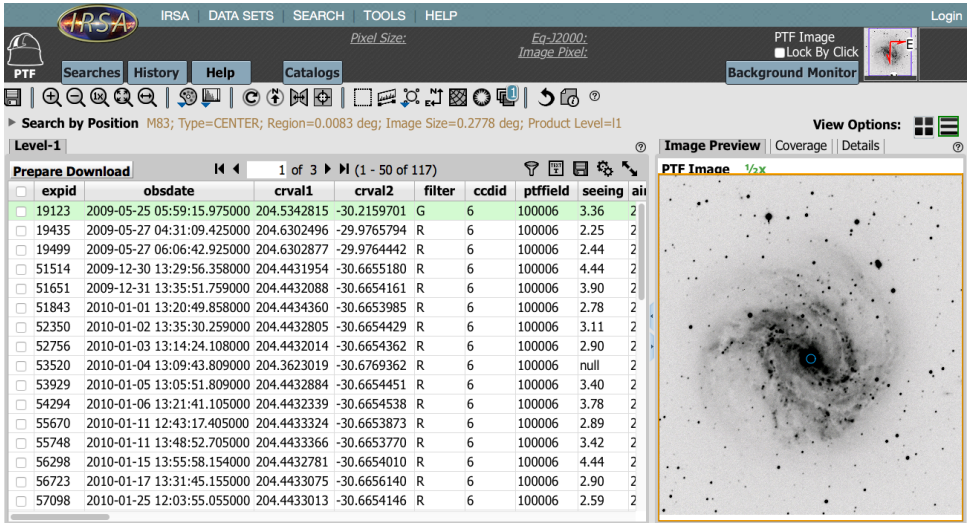
- 32 machines × 16 cores each
- 45 sec between exposures
- 630 input exposures: ~ 7.6 hrs
- from simulated summer night that crosses galactic plane

- exercises full realtime pipeline with alert packets containing 30 day histories per event, streaks, ML-vetting, all ztf-depot & archive products

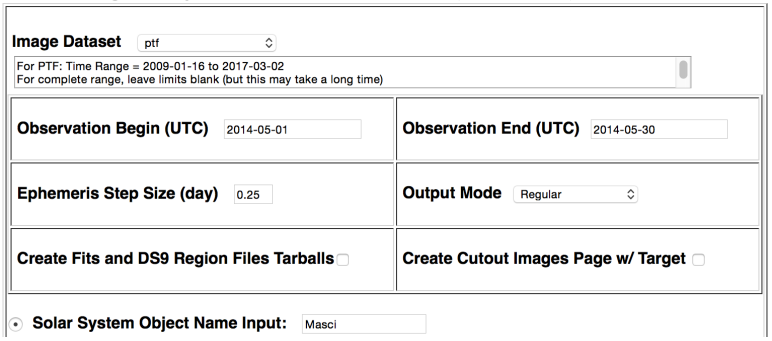
User Interfaces to archive products

- Will allow search by position, time-windows, filtering on metadata, object name, interactive manipulation, catalog overlays, visualization, basic analysis of lightcurves with periodogram service.
- Accompanying APIs (command-line driven retrieval) also available

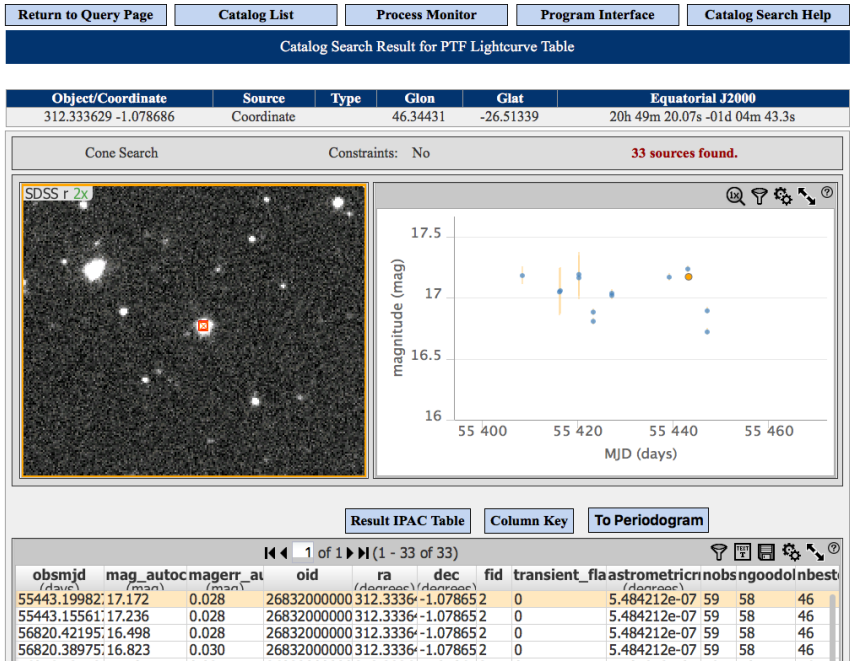
Image viewer and file-product retrieval



Moving Object Search Tool (MOST)



Lightcurve viewer/analyzer and retrieval



Data Access Policy

- Observing time during science operations will be split between three categories:
  - **Public** (NSF-funded MSIP survey: 40%)
  - **Private collaboration** (40%)
  - **Caltech TAC** (20%)
- Managed per exposure (epoch) using a *programID* propagated from scheduler to raw-image metadata

---

- Private/Caltech observers can access their data in near-realtime, soon after archive ingestion. This includes all calibration products and lightcurves from epochs tagged by their respective *programID*s queried via archive GUI.
- Public data will only be available at the public release times for general access by all.
  - raw images, processed epochal images, accompanying source-catalog files, difference images
  - reference images and catalog files
  - lightcurves constructed from public epochal data only
  - calibration data products
- Public alert packets (triggered from events detected in public exposures) will only contain public data. This includes their 30 day event histories.
- Private alert packets (triggered from events detected in private exposures) can contain unreleased public data in their 30 day event histories.
- Caltech alert packets (triggered from events detected in Caltech exposures) can contain unreleased public data in their 30 day event histories.
- No restriction on input data used to generate products for Solar System science: streaks & moving-object tracks; selected (human-vetted) products will be delivered to MPC.
- No restriction on input data used to generate reference image (co-add) products.
- No restriction on input data used to generate source match-files (lightcurve files):
  - MOU in place with the only customer of these products: Galactic Marshal
  - only privately-tagged and *already-released* public data therein to be ingested by Marshal

Data Access Policy
Implementation Status

Below is a repeat of the policies separated according to where implemented, or where access-control takes place:

pipeline ops side (implemented & ready) **user-interface to archive (in progress)**

- Private/Caltech observers can access their data in near-realtime, soon after archive ingestion. This includes all calibration products and lightcurves from epochs tagged by their respective *programID*s queried via archive GUI.
- Public data will only be available at the public release times for general access by all.
  - raw images, processed epochal images, accompanying source-catalog files, difference images
  - reference images and catalog files
  - lightcurves constructed from public epochal data only
  - calibration data products
- Public alert packets (triggered from events detected in public exposures) will only contain public data. This includes their 30 day event histories.
- Private alert packets (triggered from events detected in private exposures) can contain unreleased public data in their 30 day event histories.
- Caltech alert packets (triggered from events detected in Caltech exposures) can contain unreleased public data in their 30 day event histories.
- No restriction on input data used to generate products for Solar System science: streaks & moving-object tracks; selected (human-vetted) products will be delivered to MPC.
- No restriction on input data used to generate reference image (co-add) products.
- No restriction on input data used to generate source match-files (lightcurve files):
  - MOU in place with the only customer of these products: Galactic Marshal
  - only privately-tagged and *already-released* public data therein to be ingested by Marshal

Readiness of Sub-systems

green => in progress
red => requires input or development
black => high-level implementation complete

Data transfer from P48 to IPAC and interfaces

Pipeline Executive (includes Virtual Pipeline Operator tasks)
- job scheduling, prioritization; calibration generation & picking
- moving object pipeline execution; nightly coverage map generation

Pipeline details / algorithms
- instrumental calibration (astrometric & photometric & signatures):
  > new instrumental signatures; bright source artifact / ghost masking; tuning
- image subtraction, event extraction, assoc. & alert packet generation:
  > train machine-learned classifiers for point sources & streaks
  > assignment of unique alert-IDs (names) to alert packets
  > update S/G classification flags for PS1 in DB with ML probabilities

Archive with search GUIs, APIs & user tools
- image and catalog **file** product server (single epoch & co-adds)
- moving-object search tool (MOST)
- object-based lightcurve retrieval service
- user-access according to policies
- on-demand (realtime) finder-chart service

Filesystem "only-view" web server

Persistent ztf-depot for realtime prods
- streaks, stamp-image cutouts
- moving-object tracks, orbits & metadata
- pipeline QA metrics / aggregate stats
- survey coverage maps
- user authentication

Temporary ztfweb for archive products
For early access only; will go away
- view to pipeline/archive filesystem only
- select products to support analyses
- file listings only, no search GUI.
- user authentication

ZTF Alert Distribution System (ZADS)
- Kafka cluster (producer) set-up
- Interfacing / mirroring with UW

ZSDS Documentation: 85% complete

Data System schedule and activities

- **Black => Data System activities / milestones**
- ★ **Green => Key project dates**

★ **Sep 6, 2017:** first light (start of engineering commissioning) -- **Delayed**
- have access to full science and engineering observing grids; prep static calibrator catalogs in operations system
- ensure archive can ingest reference (co-add) image products with metadata
- *some* calibrations / detector parameters for pipeline (gains, distortion, astrometry parameters, bad pixel masks)
- characterize detectors; identify new artifacts; ghosting : devise correction/masking methods for pipeline

★ **Sep 18, 2017:** flat-field illuminator ship (earliest)
- high spatial-frequency responsivity maps (flats) and analysis using flat-screen images

• **Sep 20, 2017:**

✔ – moving object pipeline (point source event linking) in place and operational in test
- update PS1 table in operations DB with ML-based S/G classification probabilities

★ **Oct 2, 2017:** start of science validation
- user-interfaces to retrieve image/catalog file products from archive in place
- pipeline tuning & optimization
- optimize filtering of input photometric and astrometric calibrator catalogs
- optimal depth for reference images per filter
- full *routine* calibration generation and analysis for pipeline input (lo/hi-ν flats, biases, final masks …)
- exercise / refine / rehearse daily operational routines (P48 – IPAC – archive – customers – points of contact)
- performance analyses: astrometric / photometric precision, depth, FWHM variation: need tools from collab.
- train ML classifiers (point sources & streaks)

Data System schedule and activities (continued)

- **Black => Data System activities / milestones**
- ★ **Green => Key project dates**

- **Nov 30, 2017:**
  - user-interface to search and retrieve lightcurves from source match-files (first requires object-based DB)
  - integration of Moving Object Search Tool
- **Dec 10, 2017:**
  - alert-packet schema finalized with naming/alert-ID mechanism in place
  - Kafka "producer" interfacing / mirroring with UW

★ **Jan 2018:** start of science survey
  - finder-chart service interfacing with archive in place and operational
  - ready to generate reference-image library, meshed with image-differencing and event extraction

★ **Apr 1, 2018:** public alerts commence

★ **Jan 2019:** first public data release

Calibration & pipeline-tuning plan

- A “pipeline-needs” document was forwarded to project for integration into a grander commissioning / science-verification plan.

- This outlines experiments, detector characterization activities, and methods for deriving required pipeline calibration products and parameters, suggested stability analyses etc. For example:
  - Bias frames
  - Dynamic over-scan corrections
  - Bad pixel masks
  - Illuminated “dome-screen” exposures for flat-field derivation; all filters
  - Low spatial frequency responsivity maps
  - Read-noise, electronic gain estimates, saturation limits
  - Non-linearity check / assessment
  - Prior optical distortion map and other astrometric priors for all readout channels
  - Ghost / halo characterization and models to facilitate realtime prediction and masking
  - Scattered light characterization

- The above have been prioritized (primary vs secondary) according to pipeline needs in order to push products through science pipelines during early commissioning.

- Also listed a number of general performance and quality checks as survey proceeds: astrometric & photometric precision; sensitivity; absolute calibration accuracy; image quality.

- **We will need assistance from collaboration with the above:** analysis, derivation of calibrations, development of tools for routine monitoring.

Data System Operations Staffing

- Workforce will decrease during the transition from Development to Commissioning/Science verification (4 months)
- We have defined a Data System Verification period following this: **Jan 2018 (survey start) to mid Jul 2018**:
  - Need to accommodate additional/ad-hoc pipeline tuning, possible ad-hoc reprocessing activity received from analyses and ongoing performance monitoring
  - Development of additional functionality and archive-centric tools are anticipated into science operations; requires cost analysis before proceeding; not reflected below.

Sep 2017 → Jul 2018

Data System Workforce (FTE-Rate)	Dev	Commissioning, SV & DS Verification	Nominal Ops
Task Management and Reporting	0.50	0.40	0.30
Pipeline Development	2.70	0.35	
Archive Development	1.50		
Simulation, Analysis, Performance	0.15	0.15	
Database Administration	0.20	0.30	0.10
Ongoing PTF/iPTF reprocessing	0.20		
Pipeline Maintenance and Operations		1.50	1.50
Archive Ingest and Operations		1.00	1.00
Datacenter Operations	0.50	0.50	0.25
TOTAL	**5.75**	**4.20**	**3.15**

Typical "day-in-the-life" at the Data Center

Following an observing night or prior to forthcoming night (following end-of-night signal from P48):

1. Update all Solar-System ephemeris files in operations with latest from MPC.
2. If enough data were acquired in recent night, trigger moving-object pipeline (sliding window fashion).
3. Coverage map updates if enough data were acquired during previous night.
4. Overall ingest, processing and archive accountability check, with summary stats emailed to project.
5. Ensure there is sufficient sandbox disk space in pipeline operations, otherwise cleanup.
6. Ensure there is sufficient archive disk space and that date-dependent paths are set.
7. Check and cleanup ztf-depot file system; ensure date-dependent paths are set.
8. Deposit recent night transient-candidate metadata from DB to filesystem for fast history association
9. Check with scheduling team if any new fields are off the pre-defined "science grid". If so, prepare new calibrator files, pipeline config. files, and database content.
10. Ensure Virtual Pipeline Operator (VPO) is ready to ingest new data; review and revise pipeline triggering rules, and priorities if needed.
11. Early in survey: check if enough good-quality data were acquired for new fields and trigger reference image (co-add) pipeline.
12. Every month or more: contingent on data volume, trigger source-matching (lightcurve) pipeline.
13. Scheduled downtime for database maintenance.
14. Scheduled downtime for general systems maintenance.

- Most of the above are automated through VPO
- Not necessarily orchestrated in this order

Contingency Planning

- Need to prepare for a number of "what if" scenarios, e.g.,
  - Operations database goes down?
  - Archive-ingestion stops or starts to lag behind significantly?
  - Data link (HPWREN) issues with back-log of data at P48?
  - Other network problems?
  - Hardware issues and replacement?
  - Missing image metadata or glitches in observing system?
- We will disable specific components of the DS and rehearse recovery procedures.
- Prompt communication with project and users when issues occur.
- Need to establish points-of-contact early in project; includes back-up staff.
- Recovery procedures are currently being documented by ISG, DBA and pipeline ops staff.

Concerns (or challenges)

My personal take (mostly science-quality centric):

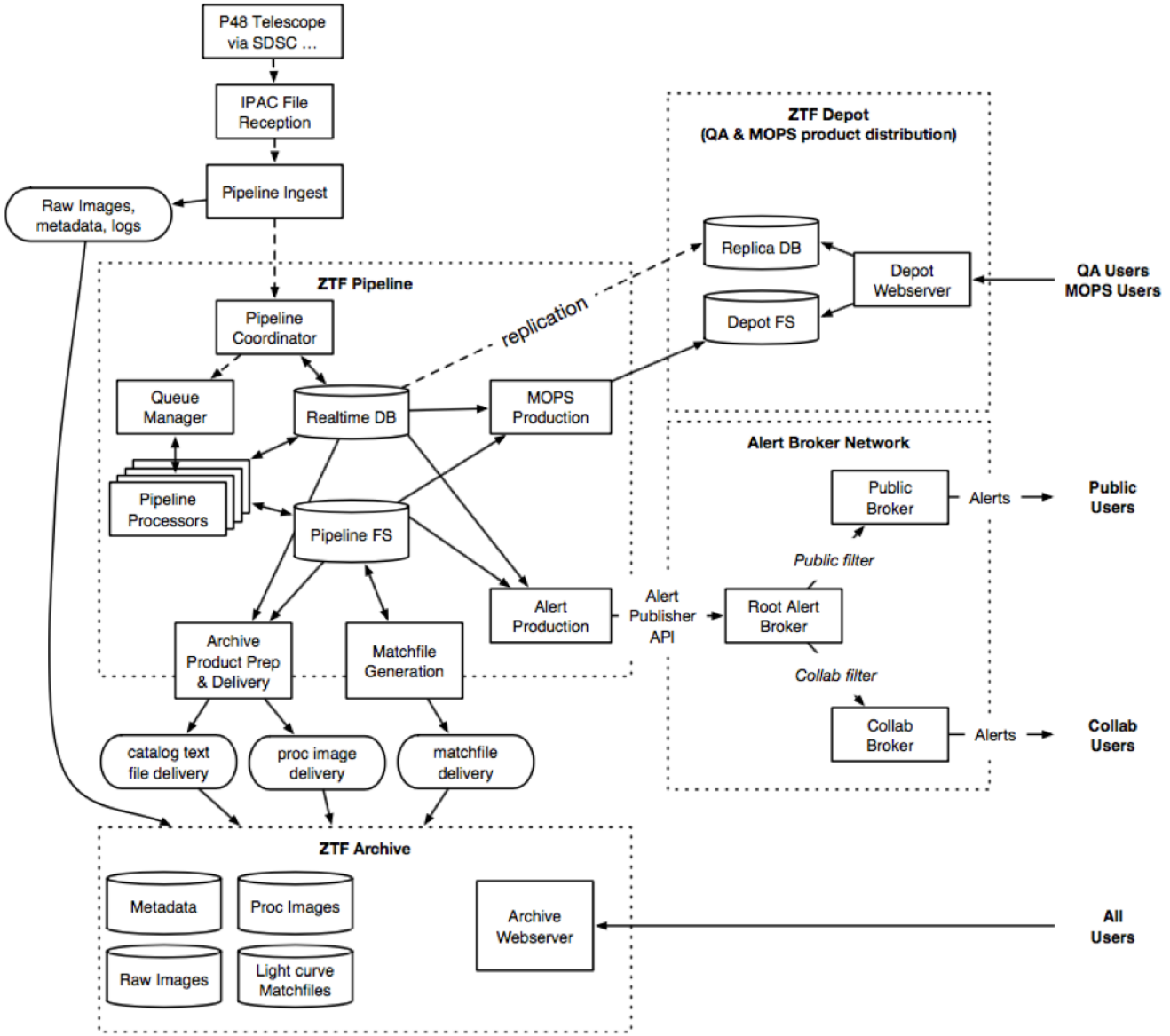
- **Flat-fielding:** not taken seriously enough (yet) by the project.
  - will impact quality of image-subtraction; relative photometric accuracy – crucial for reference-image generation, lightcurves, other post-processing tasks on archived data.
- Astrometric calibration will be a challenge(!)
- Galactic plane processing: our simulations are holding up to the expected high source densities (pipeline to archive); *realtime calibration* steps will be a challenge.
- New detector features or idiosyncrasies requiring significant pipeline (re)coding / restructuring.

Other (operational related):

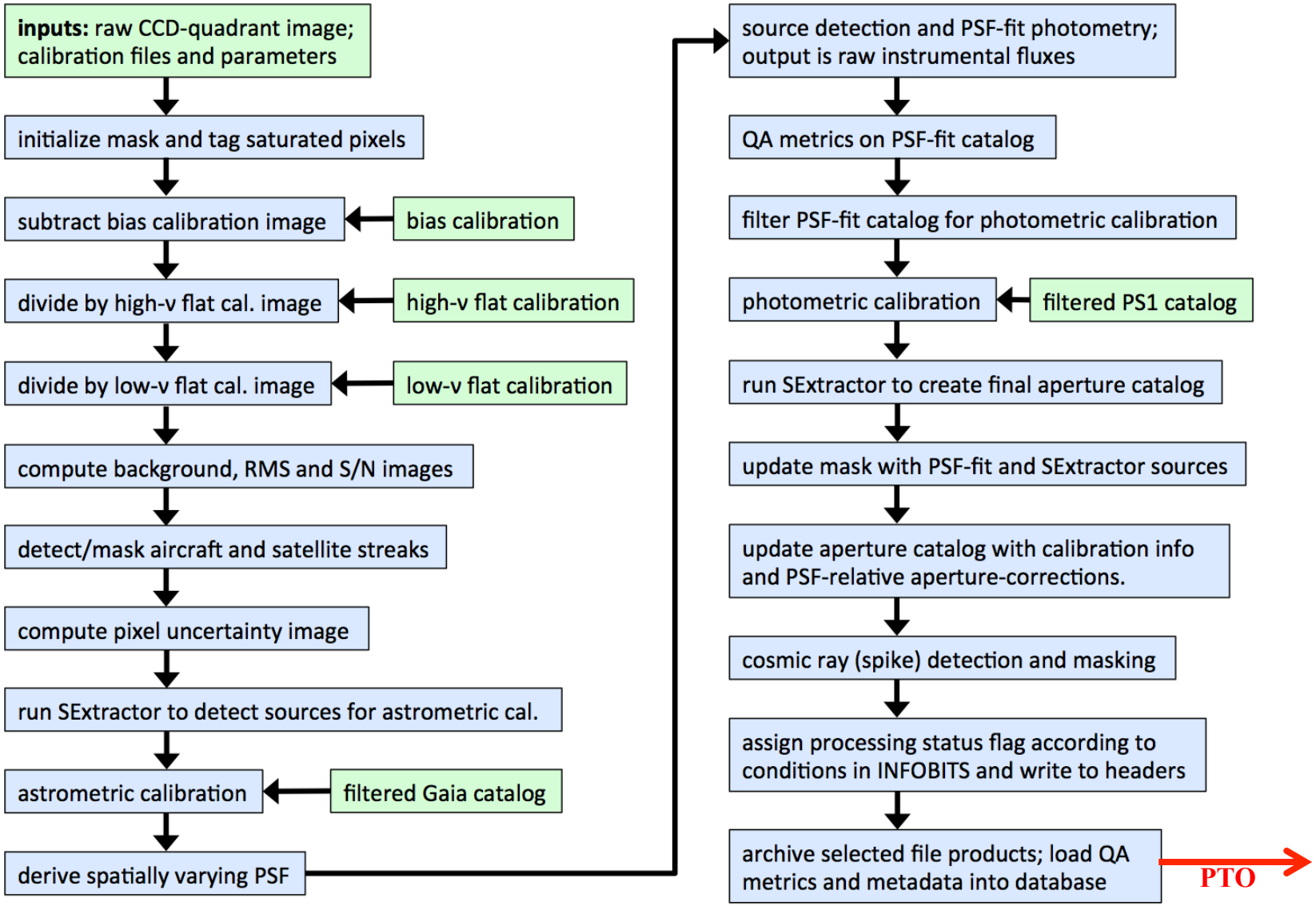
- Overall robustness to pipeline failures; catching errors early; tractability; recovery; reporting.
- Delays in data-transfer: how fast can we catch-up and resume realtime processing?
- Competition for bandwidth from Palomar: other activities, e.g., SED Machine on P60.
- Managing expectations. Will require assistance from Project Scientist (communication channel).

Back up slides

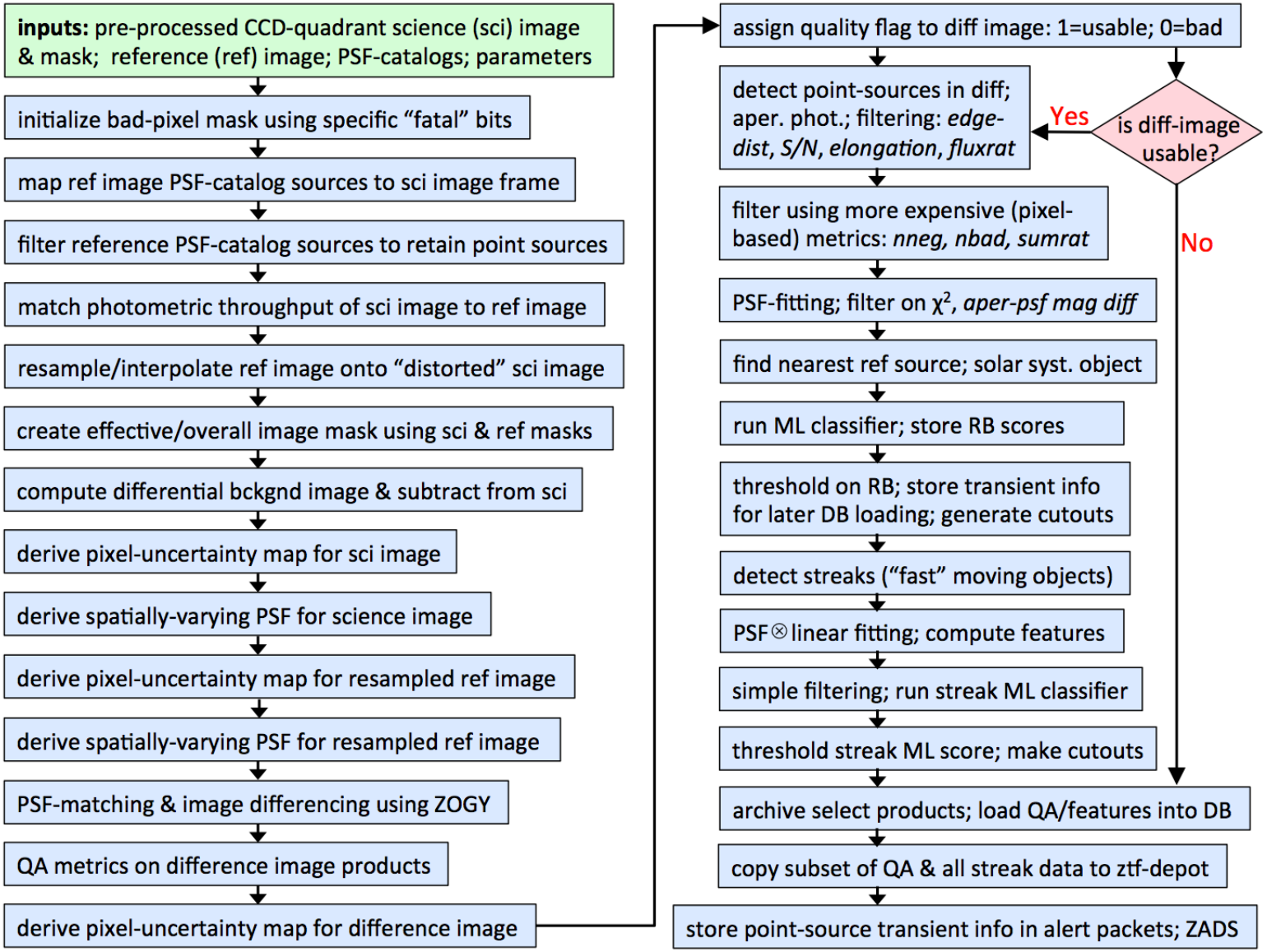
More detailed data flow



ZTF Real-time pipeline (phase 1): instrumental calibration



ZTF Real-time pipeline (phase 2): image subtraction & extraction



ZTF real-time processing throughput (naïve estimate)

- Incoming data rate (set by cadence):
  - one exposure or 64 quadrant images / 45 sec.
  - *inprate* ~ 85 quad images / minute, on average

- Processing rate (median as of today):
  - *outrate* ~ 1 quad image / 4 minutes / CPU core

- If processing was purely CPU-limited, no or negligible I/O latency, *minimum* number of CPU cores needed to keep up with input data rate is:
  N_{cores} = *inprate* / *outrate* = 340 cores

- This estimate is naïve since it ignores I/O, network speed, other interleaved processing tasks.
  Goal is to process faster than incoming data rate.

- Our currently "active" ZTF compute cluster has 16 physical cores × 32 nodes = 512 cores
  (or 16 × 2 × 32 = 1024 admissible simultaneous threads, contingent on shared resources)

Data System staff functions

Task Management

Coordinate data system activities; report to ZTF project manager; coordinate with project scientist.

Pipeline Development

Design, prototype, and implement pipeline software modules; integrate into production system, interfacing with database and archive

Archive Development

Scripts to transfer ZTF data to formats consistent with permanent storage at IRSA; adaptation of IRSA user interfaces for ZTF-special capabilities.

Simulation, Analysis, and Performance Monitoring

Provide synthetic data sets for pipeline testing; analysis of instrument performance; evaluate instrument calibration and stability with recommendations for pipeline tuning/ optimization during survey (works with Pipeline Developer)

Database Administration

Design ZTF schema; optimize database tables for quick access (both on archive and pipeline operations side); manage daily large database merges for new data.

PTF / iPTF Operations

Complete ingest and processing of PTF data; requested bulk reprocessing to support future public release

Pipeline Operations and Archive Ingest:

Execute and monitor pipelines; periodic transfers of raw and processed data to archive; daily cleanup/set-up for forthcoming night; schedule and manage reprocessing requests as needed.

Datacenter Operations

Procure and install all ZTF-related hardware: compute drones, storage arrays and disks, network gear; offsite backups; environment monitoring.