Catalina Sky Survey Operations and Processing

R. Seaman[1], E. Christensen, A. Gibbs, S. Larson, F. Shelly

12 November 2020

1. Introduction

This document provides an overview of the operations, including software, of the Catalina Sky Survey (CSS). CSS is a ground based telescopic survey project that seeks to discover Near-Earth Objects (NEOs) that potentially could impact Earth. In its current form, CSS has been in operation since 2003 and is currently funded under a NASA NEOO grant ("The Catalina Sky Survey for Near-Earth Objects") which includes the requirement to make the CSS archives available to the public. To meet this requirement, CSS will deliver the data resulting from survey operations to SBN by means of an accredited configuration-controlled pipeline. The discussion below should aid users and implementers in understanding these data.

2. The Catalina Sky Survey Archive Pipeline

The Catalina Sky Survey operates several telescopes dedicated to the discovery and astrometric follow-up of NEOs. The CSS pipeline makes no distinction between survey and targeted observations, even from the same telescope. A NEO telescope has many hardware and software subsystems and some are outside the purview of this document. Examples of these are the telescope control system (TCS), survey planning and target list generation, observation queuing and scheduler, and general purpose observing tools such as used to focus the telescope or evaluate observing conditions, etc.

Processing begins with the reduction of raw camera images (CCD or perhaps CMOS in the future). Imaging data suitable for moving object detection is subject to the usual data reduction steps of photometric and astrometric calibration. CSS has generally operated its cameras unfiltered and the photometric steps are straightforward. Currently the Gaia DR2 catalog is used to provide both photometric and astrometric references. Other catalogs have been used in the past and this will undoubtedly change again in the future.

The most notable fact about CSS data is that each field is reimaged multiple times, typically interleaved with other fields over a period of a half hour or so. Survey data has four images per field; targeted follow-up may utilize *track & stack* mode where each of these four images is further split into several very short exposures to freeze the target's motion. The four stacks are then shifted and coadded to maximize the signal before moving object detection is performed.

[1] rseaman@arizona.edu, Lunar and Planetary Laboratory, 1629 E University Drive, Tucson, AZ 85721

The CSS pipeline implements two kinds of source detection. Most CSS detections result from a comparison between the source extractor catalogs for the four images. Others are generated from an image subtraction step. These two lists are combined into one for each image, and then the lists of many thousands of possible detections are searched for linear motion. The resulting candidate moving objects are scored as to likelihood of being real as well as likelihood of their motion resembling that of near-earth objects.

The previous three paragraphs generally describe activities that are under configuration management and that create diverse controlled data products. The prioritized lists of candidates are matched against the known catalog of solar system objects and unknown objects are presented to the observer for visual validation through blinking of the zoomed-in images (other surveys may use image cut-outs). The human validation process itself is of necessity outside configuration control, but the final step of report generation is a controlled step. Historically this has used the MPC 80-column format, but Catalina has transitioned to the IAU Astrometric Data Exchange Standard (ADES).

Label generation is the final controlled step and occurs the following day. All previous steps occur on computers at the telescopes while labels are created on a server at the CSS offices at the Lunar and Planetary Laboratory (LPL) on the University of Arizona campus. A scripted data transfer process replicates the mountaintop data downtown with a semaphore file being written as the final step after checksum verification. A Linux cron process later copies the data into its permanent location on CSS storage. At this point a script is executed to create the PDS4 labels for the diverse data products. On the LPL server the labels are written to a separate directory with their own checksums.

An implementation detail is that CSS maintains multiple copies of our data, automatically replicating copy A (at LPL) to copy B (at the UITS campus data center). Transport of CSS data and the corresponding PDS4 labels to PSI will occur from copy B at UITS.

3. Controlled Pipeline Software

Controlled software begins with the pipeline reduction of raw images from the CCD or CMOS cameras, while software associated with controlling the survey planning, exposure queuing, telescope, camera, etc., is generally outside the scope of this document. CSS uses Subversion (SVN) for source code management of all modules described below. CSS software is being published via the NASA github and is mostly written in TCL and C. Figure 1 a-c in the appendix shows the overall flow of the processing and the resulting archival SBN data products (and temporary working files surrounded by dashed lines). Circled numbers on the figures correspond to the general processing sub-sections below, and rectangular boxes represent specific software modules in the CSS pipeline. Note that the pipeline is flexibly configurable to run on a single or on multiple hosts, and these hosts vary between telescopes.

3.1 Image Calibration

In the CSS context, image calibration means our version of normal CCD flat-fielding, de-biasing, trimming off overscan, masking vignetted pixels, etc. More generally it means whatever bespoke image processing is required prior to later steps, e.g., updating the raw header keywords from the camera software with information from the telescope control system, timekeeping infrastructure, etc. Routine configuration changes include generating library flat fields, updating CCD parameters such as binning, retuning TCS and camera parameters, OS updates, and other familiar activities of ground-based telescope operations. These steps are packaged into a single module with associated parameter and library routines that can overall be referred to as "CALIBRATE". All CSS imaging data products will be delivered in FITS tile-compressed format as generated by the third-party FPACK command.

3.2 Source Extraction

CSS relies on the community software package Source Extractor[2] to detect and measure the point sources contained in each of the many hundreds of images taken at each of its telescopes each clear night. SEXTRACTOR responds to many command line and parameter file settings and CCS will occasionally modify these for diverse reasons. SEXTRACTOR is initially executed on a per-image basis, and later interacts with SCAMP, described in the next section, in a somewhat complicated fashion that updates the output source metadata in several steps. The original SEXTRACTOR output is a FITS binary table format and later copies are an ASCII text format. Both formats are archived in SBN, and may contain different columns. A helper module called SEXB2SEXT creates the text files. Most of the larger CSS text data products will be delivered in GZIP compressed format.

3.3 Astrometric Calibration

CSS relies on the SCAMP[3] package (a relative of SEXTRACTOR) to generate astrometric solutions which ultimately become the asteroid observations submitted to the Minor Planet Center. SCAMP responds to many command line and parameter file settings and CSS occasionally changes these. SCAMP provides a calibration against an astrometric catalog, currently Gaia-DR2. When new astrometric catalogs become available, CSS will evaluate them. As shown in Figure 1a, SCAMP is executed twice and updates the calibrated image with a world coordinate system as well as generating log file and metadata output that are archived.

3.4 Photometric Calibration

The initial photometric calibration for CSS observations is also performed by SCAMP (that's what the 'P' stands for), but the zero point is adjusted according to the photometric catalog (also Gaia DR2). In normal survey operations, CSS photometry is unfiltered, but our telescopes have

[2] *https://www.astromatic.net/software/sextractor*
[3] *https://www.astromatic.net/2010/04/20/scamp-1-7-0-release*

the option of using filters. CSS occasionally tunes photometric settings and when new photometric catalogs become available, CSS will evaluate them. The final photometric calibration is performed by a module called VPHOT, and helper modules SKY2XY, STARMATCH, and SCAMPFIELD are used.

3.5 Image Subtraction

Different solar system surveys use different detection techniques, generally falling under catalog-based and image-subtraction algorithms. CSS uses both and merges the candidate moving object detections from both as below. CSS image subtraction is performed on a per-field basis and does not require the construction of master templates for this purpose. As with other modules described here, tuning occasionally occurs. Image subtraction is implemented by the IPMTD module that accepts background-subtracted images on input. These are created by the FLATSKY module and are archived in SBN along with the lists of candidate moving target detections from this branch of the pipeline.

3.6 Moving Object Detection

Each survey field is a sequence of four (or possibly more than four, or fewer) images. Stellar point sources and other stationary detections are omitted from subsequent processing, leaving hundreds or thousands of possible detections per image of objects that are moving. A moving object detection involves recognizing near-linear motions when comparing the lists of candidate detections from each image. A certain tolerance is allowed for centroiding uncertainties as well as possible non-linear motions for nearby objects. Catalog-based moving object candidates are generated, and combined with image-subtraction candidates, via the MTDLINK module.

Moving object detection is greatly dependent on rate, and sensitivity to rapid movers requires sorting through many pixels. In addition, tuning the algorithm is subject to asteroid population simulations as well as to instrumental effects.

3.7 Identifying Known Objects, Scoring Unknown Candidates

The lists of detections are compared against the list of catalog objects known to be in the field, including the predicted magnitude of each object from its ephemeris, via the MAKEEPHM and IDENTIFY modules. This generates lists of known objects.

CSS imaging is very wide field, 5 square degrees for G96 and 19 square degrees for 703 (a Schmidt telescope). Each exposure, especially near the ecliptic, may have hundreds of detectable moving objects and many thousands of point sources. These are scored according to the likely reality of a moving object, as well as by its likelihood of being a Near Earth Object via the DIGEST2 module from the Minor Planet Center. Annotated candidates are turned into the final list of candidate detections using MTDF2DETS. The most likely NEO candidates are presented to the observer for validation.

3.8 Human Validation and MPC Report Generation

At various steps in the community NEO workflow, humans monitor automated processes. Catalina implements near real time validation of all NEO candidate astrometry before submission to the Minor Planet Center. Thousands of times during a typical night at the telescope, the CSS observer will blink candidate detections and make a keystroke determination as to whether the candidate corresponds to a real object. This is implemented via the **VALIDATION** module. The **MKMPCREPORT** module generates the final reports for submission of asteroid astrometry to the Minor Planet Center. At the time of this writing this meets the "80-column" MPC 1992 format, but CSS and the larger community are transitioning to the IAU Astrometry Data Exchange Standard (ADES), developed since 2015. CSS will continue to generate the 1992 format after the transition.

4. Catalina Sky Survey Operations

The CSS pipeline software described in section 3 is very general purpose and has been applied to six diverse telescopes used directly by Catalina Sky Survey[4] over the years:

- Mount Bigelow 0.7-m Schmidt telescope, MPC code 703
- Mount Lemmon 1.5-m telescope, MPC code G96
- Siding Spring 0.5-m Uppsala Schmidt, MPC code E12
- Mount Lemmon 1.0-m telescope, MPC code I52 *(follow-up)*
- Mount Bigelow 1.54-m Kuiper telescope, MPC code V06 *(follow-up)*
- Kitt Peak 2.3-m Bok telescope, MPC code V00 *(in collaboration with Spacewatch and U. Minn.)*

Astrometric measurements of asteroids, comets, and other small solar system bodies benefit from coordinated observations between survey and follow-up telescopes. Different NEO projects implement differing observing strategies. For Catalina Sky Survey these fall into the following general classes of use cases. The same telescopes, instruments, and software support this diversity of activities.

4.1 Near-Earth Object discovery and confirmation

A Near-Earth Object (NEO) discovery starts with the detection of moving celestial sources. These can be diverse objects such as main belt asteroids, Jupiter Trojans, Trans-Neptunian Objects, artificial satellites, or they can be NEOs. Distinguishing between these different kinds of objects requires accumulating multiple observations that can be used to solve for the Keplerian orbit. Short arcs solutions are of necessity uncertain and multiple follow-up observations are needed not only to beat that uncertainty down, but to distinguish between possible near and far orbits.

---

[4] The CSS pipeline has also been applied to data from several external telescopes, but these data sets will not be included in the CSS PDS holdings.

The Catalina survey telescopes have had fields-of-view from about 1 square degree to 19 square degrees. Survey patterns are a trade-off between exposure depth and the speed of covering sky with 30 second exposures being typical for CSS. The current 10K CCD at Catalina's 1.5-m telescope on Mount Lemmon (MPC code G96) can cover the entire sky visible from Arizona about 1.5 times per month. A similar 10K camera on the Mount Bigelow 0.7-m Schmidt telescope covers the visible sky every three or four days to a brighter limiting magnitude.

Depending on multiple factors including, of course, the weather, each night CSS will submit between a few and a few dozen NEO candidates to the NEO Confirmation Page[5] of the Minor Planet Center. Professional and amateur observers around the world will select objects for astrometric follow-up from the NEOCP, ideally leading to confirmation as NEOs. CSS itself operates follow-up telescopes to perform this critical chore. Follow-up telescopes will ideally have apertures similar in size to the survey telescopes, although much productive community follow-up is accomplished with quite small apertures. Follow-up observing is also distinguished by techniques such as tracking at non-sidereal rates to match the motions of the asteroids, and/or a track-and-stack observing mode that acquires multiple exposures to coadd and freeze each object's motion.

Follow-up observations are also reported to the Minor Planet Center for posting to the NEOCP. When a sufficient arc is accumulated for an acceptable initial orbit a Minor Planet Electronic Circular (MPEC) is published[6] to convey that orbit as well as to name the newly discovered object.

4.2 NEO Arc Extensions

Observations of NEOs continue after discovery and indefinitely through future apparitions when Earth and asteroid are again well-placed for observations. Even large asteroids become invisible from a particular telescope when too near the Sun in the sky, or when their declinations are outside of those reachable from the observing site. NEOs have eccentric orbits and when the asteroids are far from the Sun they may be too faint for any telescope on Earth.[7]

Catalina Sky Survey's follow-up telescopes also take observations for the purpose of NEO arc extensions, ranking different targets by the benefits to their orbits.

4.3 NEO Precoveries

Just because an asteroid has yet to be discovered does not mean that NEO surveys have not imaged it, and archival searches for pre-discovery detections (so-called precoveries) can often extend an orbital arc backwards significantly. There are many reasons why the detection may not have been made when the data were originally acquired, such as interference from stars

[5] *https://minorplanetcenter.net/iau/NEO/toconfirm_tabular.html*
[6] *https://minorplanetcenter.net/mpec/RecentMPECs.html*
[7] See Lowell Observatory's asteroid observability tool: *https://asteroid.lowell.edu/astobs/*

along the line-of-site, imaging artifacts, or simply poor sky conditions that made a blind detection impossible. In later months or years the knowledge that an asteroid should have appeared in a particular field, at a certain magnitude, and moving at a known rate can be enough to make the detection possible. Multiple users can productively search for precoveries and CSS hopes that opening its data holdings to the public via the Planetary Data System will support many more precoveries in the future. A single precovery can substitute for multiple arc extensions, increasing the efficiency of the entire NEO community.

4.4 Incidental MBA astrometry

While observing foreground NEOs, surveys also acquire vast numbers of observations of background Main Belt Asteroids (MBAs) and other classes of objects. Many of these objects have precise or nearly precise orbits. Large objects in circular orbits will be recovered many times over the years, and matching detections against the growing catalog of known objects permits reporting astrometry in an automated fashion. Several thousand automated MBA detections are made for each NEO detected, and several known NEOs may be detected for each unknown NEO candidate.

4.5 Asteroid and comet photometry and time series

The discovery of asteroids and the improvement of their orbits is a requirement for their scientific study. Each astrometric observation provides a rough photometric measurement, but CSS telescopes can also be used to acquire more precise photometry of known asteroids. This can require many times the investment in exposure time compared to the short exposure, low signal-to-noise observations required by an efficient survey mode. Asteroids rotate and time series observations can characterize the rate of rotation. Such characterization observations amount to only a small fraction of CSS data-taking at present, but this is likely to increase in the future as additional surveys come online. Archival data sets can similarly be mined for detections of active asteroids and comets to extend photometric measurements far into the past.

4.6 Crowd-sourced validation

A hallmark of Catalina Sky Survey operations is real time validation. Our observers are presented with roughly 5,000 blinking targets each night for rapid decisions on whether each object is real or not. An algorithm, digest2[8], provides an estimate of whether a particular short tracklet represents a NEO or rather some other class of object. Our observers' eyeballs are the ultimate limiting resource for CSS.

This scales well for same-night observations since a single observer can support one or two telescopes and adding telescopes generally means adding more observers. This does not scale for archival reprocessing which may deliver many additional detections than originally found

---

[8] *https://arxiv.org/abs/1904.09188*

due principally to two effects: 1) a newer version of the CSS pipeline is used which is likely to be much more efficient at finding moving object candidates, and 2) the catalog of known object will have grown which has the side effect of allowing us to dig deeper into the noise for unknown objects such as NEOs.

Crowd-sourcing is one way to requisition additional eyeballs to this task and CSS is currently operating a Trans-Neptunian Objects citizen science project on the Zooniverse.[9] This TNO project also serves as a prototype for a later NEO project which may well be layered more directly on CSS holdings in the PDS.

4.7 Community follow-up with NEOfixer

This document has discussed not only how CSS operations work now, but also has touched on various ways Catalina may evolve in the future, perhaps also to take advantage of PDS facilities. A general trend in astronomy and planetary science is ever more close cooperation between institutions and individuals. A number of trends for CSS and the NEO community involves the advent of transient broker technology to coordinate observing, archives, and processing to increase efficiency and data mining. CSS is commissioning its NEOfixer broker to better ensure that all NEO candidates receive timely follow-up and that the highest priority arc extensions are attended to, see figure 1.

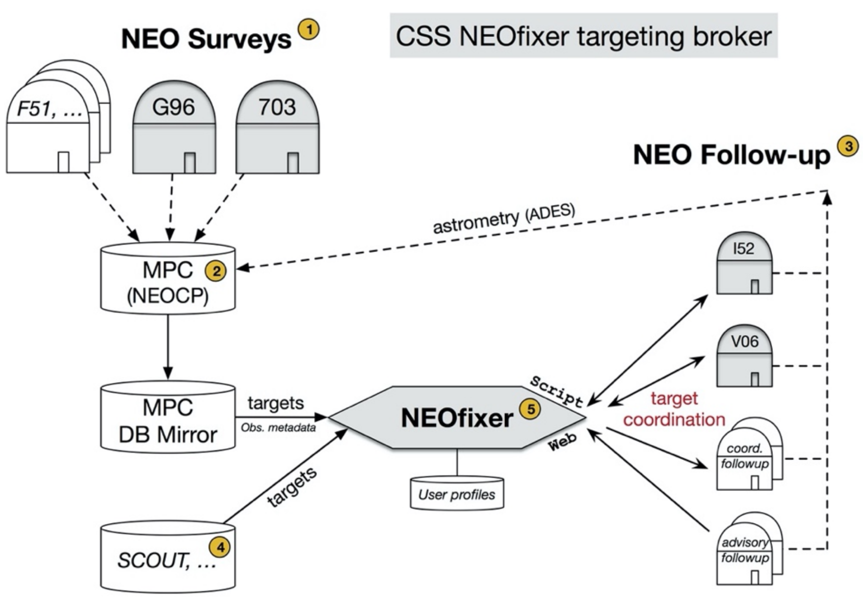

Figure 1 – NEO discovery and follow-up using NEOfixer. 1) NEO surveys submit NEO candidates to 2) the MPC. 3) NEO follow-up telescopes respond directly or indirectly, and 4) input from Scout and other sites (potentially PDS) also flows into the 5) NEOfixer broker which scores targets and provides priorities.

---

[9] *https://www.zooniverse.org/projects/fulsdavid/catalina-outer-solar-system-survey*

5. Appendix

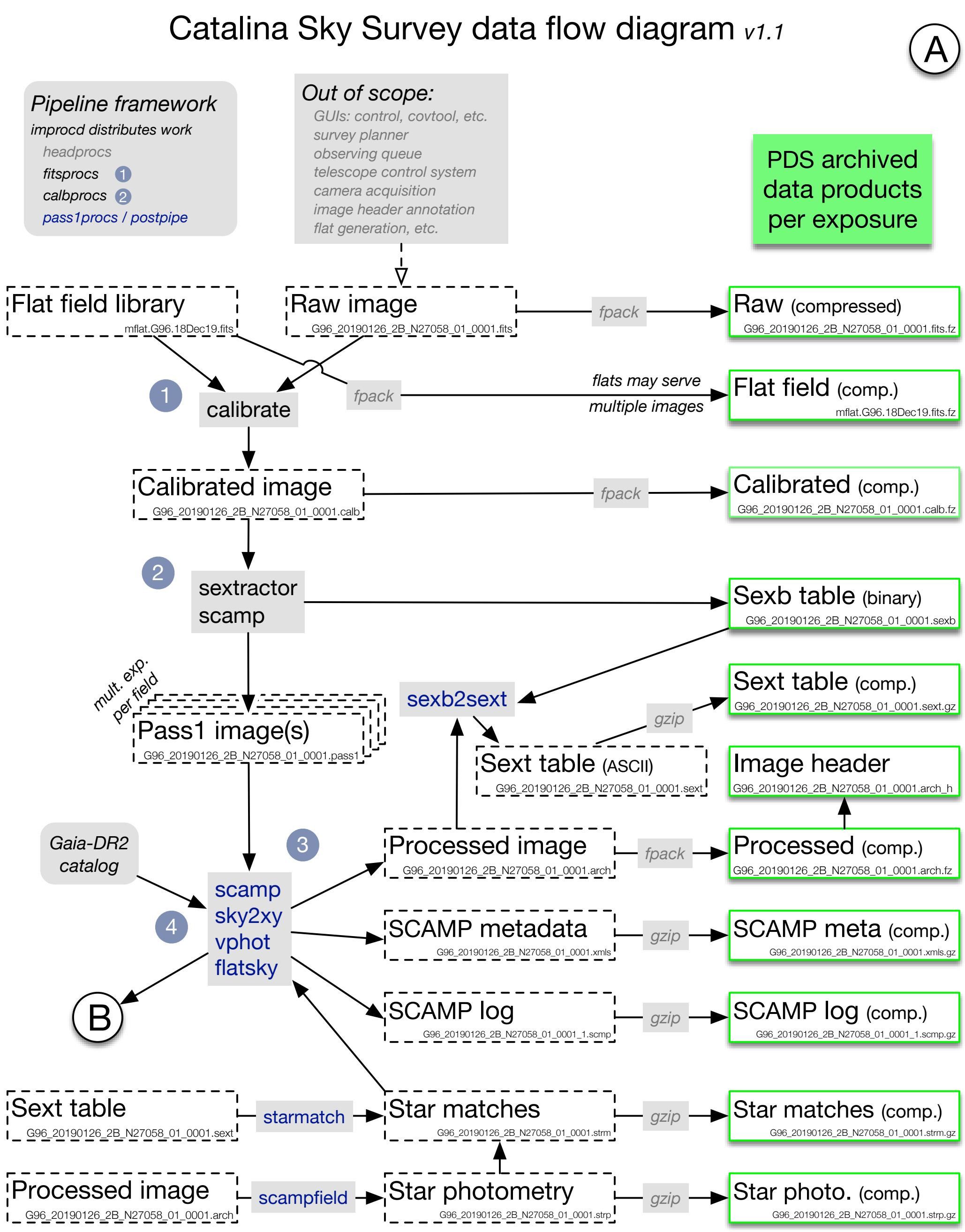

Figure 2 – CSS data flow with resulting SBN archival data products highlighted, generally in the column on the right. The CSS pipeline framework is outlined, with numbered references to module descriptions in section 4. Data product types on this page are archived one each per exposure of each telescope's camera. (A flat field may be used for multiple raw images.)

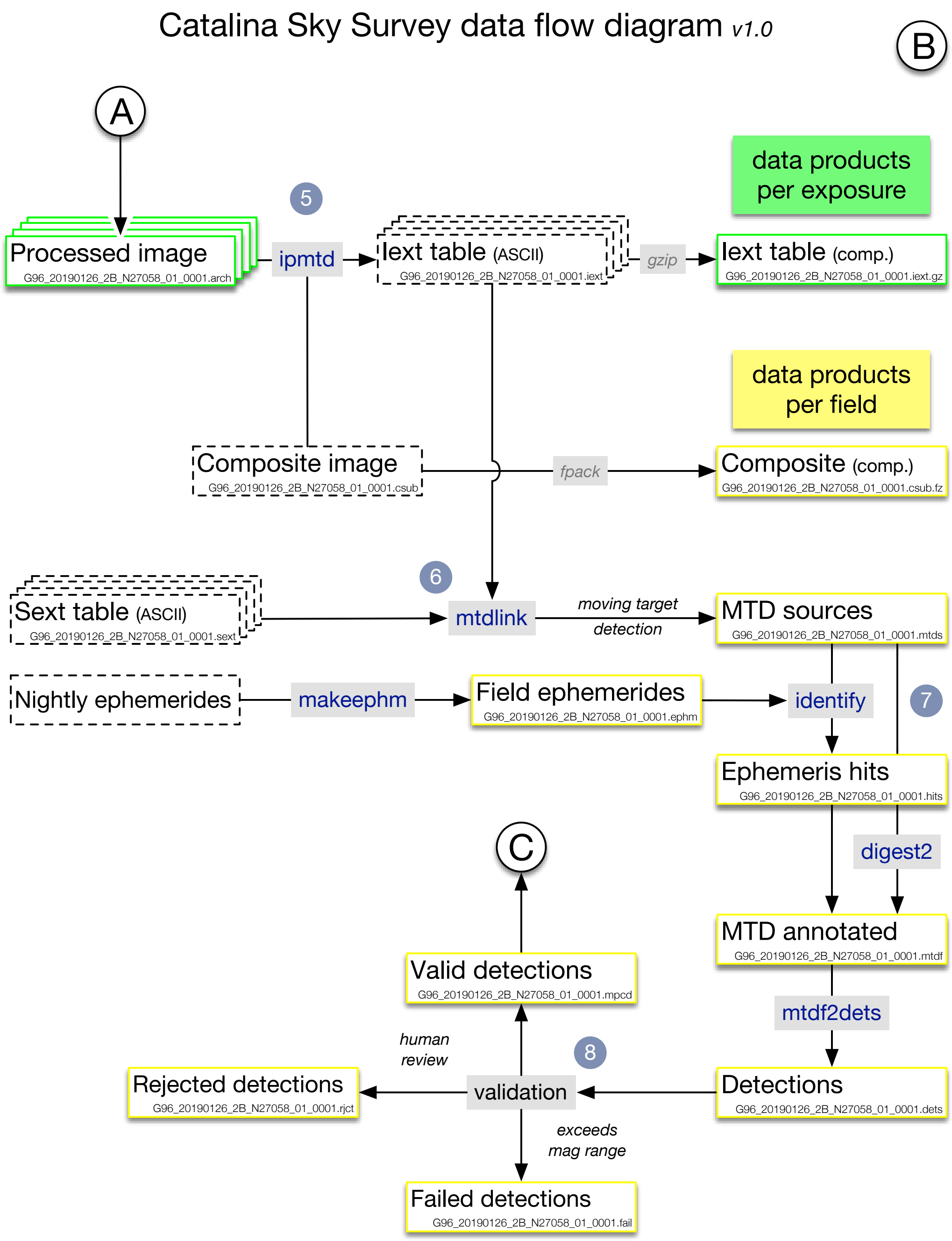

Figure 3 – CSS data flow continued. As indicated, data product types on this page are generally archived one each per multi-exposure field. Each field is reimaged four times in survey mode (three are sufficient for a moving object solution), and perhaps more times for targeted follow-up. Depending on the pipeline processing for a specific raw image, subsequent steps may be omitted due to various causes such as weather.

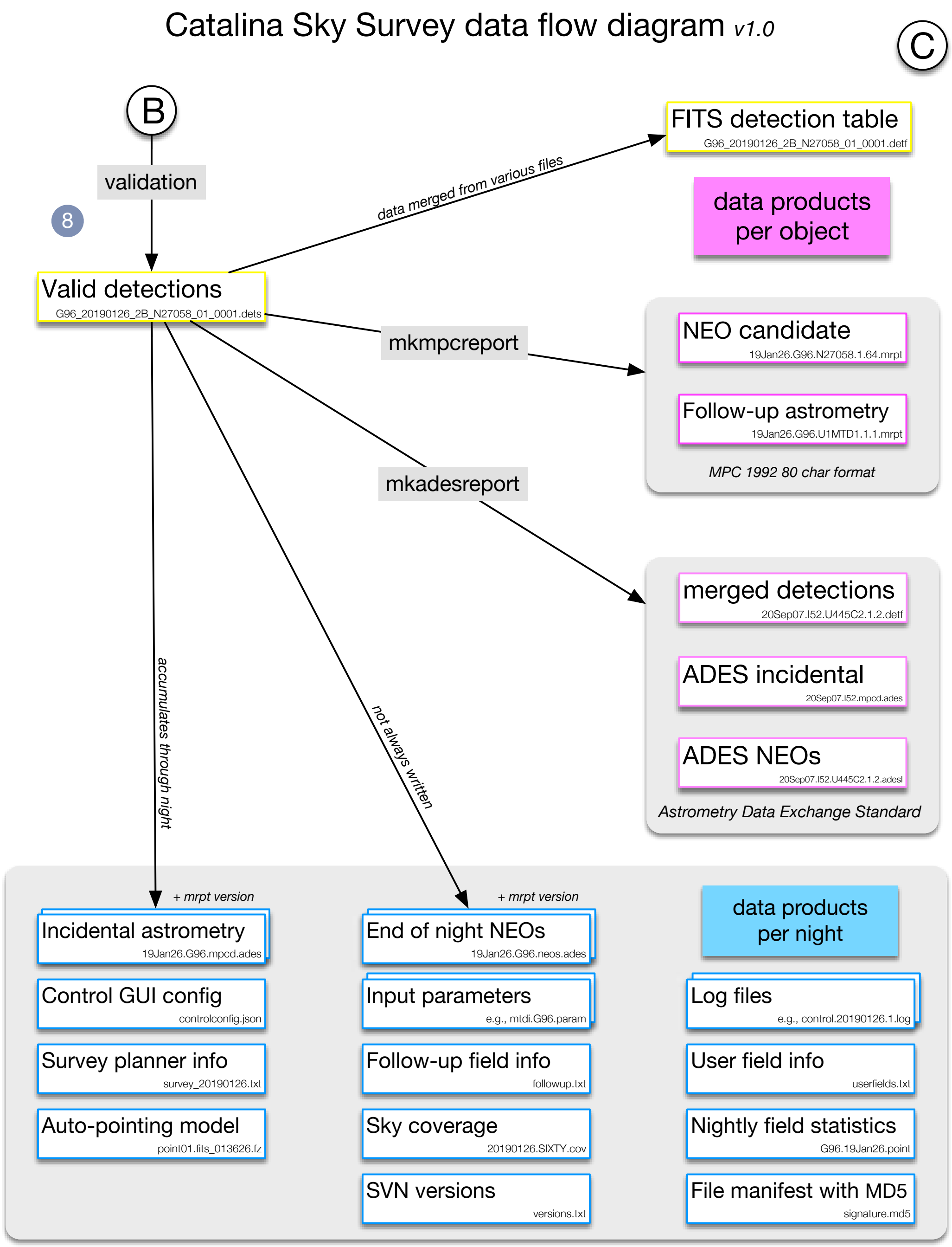

Figure 4 – CSS data flow continued. Data product types on this page may be archived one per candidate discovery or one or a few times per observing session (a night at one telescope).