

# ZTF Science Data System update

Frank Masci

Board meeting, December 11, 2017

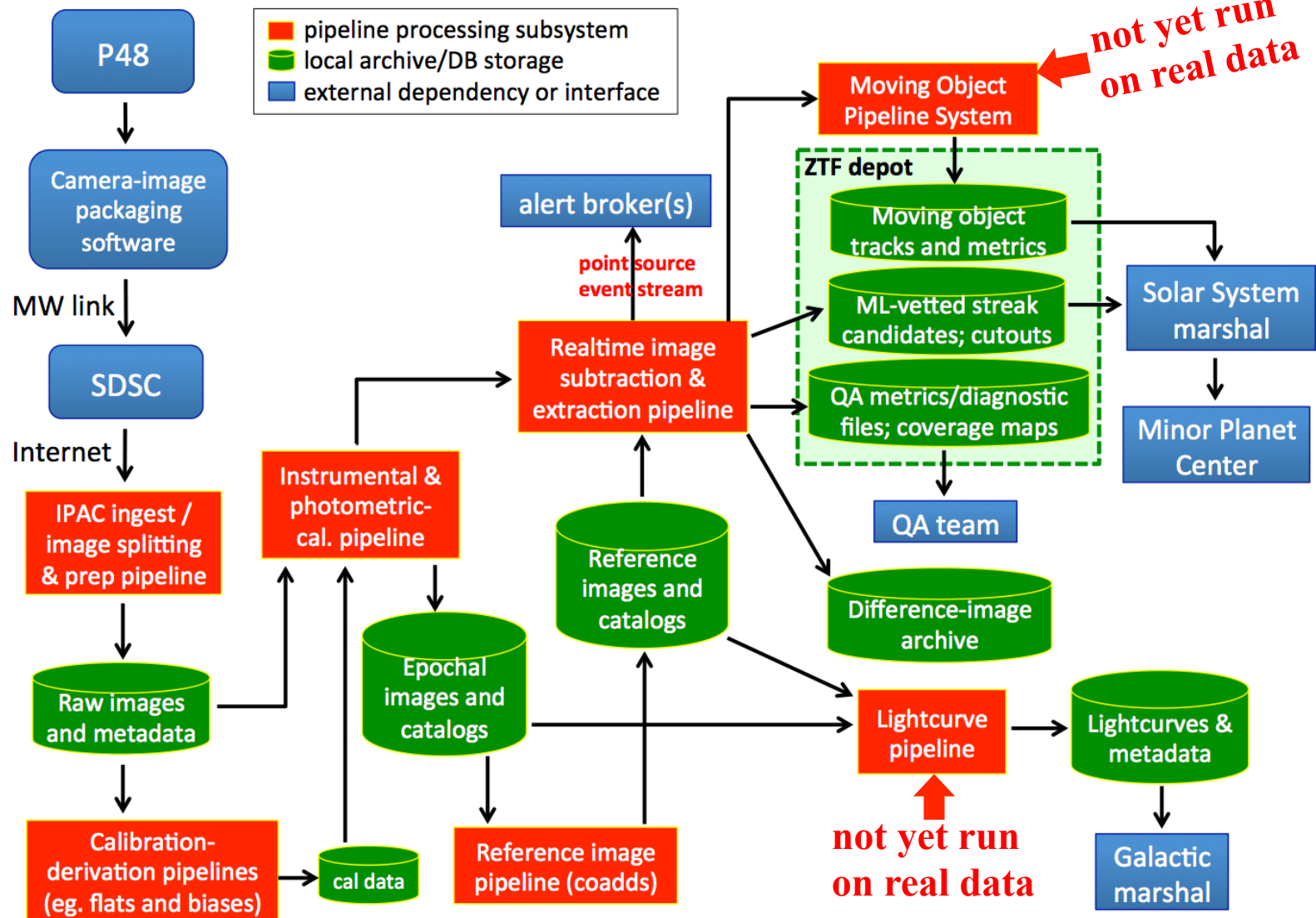


# The ZTF Science Data System (ZSDS)

---

- Housed at the Infrared Processing and Analysis Center (IPAC), Caltech.
- Responsibilities:
  - ingestion and archiving of all raw data acquired (both engineering & science)
  - science-data processing pipelines
  - long-term archiving of data products, curation, user-interfaces, and APIs for data retrieval
  - managing user-access according to data-access policies
  - near-realtime generation of flux-transient alerts and metadata
  - near-realtime generation of products to enable NEA discovery (streaks & tracklets)
  - generation and reporting of data quality metrics, survey statistics, coverage maps and diagnostics on camera/observing system
  - maintenance of pipelines, operations, databases, file servers, and archive infrastructure
  - documentation and user support

# Overall data flow & distribution



# Accomplishments

---

- Camera / data-system interface with all required metadata for pipeline processing finalized
- All core pipelines now operational and being run daily on commissioning / Science Validation data
- Products are being archived and accessed by project team members for analysis
- Archive access with user-authentication and filter/search/image-cutout capability using API service
- Instrumental calibration pipeline has been tuned (astrometric and photometric calibration steps)
- Reference (co-addition) pipeline now being exercised on calibrated epochal image products
- Tuning is adapted to current camera performance and feedback from team analyses
- Alert-packets now generated from image-differencing pipeline for internal distribution
- Streak detections (from fast-moving solar-system objects) along with metadata now generated
- Tuning continues for image-differencing and event-extraction pipeline
- Two additional pipelines that critically depend on quality of outputs from upstream pipelines have not yet been exercised on real data:
  - (i) lightcurve (source-matching) pipeline from epochal (un-differenced) extractions
  - (ii) moving-object tracklet pipeline from linking difference-image extractions

# Current development activities (from initial Data-System task plan)

---

## **Pipeline/processing-related:**

- Ongoing tuning of image-differencing pipeline
- Continuing to generating test reference images to feed image-differencing pipeline
- Continued analysis of photometry with feedback to pipeline parameter tuning
- Alert-naming infrastructure: assignment / reuse of object names for events in alert packets
- Alert-packed distribution infrastructure using Kafka interfacing with UW (completion date TBD)
- Configuring an additional set of 32 cluster nodes
- Streamlining “virtual pipeline operator” (VPO): automation of daily maintenance tasks
- Documentation; includes documenting performance of current system
- In queue: exercise and tune moving-object (tracklet-generation) and lightcurve pipelines

**Estimated completion for the above: Jan 15, 2018**

## **Archive-related, data-access tool related:**

- Refinements to data-access GUI (more filters, catalog overlays, etc.); completion: **Jan 20, 2018**
- Integration of Moving Object Search Tool and APIs for known solar system objects: **Jan 20, 2018**
- Finderchart service (customized for ZTF): **ready date TBD**
- Lightcurve retrieval GUI: **Mar 30, 2018**

# Deliverables expected from team and items requiring further analysis

---

## **Remaining deliverables for integration and testing in pipeline (drop in, no new software needed):**

- Trained ML-classification (Real-Bogus) model for point-source events
- Trained ML-classification model for streaks ('fast' moving objects)
- PS1 star/galaxy classification scores to associate with realtime event stream
- Flat-field optimization: acquisition strategy (LED config./colors); tuning application to science data

**All of the above: need no later than Jan 10, 2018**

## **Identified relatively recently from analyses – will require new pipeline software if significant:**

- Ghost-map generation, ghost-prediction and masking
- Non-linearity correction
- Exposure-time correction map (flat-field augmentation)