# Ingestion of Transient Object Detections from PanSTARRS1 and ATLAS

## *WP2(Data Access Centre)*
## *Task 2.3 (Support of Phase A Dev work)*

| | |
|---|---|
| **Project Acronym** | LUSC-A |
| **Project Title** | UK Involvement in the Large Synoptic Survey Telescope |
| **Document Number** | LUSC-A-03 <Contact George.beckett@ed.ac.uk for number> |

| | |
|---|---|
| **Submission date** | 17/3/2017 |
| **Version** | 1.0 |
| **Status** | Completed document |
| **Author(s) inc. institutional affiliation** | Ken Smith (QUB) |
| **Reviewer(s)** | Bob Mann (Edin) |

| **Dissemination level** | |
|---|---|
| Confidential | *The prime purpose is for informing other members of the consortium. In due course it may be suitable to be public.* |

# Version History

| Version | Date | Comments, Changes, Status | Authors, contributors, reviewers |
|---|---|---|---|
| 1.0 | 17/3/2017 | This is a report of work done for another project, for DAC context. It does not need updating. | K. Smith |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

# Ingestion of Transient Object Detections from Pan-STARRS1 and ATLAS

Ken Smith

Astrophysics Research Centre

Queen's University Belfast

Belfast BT7 1NN

March 2, 2017

**Abstract**

The following document describes the processing of detection files from both Pan-STARRS and ATLAS.

# Contents



# List of Figures

# 1 Introduction

## 1.1 The Pan-STARRS Telescopes

The Panoramic Survey Telescope and Rapid Response System (Pan-STARRS) 1 telescope (Chambers et al , 2016), (Magnier et al , 2016), (Waters et al , 2016), (Magnier et al , 2016), (Magnier et al , 2016), (Flewelling et al , 2016) is a 1.8 metre "wide field" instrument (Kaiser et al , 2010), capable of observing three quarters of the sky ($3\pi$ steradians, or about 30,000 square degrees) from the north celestial pole to 30 degrees south. (In fact, from 2015, it started to observe even more sky down to about -40 degrees declination.) Pan-STARRS 1 and Pan-STARRS 2 (to be commissioned in 2017) sit adjacent to each other on Haleakala, Maui.

The telescopes use 6 filters - g, r, i, z, y, w (=gri). In the current operational NEO survey, the main filters utilised are w (during dark time) and i (during bright time).

Images are captured on a 1.4 gigapixel CCD camera - currently the world's largest such device. The camera consists of 60 x 4800 x 4800 pixel chips in the focal plane. Each chip is itself an 8 x 8 array of 600 x 600 pixel CCDs. The pixel scale is 0.259 arcsecs per pixel, hence the total illuminated area is around 7.16 square degrees of sky. On any night the telescope may take 850 exposures, hence around 6,000 square degrees per night. (In actuality, the telescope takes "quads" (4 exposures, separated by around 7 minutes) at each position, meaning that only 1500 degrees is covered on any one night.) The images "warped" and resampled onto a fixed sky tessellation by the Image Processing Pipeline (IPP), with a pixel scale of 0.25 arcsec per pixel.

Over time, Pan-STARRS has built a high quality "static sky" image which is subtracted (Alard & Lupton , 1998) by IPP from the nightly warped images (after convolution to match the PSF) to produce difference images. Anything that flashes or moves should remain in the resultant Difference Image. The locations of these events are catalogued by the IPP software in the form of FITS tables ("cmf files") and ingested by the Transient Science Server (TSS) at Queen's University. The FITS catalogues are usually organised into one per "skycell". A skycell is a unit of the sky tessellation, and there are usually about 51 skycells covered by a PS1 exposure.

On any night, if all 850 exposures are good, this will result in 43,350 (850 x 51) catalogue files. On average, each catalogue file will contain up to 500 diff detections, so up to 21 million detections are ingested per night. (In fact, given the fact that the galactic plane is excluded and weather issues, on average between 5 and 10 million detections are ingested per night.)

## 1.2 The ATLAS Telescopes

The Asteroid Terrestrial impact Last Alert System (ATLAS) consists of two 0.5m Schmidt Cassegrain telescopes on the mountains of Haleakala (Maui) and Mauna Loa (Big Island) in Hawaii. The first telescope on Haleakala has been fully operational since 2015, and the Mauna Loa telescope was commissioned at the end of January 2017. The telescopes use two main filters in normal operations - "cyan" (during dark time - roughly equivalent to PS1 g+r) and "orange" (during bright time - roughly equivalent to r+i).

The detector on each telescope ("ACAM") is a STA 1600 10560 x 10560 single CCD. The pixel scale is 1.86 arcsec per pixel, meaning that each exposure covers about 29.7 square degrees. Each telescope can take about 1,000 exposures per night, covering 29,700 square degrees, meaning that the entire visible sky (north celestial pole down to about -44 degrees declination) can be imaged per night. In current operation, the telescope scans the same field 5 times per night, hence the total area covered is around 6000 square degrees per night. The telescope scans declination strips of the sky and covers the same area every 4 nights. Each exposure is astrometrically calibrated using a gnomonic projection onto the sky by the custom ATLAS reduction software.

Over time the telescope has built up a static sky "wallpaper". The ATLAS pipeline subtracts this from the nightly images (with PSF matching to the broadest PSF image), and anything that flashes or moves will remain in the resultant difference images. The ATLAS reduction pipeline ("tphot") produces catalogues of the locations of diff detections in the form of a headed ASCII table ("ddt files"). These ASCII tables, usually one per exposure, are ingested into the TSS.

On a typical night, a single ATLAS telescope produces about 1,000 catalogue files per night (one per exposure). Each catalogue file contains on average about 5,000 detections, so about 5 million detections are ingested per night.

## 1.3 The Transient Science Server

The Queen's pipeline automatically ingests the catalogued objects into a database. Detections are read by a C++ ingester which talks to a MySQL database. Detections are first checked against previously ingested objects. If a detection is within a specified proximity of an existing object, it is ingested and formally associated with that object, becoming another point on the lightcurve. If there are no objects within the specified proximity of the ingested detection, a new object is created and the detection is associated with the new object.

To speed up ingest, multiple ingester processes are run in parallel. Normally, objects within the same footprint of the sky are ingested by the same process to minimise database contention.

After ingest, selection criteria are applied to flag objects that might be interesting. The selection criteria include: multiple detections with minimum time interval between them, a certain number of detections apparent within a specified window of observations, a maximum RMS detection scatter. These thresholds help to eliminate false detections (caused by bad subtractions, chip defects, crosstalk, etc), narrowing down the number of potential transients that get flagged. Once marked as objects of interest, contextual checks are done. Depending on the detection's proximity to known objects the transient is machine-context-classified (using a simple decision tree algorithm), yielding best guess as to the object type (supernova, AGN, variable star, etc). Another check cross-matches the object with the Minor Planet database to eliminate movers. Objects context classified as variable stars are generally moved aside (since there are many). The flagged objects are also cross-matched against a database of transient objects discovered by other surveys and tagged if they are already known. Once the object has been context classified (with stars removed) small cut-out input, reference and difference images ("postage stamps") around the detections are requested from the relevant telescope. These are produced in Hawaii for Pan-STARRS and locally for ATLAS.

Once the difference images have arrived, a machine learning algorithm, trained to recognise "good" subtractions (i.e. PSF-like objects in the centre of the difference stamp) is run on the multiple (at least two) difference images associated with each object. A median score between 0 (bogus) and 1 (real) is applied to each object. All objects below a specified threshold are thrown away, with the rest passed to human scanners.

A web interface has been built atop the database to allow the human scanners (Eyeballers) look through these images. The user is presented with a triplet of postage stamps: the target (nightly image), the reference (the subtraction template) and the difference image. The user looks at these and uses a web form to decide whether or not they are real or bogus transients. Figure 1 shows a schematic of the TSS.

## 1.4 Software Stack

The ingester is written in C++. A wrapper (either Perl or Condor) splits the catalogue files across the multiple processors to facilitate parallel ingest.

The Pan-STARRS Image Processing Pipeline tags each detection with a set of flags that may indicate that the detection is faulty - e.g. the PSF fit has failed, the detection is saturated, the detection is a suspected cosmic ray, etc. If any of these flags are set, the detection is NOT ingested. This accounts for half of all Pan-STARRS detections. No equivalent scheme yet exists for ATLAS transient detections, so all ATLAS detections are ingested into the database. (The exception is that all ATLAS detections associated with an exposure are rejected if that detection file contains an unusually large number of detections - indicating a difference failure.)

Pan-STARRS detections within a galactic latitude $b = |5|$ (i.e. in the galactic

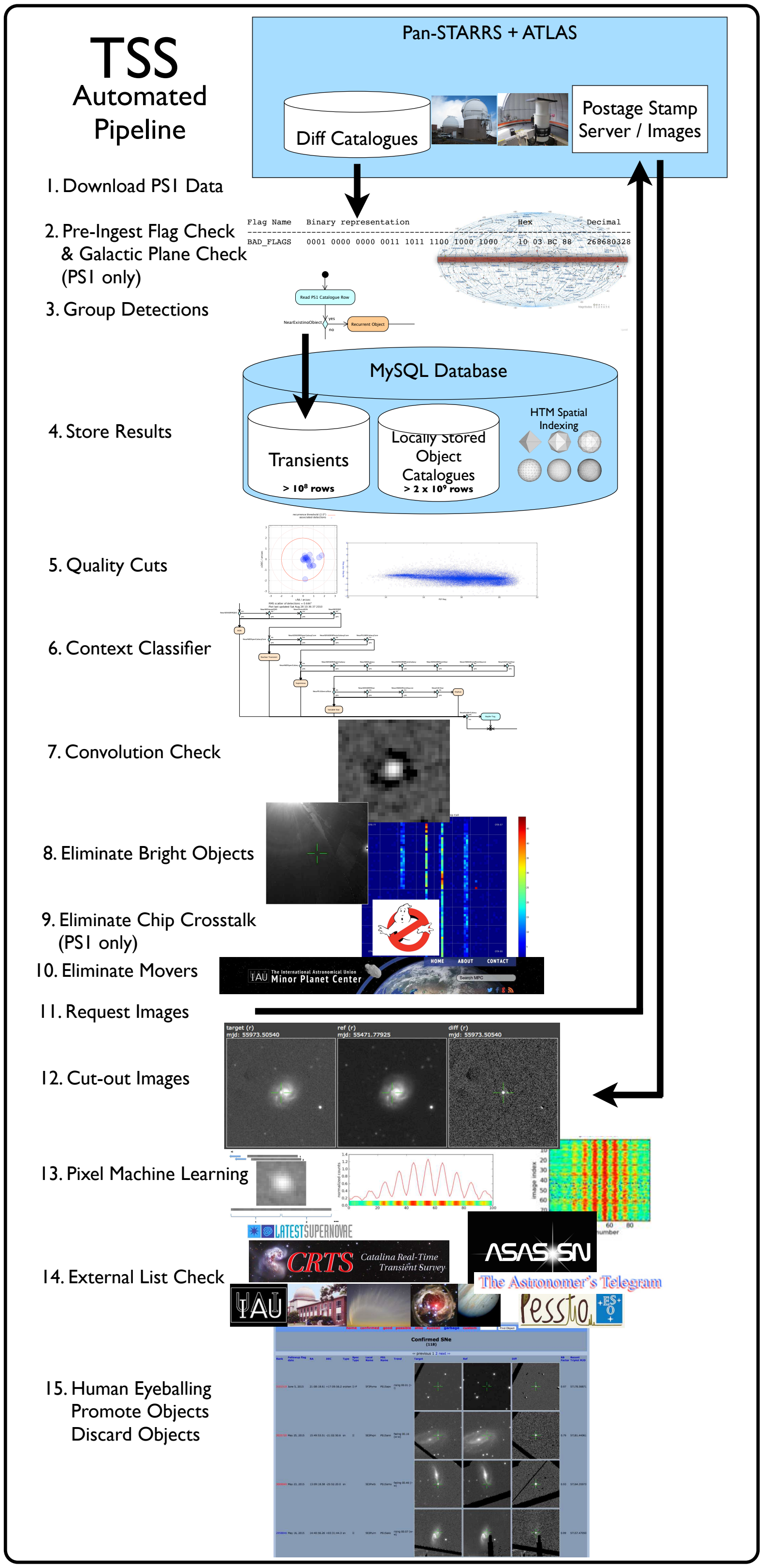


**Figure 1:** *The QUB Transient Science Server*

plane) are currently NOT ingested into the Pan-STARRS database. All ATLAS detections in the galactic plane are ingested.

The data are spatially indexed on ingest using Hierarchical Triangular Mesh. Since the vast majority of queries are cone searches of the order of a few arcsec in size, HTM Level 16 (10-15 arcsec sized triangles) is used. The C++ HTM library provides a cone searching interface which returns all triangles that overlap a cone on the sky. All objects that have those triangle IDs are returned to the requesting code. The cone search code then just needs to eliminate objects outside the requested radius.

At the point of ingest, an HTM triangle ID is calculated for a detection based on its RA and Dec. A cone search is run against the database of already ingested objects. If an object is within the cone search radius, the detection is associated with that object. If not, a new object is created. In both cases the detection is ingested into the database. The critical values ingested are RA, Dec, htmID, mag, dmag. Other calculated detection properties are also ingested to facilitate post ingest cutting. The header part of the detection file will contain values common to all the detections in that file, such as MJD, filter, zeropoint and RA and Dec of the telescope pointing. (The PS1 FITS tables ["cmf files"] contain a header component and a binary table data extension. The ATLAS ASCII files ["ddt files"] contain a header component consisting of key value pairs preceding a table keys header row followed by the rows of table values.)

Once the detections are ingested into the database, a multiprocessing Python module runs the post ingest cuts. The cuts are run for intra day checks (e.g. 3 of 5 detections required) and inter day checks (e.g. at least 2 detections intra day on one day, followed by at least another 2 detections intra day on a second day within a time window). The convolution check, bright star proximity check, context checking, machine learning, crosstalk checking, mover checking, external list crossmatching and image requesting code are all written in python. Objects that are securely context classified as variable stars are moved aside. (Note that a new version of the context checking code ("Sherlock"), which determines the suspected type of the transient is being developed, and will soon be rolled out in place of the existing decision tree code.)

The web interface is also written in python using the Django Web Framework. Users log into the system and can promote and demote objects either individually or in bulk (e.g. moving all known asteroids tagged on a particular day to an attic list).

For ATLAS data, forced difference photometry is performed for all objects promoted by users. The forced photometry code is a python wrapper for the custom C photometry code developed by the ATLAS team (tphot). Forced photometry is not available within the current Pan-STARRS processing, but this may change in the not too distant future.

All of the code is stored in a private Git repository (currently Bitbucket).

## 1.5 Hardware

The processing of data is done by two rack servers - a 16 core server with 96GB RAM for the PS1 data, and a 28 core server with 128GB RAM for the ATLAS data.

The MySQL database is hosted on a large (multi-terabyte) solid state disks to reduce read latency. The multi-billion row context check database (e.g. Guide Star, gaia, SDSS, 2MASS) currently consumes about 2.4TB.

The Pan-STARRS 1 database, which has been running since June 2013 and contains 2.63 billion detections is currently 1.4TB in size. The Pan-STARRS postage stamps, stored on spinning disk, currently occupy about 3TB of space. Hence a single Pan-STARRS telescope requires about 0.5TB per year for the database and about 1 TB per year of stamps. Several times this space for backups on spinning disk is also reserved (daily, weekly, monthly).

The ATLAS 1 database, which has been running since December 2015 and contains 1.29 billion detections is currently 0.62 TB in size. This is likely to increase greatly with improvement of Schmidt Correctors and commissioning of the second telescope. The ATLAS postage stamps currently occupy 1.2TB, but extraction of the stamps also requires download of the original input, reference and difference images - which accounts for a further 33TB of data. Hence single telescope database is growing by

0.6TB per year, stamps are growing by 1TB per year and image data (if download continues at the current rate) will require 30TB per year. Several times this space for backups on spinning disk is also reserved (daily, weekly, monthly). (Note that there are also plans to create an ATLAS billion star database which would store 1000 points on the lightcurve per year for a billion stars. The size of this database would potentially be at least 100TB per year, assuming just 100 bytes per detection and the database would need to ingest 64,000 detections per second, assuming 50% utilisation. The current infrastructure is not yet designed to cope with a database of this size.)

# References


Alard, C., Lupton, R. H., 1998 APJ 503 325

Kaiser, N., Burgett, W., Chambers, K., et al. 2010, Proc. SPIE, 7733

Chambers et al, arXiv:1612.05560

Magnier et al, arXiv:1612.05240

Waters et al, arXiv:1612.05245

Magnier et al, arXiv:1612.05244

Magnier et al, arXiv:1612.05242

Flewelling et al, arXiv:1612.05243