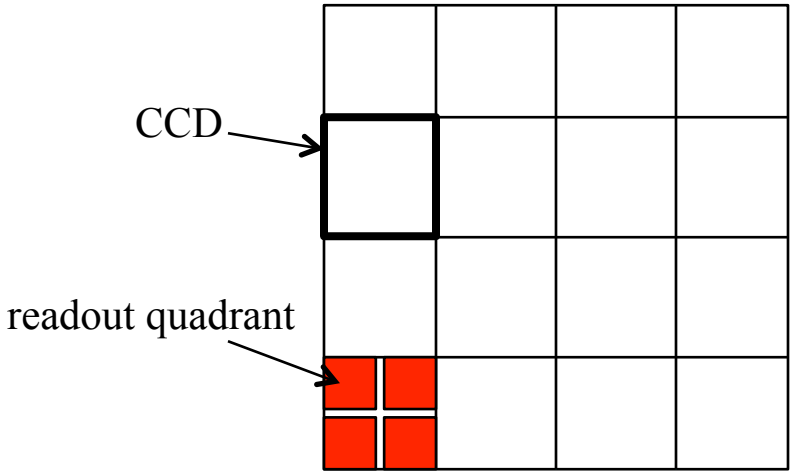
# ZTF Data System Update and Plan

Frank Masci & the IPAC-Caltech ZTF Team

March 9, 2017

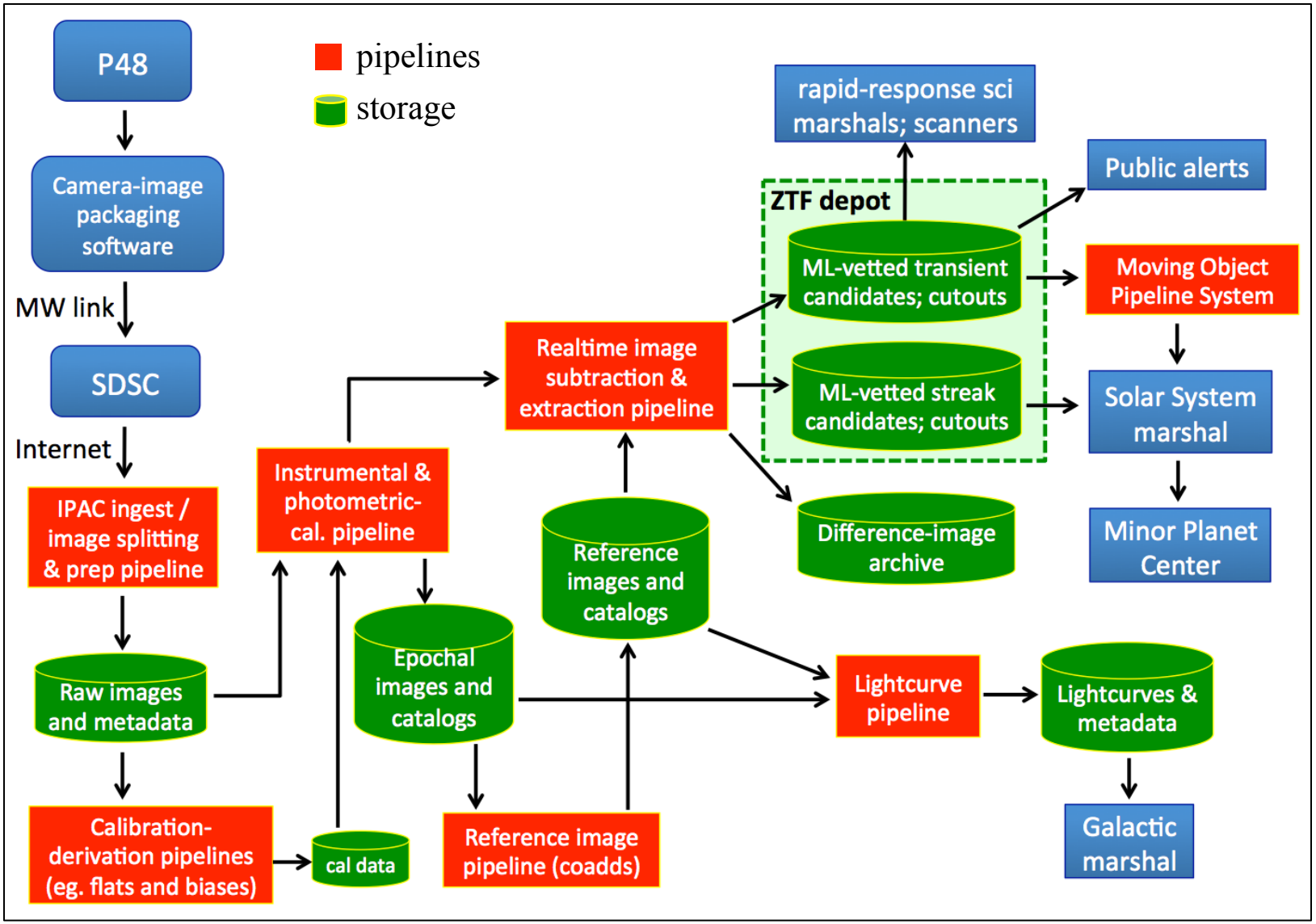
# ZTF Raw Camera Image Data

- One camera exposure: 16 CCDs; each ~ 6k x 6k pixels
- Image data packet transmitted is one CCD (= four readout-quadrant images)
- 16 CCD-based image files are transmitted ~ every 45 sec.
- Full camera exposure: ~ 1.3GB uncompressed
- Require *lossy* compression to accommodate transfer bandwidth (~ 110 - 150 Mbits/sec, variable)

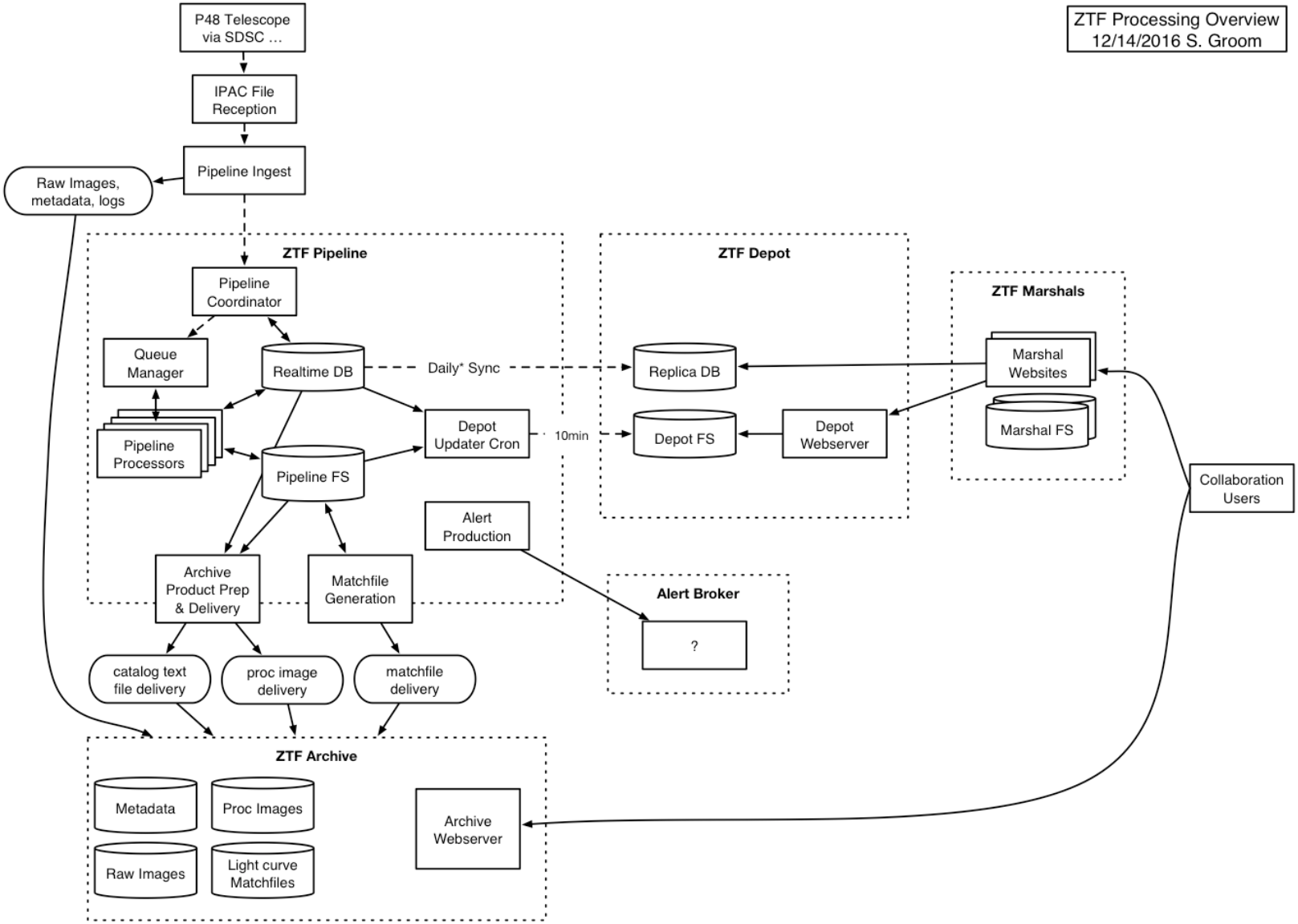



Basic image-unit for pipeline processing from which all products are derived is a ~3k x 3k readout quadrant image.

# Overview of the ZTF Data System



# Architecture Details



# ZTF Pipeline Status: ✔ => done (integrated)

Overall, there are 10 inter-dependent pipelines (one is TBD):

**Raw data ingestion/processing:**

✔ 1. Raw data ingest, archival of raw images and storage of metadata in database [*realtime*]

✔ 2. Raw-image decompression, splitting into readout-quadrant images, floating bias correction, simple QA [*realtime*]

**Calibration generation:**

✔ 3. Bias-image derivation from stacking calibration images acquired in afternoon [*made before on-sky operations*]

✔ 4. High-ν flat (pixel-to-pixel responsivity) from stacking calibration images [*made before on-sky operations*]

**5. TBD:** Low-ν flat from either long-term ZPVM or dithered-star observations [*every week, month or longer?*]

**Real-time:**

✔ 6. Instrumental calibration of readout-quadrant images: astrometry and photmetric cal [*realtime*]

✔ 7. Image subtraction and transient discovery (point sources / streaks), metadata and cutouts [*realtime*]

**Ensemble-based processing:**

✔ 8. Reference-image generation (co-addition of epochal images from 6) [*as needed: when good quality data available*]

9. Source-matching with relative photometric refinement for lightcurves; inputs from 6 [*every two weeks or longer?*]

10. Moving object pipeline system (MOPS): tracklets from linking transients from 7 [*every 3 or 4 hours during night*]

# Deliverables and Products (reminder)

1. **Instrumentally calibrated, <u>readout-quadrant based</u> epochal image products:**
   - images with photometric zero-points derived from STF-fit photometry
   - bit-mask images
   - two source catalogs per image: PSF-fitting and aperture photometry:
   - difference images with QA metadata

2. **Reference images (co-adds), coverage, unc maps, and two source catalogs per image**: PSF-fitting and aperture

3. **Match-files per readout-quadrant from source-matching of epochal extractions:**
   - based on epochal PSF-fit photometry catalogs: to support "object-based" lightcurve database:
   - object cone searches via user interface ➔ LC + LC-collapsed metrics extracted from source match-file

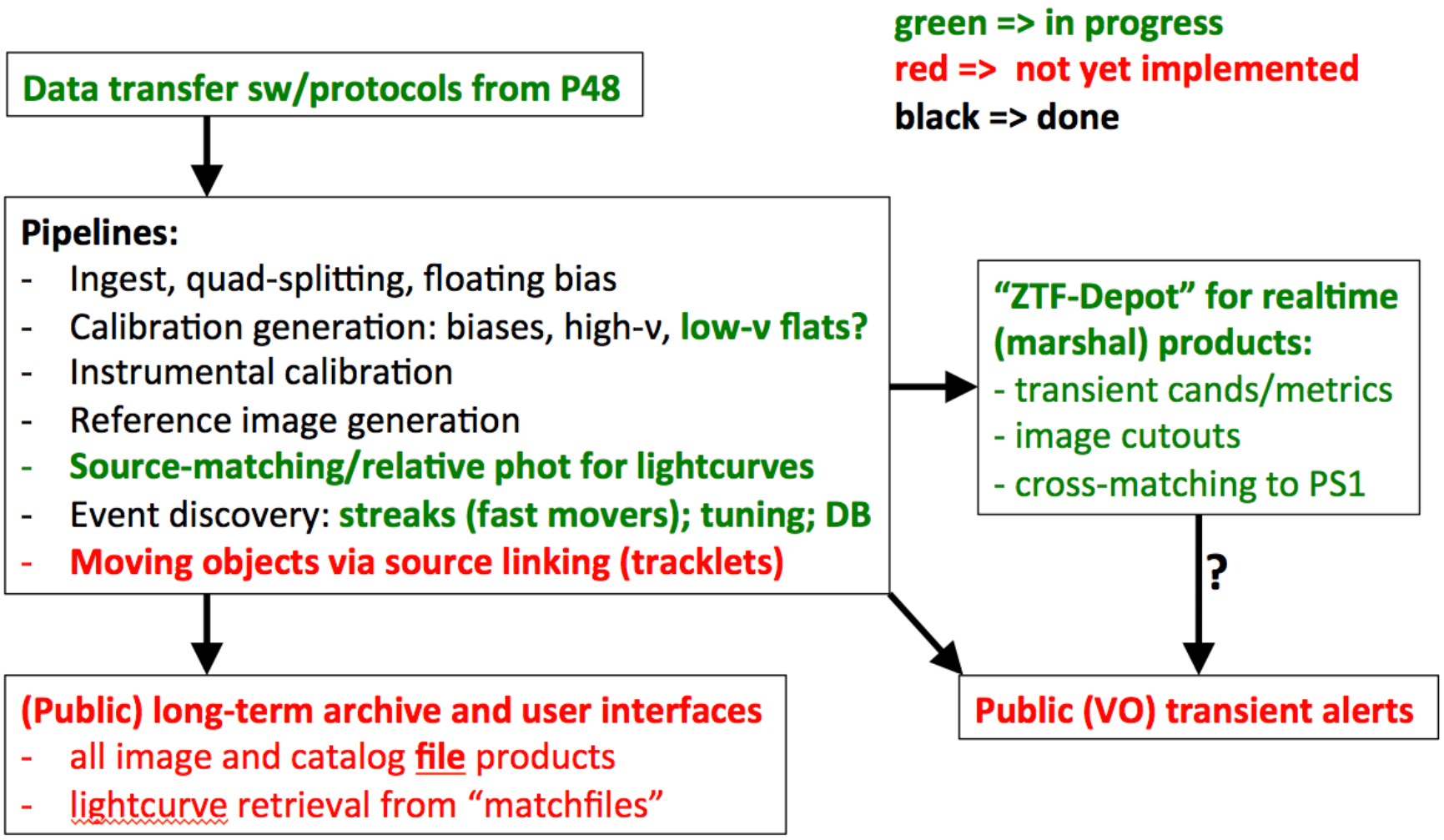
4. **Products to support near real-time discovery:** *thresholded* transient candidates (point sources and streaks) with metadata and image cutouts

5. **Historical (users) database of all transient candidates and metadata generated from real-time pipeline**

6. **To commence following survey start**: alert (event) stream extracted from real-time pipeline with metadata

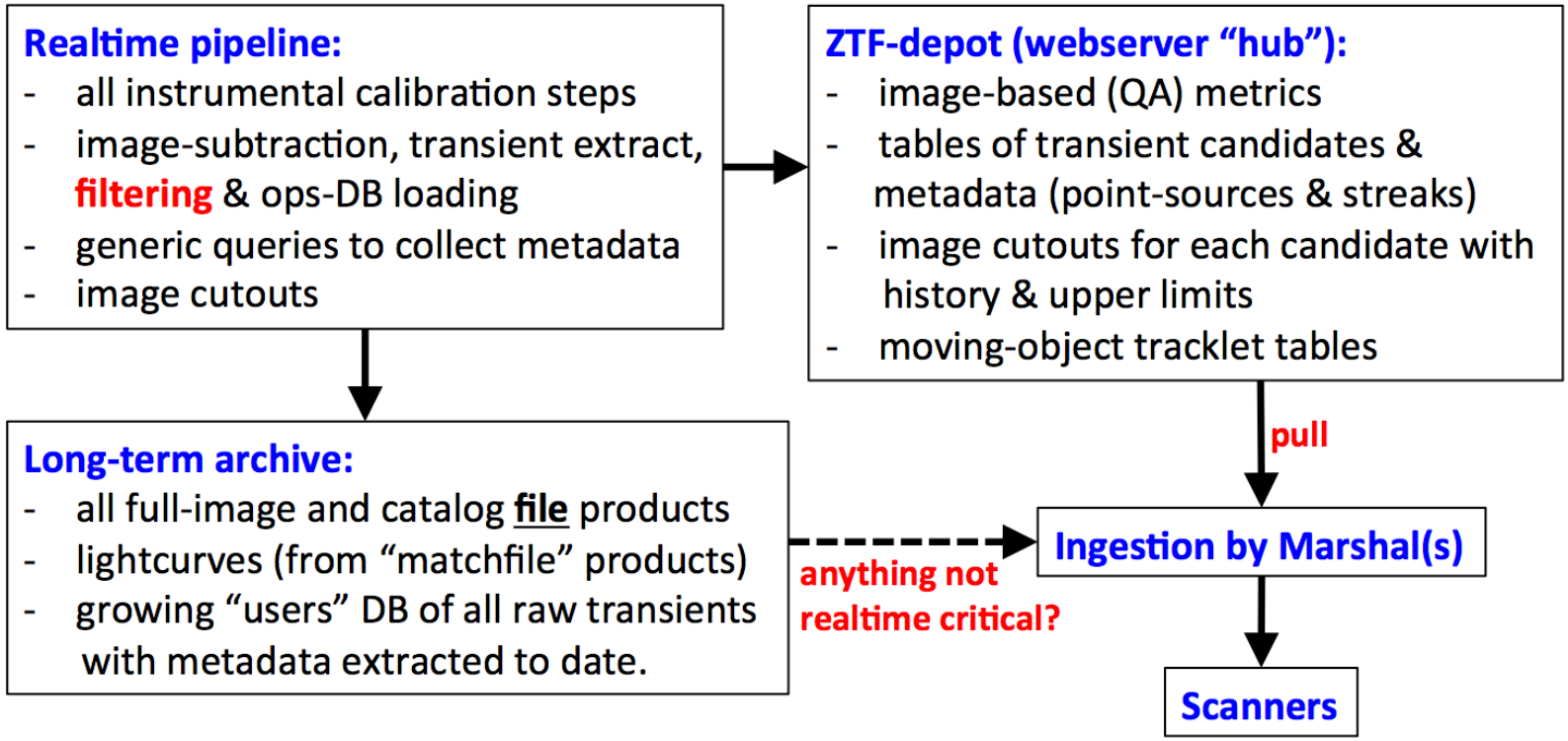
7. **Products to support Solar System/NEO discovery and characterization:**
   - moving object tracks from linking point-source transients; known objects are tagged.
   - delivered to the IAU's Minor Planet Center following human vetting.

# Development Status (high level)



# Archive and Depot development

- To support "fast response science": plan is to deliver a generic event stream (following any automated RB vetting or light filtering in pipeline) to a webserver for collection by all marshals.
- Other (historical) products, including all extracted events can be retrieved from growing archive.



# Remaining activities & milestones

## Activities

- Archive and user-interface development
- ZTF-depot details, with user-access/control, finder chart service.
- Lightcurve DB (low-risk object-based/match-file model) and retrieval interface
- Moving-object pipeline: source linking and streaks
- VO-alert infrastructure and understand interfaces
- P48-IPAC data transfer protocols and software.
- Reprocessing and preparation for PTF DR4 (soft release to collab: *public* – 4 months)
- Generate a 30-day long simulation as stress testing and enable DB performance/sizing decisions.
- Pipeline profiling / testing / tuning: hardware decisions

## Upcoming Milestones

- **Early May: transfer testing**
  - ✓ Send simulated camera data from P48 to IPAC
  - ✓ Compression tuning / bandwidth testing
  - ✓ Exercise pipeline subsystems
- **Early Aug: processing system v1.0:**
  - ✓ Data Transfer/ingest from P48 to IPAC
  - ✓ ZTF-Depot providing initial content to Marshals
  - ✓ Source linking for asteroids and streak ID/vetting
- **Mid August: First Light**
  - ✓ Static-calibrations; pipeline tuning
  - ✓ Instrumental characterization; new pipeline s/w
  - ✓ Data product access method is TBD
- **Mid September: start of science verification**
  - ✓ Performance monitoring / tuning
  - ✓ Exercise/refine operational daily routines
  - ✓ Archive and user-interfaces in place
- **Late 2017: start of science survey**
  - ✓ public alerts once better defined (TBD)

# Delayed First Light (initially Jan 2017)

- A 7.5 month delay has allowed us to streamline a more efficient data processing system
  - address unexpected challenges in requested functionality (see below)
  - much prototyping and R&D involved before committing to production implementation

- Labor to implement a data system to handle a bigger/complex survey was, in our opinion underestimated. “Devil is in the details”.

- Functionality that has presented some challenges (some still in design phase):
  - ZTF-depot details; DB sizing to handle I/O loads; user restrictions..
  - Image-subtraction pipeline details
  - Archive light-curve DB design decisions and infrastructure: match-file model redeisgned
  - Asteroids: streak finding/vetting pipeline (not MSIP)
  - Asteroids: point-source transient linking and MPC interface (not MSIP)
  - Public VO alerts: advanced to start of science operations
  - Synthetic transient-injection pipeline (a late addition, but recognize importance)

- Ongoing support for PTF:
  - Supported DR3 with Lightcurve DB release: underestimated work effort for LC DB
  - H-alpha code updates and reprocessing
  - Other ad-hoc reprocessing requests
  - Bulk reprocessing now in progress to support PTF DR4 – MSIP deliverable

# Staffing plan (roles): Development vs. Operations

*Staffing to accomplish the Baseline Scope. Additional scope (e.g., advanced light-curve database) would require additional staffing resources beyond these.*

**Development (5.75 FTE)**

- Task Leadership (0.5)
- Pipeline Development (2.7)
- Database Administration (0.2)
- Archive Development (1.5)
- Datacenter Operations (0.5)
- Simulation and Analysis (0.15)
- PTF/iPTF/ZTF Operations (0.2)

During "Steady-state operations", starting in ZY5, a further reduction in staffing to ~3.15 FTE should be possible.

**First-Year Operations (4.45 FTE)**

- Task Leadership (0.5)
- Pipeline Operations (0.5)
- Pipeline Maintenance (0.9)
- Pipeline Additional Development (0.6)
- Database Administration (0.3)
- Archive Ingest Operations (1.0)
- Datacenter Operations (0.5)
- Analysis and Performance Monitoring (0.15)

# Plan for ZTF years 3 – 6

| | 2017 | 2018 | 2019 | 2020 | | |
|---|---|---|---|---|---|---|
| | **ZY3** | **ZY4** | **ZY5** | **ZY6** | | **ZY3–6** |
| **Labor (WY)** | 5.54 | 4.45 | 3.15 | 3.15 | | |
| Labor | $ 921,126 | $ 759,254 | $ 557,738 | $ 557,738 | | |
| Procurement & Travel | $ 482,626 | $ 370,876 | $ 344,673 | $ 43,102 | | |
| **Total** | $ 1,403,751 | $ 1,130,130 | $ 902,412 | $ 600,841 | | **$ 4,037,134** |

ZY = FY shifted earlier by one month; ZY3 ~ FY17.
This assumes a full year of operations in ZY6.