Summary of Pipeline Data Flow in the ZSDS

F. Masci, 4 / 12 / 2016, v2.1

This document expands on the data flow for the pipelines described in Section 7 of the Software Definitions, Deliverables & Services document:
https://ipacwiki.ipac.caltech.edu/@api/deki/files/4547/=ztfsds_swreq_v1.3.pdf

Data units and definitions

- raw CCD-image file: raw compressed MEF file corresponding to one CCD image from camera with metadata; consists of eight FITS extensions storing four readout-channel images and another four for their corresponding bias strips. This file will be tagged for either calibration or science-survey purposes.

- readout-channel image: individual readout-channel FITS image, i.e., a quadrant of the CCD.

- readout-channel bias image: individual readout-channel bias strip corresponding to readout channel image above.

- ensemble-based: a pipeline that operates on a collection of preprocessed readout-channel images from the *same quadrant* of a CCD at once.

Below is an overview of the execution frequency and inputs/outputs of the core ZSDS pipelines. **The numbers beside each processing step are labeled in Figure 1.**

(1) **Ingest & preparation** (Section 7.2): *realtime and continuous*; receive raw data from SDSC, either calibration or science-survey image data with metadata. Delivered to inbox.
Inputs: read raw CCD-image file [and manifest summary file] from inbox.
Outputs: raw CCD-image file copied to raw archive with metadata stored in ops DB.

(2) **Raw-image splitting** (Section 7.2): *realtime and continuous*; read raw CCD-image file from raw archive, decompress and split into eight FITS extensions.
Inputs: raw CCD-image file from raw archive (can be either calibration or survey based).
Outputs: eight FITS images: four readout-channel images and four corresponding readout-channel bias images. These files are copied/distributed across sandbox file-systems. Readout-channel and bias image pairs should have filenames to enable simple association. Simple raw-image QA metrics are computed and stored in ops DB.

(3) **Super-bias derivation** (Section 7.3): *late afternoon/early evening*; **ensemble-based**
Inputs: readout-channel images and accompanying bias images output from (2) corresponding to the *same* CCD quadrant intended for this calibration; static bad-pixel mask for that quadrant.
Outputs: super-bias correction image for this CCD quadrant; stack standard-deviation image; bad-pixel mask; simple QA. All stored in calibration archive with metadata in ops DB for later retrieval that night.

(4) **High-ν flat or pixel-to-pixel responsivity map** (Section 7.4): *late afternoon/early evening*; **ensemble-based**
Inputs: readout-channel images and accompanying bias images output from (2) corresponding to the *same* CCD quadrant intended for this calibration; super-bias correction image and accompanying bad-pixel mask for that quadrant output from (3)
Outputs: high-ν flat-field correction image for this CCD quadrant with accompanying uncertainty image; simple QA. All stored in calibration archive with metadata in ops DB for later retrieval that night.

(5) **Low-ν flat from either dithered-star observations or long-term ZPVM** (Section 7.5): *weekly, monthly or longer* (TBD); **ensemble-based**
Inputs: processed readout-channel images from (6) corresponding to the *same* CCD quadrant from on-sky survey; original low-ν flat used to process these images in (6).
Outputs: low-ν flat-field correction image for this CCD quadrant with simple QA. All stored in calibration archive with metadata in ops DB for later retrieval.

(6) **Instrumental and photometric calibration** (Section 7.6): *realtime and continuous*
Inputs: on-sky survey readout-channel image and accompanying bias image output from (2); super-bias calibration product from (3); high-ν flat-field product from (4); low-ν flat-field product from (5).
Outputs: calibrated science image; accompanying mask; two catalog files (aperture & PSF-fitting); simple QA metrics and metadata in ops DB.

(7) **Reference image generation/co-addition** (Section 7.7): *weekly, biweekly or longer, contingent on survey plan and data availability*; **ensemble-based**
Inputs: processed readout-channel images from (6) corresponding to the *same* CCD quadrant from survey that fall in a pre-defined tile of the sky grid
Outputs: co-add products: primary reference image; depth-map; uncertainty image; PSF-file; two catalog files (aper. & PSF-fitting); simple QA metrics and metadata in ops DB.

(8) **Realtime transient discovery** (Section 7.8): *realtime and continuous; immediate follow-on from pipeline (6) above.*
Inputs: processed on-sky survey readout-channel image from (6); reference image products from (7).
Outputs: positive and negative difference image, uncertainty image, log file, PSF file, QA metrics file; catalogs of transient candidates containing source features, file with RealBogus scores; all to be loaded into Transients DB; streak detection files also with RealBogus scores for loading into ops DB; copy select products to public web-server to support retrieval of image-cutouts on queried transients for near-realtime discovery.

(9) **Source-matching & lightcurve generation** (Section 7.9): *weekly, biweekly or longer* (TBD); **ensemble-based**
Inputs: epochal source catalogs from processed readout-channel images from (6) corresponding to the *same* CCD quadrant from survey that fall in a pre-defined tile of the sky grid; reference image source catalog from (7).
Outputs: three database load-files: (i) objects; (ii) association indices; (iii) epochs. These will accompany (independently) load-files generated from the epochal catalogs made from (6).

(10) **Moving object pipeline system** (MOPS; Section 7.10): *every three or four hours throughout night* (TBD); **ensemble-based**
Inputs: machine-learned vetted candidates from (8), read from Transients DB, spanning ~ last two nights (maximum three).
Outputs: moving-object tracklet candidates in ascii tables; ready for human vetting; final report manually submitted to MPC.

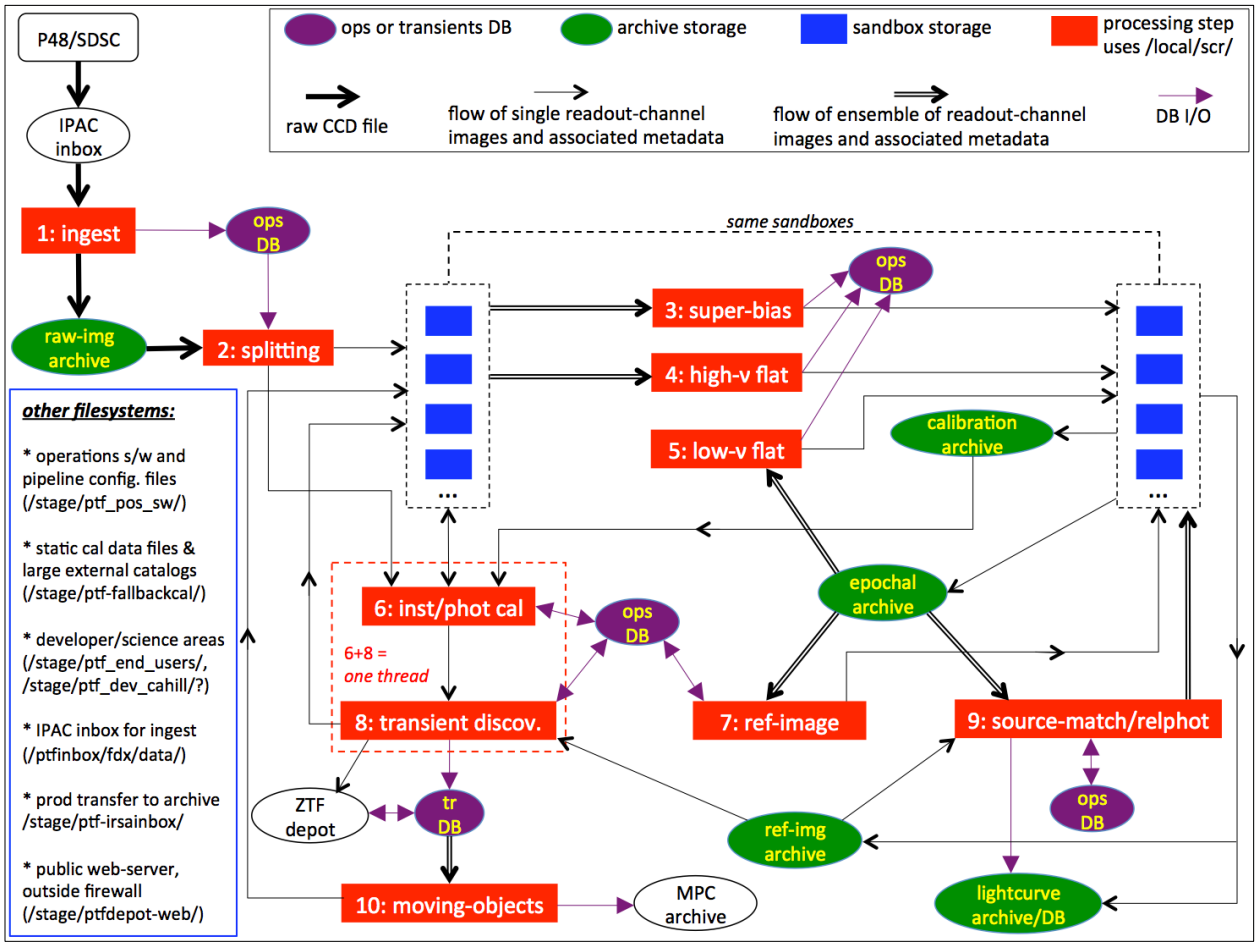

Figure 1: data flow in the ZSDS; red boxes with numbers are expanded above.