ZTF pipeline products and *depot* interface to support *real-time* discovery

F. Masci, 03 / 01 / 2017, v1.4

1. Introduction and scope

This document describes an interface for delivering transient candidates (or events) with contextual information to the various ZTF marshals as generated by the realtime pipeline.

The access portal for *realtime* products will probably be a webserver. This will be referred to as *ztf-depot*. We outline the specific products, formats, and directory layout on ztf-depot. The delivered event stream will be generic in the sense that it will need to support all science use cases. Therefore, consider ztf-depot as a distribution hub for all *discovery-based* science, i.e., science applications requiring a fast-response and decisions for follow-up. This should not be confused with the long-term archive hosted by IRSA, where products from a night's processing will typically be available later that day.

It is envisaged that the individual marshals will pull this data and apply additional filtering according to their science needs prior to scanning.

2. Overview of process

Every **TBD** minutes throughout the night, a cron-job at IPAC will:

1. Query the pipeline operations database for transient candidates using *generic* queries that will *only* apply light filtering to weed out obvious false-positives. This filtering provides a means to control the data load and latency of deliverables. Various table-joins will be performed to return metadata for each transient. The realtime pipeline will also perform some basic cross-matching to static catalog tables (e.g., the "Stars" and "LU" tables that store Star/Galaxy probability scores and metadata for local galaxies respectively).
2. For each target candidate transient, image cutouts will be generated from the science exposure, reference, and difference images. For the science and difference images, a history of cutouts centered on the target transient position going back 3 (**TBD**) days will also be generated, regardless of any prior detected candidate at the target position. I.e., these are "forced" cutouts. Flux upper limits (and any prior detections on/near the target position) will accompany the historical cutouts. See Section 3 on how you would access longer histories if needed.
3. This functionality will also apply to streak candidates to support Solar System science.
4. QA metrics at the image level (calibration, science, and difference image products), and aggregate statistics on the number of exposures received/processed as known at query-time are also generated.
5. The database metadata for each transient (with any detection history) are written to ASCII tables. Image cutouts from the science, reference, and difference images will be delivered as JPEGs, with filenames indexed by image and candidate IDs. These products are copied to specific directories on ztf-depot (see Figure 2).

6. Depending on the *programID* from scheduling, file products on ztf-depot will be separated into three categories (subdirectories in Figure 2): **public/** (MSIP deliverables), **collab/** (collaboration), or **caltech/** (Caltech observing time). Individual users will be assigned an account to ztf-depot, with access privileges to one or more of these categories. A user-to-category map table will be used to assign privileges. Here are the access rules for each category and any prior historical observations:
    - Target transients or events discovered/observed in **public** exposures can only access associated historical information that was also tagged as **public** and none other.
    - Target transients or events discovered/observed in **collab** exposures can only access associated historical information that was observed under either **collab** or **public**.
    - Target transients or events discovered/observed in **caltech** exposures can only access associated historical information that was observed under either **caltech** or **public**.
7. If a user identifies an interesting candidate for spectroscopic follow-up, they will need to generate a *finder-chart* on demand. This is a 1000k x 1000k pixel cutout (in JPEG or PNG format) centered on the target position from the science image exposure. This will *need to be generated on demand in realtime*. Given the size of this image and number of potential candidates, the realtime pipeline will not be able to handle the load. The plan is to query the IRSA archive and generate a finder-chart on-the-fly. This requires the science image exposure to be archived and immediately available *close to realtime*.

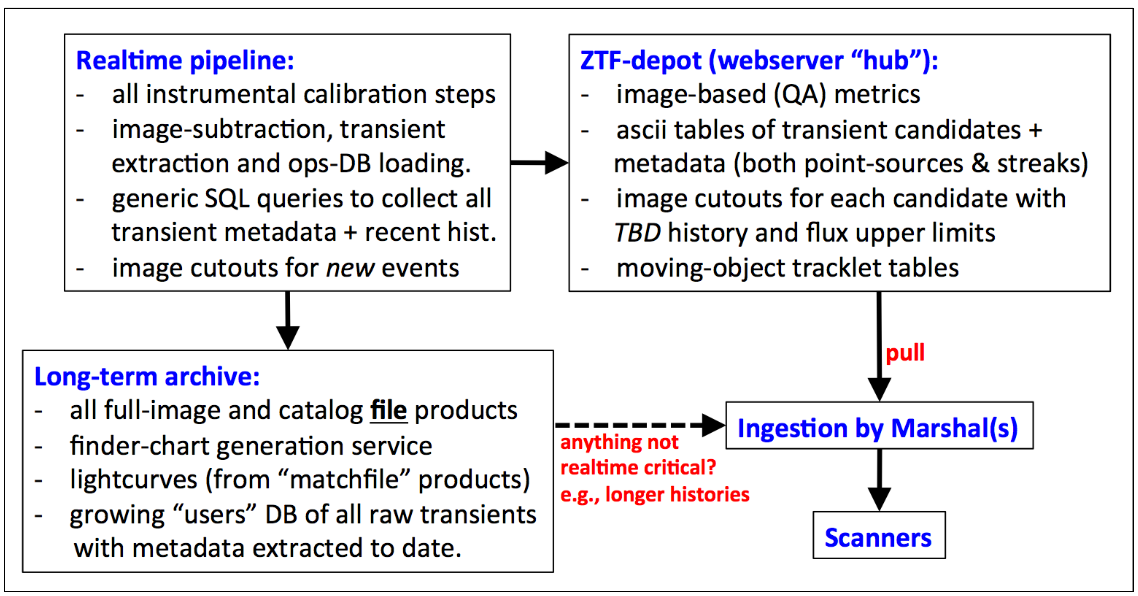

Figure 1: data-flow to support discovery and archival (longer-term) archival science

3. Restrictions and general archive access

The ZTF Pipeline Operations Database will be off-limits to everyone during routine operations. The intent is to insulate the realtime pipeline (which includes the database) from additional external user-queries. The latest contents of the operations database will be copied to a separate "users" database (on a different server) following completion of nightly processing. This users database can be accessed by anyone in the collaboration and will contain the full history of all ZTF data, transient candidates, and QA metrics. Its intent is to support archival studies as well as ad-hoc queries to tune the realtime event stream delivered to ztf-depot.

Coupled with the users database, image cutouts on transient candidates for longer histories than those available on ztf-depot will be accessible from IRSA. There will be an image-cutout service to enable cutouts on the archived images (either science, reference, or subtraction images).

4. Directory structure and files on ztf-depot

Figure 2 shows a provisional layout of the products. Directories and filenames with angled brackets denote variable quantities.

Key for Figure 2:

adhoc/ – directory for storing additional non-realtime-critical products requested by users, i.e., to support external analyses. Such products may not be available in the archive.

readme_v<X>.txt – description of all files and columns/metadata therein for products under <YYYYMMDD>/. The v<X> is a version number pertaining to the depot design.

<YYYYMMDD>/ – date of night being processed: based on UT. At most, two date directories will be retained (previous and current). Not to be confused with the product history for scanners. Intent is to enable recovery from server and/or client-side hardware outages.

MOPSprods/ – directory for products from moving-object pipeline system (object tracks from linking point source transient candidates).

CovMaps/ – all-sky exposure depth-of-coverage maps following *end-of-night* processing for (i) only that night that was processed; (ii) cumulative over the course of the survey up to and including that night. These are generated for different filters in *equ* and *gal* coordinate systems in a Hammer-Aitoff projection. Format is both FITS and PNG.

CalQA/ – directory containing QA metadata files for image-based calibration files, i.e., for bias and dome-screen flat frames after splitting into quadrants and actual calibration (ensemble) products made from products acquired on <YYYYMMDD>/.

AggregateStats.txt – file summarizing number of images observed versus processed etc. as available at query time (See Section 6.1).

<ObsTimeRange>/ – directory name for realtime *science products* corresponding to query interval; one choice is the UT-based exposure-observation time range falling in <YYYYMMDD>, rounded to nearest second. E.g., 21h33m50s-21h43m50s/, 21h43m51s-21h53m53s/ …

collab/, public/, caltech/ – directories separating survey modes; also to facilitate specific-user access.

ccd<ID>/ – subdirectory containing *discovery-based* products for a specific CCD ID [01..16]. By discovery-based, we mean those images from which the *target* transient candidates were initially extracted (see below). A breakdown by CCD ID will keep the number of product files per directory at a manageable level (a factor of 16 less).

<rootfname> – unique root filename of chip-quadrant (processing unit in pipeline), e.g., *ztf_2016032039853_000630_sr_c01_o_q2*, where fields are exposure date-time string; field ID [~ 000001 – 002000]; filter name; CCD ID [01..16]; quadrant ID therein [1..4].

<rootfname>_qa.txt – chip-quadrant-image based QA metrics: raw images, calibration products (under CalQA/ above), instrumentally calibrated science and subtraction images (under specific <ObsTimeRange>/).

<rootfname>_cands.txt – IDs and metrics for the *target* point source transient candidates. Includes distance to nearest *known* Solar System object, its name, and associated metadata.

<rootfname>_streaks.txt – IDs and metrics for streak-transient candidates.

candid<ID>_pid<ID>_targ_{sci, scimref}.jpg – JPEG image-cutouts for *target* point source candidate with <ID> discovered in image with product ID (pid) <ID>. Metadata and ID information is stored in <rootfname>_cands.txt. All point source cutouts are 51 × 51 pix^2 in size.

candid<ID>_pid<ID>_hist_{sci, scimref}.jpg – for a given *target* candid<ID>, these are "forced" cutouts on any available historical image exposures, i.e., going back **Ndhist** days from the target transient discovery epoch (see Section 5.2). Only the pid<ID> will differ between these historical cutouts while candid<ID> remains the same.

candid<ID>_ref.jpg – JPEG image-cutouts from reference image used to discover the *target* transient candidate with candid<ID>. **Note:** due to the two overlapping sky-tiling grids, the historical *scimref* cutouts may have been generated using reference images pertaining to a different CCD ID.

candid<ID>_history.txt – summary file containing historical metadata for target candid<ID> going back **Ndhistcand** (≤ 30) days; specifically: (i) *any* spatially coincident historical detections on/around the target position (see Section 5.3); and (ii) difference image flux upper limits where there are no spatially coincident detections.

strkid<ID>_pid<ID>_scimref.jpg – JPEG image-cutout for streak candidate with <ID> and other information stored in <rootname>_streaks.txt. All streak cutouts are 65 × 65 pix^2 in size.

Example outputs using simulated data and following the directory structure in Figure 2 are in: *http://web.ipac.caltech.edu/staff/fmasci/home/junk/ztf-depot/*

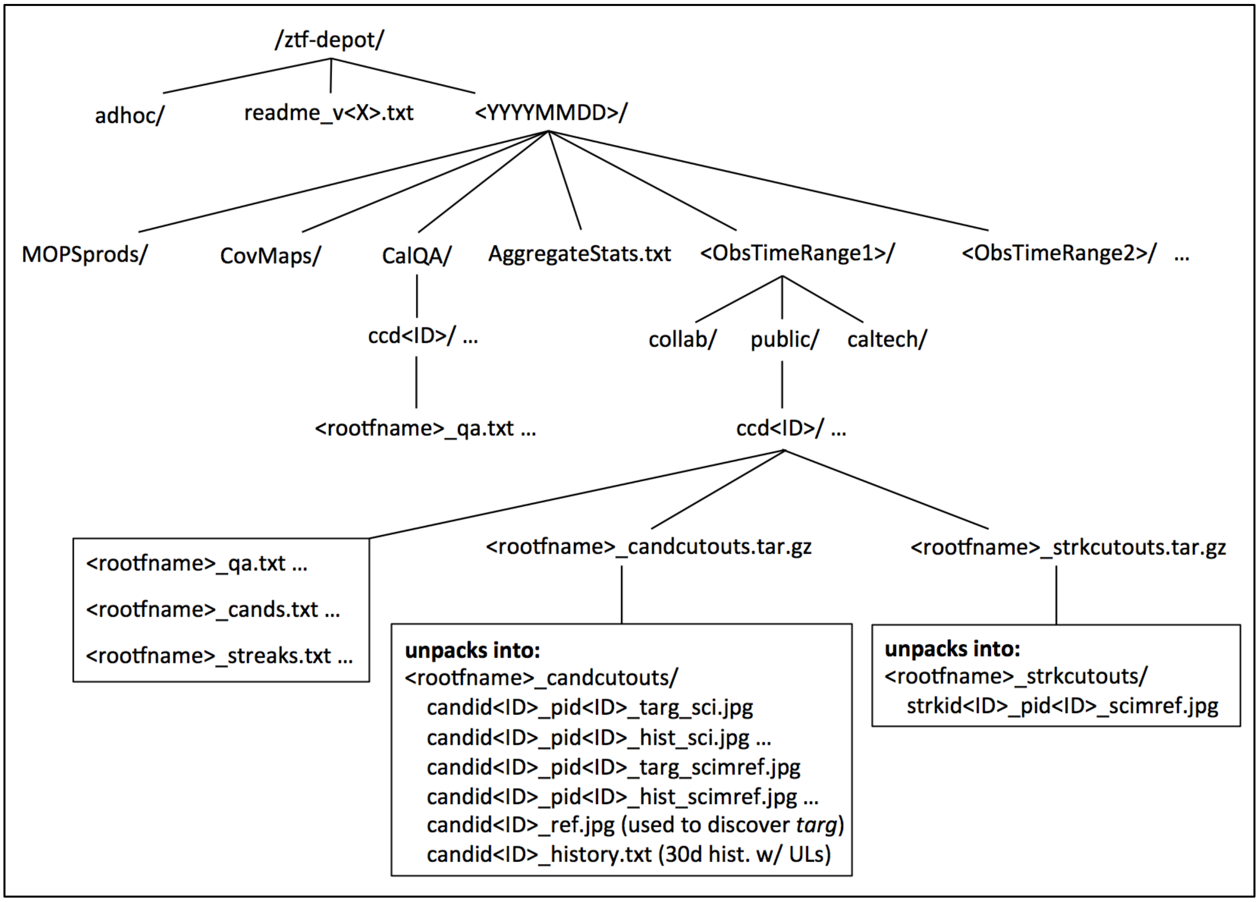

Figure 2: proposed directory layout and files for ztf-depot

5. Generic SQL queries

5.1. Initial Query for point-source transient candidates

This will be the first query executed for some input observation time interval.

The output from this query will be written to **<rootfname>_cands.txt**. See Figure 2 and the file definitions above that figure.

<u>Input parameters:</u>
rb_cut – *RealBogus* quality cut. It is not yet known if machine-learned vetting (*RealBogus*) will be implemented for *point source* transient candidates. The author has identified several alternative candidate metrics for retaining *reliable* candidates, i.e., as a proxy for *rb_cut*.

jd_start, jd_end – start/end observation JDs, i.e., range of image exposure JDs corresponding to the specific query interval as encoded in the directory name in Figure 2: **<ObsTimeRange>/**.

```
e.g., jd_start=2457743.9690356; jd_end=2457743.97598; rb_cut=0:

select s.jd, s.fid, s.pid, s.diffmaglim, s.pdiffimfilename, i.programpi,
       i.programid, c.*, st.strcat, st.mag as sgmag, st.score as sgscore, l.*
from subtractions s, immeta i, candidates c
     left outer join candmatches cm on c.candid = cm.candid
     left outer join stars st on cm.strid = st.strid
     left outer join lu l on cm.luid = l.luid
where c.pid = s.pid and
      i.pid = s.pid and
      c.isdiffpos = 't' and
      c.rb >= 0 and
      s.jd > 2457743.9690356 and
      s.jd <= 2457743.97598
order by s.jd, c.magap asc
```

5.2. Query image filenames to enable *forced* image-cutouts

For each target transient candidate returned by the query in 5.1, a query is submitted to find all the overlapping images going back **Ndhist** days, including images containing the input targets. This is regardless if a prior record (detection) in the candidates table existed at/near the input target position. Filenames corresponding to the positive difference images are returned. Other image types (e.g., for the science and reference images) have the same root name but different suffixes.

The realtime image-subtraction pipeline already generates science, reference, and difference image-cutouts for all extracted *target* transient candidates. These are the same candidates (with candid<ID>) that are returned by the query in Section 5.1. One needs to collect these files and rename them to **candid<ID>_pid<ID>_targ_sci.jpg, candid<ID>_pid<ID>_targ_scimref.jpg,** and **candid<ID>_ref.jpg** as per Figure 2.

Note: the forced (historical) image-cutouts are not pre-generated by the image-subtraction pipeline. A separate script is needed to generate these using results returned by the query below (if any are returned). See Figure 2 for the naming format of these historical cutouts.

Input parameters:
ra_inp, dec_inp, jd_inp – position and obs JD of target transient returned by the query in 5.1.

Ndhist – go back this number of days when searching for available overlapping images.

```
e.g., Ndhist=3d; jd_inp=2457743.97598; ra_inp=106.057199; dec_inp=11.8764987:

select s.jd, s.pid, s.fid, s.diffmaglim, s.pdiffimfilename, i.programpi,
       i.programid
from subtractions s, immeta i where
q3c_poly_query(cast(106.057199 as double precision),
               cast(11.8764987 as double precision),
               array[s.ra1,s.dec1,s.ra2,s.dec2,s.ra3,s.dec3,s.ra4,s.dec4])
and i.pid = s.pid and
    s.jd <= 2457743.97598 and
    (2457743.97598 - s.jd) <= 3
order by s.jd
```

5.3. Search for *any* historical spatially-coincident candidates or upper limits otherwise

For each target transient candidate returned by the query in 5.1, another query is executed to search the candidates table for any historical spatially-coincident candidates on/near the target position. If found, all metrics for the historical candidates are returned. If no spatially-coincident candidates are found from an image epoch falling in the query timespan, difference-image based flux upper limits are returned.

The output from this query will be written to **cand<ID>_history.txt**, where cand<ID> refers to the input target transient. See Figure 2 and the file definitions above that figure.

Input parameters:
ra_inp, dec_inp, jd_inp – position and obs JD of target transient returned by the query in 5.1. These are the same inputs used for the historical image query in 5.2.

Ndhistcand – go back this number of days when searching the candidates table for any prior records.

rad_tol – radial tolerance [deg] to support spatial search.

e.g., `Ndhistcand=30d; jd_inp=2457746; ra_inp=62.0125062; dec_inp=16.0619949; rad_tol=0.0004167 deg (=1.5″):`

```
select a.jd, a.fId, a.pId, a.diffmaglim, a.pdiffimfilename,
       b.programpi, b.programId, c.*
from Subtractions a
INNER JOIN ImMeta b
ON a.pId = b.pId
LEFT OUTER JOIN Candidates c
ON a.pId = c.pId
and q3c_radial_query(cast(62.0125062 as double precision),
                     cast(16.0619949 as double precision),
                     c.ra,
                     c.dec,
                     cast(0.0004167 as real))
and isdiffpos = 't'
where q3c_poly_query(cast(62.0125062 as double precision),
                     cast(16.0619949 as double precision),
               array[a.ra1,a.dec1,a.ra2,a.dec2,a.ra3,a.dec3,a.ra4,a.dec4])
and a.jd < 2457746
and (2457746 - a.jd) <= cast(30 as double precision)
order by a.jd, magap
```

5.4. Query for streaking-object candidates

Query the "streaks" table for streaking objects. There are no additional queries to search for spatially-coincident historical candidates or to generate historical cutouts. Image cutouts are only generated on the detected streaking targets returned by the query below, but only using positive-difference image products.

The output from the query below will be written to **<rootfname>_streaks.txt.** The realtime image-subtraction pipeline already generates difference image cutouts for all extracted *streak*-transient candidates. One will need to collect these files and rename them to **strkid<ID>_pid<ID>_scimref.jpg**, as per Figure 2.

Input parameters:
rb_cut – *RealBogus* quality cut. It is not yet known if machine-learned vetting (*RealBogus*) will be implemented for *streaking* object candidates.

jd_start, jd_end – start/end observation JDs, i.e., range of image exposure JDs corresponding to the specific query interval as encoded in the directory name in Figure 2: **<ObsTimeRange>**/.

```
e.g., jd_start=2457743.9690356; jd_end=2457743.97598; rb_cut=0:

select s.jd, s.fid, s.pid, s.diffmaglim, s.pdiffimfilename, i.programpi,
       i.programid, str.*
from subtractions s, streaks str, immeta i
where str.pid = s.pid and
      i.pid = s.pid and
      str.rb >= 0 and
      s.jd > 2457743.9690356 and
      s.jd <= 2457743.97598
order by s.jd
```

6. Image-based QA metrics and aggregated statistics

The specifications below pertain to product files named as **AggregateStats.txt** and **<rootfname>_qa.txt** in Figure 2. The former stores summary (aggregated) statistics throughput the night, as known at query time when populating the latest **<ObsTimeRange>/** directory, while the latter stores specific image-based statistics.

6.1. Aggregation Statistics

Statistics are accumulated throughout a night (starting at noon PT) following each incremental database query. The very last query of the night will therefore include summary statistics for the entire night.

For example:

```
- number of raw CCD image files ingested and archived at IPAC (not unpacked);

- number of raw CCD images successfully uncompressed and split into their
  readout channels (the basic processing unit).

- number of science images "successfully" instrumentally calibrated
```

(or labelled as "usable") according to their "status" flag, defined via specific conditions in their "infobits" (see below).

- number of "positive" (sci-ref) subtraction images tagged with "statusdif=1" (=> transient candidates *were* extracted therefrom and loaded into DB).

- number of "positive" (sci-ref) subtraction images tagged with "statusdif=0" (=> *no* transient candidates extracted therefrom and loaded into DB).

- number of machine-learned vetted transient candidates extracted from all "positive" (sci-ref) difference images and loaded into DB.

- number of machine-learned vetted transient candidates extracted from all "negative" (ref-sci) difference images and loaded into DB.

- number of unique solar-system objects associated with transient extractions.

6.2. Image-based QA metrics and processing status flags

- The metrics below will be collected from five pipeline subsystems. These will be queried from the pipeline operations DB every TBD minutes.
  (i) ingest
  (ii) splitting
  (iii) calibration generation (e.g., bias and flat products)
  (iv) instrumental calibration
  (v) image subtraction and transient extraction

- All image-based metrics (unless otherwise stated) are readout-channel based.

- Only those metrics derived from end-to-end processing of specific image types (value of "imgtype" below) will be written to their specific **<rootfname>_qa.txt** files. At the time of writing, metrics for five imgtypes will be reported: science (or "o"); raw bias frames ("b"); raw flatfield frames ("f"); bias calibration products ("bias"); and flatfield calibration products ("flat").

- According to the directory structure in Figure 2, image types labeled as imgtype="b", "f", "bias", or "flat" will have their **<rootfname>_qa.txt** file reside under CalQA/ccd<ID>/…, while imgtype="o" (science), these will reside under the specific-query interval: <ObsTimeRange>/<surveymode>/ccd<ID>/…

Generic, for all image types: imgtype = "o", "b", "f", "bias", or "flat":

imgtype -- either science, bias (raw), flat (raw), bias (product), flat (product)
field -- field ID
ccdid -- CCD ID: 1..16
fid -- filter ID: 0..3 (0=no filter, i.e., for bias frames; 1=g; 2=r; 3=i)
rcid -- readout-channel ID in field: 0..63
qid -- readout-channel ID in CCD: 1..4
nid -- night ID counter
nightdate -- date corresponding to nid

For image types: imgtype = "o", "b", "f" (following quadrant-splitting):

obsdatetime -- corresponding to camera exposure
obsjd -- JD corresponding to obsdatetime
createddateraw -- date/time when raw CCD image file was archived at IPAC
xfertime -- difference: "createddateraw - obsdatetime" [minutes]
overscanmed -- median of overscan pixels [DN]
overscanrms -- RMS from average of overscan pixels [DN]
overscanpct -- Data scale using percentiles [DN]
rawgmed -- global median of raw CCD pixels [DN]
rawgrms -- global RMS from average of raw CCD pixels [DN]
rawgpct -- global data scale of raw CCD pixels using percentiles [DN]
fbresidmed -- median of fit residuals for floating-bias correction [DN]
fbresidrms -- RMS of fit residuals for floating-bias correction [DN]

For calibration (ensemble-based) image products: imgtype = "bias" or "flat":

nframescal -- number of images used to generate calibration product
nbadpixcal -- number of bad pixels

```
gmediancal -- global median of pixel values
gstddevcal -- global stddev of pixel values
gpctdiffcal -- global robust measure of pixel RMS based on percentiles
medstackpixunccal -- median of pixel-stack stddev/sqrt(nframes) over all pixels
stddevmedcolcal -- stddev of median-collapsed pixel columns (~ x-axis variation)
stddevmedcolcal -- stddev of median-collapsed pixel rows (~ y-axis variation)
```

For imgtype = "o" only (following instrumental calibration and image-differencing):

```
Instrumental calibration based
anmatches -- total #external-cat matches for computing astrometric cal metrics
anmatches11 -- #external-cat matches in grid partition 1,1
anmatches12 -- #external-cat matches in grid partition 1,2
anmatches13 -- #external-cat matches in grid partition 1,3
anmatches21 -- #external-cat matches in grid partition 2,1
anmatches22 -- #external-cat matches in grid partition 2,2
anmatches23 -- #external-cat matches in grid partition 2,3
anmatches31 -- #external-cat matches in grid partition 3,1
anmatches32 -- #external-cat matches in grid partition 3,2
anmatches33 -- #external-cat matches in grid partition 3,3
admed1 -- median of reconst - ref posns along RA (arcsec)
admed2 -- median of reconst - ref posns along Dec (arcsec)
admedrad -- median of reconst - ref radial sepns (arcsec)
adpctdif1 -- pctdif of reconst - ref posns along RA (arcsec)
adpctdif2 -- pctdif of reconst - ref posns along Dec (arcsec)
adminmed -- minimum local admedrad over grid partitions (arcsec)
admaxmed -- maximum local admedrad over grid partitions (arcsec)
arefmatchpct -- percentage of reference catalog matches
adetmatchpct -- percentage of detected catalog matches
andegref -- SCAMP pass2 NDeg_Reference (degrees of freedom in solution)
anstarsdet -- SCAMP pass2 NDetect (number of stars detected)
anstarsref -- SCAMP pass2 n_catalog (number of reference stars)
ascmp1sigma1 -- SCAMP pass1 sigma along axis 1 (arcsec)
ascmp1sigma2 -- SCAMP pass1 sigma along axis 2 (arcsec)
ascmp1chi2 -- SCAMP pass1 chi2
ascmp2sigma1 -- SCAMP pass2 sigma along axis 1 (arcsec)
ascmp2sigma2 -- SCAMP pass2 sigma along axis 2 (arcsec)
ascmp2chi2 -- SCAMP pass2 chi2
awmeanscale -- Mean percentage diff of final - prior scale
awminscale -- Min percentage diff of final - prior scale
awmaxscale -- Max percentage diff of final - prior scale
aboreshiftra -- Shift of boresight RA value relative to prior input [arcsec]
aboreshiftdec -- Shift of boresight Dec value relative to prior input [arcsec]
aimshiftra -- Shift of readout-channel image center RA value [arcsec]
aimshiftdec -- Shift of readout-channel image center Dec value [arcsec]
gmedian -- Global median of sci-image pixel values [DN]
gstddev -- Robust measure of global std-deviation of sci-image pix values [DN]
npsfcat -- Number of sources in PSF-fit catalog
minsnr -- Minimum source signal-to-noise ratio in PSF-fit catalog
maxsnr -- Maximum source signal-to-noise ratio in PSF-fit catalog
minmag -- Minimum (brightest) source magnitude in PSF-fit catalog [mag]
maxmag -- Maximum (faintest) source magnitude in PSF-fit catalog [mag]
npsfcatmlim -- Number of sources used for maglimcat and medchilosnr computations
medchitot -- Median chi-metric of all sources in PSF-fit catalog
medchilosnr -- Median chi-metric near 5-sigma magnitude limit
pnmatches -- Number of source matches to support photometric calibration
pabszp -- Absolute photometric calibration zero point [mag]
pabszpunc -- Uncertainty in absolute photometric calibration zero point [mag]
pcolterm -- Color coefficient for absolute photometric calibration
pabszprms -- RMS in ZP (mag-difference) fit residuals [mag]
maglimcat -- Magnitude limit of PSF-fit catalog based on photometric uncs [mag]
maglimit -- Magnitude limit of PSF-fit catalog based on semi-empirical formula
nsexcat -- Number of sources in SExtractor catalog
fwhm -- Median FWHM of sources in SExtractor catalog [pixels]
ellip -- Median ellipticity of sources in SExtractor catalog
peakdist -- Median sepn: "source peak - centroid" in SExtractor catalog [pix]
npixgood -- Number of good (unmasked) pixels
npixbad -- Total number of masked pixels
npixnoisy -- Number of noisy pixels
npixsat -- Number of saturated pixels
nnanpix -- Number of NaN'd pixels
gpctdif -- Spread in sci-image pixel values based on percentiles [DN]
infobits -- bit-string encoding conditions/anomalies from instrumental cal.
```

bitsinfobits -- comma-separated list of bits encoded in infobits
status -- overall science image quality flag according to specific infobits:
0 => processed readout-channel image is unusable, 1 otherwise

Difference-image based

fluxrat -- scale factor for gain matching of sci and ref image pixel values
scigain -- Effective electronic gain of rescaled sci image [e-/DN]
scisat -- Pixel saturation value of rescaled sci image [DN]
scibckgnd -- Robust estimate of background level in scaled science image [DN]
scisigpix -- Robust estimate of sigma/pixel in scaled science image [DN]
sciinpseeing -- Refined seeing FWHM of rescaled sci image [pixels]
refimfilename – Name of archived reference image used
refid – reference image database identifier
refsat -- Pixel saturation value of resampled ref image [DN]
refbckgnd -- Robust estimate of bckgnd level in resampled reference image [DN]
refsigpix -- Robust estimate of sigma/pixel in resampled reference image [DN]
refinpseeing -- Refined seeing FWHM of resampled ref image [pixels]
pdiffbckgnd -- median background level in final positive difference image [DN]
diffsigpix -- robust sigma/pixel in final difference images [DN]
diffpctbad -- percentage of difference image pixels tagged as bad/unusable [%]
diffmaglim -- approx. 5-sigma mag limit in final difference image based on
PSF-fit photometry [mag]
difffwhm -- Effective point-source FWHM in final difference images [pixels]
diffavgsqbef -- average of squared diff-image pixel values before PSF-matching
diffavgsqaft -- average of squared diff-image pixel values after PSF-matching
diffavgsqchg -- % change: '100 x (diffavgsqaft - diffavgsqbef)'/diffavgsqbef
ncandscimreffilt -- number of candidates extracted from sci minus ref image
ncandrefmscifilt -- number of candidates extracted from ref minus sci image
nsolarsystobj -- number of unique Solar System objects associated with
transient candidates
infobitsref -- image InfoBits string for input reference image
bitsinfobitsref -- comma-separated list of bits encoded in infobitsref
statusdif -- overall difference-image quality flag; 0 = bad/unusable,
1 = usable for transient extraction

7. Tables associated with Transient Discovery in the Operations Database

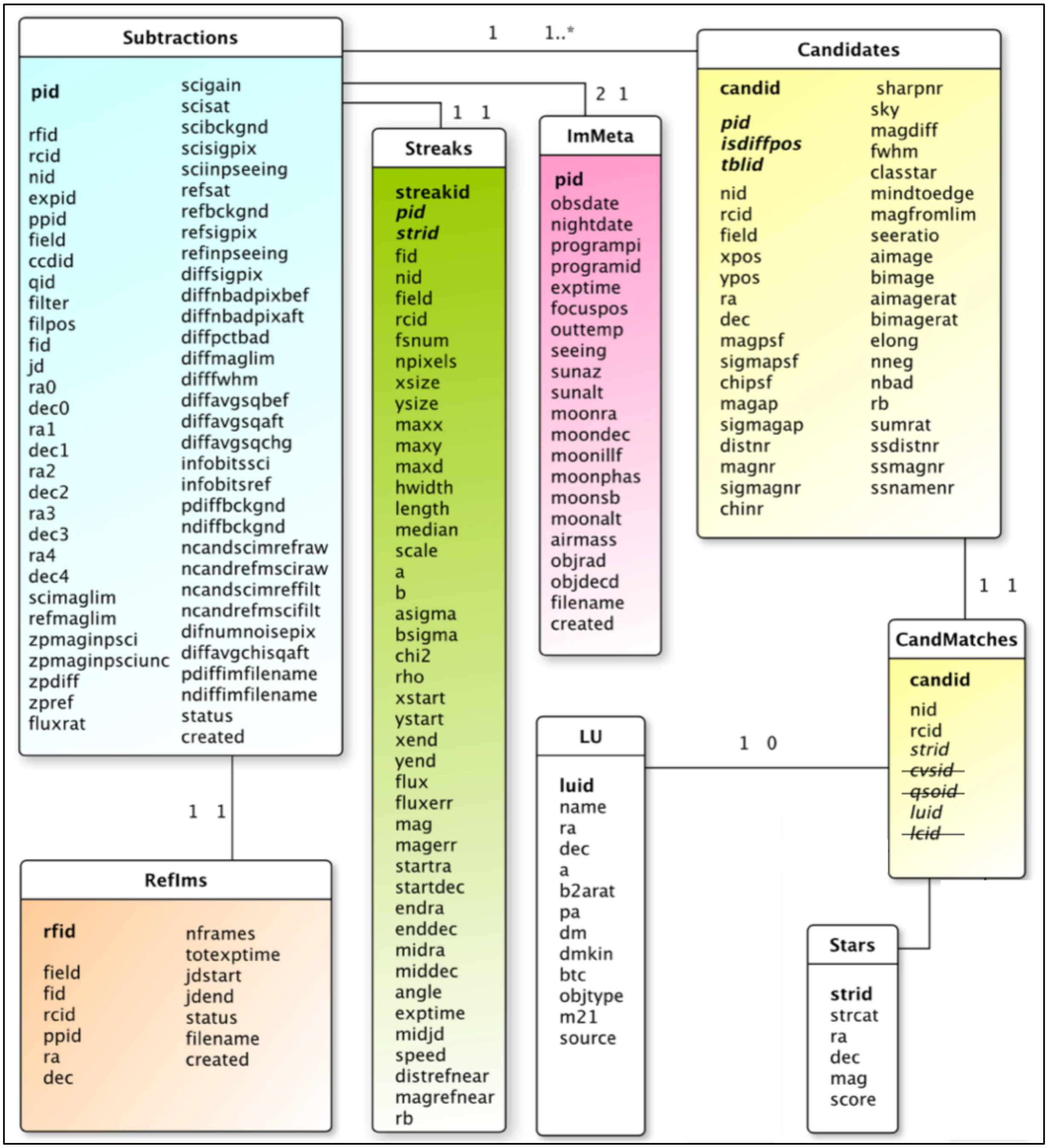

Figure 3: Schema for ZTF Transients Database (subset of ZTF Operations DB). v.11/9/16

8. ZTF-depot sizing estimates (image cutouts only)

- Assume (following internal real-bogus filtering) ~ 20 point-source candidates per readout-channel image.

- With ~ 700 exposures / night resulting in 44,800 positive subtraction images, this leads to ~ 1 million *target* point-source candidates / night.

- These exposures (and candidates) are ~ equally apportioned across multiple filters and two epochs (coverages) / night. This point is not relevant to the overall sizing estimate.

- For each *target* point-source candidate, there will be three image cutouts (*sci, ref, and sci-minus-ref*), yielding ~ 3 million *target* candidate image cutouts / night.

- We also expect ~ 200,000 *streak*-candidates / night, from which only *sci-minus-ref* image cutouts will be generated.

- From the last two bullets, that's ~ 3.2 million image-cutouts per night (or ~ 5,300 cutouts / minute) *for target candidates only.*

- At the time of writing, each JPG cutout is ~ 1.3kB in size. This amounts to ~ 4.2 GB in cutouts alone per night (or ~ 7 Mbytes / minute) *for target candidates only.*

- Assuming **Ndhist** = 3 days when generating *forced* historical image cutouts on each *point-source target* candidate only, the above numbers are multiplied by three, to yield potentially ~ 9 million cutouts per night (or ~ 15,000 / minute).