# Overview of Catalina Sky Survey PDS Archive

*R. Seaman[1], C. Neese[2], J. Stone, E. Christensen*

13 November 2020

## 1 Catalina Sky Survey Archive Bundle Structure

The basic structure of the Catalina Sky Survey PDS bundle is outlined in figure 1. The subdirectories listed are for CSS telescopes (giving their MPC site codes) currently operating.

| Bundle | Collection | Subdirectory |
|---|---|---|
| Catalina | Data_Raw | 703<br>G96<br>I52<br>V00<br>V06 |
| | Data_Partially_Processed | 703<br>G96<br>I52<br>V00<br>V06 |
| | Data_Calibrated | 703<br>G96<br>I52<br>V00<br>V06 |
| | Data_Derived | 703<br>G96<br>I52<br>V00<br>V06 |
| | Calibration | |
| | Document | |

*Figure 1 – PDS bundle design for Catalina Sky Survey data holdings.*

The legacy holdings for the Siding Spring Survey, MPC code E12, will be ingested into the PDS Small Bodies Node at a later phase of the CSS archive reprocessing project. Future telescopes such as the Schulman telescope, MPC code W84, of the Mount Lemmon SkyCenter will be

[1] rseaman@arizona.edu, Lunar and Planetary Laboratory, 1629 E University Drive, Tucson, AZ 85721
[2] Planetary Science Institute, 1700 Fort Lowell Road, Suite 106, Tucson, AZ 85719

considered for inclusion in the SBN holdings of the Catalina Sky Survey as those projects evolve and are implemented. Precise details are yet to be settled.

# 2 Catalina Bundle Contents

In the tables below the subdirectory nomenclature is "tel" for the MPC code of the telescope and "yyyy" and "yyMmmdd" are the UTC of the next morning, eg, "G96/2020/20Nov13". See section 4 below for the file naming convention. The "#" column is the number of each data product to expect, though various pipeline exceptions can result in files being missing. Most image classes are contingently optional, at least occasionally. The "#" code letters are:

- E = number of exposures in the night, typically 0-800
- F = number of fields observed, typically 0-200
- D = number of NEO discoveries, varies night to night
- G = variable depending on several factors

## 2.1 Catalina Bundle

| Bundle | LID |
|---|---|
| Catalina | urn:nasa:pds:gbo.ast.catalina.survey |

| Product | Format | Description | # | Pipeline? |
|---|---|---|---|---|
| Product_Bundle.xml | XML | Bundle XML Label | 1 | no |

## 2.2 Raw Data

Raw data for the Catalina Sky Survey are large numbers of short exposure FITS images from CCD (or perhaps in the future, CMOS) cameras. CSS survey cameras are widefield from about 1 square degree up to 19 square degrees. CSS follow-up cameras are typically a fraction of a square degree, though all cameras and telescopes can serve both functions, eg, a widefield survey telescope may be used for a targeted follow-up of a short-arc object whose location is highly uncertain. Survey exposures are taken at sidereal rates with trailing asteroids, while follow-up exposures are often at asteroid rates with trailing stars. Multiple exposures separated by seconds or minutes are used to detect moving objects.

| Collection | LID |
|---|---|
| Data_Raw | urn:nasa:pds:gbo.ast.catalina.survey:data_raw |

| Product | Format | Description | # | Pipeline? |
|---|---|---|---|---|
| Collection_Data_Raw.xml | XML | Collection XML Label | 1 | TBD |
| Collection_Data_Raw.tab | CSV | Collection inventory | 1 | TBD |

| Subdirectory: | tel/yyyy/yyMmmdd | | | |
|---|---|---|---|---|
| **Product** | **Format** | **Description** | **#** | **Pipeline?** |
| Filename.fits | FITS | Raw camera exposures | E | yes |

### 2.3 Partially Processed Data

The CSS pipeline produces a variety of partially processed data products that are input to all later algorithms. These may be directly useful for diverse purposes and also support later reprocessing.

| Collection | LID |
|---|---|
| Data_Partially_Processed | urn:nasa:pds:gbo.ast.catalina.survey:data_partially_processed |

| Product | Format | Description | # | Pipeline? |
|---|---|---|---|---|
| Collection_Data_Partially_Processed.xml | XML | Collection XML Label | 1 | TBD |
| Collection_Data_ Partially_Processed.tab | CSV | Collection inventory | 1 | TBD |

| Subdirectory: | tel/yyyy/yyMmmdd | | | |
|---|---|---|---|---|
| **Product** | **Format** | **Description** | **#** | **Pipeline?** |
| Filename.calb | FITS | CCD calibrated | E | yes |
| Filename.pass1 | FITS | Pass1 images | E | yes |
| Filename.csub | FITS | Difference images | E | yes |
| Filename.avgs | FITS | T&S coadd, sidereal | E | yes |
| Filename.avgr | FITS | T&S coadd, asteroid rate | E | yes |

### 2.4 Pipeline Calibrated Data

The pipeline-calibrated images supply the pixels for source detection and all that follows. Legacy data sets may begin with the calibrated arch files as raw data were not retained in the early years of the survey.

| Collection | LID |
|---|---|
| Data_Calibrated | urn:nasa:pds:gbo.ast.catalina.survey:data_calibrated |

| Product | Format | Description | # | Pipeline? |
|---|---|---|---|---|
| Collection_Data_Calibrated.xml | XML | Collection XML Label | 1 | TBD |
| Collection_Data_Calibrated.tab | CSV | Collection inventory | 1 | TBD |

| Subdirectory: | tel/yyyy/yyMmmdd | | | |
|---|---|---|---|---|
| **Product** | **Format** | **Description** | **#** | **Pipeline?** |
| Filename.arch | FITS | Calibrated exposures | E | yes |

## 2.5 Derived Data Products

Derived data products are very diverse in the Catalina Sky Survey workflow. The basic strategy of the CSS pipeline is founded on building a catalog of point sources for each image using the third party Sextractor software. See the document CSS_operations.pdf for details. Various techniques correlate and classify the point source catalogs to identify candidate moving objects and match them against known asteroid ephemerides. Human validation and machine heuristics identifies candidate Near-Earth Asteroids.

| Collection | LID |
|---|---|
| Data_Derived | urn:nasa:pds:gbo.ast.catalina.survey:data_derived |

| Product | Format | Description | # | Pipeline? |
|---|---|---|---|---|
| Collection_Data_Derived.xml | XML | Collection XML Label | 1 | TBD |
| Collection_Data_Derived.tab | CSV | Collection inventory | 1 | TBD |

**Subdirectory:** **tel/yyyy/yyMmmdd**

| Product | Format | Description | # | Pipeline? |
|---|---|---|---|---|
| Filename.sext | ASCII | Bright sources | E | yes |
| Filename.sexb | FITS (bintable) | Deep sources | E | yes |
| Filename.iext | ASCII | Diff. sources | E | yes |
| Filename.strp | ASCII | Field catalog | E | yes |
| Filename.strm | ASCII | Field catalog | E | yes |
| Filename.scmp | ASCII | SCAMP output | E | yes |
| Filename.ephm | ASCII | Ephemeris | F | yes |
| Filename.mtds | ASCII | MTD objects | F | yes |
| Filename.mtdf | ASCII | Filtered objects | F | yes |
| Filename.dets | ASCII | Asteroid detections | F | yes |
| Filename.hits | ASCII | Hits | F | yes |
| Filename.rjct | ASCII | Rejects | F | yes |
| Filename.mpcd | ASCII | MPC batch | F | yes |
| Filename.neos | ASCII | NEOs and NEO candidates | D | yes |
| Filename.fail | ASCII | Failed astrometry | D | yes |
| Filename.ast | ASCII | Astrometry | D | yes |
| Filename.arch_h | ASCII | Text FITS header | E | yes |
| Filename.tssexb | FITS (bintable) | T&S sextractor output | G | yes |
| Filename.avgrsexb | FITS (bintable) | Rate coadd sexb output | G | yes |
| Filename.sexs | ASCII | rate coadd sext output | G | yes |
| Filename.detl | ASCII | input detection list | D | yes |
| Filename.detf | FITS (bintable) | merge detection metadata | F | yes |
| Filename.detb | FITS (bintable) | batch detection metadata | D | yes |
| Filename.mrpt | ASCII | IAU 80-column astrometry submission | D | yes |
| Filename.ades | XML | ADES astrom submission | D | yes |

| Subdirectory: | tel/yyyy/yyMmmdd | | | |
|---|---|---|---|---|
| **Product** | **Format** | **Description** | **#** | **Pipeline?** |
| Filename.param | ASCII | Parameters | G | no |
| Filename.params | ASCII | Input parameters | G | no |
| Filename.outparams | ASCII | Output parameters | G | no |
| Filename.json | ASCII | JSON | D | no |
| Filename.log | ASCII | Logs | D | no |
| Filename.point | ASCII | Pointing | 1 | no |
| Filename.signature | ASCII | Signature | D | no |
| Filename.xmls | ASCII | SCAMP params | E | no |
| Filename.cov | ASCII | Survey coverage | 1 | no |
| Filename.txt | ASCII | TBD | G | no |
| Filename.md5 | ASCII | MD5 file hashes | 1 | no |

### 2.6 Calibrations

CCD data reductions in the Catalina Sky Survey are similar to other telescopes, but NEO data in general are distinguished by large pixels and a requirement for efficient logistics above almost everything else. CSS relies on pragmatic focusing, queue-based data-taking, flat fields selected exposure-by-exposure from a flat field library, and in general anything that streamlines operations. Flat fields from the library that were used on a given night are copied into the nightly data directory.

| Collection | LID |
|---|---|
| CALIBRATION | urn:nasa:pds:gbo.ast.catalina.survey:calibration |

| Subdirectory: | tel/yyyy/yyMmmdd | | | |
|---|---|---|---|---|
| **Product** | **Format** | **Description** | **#** | **Pipeline?** |
| *flat*.fits | FITS | Flat field images | varies | yes |

### 2.7 Documents

A variety of documents round out the bundle to capture input parameters, pointing information, survey coverage, MD5 checksums, etc.

| Collection | LID |
|---|---|
| DOCUMENT | urn:nasa:pds:gbo.ast.catalina.survey:document |

| Product | Format | Description | # | Pipeline? |
|---|---|---|---|---|
| Collection_Document.xml | XML | Collection XML Label | 1 | TBD |
| Collection_Document.tab | CSV | Collection inventory | 1 | TBD |
| CSS_bundle_overview.pdf | PDF/A | This document | 1 | No |
| CSS_history.pdf | PDF/A | History of the Catalina Sky Survey | 1 | No |

| | | | | |
|---|---|---|---|---|
| CSS_telescopes.pdf | PDF/A | CSS Telescopes and Instrumentation | 1 | No |
| CSS_operations.pdf | PDF/A | CSS Operations and Processing | 1 | No |

# 3 Catalina Sky Survey Documents

- *Overview of Catalina Sky Survey PDS Archive*, Seaman, Neese, Stone, and Christensen, 13 Nov 2020 (this document)
- *History of the Catalina Sky Survey*, CSS team, 12 Nov 2020
- *Catalina Sky Survey Telescopes and Instrumentation,* CSS team, 12 Nov 2020
- *Catalina Sky Survey Operations and Processing*, CSS team, 12 Nov 2020

# 4 CSS Data Holdings and File Naming

The Catalina Sky Survey (CSS) acquires optical imaging data of the night sky of both standardized survey fields and targeted follow-up to support the discovery of Near-Earth Objects (NEOs) as well as incidental astrometric observations of asteroids, comets, and other solar system objects. Fields are imaged multiple times (almost always four times) in support of moving object detection against the background stars. All subsequent data products, and their names, derive from four initial raw CCD images:

```
G96_20200927_2B_N20001_01_0001.fits.fz[3]
G96_20200927_2B_N20001_01_0002.fits.fz
G96_20200927_2B_N20001_01_0003.fits.fz
G96_20200927_2B_N20001_01_0004.fits.fz
```

In survey mode these are typically 30 second exposures, acquired about 7 minutes apart, interleaved with observations of other adjoining fields. Follow-up observations vary in exposure time depending on the expected brightness of the targets and are often taken in rapid sequence.

An alternate observing mode for follow-ups is to "track and stack" in which each of the four frames is acquired as separate frames that are coadded at the expected rate of motion of the targeted object. In that case there will be additional raw images – two dozen or more are not unusual. Several of the file types discussed below will only appear for track and stack observations.

Archival holdings at CSS's University of Arizona headquarters are universally in a directory structure: /archive/TEL/YYYY/NIGHT[4], where TEL is one of the MPC codes listed below, YYYY is the calendar year, and NIGHT is of the form: YYMmmDD, for example:

```
/archive/G96/2020/20Sep27
```

CSS image names and derived data products within these directories are comprised of several segments separated by underscores with an appended extension:

[3] FITS images are FPACK compressed at the telescope during the night. This leaves the headers readable.

[4] In practice, annual datasets are hosted in ZFS pools in a growing mass datastore that is replicated between the CSS offices at the Lunar and Planetary Laboratory and the campus data center, UITS.

- Telescope – three letter MPC[5] site code
- Night – UT date at the end of the night, YYYYMMDD
- Binning – CCD pixel binning, eg, 1B, 2B, or 4B
- Field – field ID, Xnnnnn where X is N, S, F, or U
- Repetition – 1 for the first time a field is imaged per night, incremented if repeated
- Sequence – sequence number of exposures, 1-4 (or more for track-and-stack)
- Extension – "fits" for the raw data or many other options (see below)
- Compression – "fz" for FITS tile compression, "gz" for gzip, or absent (optional)

The enumerated CSS MPC codes are our four survey telescopes:

- 703 – the original Catalina Sky Survey (CSS) 0.7m Schmidt on Mount Bigelow
- G96 – the Mount Lemmon Survey (MLS) 60" telescope
- E12 – the Siding Spring Survey (SSS) using the 0.5m Uppsala Schmidt (retired in 2013)
- V00 – the Kitt Peak 90" Bok telescope, when used for NEO surveying in partnership with Spacewatch and the University of Minnesota

And two follow-up[6] telescopes:

- I52 – 40" follow-up telescope next to G96 on Mount Lemmon
- V06 – the 61" Kuiper telescope on Mount Bigelow, when used by CSS for follow-up

CSS survey field nomenclature is N for fields north of the celestial equator and S for southern fields. The first two digits are rounded to a grid in degrees of declination. The final three digits are a grid in degrees of right ascension. Each telescope has a different set of standard fields depending on their field of view. For example, the Schmidt has a larger field of view so there are fewer standard field centers.

Follow-up nomenclature has evolved and for most purposes can be regarded as "F" (for follow-up) and "U" (for user) followed by 5 random but unique characters. The distinction between follow-up and user fields has largely vanished, and currently the 5 characters are assigned as a subset of the temporary object designation as an aid to the observers. Temporary designations have also evolved, sometimes rapidly, and can be best regarded as random unique strings, generally 7 characters, assigned by the different NEO surveys using various schemes.[7]

There is a lengthy list of CSS file extensions, generated for a variety of purposes leading ultimately to submitting asteroid and comet astrometry to the IAU Minor Planet Center. These may or may not be encountered in gzip compressed form (with an appended ".gz" extension) or FITS tile compressed (FPACK) format (with a ".fz").

[5] The IAU Minor Planet Center, https://minorplanetcenter.net

[6] All CSS telescopes can operate in both survey and follow-up mode, the division of labor depends on several factors, most critically the field of view of the camera. NEO survey cameras are very wide field.

[7] CSS designations currently begin with "C" followed by a 5-character high-order base running number and a trailing digit indicating the telescope: 1 for 703, 2 for G96, and so forth. These are arbitrary 7-character strings for the purposes of this document.

| *Ext* | *Format* | *Collection* | *Description* |
|---|---|---|---|
| *fits* | FITS | *Data_Raw* | Raw CCD exposures (not for legacy data) |
| *calb* | FITS | *Data_Partially_Processed* | Basic CCD calibrated |
| *pass1* | FITS | | Pass 1 images (will not usually appear) |
| *csub* | FITS | | Difference images |
| *avgs* | FITS | | Sidereal track-and-stack coadds |
| *avgr* | FITS | | Asteroid rate T-and-S coadds |
| *arch* | FITS | *Data_Calibrated* | Fully calibrated images |
| *sext* | ASCII | *Data_Derived* | Bright sextractor[8] source lists |
| *sexb* | bintable[9] | | Deep sextractor source lists (binary) |
| *iext* | ASCII | | Sources from difference imaging |
| *strp* | ASCII | | Field catalog list |
| *strm* | ASCII | | Field matches |
| *ephm* | ASCII | | Ephemeris predictions for field |
| *mtds* | ASCII | *Data_Derived* | Moving Target Detection sources |
| *mtdf* | ASCII | | Filtered and annotated MTD list |
| *dets* | ASCII | | Candidate asteroid detections |
| *hits* | ASCII | | Hits of detections against ephemeris |
| *rjct* | ASCII | | Objects rejected during validation |
| *mpcd* | ASCII | | Non-NEO asteroid astrometry |
| *neos* | ASCII | | NEO asteroid astrometry |
| *fail* | ASCII | | Failed astrometry (various reasons) |
| *arch_h* | ASCII | | Standalone FITS headers |
| *tssexb* | bintable | | T-and-S sextractor output |
| *avgrsexb* | bintable | | Rate coadd sextractor output (binary) |
| *sexs* | ASCII | | Rate coadd sextractor output (text) |
| *detf* | bintable | | Merged per-detection metadata |

Files with the above extensions will have names meeting the criteria in the first paragraph, but not all file extensions are created for all four (or more) raw FITS images.[10] Per-image processing includes image calibration and source extraction, while per-field processing applies to steps after the initial moving target detection, for example the merged per-detection metadata results in one FITS binary table per four raw input images:

```
G96_20200927_2B_N20001_01_0001.detf
```

[8] https://www.astromatic.net/software/sextractor

[9] FITS binary table format. These will not be compressed.

[10] See bundle spread sheet.

In addition, five of the file types from the Data_Derived collection have slight name variations:

| *Ext* | *Format* | *Collection* | *Description* |
|---|---|---|---|
| *scmp* | ASCII | *Data_Derived* | SCAMP output |
| *detl* | ASCII | | Input detection list for ADES generation |
| *detb* | bintable | | Batched per-detection metadata |
| *mrpt* | ASCII | | Old ("80-column") astrometry submission |
| *ades* | XML | | ADES astrometry submissions |

The CSS pipeline relies on SCAMP (Software for Calibrating AstroMetry and Photometry)[11] to compute FITS world coordinates to each image individually, and as needed to coadded images. SCAMP also is the first step in CSS photometric calibration. SCAMP appends an "_1" to input files names since CSS images have a single FITS extension.

```
G96_20200927_2B_N20001_01_0001_1.scmp.gz
```

The final four files are only created by the pipeline if NEOs are discovered (or recovered) in a field. In this case the detection number (out of many candidates) is appended along with "neos" since this detection comes from the human-validated NEO astrometry in the neos file:

```
G96_20200927_2B_N20001_01_0001.29.neos.detl
G96_20200927_2B_N20001_01_0001.29.neos.detb
G96_20200927_2B_N20001_01_0001.29.neos.mrpt
G96_20200927_2B_N20001_01_0001.29.neos.ades
```

At the very end of the night up to two batches[12] (sometimes very large) are created of "incidental" astrometry (automated discovery of non-NEOs or comets, generally). In this case the naming scheme is truncated because the discoveries / recoveries pertain to the night as a whole:

```
G96_20200927_2B.mpcd.detl     G96_20200927_2B.neos.detl
G96_20200927_2B.mpcd.detb     G96_20200927_2B.neos.detb
G96_20200927_2B.mpcd.mrpt     G96_20200927_2B.neos.mrpt
G96_20200927_2B.mpcd.ades     G96_20200927_2B.neos.ades
```

In addition to the per-image, per-field, and per-detection files described above, there are a variety of mostly text documents that describe the nightly observing and data processing as a whole. The complete list of these has varied over the years and CSS will archive such files as individual documents. On the other hand, CSS image calibrations are on the extreme end of pragmatic and flat fields will generally be the only calibration files archived.

[11] https://www.astromatic.net/software/scamp

[12] The incidental neos and/or mpcd files are only generated if each such kind of detections were made. There are various reasons for both survey and follow-up mode why this may not happen.

| *Ext* | *Format* | *Collection* | *Description* |
|---|---|---|---|
| *fits* | FITS | *Data_Calibration* | Flat fields |
| *param* | ASCII | *Document* | Pipeline and other parameters |
| *params* | ASCII | | Input parameters |
| *outparams* | ASCII | | Output parameters |
| *json* | ASCII | | JSON parameters |
| *log* | ASCII | | Processing logs |
| *point* | ASCII | | Telescope pointing (infrequent) |
| *signature* | ASCII | | Signature |
| *xmls* | ASCII | | SCAMP parameters |
| *cov* | ASCII | | Survey coverage |
| *md5* | ASCII | | MD5 file hashes |