

Kepler Data Products Overview

The Kepler project produced a wide range of data products and documentation to assist the community in using data from the mission to help study both exoplanets and other astrophysical areas. These data range from engineering data on the detectors to lists of planet candidates, and includes many of the intermediate data products in the data processing path. The data for Kepler are hosted in two locations:

- the [Mikulski Archive for Space Telescopes \(MAST\)](#) at the [Space Telescope Science Institute \(STScI\)](#) hosts **time series data and spacecraft calibration products**, and
- the [NASA Exoplanet Archive](#) at the [NASA Exoplanet Science Institute \(NExSci\)](#) hosts **data related to the Kepler mission planet searches**.

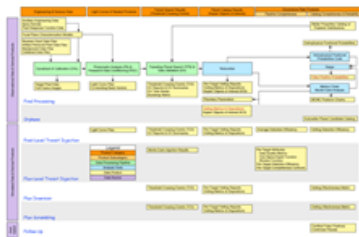
There are over 40 different kinds of data products. This document contains a brief description of the products hosted at both archives, provides a comprehensive overview of the products, and provides a single location with links to all products and their documentation. You can also [download this page as PDF](#).

How to Use This Page

The first six sections are broad categories of data, while the remaining sections are individual data products. Each one of these sections includes links to both the data and the documentation. If you know the name of the product you are interested in, use the navigation menu to the right to jump directly to that section. To find all the products related to a certain area (e.g., injection), you can use your browser's Find function (Ctrl-F or Command-F) to locate all relevant products.

The following diagram illustrates the categorization of the data products:

[\(Click to enlarge\)](#)



Broad Data Categories

[Top](#)

Engineering & Science Data

Definition: The engineering and science data category includes all *raw* and *calibrated pixel-level data* collected during the Kepler mission, as well as some *navigational information*, *engineering* and *commissioning* data, and specialized data sets used for calibration (i.e., reverse clock, artifact removal pixel, background, and collateral).

Skip to a section:

Broad Data Categories

[Engineering & Science Data](#)
[Light Curves & Related Products](#)
[Transit Search Results \(TCEs\)](#)
[Planet Candidate Results \(KOI\)](#)
[Occurrence Rate Products for Pipeline Completeness](#)
[Occurrence Rate Products for Catalog Completeness & Reliability](#)

Individual Data Products

[Ancillary Engineering Data](#)
[SPICE Kernels](#)
[Pixel Response Function Data](#)
[Focal-Plane Characterization Models](#)
[Reverse Clock Data Files](#)
[Artifact Removal Pixel Data Files](#)
[Background Data Files](#)
[Collateral Data Files](#)
[Target Pixel Data](#)
[Full-frame Images](#)
[Light Curves Files: Nominal Processing](#)
[Co-trending Basis Vectors](#)
[Light Curve Files: Pixel-level Transit Injection](#)
[Threshold-Crossing Events \(TCE\): Nominal Processing](#)
[DV Reports & DV Summaries: Nominal Processing](#)
[DV Time Series](#)
[Statistical Bootstrap Metric](#)
[Threshold-Crossing Events \(TCE\): Pixel-level Transit Injection](#)
[DV Reports & DV Summaries: Pixel-level Transit Injection](#)
[Threshold-Crossing Events \(TCE\): Flux Inversion](#)
[Threshold-Crossing Events \(TCE\): Flux Scrambling](#)
[TCERT Vetting Forms: Nominal Processing](#)
[TCERT Metrics & Dispositions: Nominal Processing](#)
[Kepler Objects of Interest \(KOI\): Nominal Processing](#)
[Planetary Parameters: Nominal Processing](#)
[TCERT Metrics & Dispositions: Supplemental Data Validation](#)
[Kepler Objects of Interest \(KOI\): Supplemental Data Validation](#)
[TCERT Vetting Forms: Pixel-level Transit Injection](#)
[TCERT Metrics & Dispositions: Pixel-level Transit Injection](#)
[TCERT Vetting Forms: Flux Inversion](#)
[TCERT Metrics & Dispositions: Flux Inversion](#)
[TCERT Vetting Forms: Flux Scrambling](#)
[TCERT Metrics & Dispositions: Flux Scrambling](#)
[Pipeline Detection Efficiency](#)
[Analytic Completeness Model](#)
[Numerical Completeness Model](#)
[Completeness Model Comparison & Limitations](#)
[Stellar Properties Catalog](#)
[Autovetter Planetary Candidate Catalog](#)
[Astrophysical Positional Probabilities](#)
[Astrophysical False Positive Probabilities](#)
[MCMC Posterior Chains](#)
[TCERT Detection Efficiency](#)
[TCERT Reliability Metric](#)
[Certified False Positives](#)
[Confirmed Planets](#)

For most users, the calibrated pixels in the [target pixel files](#) and [full-frame images](#) are likely to be the products of most interest. The other products in this category are for advanced users who wish to perform their own pixel-level calibrations starting with the raw data. The [focal-plane characterization models](#), [Kepler Data Processing Handbook](#), and the [data processing pipeline code](#) will provide guidance for such endeavors.

[Top](#)

Light Curves & Related Products

Definition: The light curves and related products category includes the uncorrected (i.e., simple-aperture photometry) and corrected (i.e., pre-search data conditioned) light curves obtained during the Kepler mission. These long- and short-cadence [light curve files](#) are the primary data products for many scientific investigations, particularly those unrelated to the search for exoplanets.

This category also includes [co-trending basis vectors](#), which may prove useful in removing systematic trends from light curves created from the [target pixel data](#) and a set of [light curves with injected transit signatures](#). These injected light curves should not be used for general science investigations, but are provided for evaluating individual objects and for characterizing the performance of transit search pipelines.

[Top](#)

Transit Search Results (Threshold-Crossing Events)

Definition: The transit search results category includes the products associated with Kepler's search of the pre-conditioned light curves for transit-like signatures. The signatures, which are known as "[Threshold-Crossing Events](#)," are evaluated in the [DV reports & DV summaries](#). The quarter-stitched, gap-filled [time series](#) that were searched by TPS & DV are available, as are the results of the [bootstrap statistical test](#).

In addition to a complete set of products for the nominally processed flight data, this category includes DV reports and DV summaries for the evaluation of [transits injected](#) at the pixel-level and for those [Kepler objects of interest](#) that were found by earlier versions of the Kepler data processing pipeline, but were not found by later versions (i.e., the orphans). This category also provides the Threshold-Crossing Events identified by pipeline searches of several synthetic data sets (i.e., [pixel-level transit injection](#), [flux inversion](#), and [flux scrambling](#)) that were used to assess our ability to distinguish planet candidates and false positives.

[Top](#)

Planet Candidate Results (Kepler Objects of Interest)

Definition: The planet candidate results category includes the products associated with Kepler's planet candidate catalogs that were created by dispositioning the [transit search results](#). Specifically, those Threshold-Crossing Events that appear sufficiently transit-like to be astrophysical in origin are promoted to become "[Kepler objects of interest](#)" and are then subjected to further scrutiny to determine their true nature (i.e., their [disposition](#)). For each Kepler object of interest, some additional metrics are computed for use in dispositioning them as "planet candidates" or "false positives," and additional fits are performed to infer their [planetary parameters](#). For convenience, many of the metrics are concatenated into [vetting forms](#) that permit manual examination of individual objects and [electronic form](#) for use as input to the Robovetter.

In addition to a complete set of results for the nominally processed flight data, this category includes similar products for several synthetic data sets (i.e., [pixel-level transit injection](#), [flux inversion](#), and [flux scrambling](#)) that were used to characterize [pipeline](#) and [Robovetter](#) performance.

[Top](#)

Occurrence Rate Products for Pipeline Completeness

Definition: The first category of occurrence rate products provides estimates of pipeline completeness for use in occurrence rate calculations. This category describes how to combine the [average detection efficiency](#) with either analytic or numerical models of the noise properties and window function effects to produce two-dimensional contour plots of detection efficiency for individual stars as a function of their planetary radius and orbital period. The other products in this category include [empirical completeness results](#), which provide ground-truth for selected stars, the [Kepler PORTs](#) tools, which can be used to generate contour plots, and a discussion of the limitations of the analytic and numerical models based on comparisons with the empirical results.

[Top](#)

Occurrence Rate Products for Catalog Completeness & Reliability

Definition: The second category of occurrence rate products contains a menagerie that may be used to compute occurrence rates. It includes essential ingredients, such as the [stellar properties catalog](#), which is propagated into the [DV reports & DV summaries](#) and the

on-line [Kepler objects of interest](#) catalog. Also included are the [TCERT detection efficiency](#), which measures the Robovetter pass rate for injected transits (i.e., the completeness), the [TCERT effectiveness metric](#), which measures the Robovetter fail rate for inverted transits (i.e., the effectiveness), and the [MCMC posterior chains](#), which allow the errors associated with the planetary fits to be propagated forward to produce errors on the resulting occurrence rates.

It is worth noting that these products are mature and their use in occurrence rate calculations is straightforward. However, the utilization of the remaining products—[autovetter planet candidate catalog \(DR24 only\)](#), [astrophysical positional probabilities](#), [false positive probabilities](#), [certified false positives](#), and [planetary systems/confirmed planets](#)—is more complex. Though each provides insight into the completeness and reliability of the final Kepler object of interest catalog, the best way to interpret and apply them to deduce occurrence rates was still an active area of research at the time of Kepler mission close out.

Individual Data Products

[Top](#)

Ancillary Engineering Data

Definition: Ancillary engineering data are a subset of the engineering telemetry collected by the Kepler spacecraft that has been archived for flight system characterization and science data analysis.

Location: Go to [Kepler Full Frame Image \(FFI\) & Engineering Data Search](#) (MAST) and specify an Archive Class of “Ancillary Engineering Data (ENG).” The same search interface is available from the common entry point for Kepler ancillary files at [Kepler Ancillary Files](#) (MAST).

Format: There is one FITS file per long-cadence data collection interval. The files have roughly 200 extensions, one for each reported quantity.

Inputs: On-orbit engineering data collected by the spacecraft.

Outputs: These data are no longer used by Kepler data processing pipeline, as the *co-trending basis vectors* are more effective at removing systematics.

Documentation: Kepler Archive Manual (KDMC-10008-006), §2.3.10, with additional on-line material at [Ancillary Engineering Data Files](#) (MAST).

Data Releases: The ancillary engineering data are downloaded once each quarter, but are not processed, so the data-release concept does not apply.

[Top](#)

SPICE Kernels

Definition: The SPICE kernels contain data specific to Kepler that permit precise calculations of the spacecraft ephemerides. More generally, SPICE is a geometry information system used for planning and interpreting scientific observations from space-based instruments aboard robotic planetary spacecraft.

Location: [Kepler SPICE Index](#) (MAST) or the common entry point for Kepler ancillary files at [Kepler Ancillary Files](#) (MAST).

Format: bsp and tsc binary files

Inputs: On-orbit measurements of spacecraft position

Outputs: Accurate Barycentric times for cadences

Documentation: See [JPL NAIF page](#)

Data Releases: Provided files apply through end of mission.

[Top](#)

Pixel Response Function Data

Definition: The Pixel Response Function (PRF) Data are 121 dithered, long cadences collected during spacecraft commissioning to measure the PRF. The PRF combines the optical point spread function (PSF), jitter power spectral density (PSD), variations in focus across the field, CCD response function, and electronic impulse response ([KIH §3.5](#)).

Location: See [Commissioning Data](#) at MAST

Format: One FITS file per target in the same format as the target pixel data ([KAM §2.3.2](#)); note that this data set is unusual in that it is intentionally dithered, so simple aperture photometry does not produce a traditional light curve.

Inputs: On-orbit data collected once during commissioning

Outputs: A piecewise continuous PRF model used by the Kepler data processing pipeline and approximated by a discrete PRF model (*KAM §2.3.5.17*) created using SOC 5.0.

Documentation: *KIH §3.5*, *KDCH §5.2*, and *KSCI-19112-001*.

Data Releases: DR25 (though the data was taken in April, 2009, the released processing uses the final Kepler data processing pipeline)

[Top](#)

Focal-Plane Characterization Models

Definition: The focal-plane characterization (FC) models describe the properties of the focal plane, thereby providing calibration information for the Kepler data processing pipeline and descriptions of phenomena that occur in the data.

Location: *Kepler Focal Plane Characteristics* (MAST) or the common entry point for Kepler ancillary files at *Kepler Ancillary Files* (MAST)

Format: Varied (see table below for details and *KAM Table 2-2* for file version number)

FC Model	Section (§)	File Name
clock-state-mask	1	clock_state_mask_KADN-26205_20081218.fits
2-D black	2	kplr2015238_2dblack.fits, kplr2015238_2dblack-uncert.fits
dynamic black	3	black-retrieval-tool directory, kplr<YYYYDOYHHMMSS>-q<##>-<mmo>-dr<##>_dynablack.mat ("blob file")
1-D black	4	extract_one_d_black_from_blob.m, kplr<YYYYDOYHHMMSS>-q<##>-<mmo>-dr<##>_1dblack.mat
scene-dependent rows	5	is_scene_dependent.m kplr2015279231411-dr25_blackrowmask.txt
bad smear columns	6	is_saturated_smear_column.m
read noise	7	kplr2008102416_read-noise.txt
gain	8	kplr2008072318_gain.txt
linearity	9	kplr2009060215_linearity.txt
undershoot	10	kplr2008102809_undershoot.txt
large-scale flat	11	kplr2015236_largeflat.fits
small-scale flat	12	kplr2015236_smallflat.fits, kplr2015236_smallflat-uncert.fits
invalid pixels	13	n/a
pointing	14	kplr2013050211_pointing.txt
roll-time	15	kplr2013040800_rolltime.txt
focal-plane geometry	16	kplr2009060418_geometry.txt
simplified (RA, DEC) to pixel mapping	14-16	morc_2_ra_dec_4_seasons.xls

pixel response function	17	kplr<module>.<output>_<yyyyddd>.fits
video crosstalk	18	kplr_2015111701_video_crosstalk_rec_src.csv
spectral response	19	kplr_2015120120_spectral_response.csv

Inputs: Ground-based pre-launch test data and on-orbit commissioning data

Outputs: Calibrated target pixel data, full-frame images, and light curves

Documentation: [KAM §2.3.5](#)

Data Releases: The focal-plane characterization models were released once at the end of the mission. They have no specific release number because they apply to all data releases.

[Top](#)

Reverse Clock Data Files

Definition: The reverse clock data files periodically capture on-orbit bias measurements using the science CCDs with their parallel clocks reversed so as to sweep current away from the output amplifiers (recall that there is no shutter on Kepler).

Location: Go to [Kepler Full Frame Image \(FFI\) & Engineering Data Search](#) (MAST) and specify an Archive Class of “Reverse Clock (RVC)”.

Format: A few FITS per month before/after science data collection

Inputs: On-orbit data collected by the spacecraft

Outputs: Used by dynablock to calibrate the pixel-level data

Documentation: [KAM §2.3.8](#), [KIH §6.8](#), and Kolodziejczak, J. J., et al. 2010, *Proc. SPIE* 7742, High Energy, Optical, and Infrared Detectors for Astronomy IV, 77421G; doi:10.1117/12.857637

Data Releases: Since the reverse clock data are packaged after download from the spacecraft, but not processed, the concept of data release numbers does not apply.

[Top](#)

Artifact Removal Pixel Data Files

Definition: The artifact removal pixel data files contain flux values for a set of individual collateral pixels read-out at the long cadence rate to monitor the crosstalk levels in both the pre-clocked bias region (i.e., “leading black”) and the over-scanned bias region (i.e., “trailing black”).

Location: Go to [Kepler Full Frame Image \(FFI\) & Engineering Data Search](#) (MAST) and specify an Archive Class of “Artifact Removal Pixel.”

Format: FITS (see [KAM Table 2-1](#) for file naming convention and [KAM Table 2-2](#) for file version number). Because of the failure of module 3 (see [KDCH §4.6](#)), there are 84 files/quarter for Q0-Q4 and 80 files/quarter for Q5-Q17 for a total of 1460.

Outputs: Used by dynablock to calibrate the [target pixel data](#).

Documentation: [KAM §2.3.8](#), [KIH §6.8](#), [KDPH §5](#), and Kolodziejczak, J. J., et al. 2010, *Proc. SPIE* 7742, High Energy, Optical, and Infrared Detectors for Astronomy IV, 77421G; doi:10.1117/12.857637

Data Releases: 12–25, although only DR25 is directly accessible through the end-of-mission archive. This product has improved over time, so later releases are preferred. Only DR24 and DR25 provide uniform reprocessing of all Kepler data.

[Top](#)

Background Data Files

Definition: The background data files contain both raw and calibrated long-cadence pixel time series for a grid of 4464 background pixels on each channel. They are used to remove the zodiacal light and unresolved background stars from the data.

Location: Go to [Kepler Full Frame Image \(FFI\) & Engineering Data Search](#) (MAST) and specify an Archive Class of “Background Pixel (BKG)”.

Format: FITS (see [KAM Table 2-1](#) for file naming convention and [KAM Table 2-2](#) for file version number). Because of the failure of module 3 (see [KDCH §4.6](#)), there are 84 files/quarter for Q0-Q4 and 80 files/quarter for Q5-Q17 for a total of 1460.

Inputs: On-orbit data collected by the spacecraft.

Outputs: Calibrated [target pixel data](#)

Documentation: [KAM §2.3.6](#), [KIH §2.6.2.4](#), [KDPH §3.3.3](#), and [KDPH §6.3.2](#).

Data Releases: DR12 to DR25, although only DR25 is directly accessible through the end-of-mission archive. This product improved over time, so later releases are preferred. Only DR24 and DR25 provide uniform reprocessing of all Kepler data.

[Top](#)

Collateral Data Files

Definition: The collateral data files contain the flux time series for the masked and virtual smear columns and trailing black rows surrounding the science portion of each CCD channel (see [KIH Figure 25](#)). These data are available at both long and short cadence; the latter also includes the overlap summed-pixels (black/masked and black/smear; see [KIH Figure 7](#)).

Location: Go to [Kepler Full Frame Image \(FFI\) & Engineering Data Search](#) (MAST) and specify an Archive Class of “Long Cadence Collateral (CLL)” or “Short Cadence Collateral (CLS)”.

Format: FITS (see [KAM Table 2-1](#) for file naming convention and [KAM Table 2-2](#) for file version number). Because of the failure of module 3 (see [KDCH §4.6](#)), the long-cadence data has 84 files/quarter for Q0-Q4 and 80 files/quarter for Q5-Q17 for a total of 1460 files. Similarly, the short-cadence data set consists of 3917 files (out of a possible 3956) because some channels contained no short-cadence targets during some observing months.

Inputs: On-orbit data collected by the spacecraft

Outputs: Calibrated target pixel data

Documentation: [KAM §2.3.7](#), [KIH §2.6.3](#), [KDPH §5.4](#)

Data Releases: DR12 to DR25, although only DR25 is directly accessible through the end-of-mission archive. This product improved over time, so later releases are preferred. Only DR24 and DR25 provide uniform reprocessing of all Kepler data.

[Top](#)

Target Pixel Data

Definition: The target pixel data are the raw- and calibrated-pixel time series for each individual target observed in a given long-cadence quarter or short-cadence month.

Location: See [Download Options for Kepler Data and Catalogs](#) (MAST) for several download options. Some target pixel data have no corresponding light curve; see [KAM §2.1.1](#); the converse should not be true.

Format: FITS (see [KAM Table 2-1](#) for file naming convention and [KAM Table 2-2](#) for file version number).

Inputs: On-orbit data collected by the spacecraft

Outputs: Light curve files

Documentation: [KAM §2.3.2](#) for file format and keywords, [KDRN](#) for release notes, entire [KDCH](#) for data characteristics, [KDPH §6](#) for photometry, [KDPH §7](#) for data conditioning, and [KIH](#) for instrument details. A list of available long- and short-cadence months is available in the [KDCH](#); those processed in each data release are listed in the associated [KDRN](#).

Data Releases: DR9 to DR25, although only DR25 is directly accessible through the end-of-mission archive. This product improved over time, so later releases are preferred. Only DR24 and DR25 provide uniform reprocessing of all Kepler data.

[Top](#)

Full-frame Images

Definition: The full-frame images contain all 1132 x 1070 pixels for all 84 channels in the focal plane for the equivalent of one long-cadence time interval (i.e., 29.4 minutes). Several FFIs were taken each quarter and used to confirm the proper orientation and placement of the detector on the sky and to assess photometer health.

Location: Go to [Kepler Full Frame Image \(FFI\) & Engineering Data Search](#) (MAST) and specify an Archive Class of “Full-Frame Images (FFI).” Or, go to the or the [MAST portal](#). Additional download options are located at [Download Options for Kepler Data and Catalogs](#) (MAST).

Format: FITS (see [KAM Table 2-1](#) for file naming convention and [KAM Table 2-2](#) for file version number). The raw, calibrated, and uncertainty data are archived in separate files.

Inputs: On-orbit data collected by the spacecraft

Outputs: Used by dynablock to calibrate the target pixel data

Documentation: [KAM §2.3.3](#) for file format and keywords, [KIH §2.8](#) for sample images, KDRN for release notes, entire KDCH for data characteristics, and [KDPH §5](#) for pixel calibration. The full set of collected FFIs is listed in the KDCH; those processed in each data release are listed in the associated KDRN.

Data Releases: DR1 to DR25, although only DR25 is directly accessible through the end-of-mission archive with one exception: the eight Q0 (golden) FFIs and the first Q2 FFI were last processed in DR21 and remain available (see [KDRN 24 §A.1.7](#) for an explanation).

[Top](#)

Light Curve Files: Nominal Processing

Definition: Light curve files are produced for individual targets using simple aperture photometry (SAP_FLUX) and systematic errors are removed (PDCSAP_FLUX) ([KAM §2.3.1](#)).

Use Case: Light curve files are posted to the archive for two fundamentally different use cases: (a) nominal processing and (b) pixel-level transit injection.

These particular light curve files are from Use Case (a)—the result of nominal end-to-end operations processing of Kepler flight data with a particular release of the Kepler data processing pipeline. These [light curves](#) and the associated [target pixel data](#) are the primary products for investigation of astrophysical phenomena associated with Kepler-observed targets.

Location: See [Kepler Target Search](#) (MAST) or the [MAST portal](#) for browsing and downloading small sets of light curve files; for bulk download, consult [KAM §3.9](#).

Format: FITS (see [KAM Table 2-1](#) for the file naming convention and [KAM Table 2-2](#) for the file version numbers). For long-cadence targets, there is one file per quarter, and for short-cadence targets, there is one file per month.

Inputs: Raw pixel-level data (i.e., reverse clock data, artifact removal pixels, background pixels, collateral pixels, raw science pixels, raw full-frame images), focal-plane characterization models, and [Kepler Input Catalog](#)).

Outputs: Archived for public use and used by the transiting planet search (TPS) module of the Kepler data processing pipeline to identify potential transit-like events (a.k.a. Threshold-Crossing Events (TCE)).

Documentation: [KAM §2.3.1](#) for file format and keywords, [KDRN](#) for release notes, entire [KDCH](#) for data characteristics, [KDPH §6](#) for photometric analysis (PA), and [KDPH §7](#) for presearch data conditioning (PDC). The full set of collected long and short cadences is listed in the KDCH; those processed in each data release are listed in the associated [KDRN](#).

Data Releases: Q1-Q17 DR25 (only the latest release is generally available)

[Top](#)

Co-trending Basis Vectors

Definition: The co-trending basis vectors (CBVs) are a set of orthonormal basis vectors that represent the correlated features (or systematic trends) in a reference ensemble of light curves for a given quarter and output channel. These CBVs may be used to remove common-mode instrumental effects when the standard PDC light curves are insufficient for a particular science goal or when the [target pixel data](#) are used to generate light curves.

Location: [Kepler CBV Files](#) (MAST) or the common entry point for Kepler ancillary files at [Kepler Ancillary Files](#) (MAST). Older CBVs that are applicable to earlier data releases are at [Kepler CBV Index](#) (MAST).

Format: FITS (see [KAM §2.3.4](#) for the file naming convention and [KAM Table 2-2](#) for the file version numbers). There is one file per quarter with a primary header and 84 data extensions, one for each channel (or mod.out).

Inputs: Raw light curves created using simple aperture photometry (i.e., SAP_FLUX; see [KAM §2.3.1](#))

Outputs: Systematic-error-corrected light curves (i.e., PDCSAP_FLUX; [KAM §2.3.1](#))

Documentation: [KAM §2.3.4](#), [KDCH §7](#), [KDRN 11-25](#), and [KDPH §8.5.1](#)

Data Releases: First released to support DR4 (Q3) and first documented in KDRN 11, these older CBVs remain valid when applied to light curves (or target pixel data) from the same data release. However, this set of products has improved dramatically over time, so later releases are preferred. Only DR24 and DR25 provide uniform reprocessing of all Kepler data.

[Top](#)

Light Curve Files: Pixel-level Transit Injection

Definition: Light curve files are produced for each target using simple aperture photometry (SAP_FLUX) and systematic errors are removed (PDCSAP_FLUX) ([KAM §2.3.1](#)).

Use Case: Light curve files are posted to the archive for two fundamentally different use cases: (a) nominal processing and (b) pixel-level transit injection.

These particular light curve files are from Use Case (b)—the result of injecting simulated transit signatures into the calibrated pixel-level data. These injected light curves are meant to provide ground truth for testing, characterization, and comparison of transit-search pipelines. Note that they contain both astrophysical and synthetic transit events, so they should not be used to identify or characterize astrophysical phenomena, as the injections are likely to perturb the data in ways that are hard to characterize.

Location: [Kepler Simulated Data](#) (NASA Exoplanet Archive) These injected light curves are not available in a searchable table to avoid confusion with the nominally processed light curves from Use Case (a). Instead, they are available as a tarball, assuming that those who want one injected light curve, will want them all.

Format: FITS (see [KAM Table 2-1](#) for the file naming convention and [KAM Table 2-2](#) for the file version numbers). There is one FITS file per quarter for each long-cadence target; no analogous short-cadence product is available.

Inputs: Synthetic transits are spatially distributed using the pixel response function (PRF) and injected into the calibrated pixels (i.e., the output of CAL) for subsequent processing through the photometric pipeline

Outputs: [Threshold-Crossing Events \(TCE\)—pixel-level transit injection](#). Since the injection was done “on the fly,” the associated [DV reports & DV summaries](#) include all the normal pixel-level tests, but the injected target pixel data are not archived. Since only those targets with TCEs are passed to DV, the set of DV time series is necessarily incomplete and, therefore, not archived.

Documentation: [KSCI-19110-001](#) These files comply with all Kepler standards for light curves, except for a few modest changes to emphasize the fact that they include injected transits and should not be used for general scientific purposes. These changes are:

1. Renamed from `kplr<kepler_id>-<stop_time>_llc.fits` to `kplr<kepler_id>-<stop_time>_INJECTED-llc.fits`.
2. Changed EXTNAME from “PRIMARY” to “INJECTED LIGHTCURVE”
3. Recomputed the checksum in the primary header keyword “CHECKSUM.”

Data Releases: Q1-Q17 DR25 (a one-time, end-of-mission release)

[Top](#)

Threshold-Crossing Events (TCE): Nominal Processing

Definition: A Threshold-Crossing Event (TCE) is a sequence of period flux decrements in the light curve (or flux time series) of a given target that resembles the signature of a transiting planet to a sufficient degree that it is passed by TPS, evaluated by DV, and subjected to further analysis in an effort to determine its nature.

Use Case: Transiting planet search (TPS) is used to produce TCEs for five different use cases: (a) nominal processing, (b) supplemental data validation (DV), (c) pixel-level transit injection, (d) flux inversion, and (e) flux scrambling.

These particular TCEs are from Use Case (a)—the result of nominal end-to-end operations processing of Kepler flight data in an effort to identify real planets in real flight data.

Location: Go to [Kepler TCE Documentation](#) (NASA Exoplanet Archive) and select the data release of choice.

Format: Interactive table available for download in various formats

Inputs: Presearch-data-conditioned light curves (i.e., PDCSAP_FLUX; [KAM §2.3.1](#)).

Outputs: Kepler Objects of Interest (KOI)—Nominal Processing

Documentation: Column definitions, query interfaces, and references can be found at [Kepler TCE Documentation](#) (NASA Exoplanet Archive), but the main documents describing these TCE releases are the associated science papers:

Q1-Q12: Tenenbaum et al. 2013, ApJS 206, 5

Q1-Q16: Tenenbaum et al. 2014, ApJS 211, 6

Q1-Q17 DR24: Seader et al. 2015, ApJS 217, 18

Q1-Q17 DR25: Twicken et al. 2016, ApJS, 152, 158

Data Releases: Q1-Q12, Q1-Q16, Q1-Q17 DR24, Q1-Q17 DR25 (the last four transit searches using TPS/DV). Notice that these searches occurred roughly annually as the volume of Kepler data grew, and then twice at end of mission.

Notes:

1. Earlier pipeline runs produced TCEs, but they are not archived (e.g., Tenenbaum et al. 2012, ApJS 199, 24).
2. The last two transit searches operated on uniformly processed light curves for all quarters, so the resulting outputs (i.e., TCEs, KOIs, pipeline completeness products, etc.) are tagged with the same data release number as the input light curves.

[Top](#)

DV Reports & DV Summaries: Nominal Processing

Definition: The Data Validation (DV) report is a multi-page document generated for every Kepler-observed target that contains at least one transit-like signature (a.k.a. Threshold-Crossing Event (TCE)). For each TCE associated with a given target, the DV summary provides a one-page overview of the information contained in the more extensive DV report (Wu et al. 2010; *KDPH*). These reports and summaries contain a variety of diagnostic parameters that help determine if a TCE is an instrumental artifact, an astrophysical false positive, or a planetary candidate.

Use Case: Data validation (DV) is used to produce reports and summaries for two different use cases: (a) nominal processing and (b) pixel-level transit injection.

These particular DV Reports & DV Summaries are from Case (a)—the result of nominal end-to-end operations processing of Kepler flight data in an effort to identify real planets in real flight data.

Location: Go to [Kepler Mission Summary Page](#) (NASA Exoplanet Archive) and select the TCE or KOI interactive table of choice. These reports/summaries can then be viewed from the pop-up bubbles by clicking on the information icon (i) next to an object name. Note that all reports/summaries for a given search are available through the TCE tables, but only for those TCEs promoted to KOI status through the KOI tables. For bulk download options, see [Bulk Data Download](#) (NASA Exoplanet Archive).

Format: Multi- or single-page PDF; many of key parameters are also available in the interactive tables (see box labeled “Select Columns” in the upper left and select/deselect parameters for display).

Inputs: Threshold-Crossing Events (TCE)—Nominal Processing

Outputs: Kepler Objects of Interest (KOI)—Nominal Processing

Documentation: Documentation is available at [Kepler TCE Documentation](#) (NASA Exoplanet Archive) and [Description of Kepler Data Validation One-Page Summary Reports](#) (NASA Exoplanet Archive).

The best reference for the DV Reports is the *KDPH*, as Wu et al. 2010 is out of date.

Data Releases: Q1-Q12, Q1-Q16, Q1-Q17 DR24, Q1-Q17 DR25 (the last four transit searches using TPS/DV). Notice that these searches occurred roughly annually as the volume of Kepler data grew, and then twice at end of mission.

[Top](#)

DV Time Series

Definition: The DV Time Series provide the quarter-stitched, gap-filled flux time series (or light curve) used by TPS/DV to find and evaluate Threshold-Crossing Events (TCEs).

Use Case: Data validation (DV) is used to produce reports and summaries for two different use cases: (a) nominal processing and (b) pixel-level transit injection.

However, DV Time Series are only available for Use Case (a)—Nominal Processing.

Location: Go to [Kepler TCE Documentation](#) (NASA Exoplanet Archive) and select the data release of choice. Bulk download is available at [Bulk Data Download](#) (NASA Exoplanet Archive). These files are also available via the [MAST portal](#) when searching for specific targets.

Format: ASCII for Q1-Q16 and Q1-Q17 DR24; FITS for Q1-Q17 DR25

Inputs: Light curve files—nominal processing (specifically, the PDCSAP_FLUX and its associated error)

Outputs: Threshold-Crossing Events (TCE)—nominal processing

Documentation: Online at [Kepler Data Validation Documentation](#) (Q1-Q16, Q1-Q17 DR24) and [KSCI-19079-001](#) (Q1-Q17 DR25). Twicken et al. (2016) describe the algorithms, parameters, and other nuances of the SOC 9.3 data processing pipeline (see [KDPH](#)).

Data Releases: Q1-Q16, Q1-Q17 DR24, and Q1-Q17 DR25

[Top](#)

Statistical Bootstrap Metric

Definition: A value calculated for each TCE that indicates the false alarm probability given the amount of noise present in the light curve.

Use Case: This metric can be used to evaluate the statistical significance of a signal.

Location: Found as columns in the TCE table

Format: Floating point columns

Inputs: TPS harmonic-removed, outlier-removed time series

Outputs: Four values for each TCE: boot_fap, boot_mesthresh, boot_mesmean, boot_messtd

Documentation: [KSCI-19086-004](#)

Data Releases: Q1-Q16, Q1-Q17 DR24 and Q1-Q17 DR25

[Top](#)

Threshold-Crossing Events (TCE): Pixel-level Transit Injection

Definition: A Threshold-Crossing Event (TCE) is a sequence of transit-like features in the flux time series of a given target that resembles the signature of a transiting planet to a sufficient degree that the target is passed on for further analysis.

Use Case: Transiting planet search (TPS) is used to produce TCEs for five different use cases: (a) nominal processing, (b) supplemental data validation (DV), (c) pixel-level transit injection, (d) flux inversion, and (e) flux scrambling.

These particular TCEs are from Use Case (c)—the result of injecting simulated planet transit signatures into the calibrated pixel-level data (i.e., the output of CAL) with subsequent processing through the remaining pipeline modules to create light curves (PA/PDC), conduct a transit search (TPS), and evaluate the identified events (DV). These injected TCEs are meant to provide ground truth for testing and characterization of the Kepler pipeline and Robovetter.

Location: These “injected TCEs” are not available in a searchable table to avoid confusion with the real events originating from Use Case (a). However, they are available in the form of a [Robovetter Results Table](#).

Format: ASCII table of injected planets and TPS results

Input: Pixel-level data.

Output: DV Reports & DV Summaries—Pixel-level Transit Injection. These injected TCEs are used to (a) measure the end-to-end detection efficiency for the Kepler data processing pipeline, including the formation and correction of the light curves (i.e., PA/PDC) and all aspects of the subsequent transit search and validation (i.e., TPS/DV) and (b) measure the TCERT/Robovetter detection efficiency by serving as ground truth for the Robovetter.

Documentation:

Data Release	Pipeline Release	Documentation
Q1-Q17 DR24	SOC 9.2	KSCI-19094-002
Q1-Q17 DR25	SOC 9.3	KSCI-19114-002 KSCI-19110-001

The Q1-Q17 DR25 injections are more tightly concentrated in MES than those for Q1-Q17 DR24, in order to better probe the transition region between detection and non-detection. DR25 also limits the M-dwarf injections to orbital periods < 100 days and includes approximately 5000 weak and 5000 near-equal-depth secondaries on selected sky groups.

Data Releases: Q1-Q17 DR24, Q1-Q17 DR25

Notes:

1. Neither these injected TCEs, nor the injected light curves from which they were derived, should be used to identify or characterize astrophysical phenomena, as these injections significantly perturb the data.

Each data release characterizes a unique version of the pipeline (see table above), so the products and results from one cannot be applied to another in straightforward fashion.

[Top](#)

DV Reports & DV Summaries: Pixel-level Transit Injection

Definition: The Data Validation (DV) report is a multi-page document generated for every Kepler-observed target that contains at least one transit-like signature (a.k.a. Threshold-Crossing Event (TCE). For each TCE associated with a given target, the DV summary provides a one-page overview of the information contained in the more extensive DV report (Wu et al. 2010; *KDPH*). These reports and summaries contain a variety of diagnostic parameters that help determine if a TCE is an instrumental artifact, an astrophysical false positive, or a planetary candidate.

Use Case: Data validation (DV) is used to produce reports and summaries for two different use cases: (a) nominal processing and (b) pixel-level transit injection.

These particular DV Reports & DV Summaries are from Use Case (b)—the result of injecting simulated planet transit signatures into the calibrated pixel-level data (i.e., the output of CAL) with subsequent processing through the remaining pipeline modules to create light curves (PA/PDC), conduct a transit search (TPS), and evaluate the identified events (DV). These injected DV reports/summaries are meant to provide ground truth for testing and characterization of the Robovetter.

Location: These “injected reports/summaries” are not available from a searchable table to avoid confusion with the real events originating from Use Case (a). However, they are available for bulk download (*Kepler Simulated Data*: NASA Exoplanet Archive).

Format: Multi- or single-page PDF

Inputs: Threshold-Crossing Events (TCE)—pixel-level transit injection

Outputs: TCERT vetting forms – pixel-level transit injection

TCERT metrics and dispositions—pixel-level transit injection

Documentation: These are standard DV products. See Threshold-Crossing Events—Pixel-level Transit Injection for details of the injection technique and subsequent data processing.

Data Releases: Q1-Q17 DR25 (a one-time, end-of-mission release)

[Top](#)

Threshold-Crossing Events (TCE): Flux Inversion

Definition: A Threshold-Crossing Event (TCE) is a sequence of transit-like features in the flux time series of a given target that resembles the signature of a transiting planet to a sufficient degree that the target is passed on for further analysis.

Use Case: Transiting planet search (TPS) is used to produce TCEs for five different use cases: (a) nominal processing, (b) supplemental data validation (DV), (c) pixel-level transit injection, (d) flux inversion, and (e) flux scrambling.

These particular TCEs are from Use Case (d)—the result of inverting the presearch-data-conditioned fluxes (i.e., PDCSAP_FLUX, *KAM §2.3.1*), conducting a transit search (TPS), and evaluating the identified events (DV). These inverted TCEs are meant to provide false alarms for testing and characterization of the Kepler pipeline and Robovetter.

Location: These “inverted TCEs” are not available in a searchable table to avoid confusion with the real events originating from Use Case (a). However, they are available in the form of a Detailed Results Table (*Kepler Simulated Data*: NASA Exoplanet Archive).

Format: ASCII table of TPS results

Input: Light Curve Files – Nominal Processing. Since flux inversion does not involve the pixel level data, no separate product equivalent to the Injected Light Curves is necessary.

Output: TCERT vetting forms—flux inversion,

TCERT metrics and dispositions—flux inversion, and

TCERT effectiveness metric.

Documentation: [KSCI-19114-002](#)

Data Releases: Q1-Q17 DR25

[Top](#)

Threshold-Crossing Events (TCE): Flux Scrambling

Definition: A Threshold-Crossing Event (TCE) is a sequence of transit-like features in the flux time series of a given target that resembles the signature of a transiting planet to a sufficient degree that the target is passed on for further analysis.

Use Case: Transiting planet search (TPS) is used to produce TCEs for five different use cases: (a) nominal processing, (b) supplemental data validation (DV), (c) pixel-level transit injection, (d) flux inversion, and (e) flux scrambling.

These particular TCEs are from Use Case (e)—the result of scrambling the presearch-data-conditioned fluxes (i.e., PDCSAP_FLUX, [KAM §2.3.1](#)) in quarter- or year-long blocks, conducting a transit search (TPS), and evaluating the identified events (DV). Three separate runs with different scrambling orders were performed (see [§2.4 of KSCI-19114](#) for details.)

Like flux inversion, flux scrambling should remove the legitimate transit events, while preserving the systematics that trigger false alarms. To the extent that this is true, the scrambled TCEs provide additional estimates of the Robovetter false-positive rate.

Location: These “scrambled TCEs” are not available in a searchable table to avoid confusion with the real events originating from Use Case (a). However, they are available in the form of a Detailed Results Table ([Kepler Simulated Data](#): NASA Exoplanet Archive).

Format: ASCII tables of TPS results

Input: Light Curve Files—Nominal Processing. Since flux scrambling does not involve the pixel level data, no separate product equivalent to the Injected Light Curves is necessary.

Output: Vetting Metrics & TCERT Reports—Flux Scrambling. These scrambled TCEs are used to measure the reliability of the planet catalog that enables a measurement of the reliability of the planet catalog, see the TCERT/Robovetter Reliability Metric.

Documentation: [KSCI-19114-002](#)

Data Releases: Q1-Q17 DR25 (a one-time, end-of-mission release)

Notes:

1. Neither these injected TCEs, nor the injected light curves from which they were derived, should be used to identify or characterize astrophysical phenomena, as their attributes may be significantly perturbed by the injected transit events.
2. Each data release characterizes a unique version of the pipeline (see table above), so the products and results from one cannot be applied to another in straightforward fashion.

[Top](#)

TCERT Vetting Forms: Nominal Processing

Definition: The TCERT Vetting Forms include the DV Summaries, major results from the DV Reports, and additional vetting metrics computed outside the Kepler data processing pipeline. They are compiled into a multi-page document for manual inspection and analysis.

Use Case: TCERT Vetting Forms are generated for four different use cases: (a) nominal processing, (b) pixel-level transit injection, (c) flux inversion, and (d) flux scrambling.

These particular TCERT Vetting Forms are from Use Case (a)—the result of nominal end-to-end operations processing of Kepler flight data in an effort to identify real planets in real flight data.

Location: Go to [Kepler Mission Summary](#) (NASA Exoplanet Archive) and select the KOI table of choice. These forms can then be viewed from the pop-up bubbles by clicking on the information icon (i) next to an object name. For bulk download options, see TCERT Reports at [Bulk Data Download](#) (NASA Exoplanet Archive).

Format: One multi-page PDF per star.

Inputs: Threshold-Crossing Events (TCE)—Nominal Processing

DV Reports & DV Summaries—Nominal Processing

Light Curve Files—Nominal Processing

Target pixel data—Nominal Processing

Outputs: Planet-candidate catalogs suitable for occurrence rate calculations.

Documentation: KSCI-19103 (Q1-Q16), KSCI-19104 (Q1-Q17 DR24), and [KSCI-19105-001](#) (Q1-Q17 DR25) at [Kepler TCE Documentation](#) (NASA Exoplanet Archive)

Data Releases: Q1-Q16, Q1-Q17 DR24, and Q1-Q17 DR25

TCERT Metrics & Dispositions: Nominal Processing

Definition: The TCERT Metrics & Dispositions include all inputs (i.e., metrics) used by the Robovetter to produce its outputs (i.e., dispositions).

Use Case: TCERT Metrics & Dispositions are generated for five different use cases: (a) nominal processing, (b) supplemental data validation (DV), (c) pixel-level transit injection, (d) flux inversion, and (e) flux scrambling.

These particular TCERT Metrics & Dispositions are from Use Case (a)—the result of nominal end-to-end operations processing of Kepler flight data in an effort to identify real planets in real flight data.

Location: See RoboVetter-Input.txt at [Jeff Coughlin's Robovetter GitHub repository](#) for DR24; [Kepler Simulated Data](#) (NASA Exoplanet Archive) or [NASA Kepler Robovetter GitHub repository](#) for DR25. Many of key parameters are also available in the interactive TCE tables (see box labeled “Select Columns” in the upper left and select/deselect parameters for display).

Format: Flat ASCII text file using Robovetter input format

Inputs: Threshold-Crossing Events (TCE)—Nominal Processing

DV Reports & DV Summaries—Nominal Processing

Light Curve Files—Nominal Processing

Outputs: Robovetter dispositions yielding planet-candidate catalogs suitable for occurrence rate calculations

Documentation: See [Jeff Coughlin's Robovetter GitHub repository](#) for DR24 and [NASA Kepler Robovetter GitHub repository](#) for DR25.

Data Releases: Q1-Q17 DR24 and Q1-Q17 DR25

Kepler Objects of Interest: Nominal Processing

Definition: Kepler Objects of Interest (KOI) are Threshold-Crossing Events (TCE), which after further analysis, are judged to be astrophysical transiting or eclipsing objects.

Use Case: The Robovetter (or the previous manual TCERT process) is used to promote TCEs to KOI status for five different use cases: (a) nominal processing, (b) supplemental data validation (DV), (c) pixel-level transit injection, (d) flux inversion, and (e) flux scrambling.

These particular KOIs are from Use Case (a)—the result of nominal end-to-end operations processing of Kepler flight data in an effort to identify real planets in real flight data.

Location: Go to [Kepler KOI Documentation](#) (NASA Exoplanet Archive) and select the data release of choice, or the cumulative table, which includes all KOIs from all data releases.

Format: Interactive online tables which can be exported in various ASCII formats.

Inputs: Threshold-Crossing Events (TCE)—Nominal Processing

Vetting Metrics & TCERT Reports—Nominal Processing

Outputs: Robovetter dispositions yielding planet-candidate catalogs suitable for occurrence rate calculations

Documentation: Column definitions, query interfaces, and references can be found at [Kepler KOI Documentation](#) (NASA Exoplanet Archive), but the main documents describing these KOI releases are the associated science papers:

- [Q0-Q1: Borucki et al. 2011, ApJ 728, 117](#)
- [Q0-Q5: Borucki et al. 2011, ApJ 736, 117](#)
- [Q1-Q6: Batalha et al. 2013, ApJS 204, 24](#)
- [Q1-Q8: Burke et al. 2014, ApJS 210, 19](#)
- [Q1-Q12: Rowe et al. 2015, ApJS 217, 16](#)
- [Q1-Q16: Mullally et al. 2015, ApJS 217, 31](#)
- [Q1-Q17 DR24: Coughlin et al. 2016 ApJS 224, 12](#)
- [Q1-Q17 DR25: Thompson et al. 2018 ApJS 235, 38](#)

Data Releases: Q1-Q6, Q1-Q8, Q1-Q12, Q1-Q16, Q1-Q17 DR24, Q1-Q17 DR25 are available through the archive; earlier releases through their journal publications.

Planetary Parameters: Nominal Processing

Definition: The Planetary Parameters are the properties of the planet candidates as inferred from fits to the transit photometry.

Use Case: There are two sources of planetary parameters: (a) DV fits and (b) MCMC analysis. The TCE tables state the DV fits. For most KOI activity tables, the DV fits are delivered initially and then replaced by MCMC fits if/when they become available.

Location: The planetary parameters (i.e., DV fits) are available for the TCE releases via the interactive tables at [Kepler TCE Documentation](#) (NASA Exoplanet Archive). The planetary parameters for the KOI releases (i.e., either DV or MCMC fits) are available via the interactive tables at [Kepler KOI Documentation](#) (NASA Exoplanet Archive). The bulk download options are at [Bulk Data Download](#) (NASA Exoplanet Archive).

Format: ASCII table (see options under “Download Table” in upper left of header on interactive tables page).

Inputs: Light Curve Files—Nominal Processing

Ephemerides from TPS for DV fits and from least-square fits for MCMC analysis

Outputs: Planet-candidate catalogs suitable for occurrence rate calculations

Documentation: TCE and KOI Catalog papers for DV fits and KSCIs for MCMC fits ([KSCI-19084-002](#) for DR24 and [KSCI-19113](#) for DR25).

Data Releases: Q1-Q6, Q1-Q8, Q1-Q12, Q1-Q16, Q1-Q17 DR24, Q1-Q17 DR25 are available through the archive; earlier releases through their journal publications.

KOI Table	KOI Planet Parameter Source	Planet Parameter Documentation
Batalha: Q1-Q6	DV	Batalha et al. 2013
Burke: Q1-Q8	DV	Burke et al. 2014
Rowe: Q1-Q12	MCMC	Rowe et al. 2015
Mullally: Q1-Q16	MCMC	KSCI-19084-002
Coughlin: Q1-Q17 DR24	MCMC	KSCI-19084-002
Thompson: Q1-Q17 DR25	MCMC	KSCI-19113-001

TCERT Metrics & Dispositions: Supplemental Data Validation

Definition: The TCERT Metrics & Dispositions include all inputs (i.e., metrics) used by the Robovetter to produce its outputs (i.e., dispositions).

Use Case: TCERT Metrics & Dispositions are generated for five different use cases: (a) nominal processing, (b) supplemental data validation (DV), (c) pixel-level transit injection, (d) flux inversion, and (e) flux scrambling.

These particular TCERT Metrics & Dispositions are from Use Case (b)—the result of a supplemental data validation (DV) run. Supplemental DV uses the same basic code as DV, but sequentially evaluates an input set of ephemerides, rather than TCEs passed by TPS.

Location: See RoboVetter-Input.txt at [Jeff Coughlin's Robovetter GitHub repository](#) for DR24; [Kepler Simulated Data](#) (NASA Exoplanet Archive) or [NASA Kepler Robovetter GitHub repository](#) for DR25.

Format: Flat ASCII text file

Inputs: Threshold-Crossing Events (TCE)—Nominal Processing

DV Reports & DV Summaries—Supplemental DV

Light Curve Files—Nominal Processing

Target pixel data—Nominal Processing

Outputs: A disparate set of objects that is unsuitable for occurrence rate calculations in straightforward fashion. Note that no new KOIs were created in the DR25 Supplemental table --- it has the exact same set of KOIs as the end-of-mission cumulative table.

Documentation: See [NASA Kepler Robovetter GitHub repository](#)

Data Releases: DR25 Supplemental

Kepler Objects of Interest: Supplemental Data Validation

Definition: Kepler Objects of Interest (KOI) are Threshold-Crossing Events (TCE), which after further analysis, are judged to be astrophysical in origin.

Use Case: The Robovetter (or the previous manual TCERT process) is used to promote TCEs to KOI status for five different use cases: (a) nominal processing, (b) supplemental data validation (DV), (c) pixel-level transit injection, (d) flux inversion, and (e) flux scrambling.

These particular KOIs are from Use Case (b)—the result of a supplemental data validation (DV) run. Supplemental DV uses the same basic code as DV, but sequentially evaluates an input set of ephemerides, rather than TCEs passed by TPS. Note that no new KOIs were created in the DR25 Supplemental table --- it has the exact same set of KOIs as the end-of-mission cumulative table.

Location: Go to [Kepler KOI Documentation](#) (NASA Exoplanet Archive) and select the DR25 supplemental table.

Format: Interactive online table.

Inputs: Threshold-Crossing Events (TCE)—Nominal Processing

DV Reports & DV Summaries—Supplemental DV

Vetting Metrics & TCERT Reports—Supplemental DV

Outputs: Robovetter dispositions for this very disparate set of objects that are not suitable for occurrence rate calculations. Note that no new KOIs were created in the DR25 Supplemental table --- it has the exact same set of KOIs as the end-of-mission cumulative table.

Documentation: Column definitions, query interfaces, and references can be found at [Kepler KOI Documentation](#) (NASA Exoplanet Archive).

Data Releases: DR25 Supplemental.

TCERT Vetting Forms: Pixel-level Transit Injection

Definition: The TCERT Vetting Forms include the DV Summaries, major results from the DV Reports, and additional vetting metrics computed outside the Kepler data processing pipeline. They are compiled into a multi-page document for manual inspection and analysis.

Use Case: TCERT Vetting Forms are generated for four different use cases: (a) nominal processing, (b) pixel-level transit injection, (c) flux inversion, and (d) flux scrambling.

These particular TCERT Vetting Forms are from Use Case (b)—the result of injecting simulated planet transit signatures into the calibrated pixel-level data (i.e., the output of CAL) with subsequent processing through the remaining pipeline modules to create light curves (PA/PDC), conduct a transit search (TPS), and evaluate the identified events (DV). These Vetting Forms for injected transits are useful in evaluating the performance of the Robovetter on this data set.

Location: [Kepler Simulated Data](#) (NASA Exoplanet Archive)

Format: One multi-page PDF per star

Inputs: Threshold-Crossing Events (TCE)—Pixel-level Transit Injection

DV Reports & DV Summaries—Pixel-level Transit Injection

Light Curve Files—Pixel-level Transit Injection

Outputs: Pipeline Detection Efficiency and TCERT Detection Efficiency

Documentation: [KSCI-19105-001](#)

Data Releases: Q1-Q17 DR25

TCERT Metrics & Dispositions: Pixel-level Transit Injection

Definition: The TCERT Metrics & Dispositions include all inputs (i.e., metrics) used by the Robovetter to produce its outputs (i.e., dispositions).

Use Case: TCERT Metrics & Dispositions are generated for five different use cases: (a) nominal processing, (b) supplemental data validation (DV), (c) pixel-level transit injection, (d) flux inversion, and (e) flux scrambling.

These particular TCERT Metrics & Dispositions are from Use Case (c)—the result of injecting simulated planet transit signatures into the calibrated pixel-level data (i.e., the output of CAL) with subsequent processing through the remaining pipeline modules to create light curves (PA/PDC), conduct a transit search (TPS), and evaluate the identified events (DV). These Metrics and Dispositions for injected transits are useful for running the Robovetter and evaluating its performance on this data set.

Location: Robovetter Inputs on [Kepler Simulated Data](#) (NASA Exoplanet Archive) and [NASA Kepler Robovetter GitHub repository](#)

Format: ASCII file

Inputs: Threshold-Crossing Events (TCE)—Pixel-level Transit Injection

DV Reports & DV Summaries—Pixel-level Transit Injection

Light Curve Files—Pixel-level Transit Injection

Outputs: Pipeline Detection Efficiency and TCERT Detection Efficiency

Documentation: Thompson et al. 2018 ApJS and [KSCI-19114-002](#)

Data Releases: Q1-Q17 DR25

[Top](#)

TCERT Vetting Forms: Flux Inversion

Definition: The TCERT Vetting Forms include the DV Summaries, major results from the DV Reports, and additional vetting metrics computed outside the Kepler data processing pipeline. They are compiled into a multi-page document for manual inspection and analysis.

Use Case: TCERT Vetting Forms are generated for four different use cases: (a) nominal processing, (b) pixel-level transit injection, (c) flux inversion, and (d) flux scrambling.

These particular TCERT Vetting Forms are from Use Case (c)—the result of inverting the presearch-data-conditioned fluxes (i.e., PDCSAP_FLUX, [KAM §2.3.1](#)), conducting a transit search (TPS), and evaluating the identified events (DV). These Vetting Forms for inverted transits are useful in evaluating the performance of the Robovetter on this data set.

Location: TCERT Reports on [Kepler Simulated Data Page](#)

Format: One multi-page PDF per star.

Inputs: Threshold-Crossing Events (TCE)—Flux Inversion

DV Reports & DV Summaries—Flux Inversion

Light Curve Files—Flux Inversion

Outputs: None.

Documentation: [KSCI-19105-001](#)

Data Releases: Q1-Q17 DR25

[Top](#)

TCERT Metrics & Dispositions: Flux Inversion

Definition: The TCERT Metrics & Dispositions include all inputs (i.e., metrics) used by the Robovetter to produce its outputs (i.e., dispositions).

Use Case: TCERT Metrics & Dispositions are generated for five different use cases: (a) nominal processing, (b) supplemental data validation (DV), (c) pixel-level transit injection, (d) flux inversion, and (e) flux scrambling.

These particular TCERT Metrics & Dispositions are from Use Case (d)—the result of inverting the presearch-data-conditioned fluxes (i.e., PDCSAP_FLUX, [KAM §2.3.1](#)), conducting a transit search (TPS), and evaluating the identified events (DV). These Metrics and Dispositions for inverted transits are useful for running the Robovetter and evaluating its performance on this data set.

Location: Robovetter Inputs on [Kepler Simulated Data Page](#) and [NASA Kepler Robovetter GitHub repository](#)

Format: Flat ASCII text file

Inputs: Threshold-Crossing Events (TCE)—Flux Inversion

DV Reports & DV Summaries—Flux Inversion

Light Curve Files—Flux Inversion

Outputs: Robovetter Outputs and TCERT Reliability Metric

Documentation: Q1-Q17 DR25: Thompson et al. 2018 ApJS and [KSCI-19114-002](#)

Data Releases: Q1-Q17 DR25

[Top](#)

TCERT Vetting Forms: Flux Scrambling

Definition: The TCERT Vetting Forms include the DV Summaries, major results from the DV Reports, and additional vetting metrics computed outside the Kepler data processing pipeline. They are compiled into a multi-page document for manual inspection and analysis.

Use Case: TCERT Vetting Forms are generated for four different use cases: (a) nominal processing, (b) pixel-level transit injection, (c) flux inversion, and (d) flux scrambling.

These particular TCERT Vetting Forms are from Use Case (d)—the result of scrambling the presearch-data-conditioned fluxes (i.e., PDCSAP_FLUX, [KAM §2.3.1](#)) in year-long blocks, conducting a transit search (TPS), and evaluating the identified events (DV). Three separate runs with different scrambling orders were performed (see [§2.4 of KSCI-19114](#) for details). These Vetting Forms are useful in evaluating the performance of the Robovetter on these scrambled data sets.

Location: TCERT Reports on [Kepler Simulated Data Page](#) and [NASA Kepler Robovetter GitHub repository](#)

Format: Multi-page PDF

Inputs: Threshold-Crossing Events (TCE)—Flux Scrambling

DV Reports & DV Summaries—Flux Scrambling

Light Curve Files—Flux Scrambling

Outputs: None

Documentation: [KSCI-19105-001](#)

Data Releases: Q1-Q17 DR25

[Top](#)

TCERT Metrics & Dispositions: Flux Scrambling

Definition: The TCERT Metrics & Dispositions include all inputs (i.e., metrics) used by the Robovetter to produce its outputs (i.e., dispositions).

Use Case: TCERT Metrics & Dispositions are generated for five different use cases: (a) nominal processing, (b) supplemental data validation (DV), (c) pixel-level transit injection, (d) flux inversion, and (e) flux scrambling.

These particular TCERT Metrics & Dispositions are from Use Case (e)—the result of scrambling the presearch-data-conditioned fluxes (i.e., PDCSAP_FLUX, [KAM §2.3.1](#)) in year-long blocks, conducting a transit search (TPS), and evaluating the identified events (DV). Three separate runs with different scrambling orders were performed (see [§2.4 of KSCI-19114](#) for details.) These Metrics and Dispositions are useful for running the Robovetter and evaluating its performance on these scrambled data sets.

Location: Robovetter Inputs on [Kepler Simulated Data Page](#).

Format: ASCII text file

Inputs: Threshold-Crossing Events (TCE)—Flux Scrambling

DV Reports & DV Summaries—Flux Scrambling

Light Curve Files—Flux Scrambling

Outputs: TCERT Reliability Metric and Robovetter Output

Documentation: Q1-Q17 DR25: Thompson et al. 2018 ApJS and [KSCI-19114-002](#)

Data Releases: Q1-Q17 DR25

[Top](#)

Pipeline Detection Efficiency

Definition: The pipeline detection efficiency describes the likelihood that the Kepler data processing pipeline would successfully recover a given transit signal as a function of its Multiple Event Statistic (MES; the strength of the transit signal relative to the noise).

Location:

Go to [Kepler Completeness and Reliability](#) (NASA Exoplanet Archive) and click on the data or documentation release, as desired.

Format: ASCII table

Inputs: Light Curve Files—Pixel-level Transit Injection

Threshold-Crossing Events (TCE)—Pixel-level Transit Injection

DV Reports & DV Summaries—Pixel-level Transit Injection

Outputs: Values of the pipeline detection efficiency using one-injection per target and averaging over many targets; used by the Analytic Completeness Model and the Numerical Completeness Model.

Documentation: The most thorough documentation resides in the published papers and documents:

One-quarter test: [Christiansen, J. L., Clarke, B. D., Burke, C. J., et al. 2013, ApJS, 207, 35](#)

One-year test: [Christiansen, J. L., Clarke, B. D., Burke, C. J., et al. 2015a, ApJ, 810, 95](#)

Q1-Q17 DR24: [Christiansen, J. L., Clarke, B. D., Burke, C. J., et al. 2016, ApJ, 828, 99](#)

Q1-Q17 DR25: [KSCI-19110-001](#)

For Q1-Q17 DR24, which characterizes pipeline build SOC 9.2, the ephemerides of the injected transits and the results of the pipeline search are archived (see [Christiansen, J. L. 2015b, KSCI-19094-001: Planet Detection Metrics: Pipeline Detection Efficiency at KSCI-19094-001](#)).

For Q1-Q17 DR25, which characterizes pipeline build SOC 9.3, the ephemerides of the injected transits and the results of the pipeline search are archived (see [KSCI-19110-001](#)). The associated light curves ([Kepler Simulated Data](#): NASA Exoplanet Archive) enable the sensitivities of competing pipelines to be compared or characterized.

Data Releases: Q1-Q17 DR24, Q1-Q17 DR25

The detection efficiency of the Kepler pipeline was investigated with a series of four experiments using pixel-level transit injection. These experiments use differing amounts of data processed by different versions of the pipeline. Only the last two experiments are directly applicable (i.e., DR24 and DR25) to specific versions of the pipeline (i.e., SOC 9.2 and SOC 9.3, respectively) and are available through the archive. The last experiment is the most comprehensive, involves the most injections, and characterizes the final version of the pipeline that generated the DR25 TCEs and KOIs.

[Top](#)

Analytic Completeness Model

Definition: The Analytic Completeness Model provides an estimate of the Kepler data processing pipeline's completeness for each individual target as a function of planet radius and orbital period using the targets' average values for Robust RMS CDPP, Duty Cycle, and Data Span, the Effective MES Threshold, and the Pipeline Detection Efficiency.

Location: Go to [Kepler Completeness and Reliability](#) (NASA Exoplanet Archive) for the individual model components and [Christopher Burke's KeplerPORTs GitHub repository](#) for the model generating software. Bulk download options are also available.

Format: Interactive online or ASCII tables

Inputs: Robust RMS CDPP, Duty Cycle, Data Span, Effective MES Threshold, and Pipeline Detection Efficiency

Outputs: Completeness contours suitable for use in occurrence rate calculations

Documentation: Go to [Kepler Completeness and Reliability](#) (NASA Exoplanet Archive) for column definitions, [Christopher Burke's KeplerPORTs GitHub repository](#) for a software readme file, and [Burke et al. 2015](#), ApJ 809, 8 for worked examples.

Data Releases: Q1-Q16, Q1-Q17 DR24, Q1-Q17 DR25

[Top](#)

Numerical Completeness Model

Definition: The Numerical Completeness Model provides an estimate of the Kepler data processing pipeline's completeness for each individual target as a function of planet radius and orbital period using the targets' Window Function, One-Sigma Depth Function, and Pipeline Detection Efficiency.

Location: Go to [Kepler Completeness and Reliability](#) (NASA Exoplanet Archive) for the individual model components and [Christopher Burke's KeplerPORTs GitHub repository](#) for the model generating software. Bulk download options are also available ([Kepler Completeness and Reliability](#) (NASA Exoplanet Archive)).

Format: FITS (window and one-sigma depth), ASCII (pipeline detection efficiency)

Inputs: Window Function, One-Sigma Depth Function, and Pipeline Detection Efficiency

Outputs: Completeness contours suitable for use in occurrence rate calculations

Documentation: See [KSCI-19085-001](#) for DR24 and [KSCI-19101-002](#) for DR25 (both at [Kepler Completeness and Reliability](#) (NASA Exoplanet Archive)), [Christopher Burke's KeplerPORTs GitHub repository](#) for a software readme file, and [Burke et al. 2015](#), ApJ 809, 8 for worked examples.

Data Releases: Q1-Q17 DR24, Q1-Q17 DR25

Empirical Completeness Results

Definition: The empirical completeness results are the output of ~600,000 flux-level transit injections in planet-radius vs. orbital-period space for selected targets. These results serve as ground-truth for developing, verifying, and evaluating the Analytic and Numerical Completeness Models.

Location: [Kepler Simulated Data](#) (NASA Exoplanet Archive) under INJ1, INJ2 and INJ3

Format: ASCII tables

Inputs: Light curves with injected transits (INJ1 (on-target), INJ2 (off-target) and INJ3 (eclipsing binaries)).

Outputs: Average Detection Efficiency table of injections and resulting injected TCEs.

Documentation: [KSCI-19110-001](#)

Data Releases: DR25

[Top](#)

Completeness Model Comparison & Limitations

Definition: This report compares the Empirical Completeness Results with those generated by the Analytic and Numerical Completeness Models in an effort to describe their limitations, identify stars that behave anomalously, and document the KeplerPORTs code used for this investigation. This KeplerPORTs is designed to compute completeness contours for individual targets as a function of planet radius and orbital period given the inputs appropriate for either model.

Location: NASA Exoplanet Archive

Format: PDF report

Inputs: [NASA Kepler KeplerPorts GitHub repository](#)

Outputs: Full detection contours for analysis of occurrence rates

Documentation: Detection Contour Documentation, [KSCI-19111-002](#)

Data Releases: DR25

[Top](#)

Stellar Properties Catalog

Definition: The stellar properties catalog provides estimates of effective temperature, surface gravity, metallicity, and radius for the stars observed by Kepler based on a variety of methods, including asteroseismology, spectroscopy, exoplanet transits, and photometry. This product set also includes the Kepler Input Catalog (KIC), which was used to select the observed targets based on photometric and physical data for most sources in the Kepler mission field of view.

Location: The KIC is accessible at: [Kepler Target Search](#) (MAST) or can be downloaded at [MAST GZ file download](#). The updates to this catalog are available at: [Kepler Stellar Documentation](#) (MAST).

Format: ASCII (KIC); IPAC table format (updates)

Inputs: Observational data and stellar evolution models

Outputs: Stellar properties for use in Data Validation, the MCMC analysis, and occurrence rate calculations

Documentation: [Brown et al. 2011, AJ 142, 112](#) (KIC) and [Kepler Stellar Documentation](#) (NASA Exoplanet Archive) for links to updates. The final, DR25 stellar documentation is [Mathur, S., Huber, E., Batalha, N. M., et al. 2017, ApJS, 229, 30](#).

Data Releases: Pre-launch (KIC) and Q1-Q12, Q1-Q16, Q1-Q17 DR24, and Q1-Q17 DR25 (updates)

[Top](#)

Autovetter Planetary Candidate Catalog

Definition: The Autovetter planetary candidate catalog is a classification of the Q1-Q17 DR24 TCEs (Threshold-Crossing Events) using a decision tree-based machine learning technique known as the random forest. The inputs to the Autovetter are a training data set composed of TCEs that have been dispositioned (mostly) by humans into these three classes, and a set of attributes (scalar statistics) associated with each TCE. From the training set, the Autovetter "learns a mapping between attributes and predicted class. This mapping is then applied uniformly and consistently to all TCEs to produce a catalog of planet candidates.

Location: Table in the journal

Format: Machine-readable table

Inputs: TCEs and a training set of planet candidates and false positives derived from previous lists of TCEs

Outputs: Dispositions of PC/FP for the DR24 TCEs

Documentation: [McCauliff et al. \(2015\)](#)

Data Releases: DR24

[Top](#)

Astrophysical Positional Probabilities

Definition: The astrophysical positional probabilities tables provides information about how well a star's location matches the location of the transit signal. For each KOI the following is given: the stars with the highest probability of being co-located with the source of the observed transit and the probability of the transit being on an unknown background source.

Use Case: These probabilities can be used to determine whether a KOI transit signal is coming from the target star or a nearby star on the sky.

Location: [FPP Table](#) (NASA Exoplanet Archive)

Format: Interactive table

Inputs: KIC, UKIRT catalog

Outputs: List of positional probabilities which can be used to analyze individual KOIs

Documentation: [KSCI-19092-002](#)

Data Releases: Q1-Q17 DR24, Q1-Q17 DR25

[Top](#)

Astrophysical False Positive Probabilities

Definition: This weighs the likelihood that a signal originates from an transiting planet against the likelihood that it comes from other possible astrophysical events that can mimic a transit. This is calculated using T. Morton's vespa code ([Timothy Morton's vespa GitHub repository](#)). The astrophysical false positive scenarios considered by these calculations are blended and unblended eclipsing binary stars, where the eclipsing binary may be physically associated with the primary target or may be a chance-alignment with the primary target. This work does not consider a planet orbiting an unresolved companion to be a false positive scenario.

Use Case: These results were used to statistically validate KOIs at the 99% level.

Location: [Astrophysical False Positive Probabilities Table](#) (NASA Exoplanet Archive)

Format: Interactive table

Inputs: KOI table and folded light curves

Outputs: Probability that a signal is caused by a blended or unblended eclipsing binary. PDF reports graphically demonstrating the results.

Documentation: [Morton et al. \(2016\)](#); [Data Column Definitions](#) and [DR24 and DR25 comparison document](#) (NASA Exoplanet Archive)

Data Releases: Q1-Q17 DR24 and Q1-Q17 DR25

[Top](#)

MCMC Posterior Chains

Definition: The transit fit errors provided in the KOI table are calculated using a Markov chain Monte Carlo method. The posterior of those chains are provided for each KOI.

Use Case: These MCMC chains can be used to better represent the errors associated with the transit fits for each KOI.

Location: [Kepler KOI Documentation](#) (NASA Exoplanet Archive)

Format: ASCII tables

Inputs: Light curve files (SAP_FLUX), TCE period/epoch

Outputs: Transit parameters for each chain for each fitted KOI

Documentation: [KSCI-19113](#)

Data Releases: Q1-Q16, Q1-Q17 DR24, and Q1-Q17 DR25

[Top](#)

TCERT Detection Efficiency

Definition: The TCERT Detection Efficiency measures the fraction of injected TCEs that were dispositioned as Planet Candidates in the KOI catalog.

Use Case: When combined with the Pipeline Detection Efficiency, this can give a full picture of which planets were not included as candidates in the KOI catalog for occurrence rate measurements.

Location: [Kepler Simulated Data Products Page](#) (NASA Exoplanet Archive)

Format: ASCII table and PDF reports

Inputs: TCEs from Pixel-level Transit Injection

Outputs: Table of dispositions and metrics for each injected TCE and TCERT reports showing various metrics and plots concerning each TCE.

Documentation: [KSCI-19114-002](#)

Data Releases: Q1-Q17 DR24 and Q1-Q17 DR25

[Top](#)

TCERT Reliability Metric

Definition: The TCERT Reliability Metric measures the fraction of simulated false alarms that are called planet candidates. These products determine the effectiveness of the Robovetter at identifying and eliminating false alarms.

Use Case: This information can be used to measure the reliability of the planet catalog.

Location: [Kepler Simulated Data Products Page](#) (NASA Exoplanet Archive)

Format: ASCII table and PDF reports

Inputs: TCEs identified in scrambled and inverted light curves.

Outputs: Table of dispositions and metrics for each simulated false alarm TCE. TCERT Reports, showing various plots and metrics, are provided for each TCE.

Documentation: [KSCI-19114-002](#)

Data Releases: Q1-Q17 DR25

[Top](#)

Certified False Positives

Definition: Certified False Positives is a list of KOIs examined by hand by the False Positive Working Group using all available data on the target to create a list of true False Positives with a reason for that designation.

Use Case: This list of true false positives can serve as a training set for automated vetters.

Location: [Certified False Positives interactive table](#) (NASA Exoplanet Archive)

Format: Interactive table

Inputs: KOI List, Kepler Data

Outputs: Table of dispositions and reasons

Documentation: [Data Column Definitions](#) (NASA Exoplanet Archive)

Data Releases: None, new KOIs are added as they are examined by the group.

[Top](#)

Planetary Systems (Confirmed Planets)

Definition: Confirmed Kepler planetary systems is not a Kepler mission product. Rather, it is a list maintained by the NASA Exoplanet Archive of planets reported in the refereed literature. Other astrophysical signals can mimic a transit signal, when either mass measurements or statistical validation reveal that these other possibilities are less than 1% likely, then the planet is considered confirmed and added to this list.

Use Case: This list is intended for those needing a list of planets with high fidelity.

Location: [Planetary Systems interactive table](#) (NASA Exoplanet Archive)

Format: Interactive table

Inputs: KOI table, follow-up observations, astrophysical false positive probabilities

Outputs: Table of confirmed planets and their properties.

Documentation: Each row comes from different papers in the scientific literature.

Data Releases: None

Last updated: 24 June 2024



Caltech

[Home](#)

[About Us](#)

[Data](#)

[Tools](#)

[Support](#)