Zwicky Transient Facility (ZTF)

QA database Requirements

CIN #677

Document Owner: Thomas Kupfer

Contributions by: Eric Bellm, Frank Masci

**Caltech Optical Observatories**

Cahill Center for Astronomy and Astrophysics

Caltech M/C 11-17

1200 E. California Blvd.

Pasadena, CA 91125

# Revision History

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Date** | **Author** | **Revision Description** |
| 0.0 | 2016-10-08 | TK | Initial draft |
| 0.1 | 2016-10-13 | TK | Revisions from Eric Bellm included |
| 0.2 | 2016-11-06 | TK | Revisions from Frank Masci included |

Contents

Revision History 2

1 Introduction 4

1.1 Purpose 4

1.2 Scope 4

1.3 Acronyms and abbreviations 4

1.4 Definitions 4

1.5 Related Documents 4

1.6 Document Organization 4

1.7 Points of Contact 4

2 Background 6

2.1 Calculating QA metrics for ZTF 6

2.2 Telescope schedule 6

2.3 Telescope status 6

2.4 Data processing 7

2.5 Long term behavior 7

3 QA database 7

3.1 Schedule of the telescope 7

3.2 Telescope status 7

3.3 Pipeline processing status 7

3.4 Long term trending 8

4 Functional Requirements 8

4.1 Location and accessibility 8

5 Appendix 8

5.1 Notes and QA metrics calculated by IPAC 8

# Introduction

## Purpose

This document presents the requirements for a Quality Assurance (QA) database for the Zwicky Transient Facility. It includes functional and performance requirements.

Using a QA database will improve the performance of the survey by identifying failures and aggregating performance history.

## Scope

The QA element provides the following elements:

* A QA database

## Acronyms and abbreviations

* DS – Data System (the data processing and archiving portion of the ZTF project)
* FITS – Flexible Image Transport System.
* FoV – Field of View
* GUI – Graphical User Interface
* MTBF – Mean Time Between Failure
* OS – Observing System (the data acquisition portion of the ZTF project)
* PBS – Product Breakdown Structure
* TBD – To Be Determined.
* TBC – To Be Confirmed.
* TBR – To Be Revised.
* TCS – Telescope Control System.
* ZTF – Zwicky Transit Facility.

## Definitions

In the requirements specifications, the following verbs are defined:

* Shall – denotes a requirement that is mandatory whenever the criterion for conformance with the specification requires that there be no deviation. It implies obligation.
* Should – denotes a guideline or recommendation whenever noncompliance with the specification is permissible. It expresses a contingent or conditional act or state, or a moral obligation.
* Will – denotes a simple statement of fact.

## Related Documents

* ZTF Science Requirements Document

## Document Organization

The rest of this document consists of a description of the QA database – followed by the requirements set for the database

## Points of Contact

Thomas Kupfer, Calibration Scientist tkupfer@astro.caltech.edu (626) 395-xxxx

Eric Bellm, Project Scientist ebellm@caltech.edu (626) 395-4504

Richard Walters rsw@astro.caltech.edu (626) 395-xxxx

Roger Smith, Systems Engineer rsmith@astro.caltech.edu (626) 395-8780

# Background

A quality assurance (QA) database will be used to track the performance of the ZTF survey. This includes scheduling, telescope status, observing performance, data processing, and also long-term behavior.

Currently, telescope status as well as observing performance is loaded into SQL database which can be accessed from a webpage. In PTF and iPTF observing metrics such as seeing are calculated in the processed images of one single CCD and written into the FITS header. Header information is loaded into Ricahrd Walter’s SQL database. Currently no weather/temperature information are stored in the headers.

Different metrics are calculated to track data processing and data quality. These include infobits (flags per image) Metrics are calculated for all PTF products but database is only accessible from IPAC. Currently users (e.g. scanners) report problems/errors and IPAC tracks the problems in an ad-hoc manner which is not ideal.

In PTF and iPTF no systematic monitoring of long-term behavior, e.g. readout noise, dark current or gain has been done.

With the significantly increased data volume for ZTF and hence significant larger number of possible detections a single database to track the survey performance is required. Requirements for this database will be outlined in this document.

## Calculating QA metrics for ZTF

To provide enough information to track the performance of the survey different metrics need to be calculated. This implies metrics which monitor short term as well as long term behavior of the survey.

Collecting all metrics into a database will allow long-term trending of specific metrics to characterize changes in instrument behavior and performance.

## Telescope schedule

During nightly searches for transients in real time, science monitors (“scanners”) often need to know the projected schedule of telescope pointings as well as the executed pointing history and the status of data processing.

## Telescope status

A nightly monitoring of the telescope status shall be obtained. This includes the following points regarding the telescope:

* Dome open/closed, weather, temperatures, sky brightness, seeing, tracking of the telescope

Seeing shall be calculated from all CCDs and stored into the header together with information on weather/temperature.

## Data processing

IPAC shall provide QA metrics and processing status (received, processed, archived) for all data products (bias, flats, raw and processed images). Different metrics shall be calculated (infobit and dmask).

See Appendix I for a list of the proposed metrics that will be loaded into the database

## Long term behavior

Metrics shall be calculated to monitor the long-term behavior of the CCDs. These include readout noise, dark current, and gain.

# QA database

To monitor the performance of the telescope, detectors, and the survey the calculated metrics shall be stored in a relational database.

The QA database shall be accessible through a web interface consisting of four different sections:

1. Schedule of the telescope (near future/past)
2. Telescope status
3. Pipeline/processing status
4. Long term trending

## Schedule of the telescope

This tab shall include the estimated future schedule, tonight’s schedule as executed, and past pointing history. The web interface may enable the user to produce sky plots and histograms of these schedules.

## Telescope status

The tab shall include different metrics which monitor the telescope. These includes if the dome is open or closed, weather monitor statistics, all-sky camera images, and thumbnails of the last . The tab shall be updated in real time independently from individual exposures and provides a tool to see if data is currently being taken.

This tab should be connected to an all sky camera which sends every 10 minutes a picture of the sky

## Pipeline processing status

The tab shall include metrics which monitors data processing and data quality for every single exposure. This includes photometric and astrometric calibration as well as pipeline performance. **(TBD possible as a red (not yet done), orange (problem found), green (data ok) flag if different processing steps has been done). If problem found it shows which flag was problematic.** Detailed information for every single individual exposure are uploaded and shall be accessible by clicking on the exposure. The goal should be that the database updates every **TBD** min. The metrics will be calculated and provided by IPAC. The database should provide an end of the night report which includes a summary of observations and performance.

## Long term trending

The tab shall include metrics to monitor the long term behavior of the detector and the survey. Gain, readout noise and dark current should be calculated in the end of every night and reported once every **two weeks (TBD)**

Long-term trending should provide plots which plots the different metrics over time to monitor the long-term health of the detector.

Similar metrics could be included for the dome flatfields

# Functional Requirements

## Location and accessibility

The QA database shall be located at Cahill. Metrics computed by IPAC shall be pushed to Cahill. **(TBD which machine)** The database shall be accessible by a webpage for everybody from the collaboration with username and password

# Appendix

## Notes and QA metrics calculated by IPAC

- It is envisaged that a database query script will run at regular (TBD)

 intervals throughout a night to collect QA metrics of interest on all data

 received and processed up to that point. This information will be deposited

 into a staging area for retrieval by the QA monitoring system (format is

 likely to be ASCII; TBD).

- Furthermore, aggregation statistics (cumulative throughout a night following

 each incremental DB query) will be computed. 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.

- 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

 (iv) instrumental calibration

 (v) image subtraction and transient extraction

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

- Only those metrics derived from processing pertinent to the specific "imgtype"

 will be populated, provided there were no prior pipeline failures. Successful

 processing of a single readout-channel end-to-end (from ingestion, splitting,

 image-differencing & extraction) will be indicated by "exitcode=0". The very

 last pipeline step that was executed for a specific readout-channel, i.e.,

 pertaining to "exitcode" is indicated by "ppid".

Identifiers and timing information for all data successfully ingested:

----------------------------------------------------------------------

field -- field ID

ccdid -- CCD ID: 1..16

fid -- filter ID: 1..3)

rcid -- readout-channel ID in field: 0..63

qid -- readout-channel ID in CCD: 1..4

nid -- night ID counter (a "night" starts at 12:00:00 [noon] Pacific Time and

 ends at 12:00:00 [noon] the next day)

nightdate -- date corresponding to nid, i.e., at start of night as defined above

imgtype -- either science, bias, flat, or other (precise identifiers are TBD)

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]

exitcode -- pipeline exit code from "final" processing step; this step is

 indicated by pipeline identifier "ppid" (exitcode=0 => success;

 exitcode!=0 => failure in pipeline with ID "ppid")

ppid -- pipeline processing ID pertaining to "exitcode"

Raw (readout-channel) image based:

----------------------------------

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]

Calibration-image based (bias and flats, indicated by "imgtype" above):

-----------------------------------------------------------------------

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 (~ read-noise if imgtype='bias')

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)

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

radiff -- Difference: TELRA - reconst RA, corrected for cos(dec) [arcsec]

decdiff -- Difference: TELDEC - reconst Dec [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 sciece 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]

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