STScI | SPACE TELESCOPE SCIENCE INSTITUTE

EXPANDING THE FRONTIERS OF SPACE ASTRONOMY

Time-Domain Processing for the Wide Field Instrument at the Roman Science Operations Center (SOC)

Harry Ferguson

February 8, 2022

SOC Responsibilities & Design

STScI provides the Science Operations Center (SOC) within a distributed Ground System Architecture

Planning & Scheduling
All mission observations

WFI Data processing
Details depend on mode

Archive
for all observations

Community interface for WFI imaging
User support , documentation, public outreach

Remaining Ground System functions: IPAC, GSFC, …

Roman Data Management

Downlinked science data

STScI
Science Operations Center
Data Management System

WFI Exposure-level processing

WFI High-level processing

Archive

Data Analysis

Low-level processed data

IPAC
Science Support Center

WFI high-level processing
Microlensing & spectroscopy

CGI data analysis environment

Science ready data

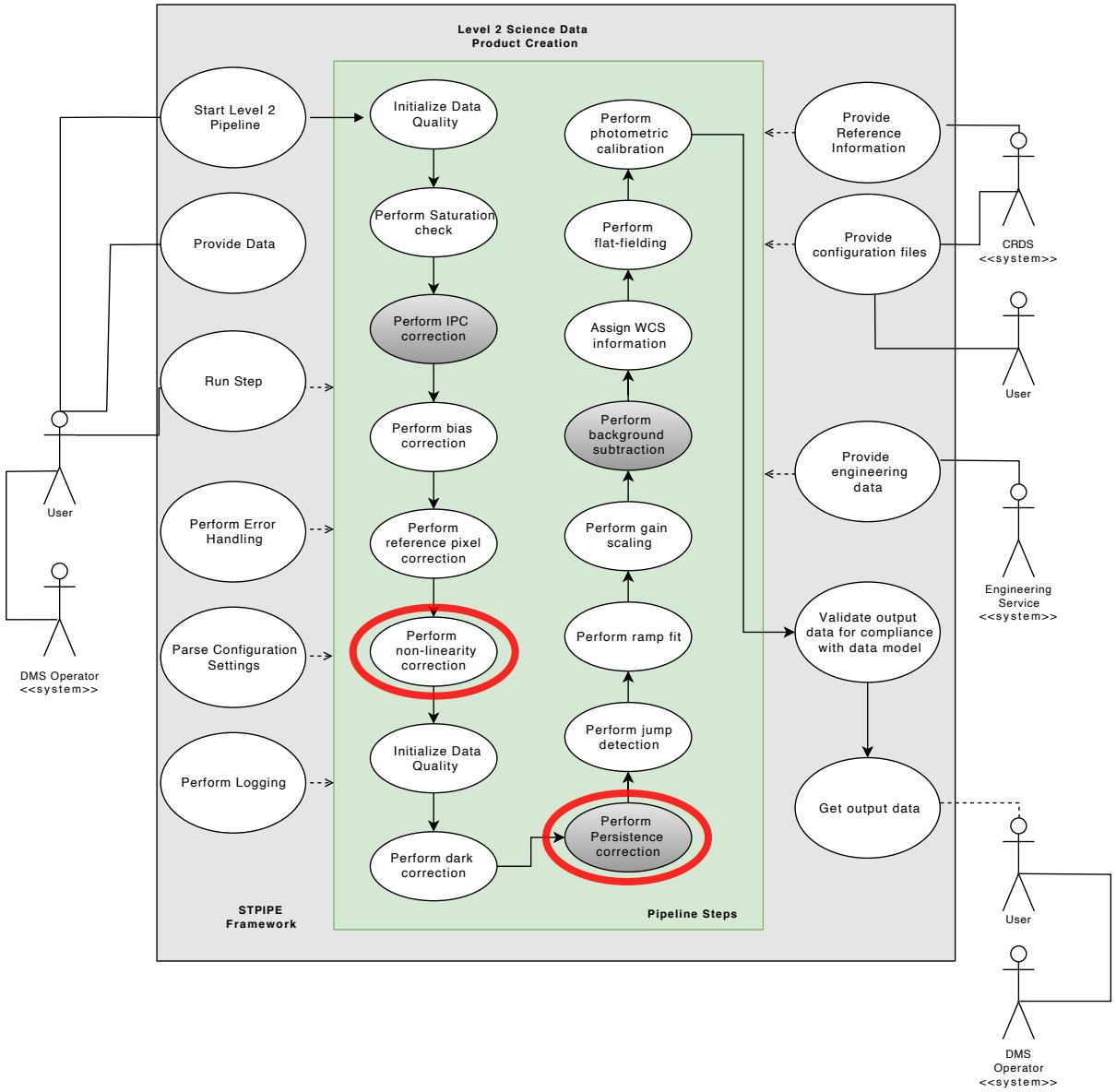
Data and software

Science community

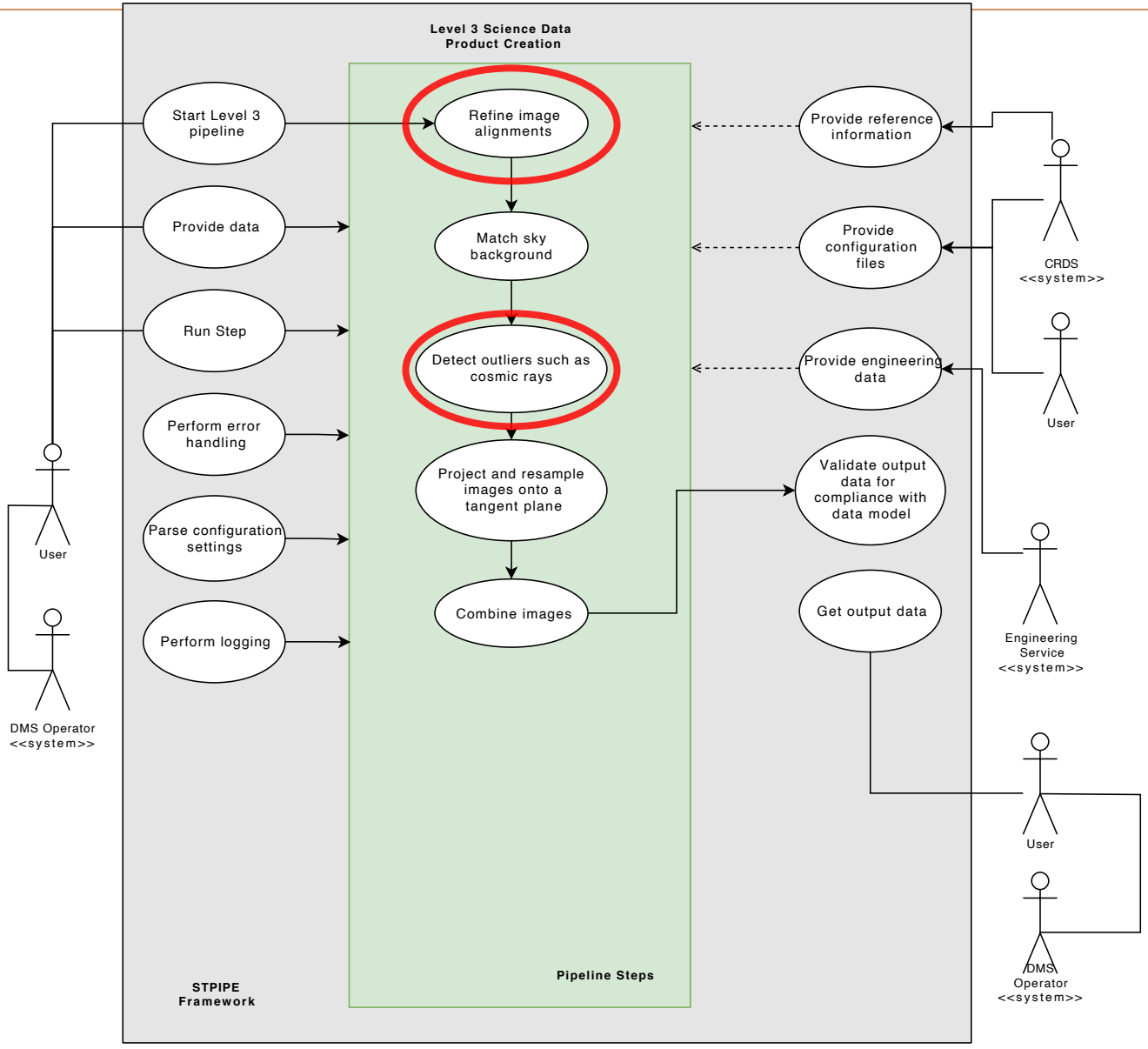
Data Products

- **Array data**
  - Level 1 – raw
  - Level 2 – instrument signatures removed; aligned to Gaia to <1.3 mas precision
  - Level 3 – rectified and co-added
  - Level 4 – segmentation maps associated with catalogs
  - Queryable Empirical PSF library
  - Level 5 – community contributed products
- **Tabular data**
  - Level 4
    - Static Object catalogs
    - Variability catalogs (from aperture photometry and difference imaging)
    - Idealized source-injection: Catalog of inputs vs. outputs (photometry, sizes, shapes)
  - Level 5
    - Community-contributed products
- **Availability (levels 1-4)**
  - At the individual FOV level within 2 days of receipt of last relevant data
  - Consistently-calibrated data of survey areas released within 6 months

Exposure Level Processing Flow

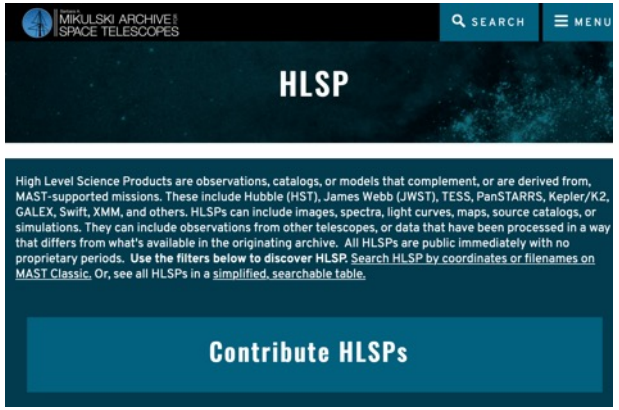


Combining images (level 3 processing)



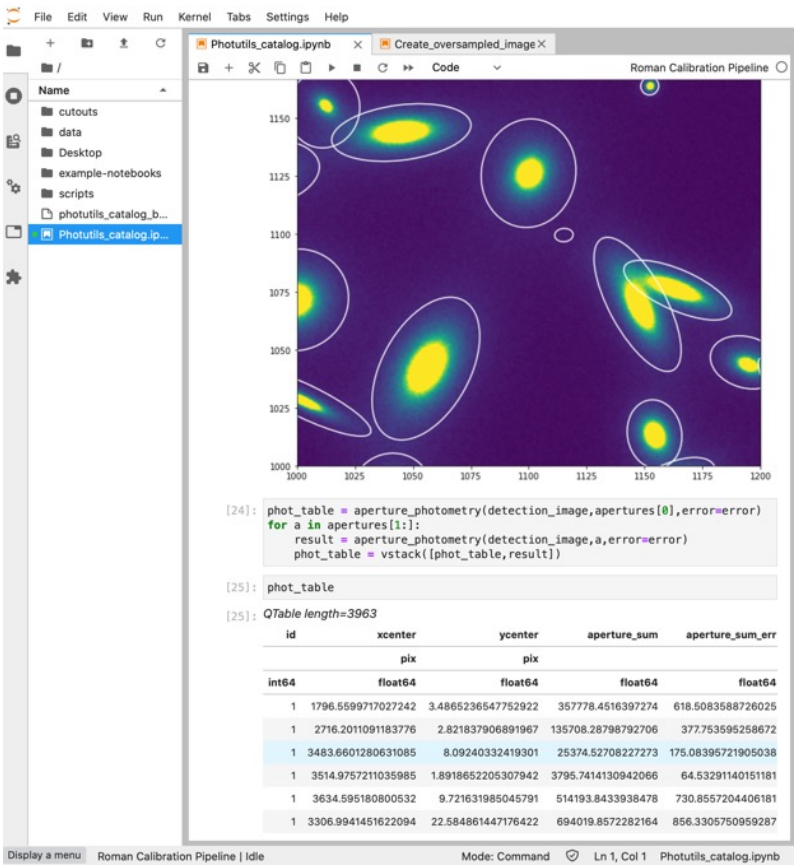
Community-Contributed Products

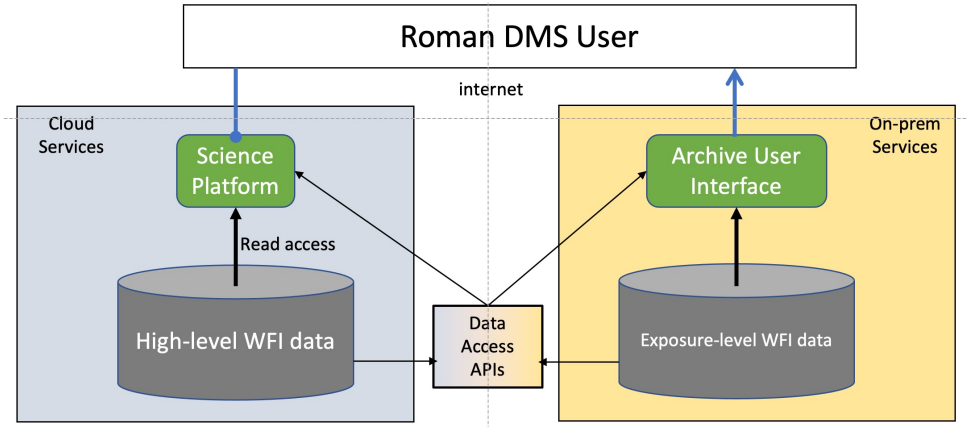
- **Public data products contributed by the science community are likely to be widely used. Examples include:**
  - Joint photometry with complementary data sets
  - Photometric redshifts that use complementary data sets
  - Value-added catalogs of derived properties (e.g. from SED fitting)
  - Hybrid spectroscopic and photometric catalogs
  - Survey-level calibrations
    - Improved astrometry & photometry after constraining for consistency across the full survey
    - Window functions, masks, PSF kernels, etc.
  - Transient-free template images
- **Details & cadence to be defined through future community engagement and opportunities**



Roman Science Platform

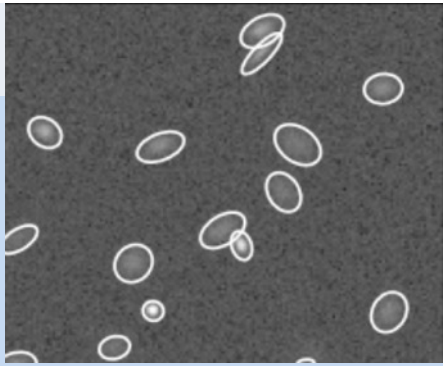
- **Cloud-based**
  - community access to CPU close to the Level 2-4 data products.
  - Roman science calibration pipeline software installed and configured.
  - Full Python + Astropy ecosystem installed and configured.
  - Ability to install other packages and your own code.
  - Easy to scale resources up or down.





Catalogs: Static

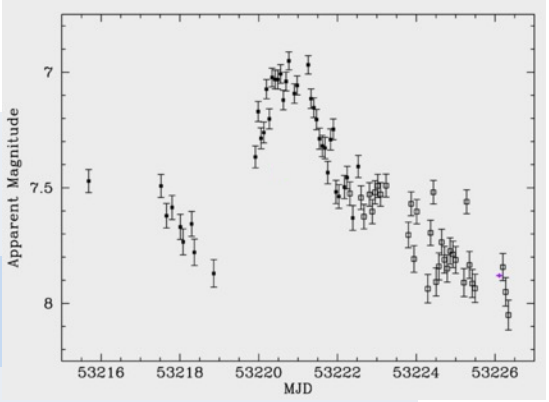
- **Primary cosmology-related goal is to enable accurate photometric redshifts**
- **Accurate, well-characterized photometry and shape measurements will also enable a wide range of general astrophysics**
- **Inputs:**
  - Level 3 images



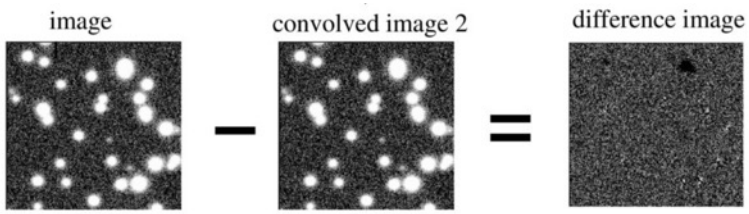
- **Operations**
  - Estimate and subtract background
  - Convolve with a detection kernel
  - Identify connected pixels above a noise-dependent threshold
  - Hierarchically de-blend overlapping sources
  - Measure fluxes through apertures
    - With and without convolution by a PSF-matching kernel to correct all photometry to one reference PSF
  - Measure shapes
  - Star-galaxy classification based on shapes (only)
  - Compute photometric redshifts from multi-band (Roman only) photometry
- **Outputs**
  - Level 4 catalogs including uncertainties computed from noise model
    - Photometry, positions, shapes, & local background estimates
    - Astrometry aligned to Gaia
  - Level 4 segmentation maps
  - Catalogs of input & output parameters for injected artificial sources

Catalogs: Variability (non-bulge fields)

- **To be run in the pipeline to satisfy the requirement for time-domain information for variable sources**
- **Photometry-based variability catalogs**
  - Inputs
    - Release-level merged survey catalogs
    - Individual images (level 2 or level 3)
  - Operations
    - Compare flux in static catalog to fluxes in individual images
  - Outputs
    - Database of the individual-image photometry for the entire survey area with variability index



- **Difference-imaging variability catalogs**
  - Inputs
    - Individual rectified images that overlap spatially
  - Operations
    - Convolve with PSF-matching kernel if needed
    - Subtract a template constructed from all but the most recent image
    - Identify point-sources in the difference above a threshold
  - Outputs
    - Level 4 catalog of sources that exceeded the threshold along with associated metadata



Bulge-microlensing processing at the SSC (IPAC)

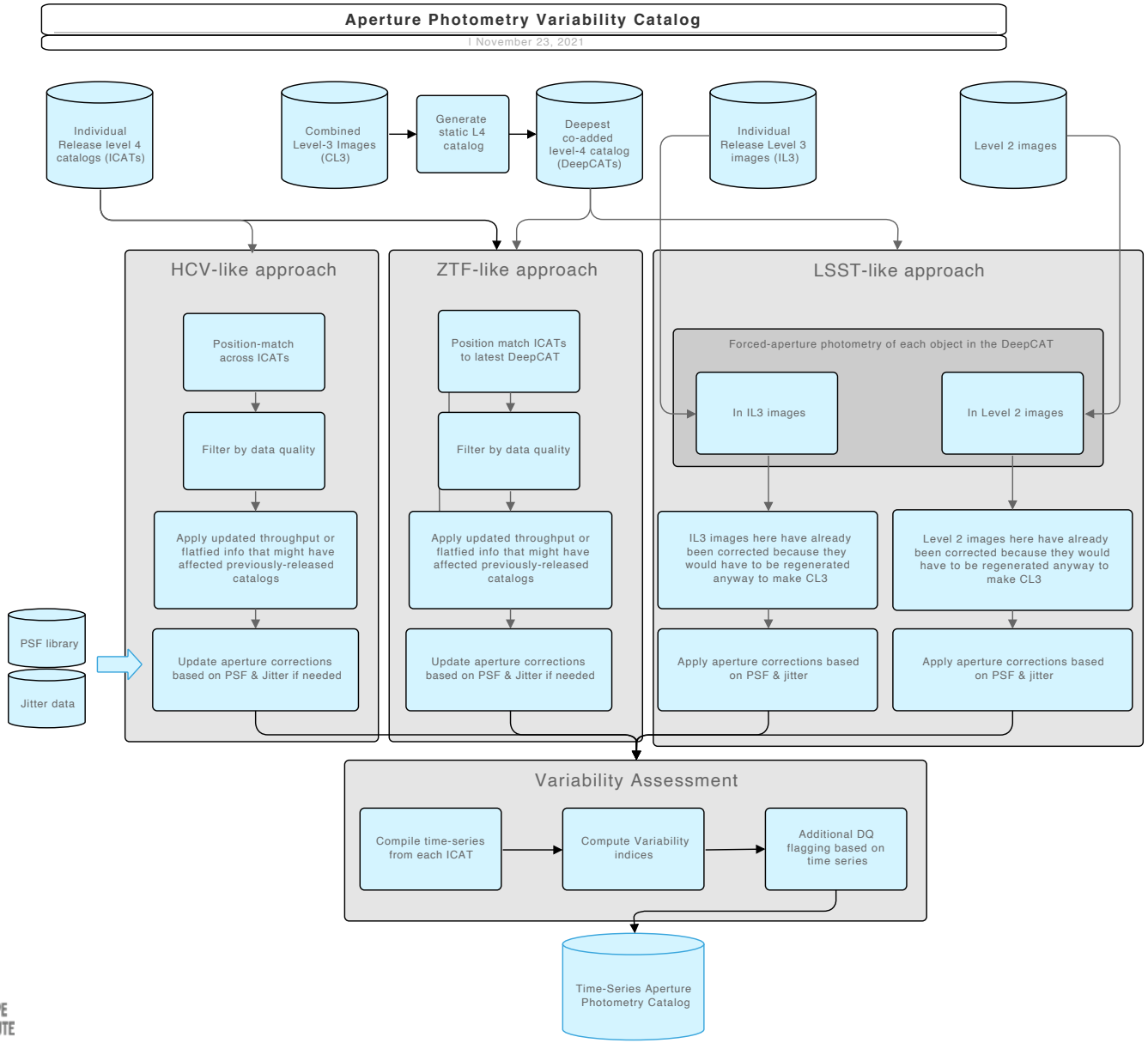
- **The bulge fields will be processed at the SSC to support the microlensing core science.**
- **SSC microlensing processing is tailored to observing strategy and science goals. Includes:**
  - Difference-imaging and PSF photometry
  - Simulations (source injection)
  - Event characterization including physical parameters of the lens system
  - Detection efficiency
  - False-negative and false-positive efficiencies
  - Reddening and extinction maps

Photometry-based Variability Catalog (non-bulge fields)

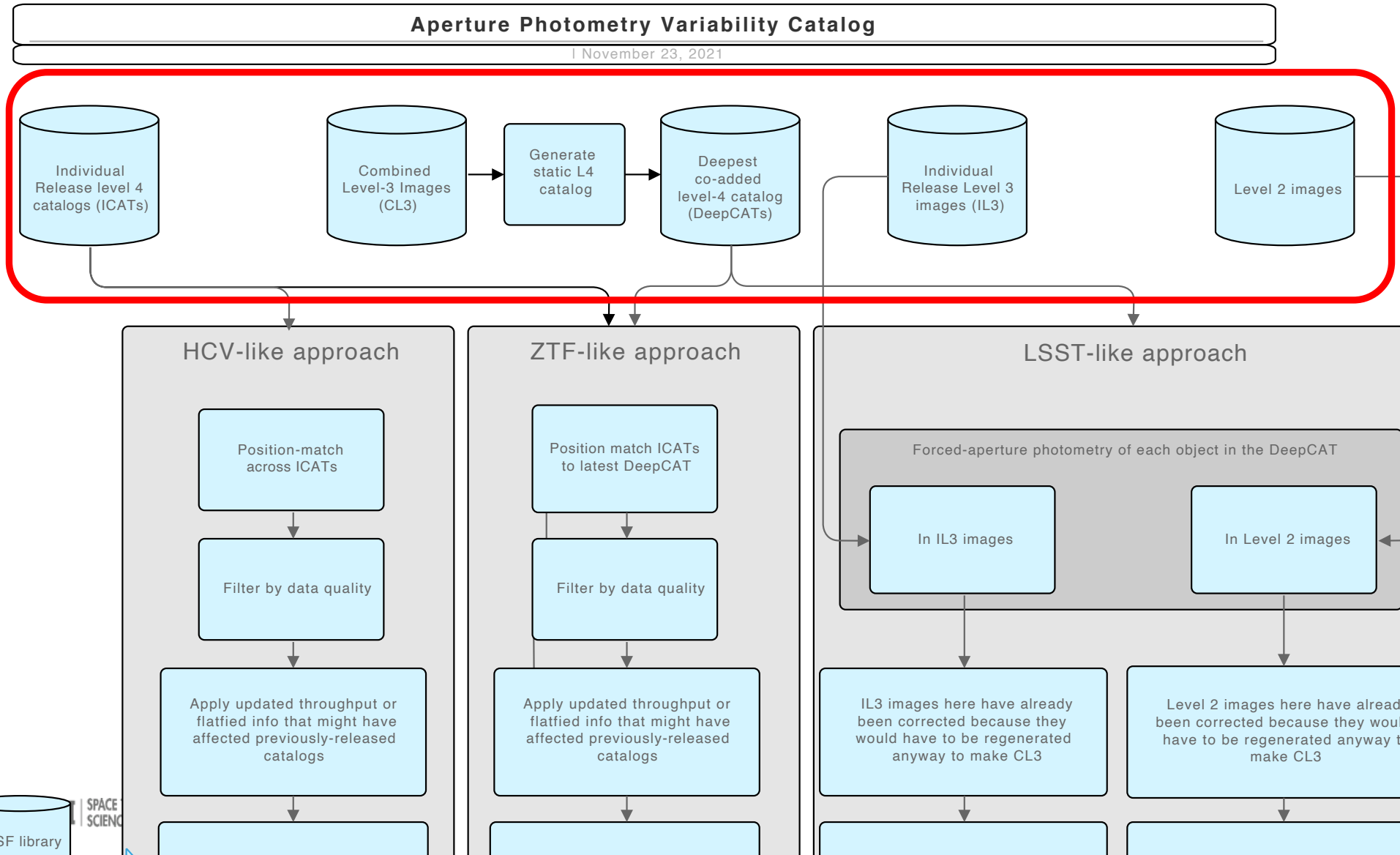
- **Different strategies adopted for different surveys**
  - Hubble Catalog of Variables (HCV)
  - Zwicky Transient Factory (ZTF)
  - Rubin Legacy Survey of Space and Time (LSST)
- **Formulation of detailed strategy for Roman is underway. Now is a good time to influence important aspects of the implementation.**

- **Trades:**
  - Catalog-match individual epochs (no forced photometry or reference to a deeper catalog) --- (HCV approach)
  - Catalog-match individual epochs to a deeper catalog (ZTF approach)
  - Forced-photometry at the positions of objects in a deep catalog (LSST approach)
    - What to do about proper motions?
  - When to do the processing?
    - Uniform survey-level catalogs have roughly 6-month cadence

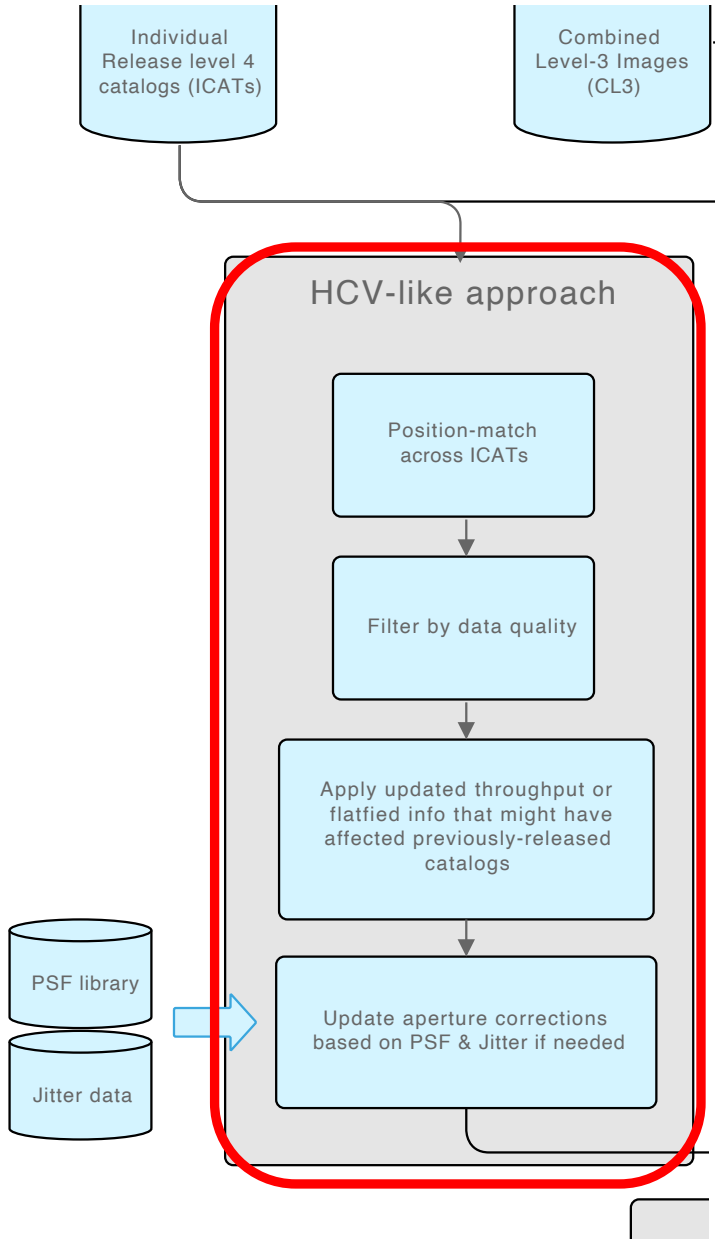
Comparing approaches from other surveys



Comparing approaches from other surveys



Comparing approaches from other surveys

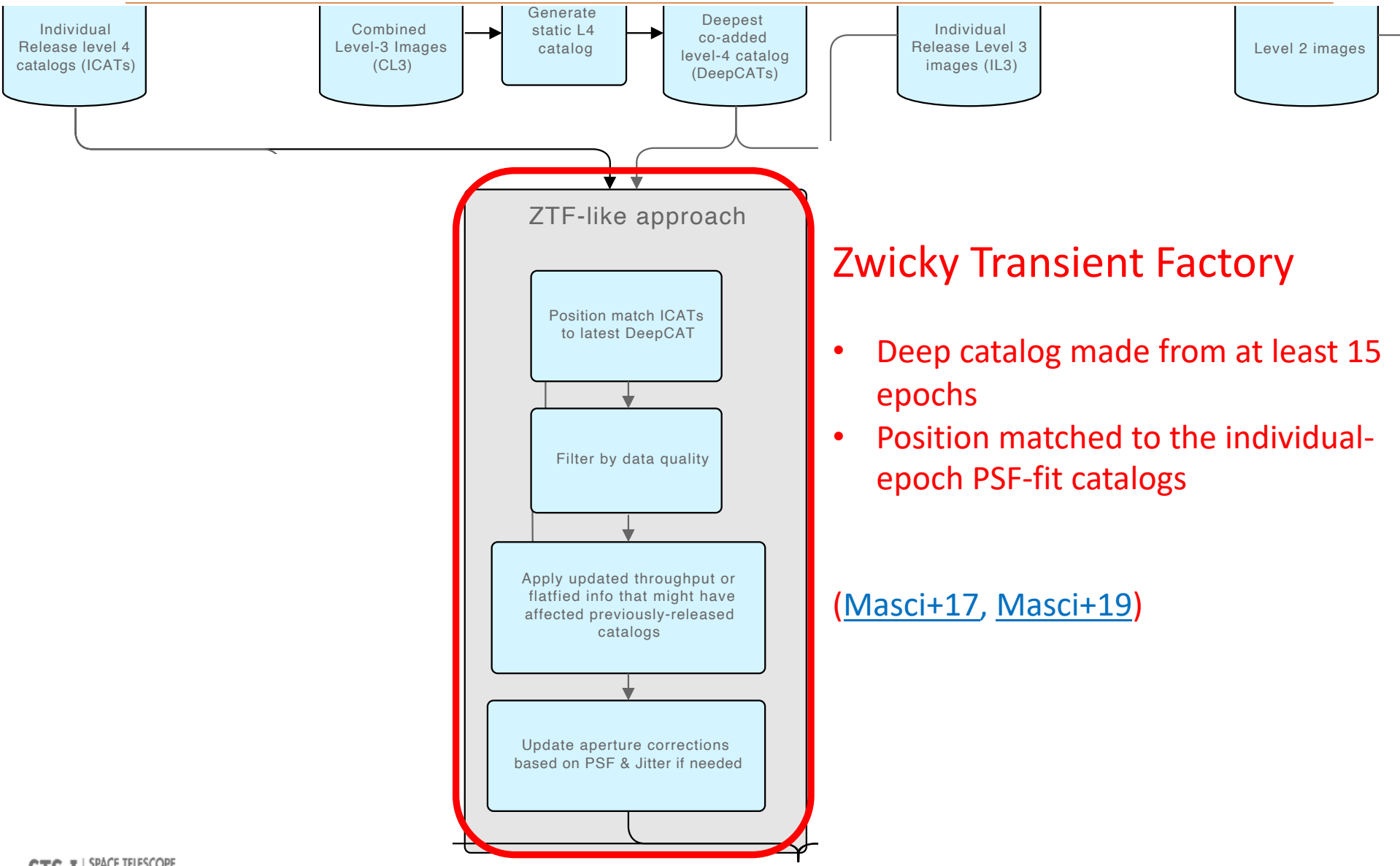


Hubble Source Catalog

- No new photometry (uses Hubble Source Catalog data)
- Restricted to fields with at least 300 sources in common.
- Catalog includes sources with at least 5 epochs in the same filter.
- Artifact rate ~ 19%

(Bonanos+19)

Comparing approaches from other surveys

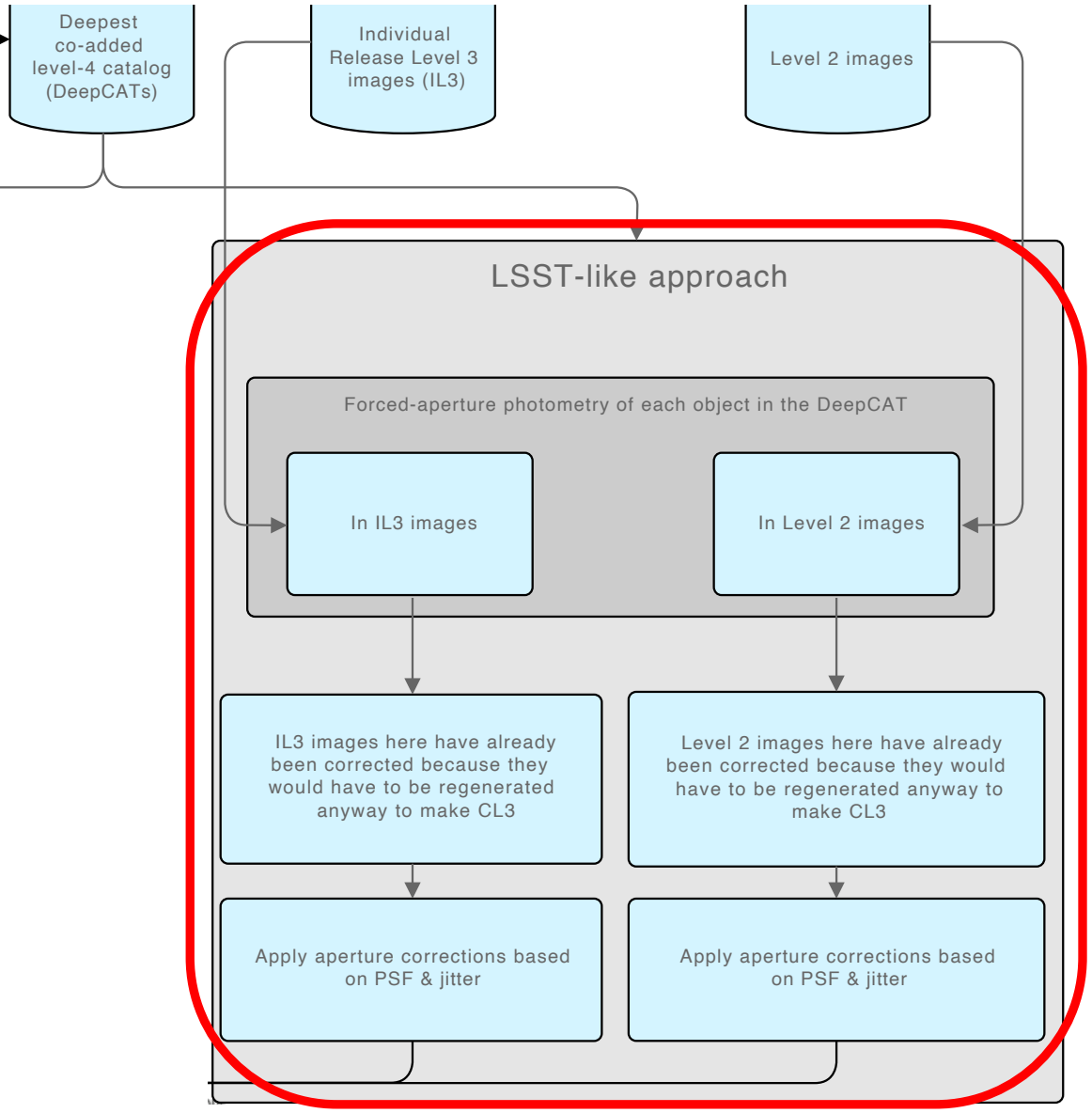


Zwicky Transient Factory

- Deep catalog made from at least 15 epochs
- Position matched to the individual-epoch PSF-fit catalogs

(Masci+17, Masci+19)

Comparing approaches from other surveys

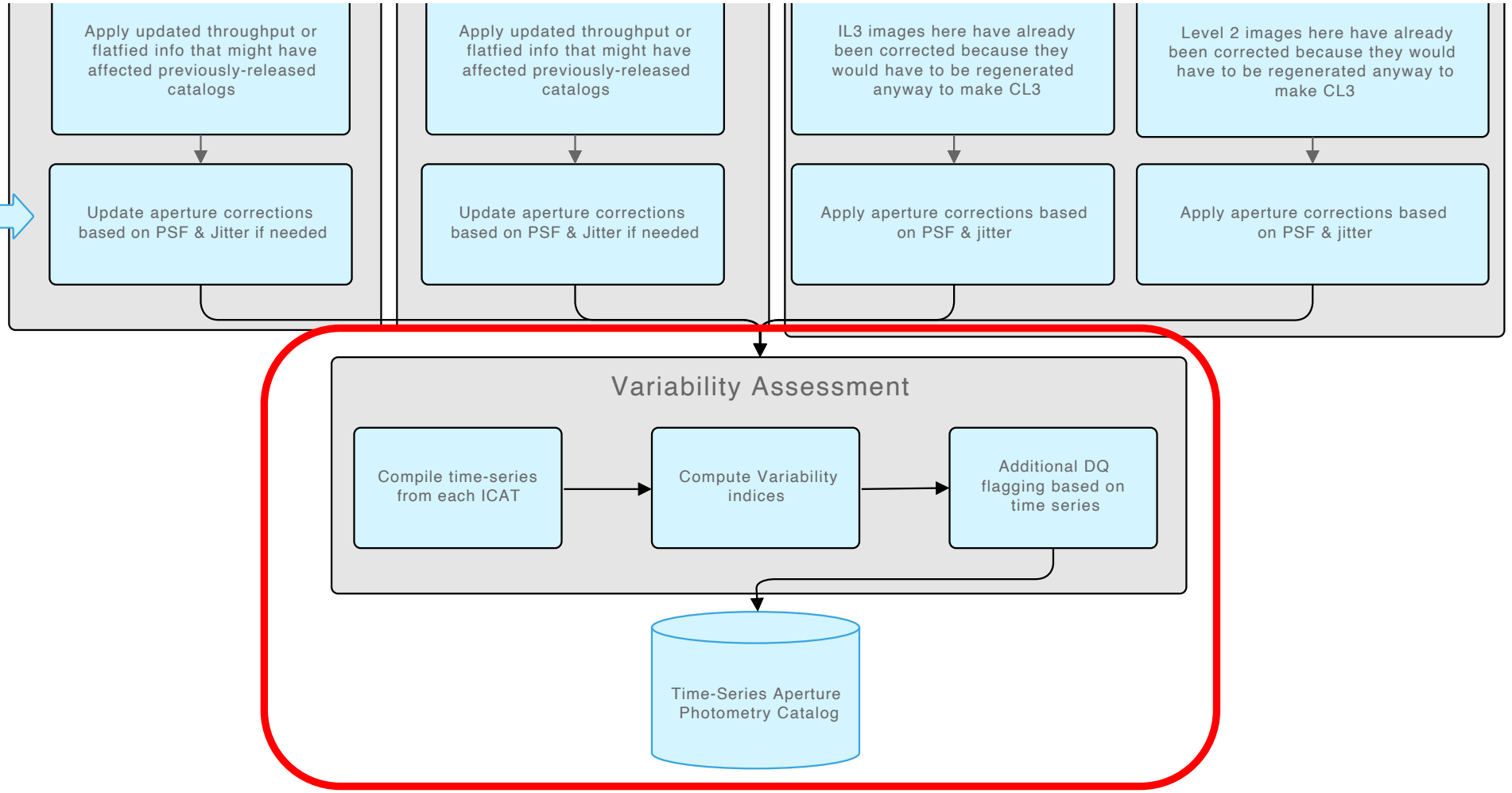


Rubin / LSST

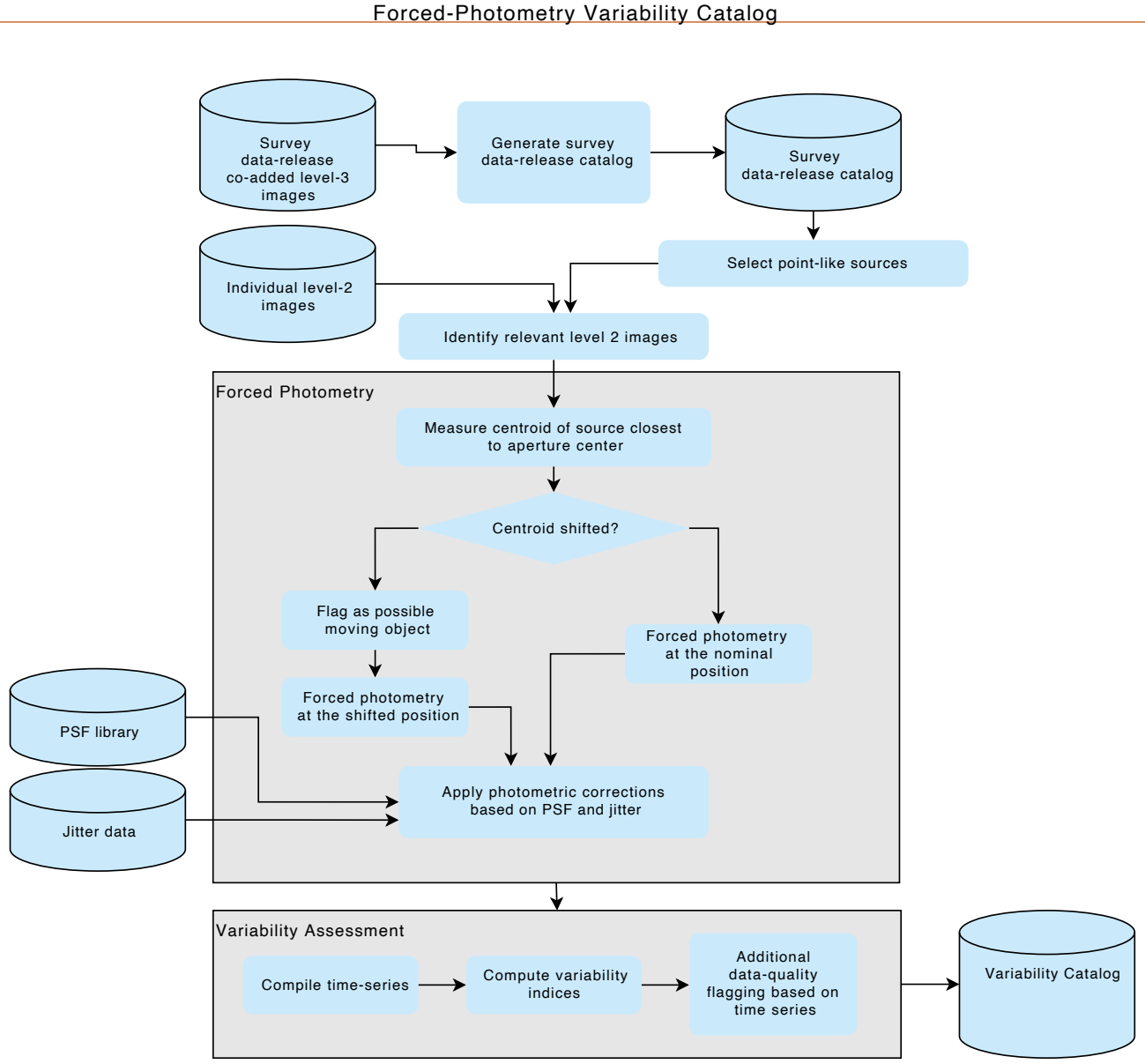
- Forced photometry catalog is a Data-Release product (1-year timescale)
  - Prompt products are only from difference imaging
- Reference images "locked" from 15-40 images
- Catalog photometry in all visits for objects detected both in co-adds and in individual visits
  - Algorithm for matching not described
- Includes proper motion, parallax, variability, periodicity & other characterizations.

(Data Products Definition Document)

Comparing approaches from other surveys

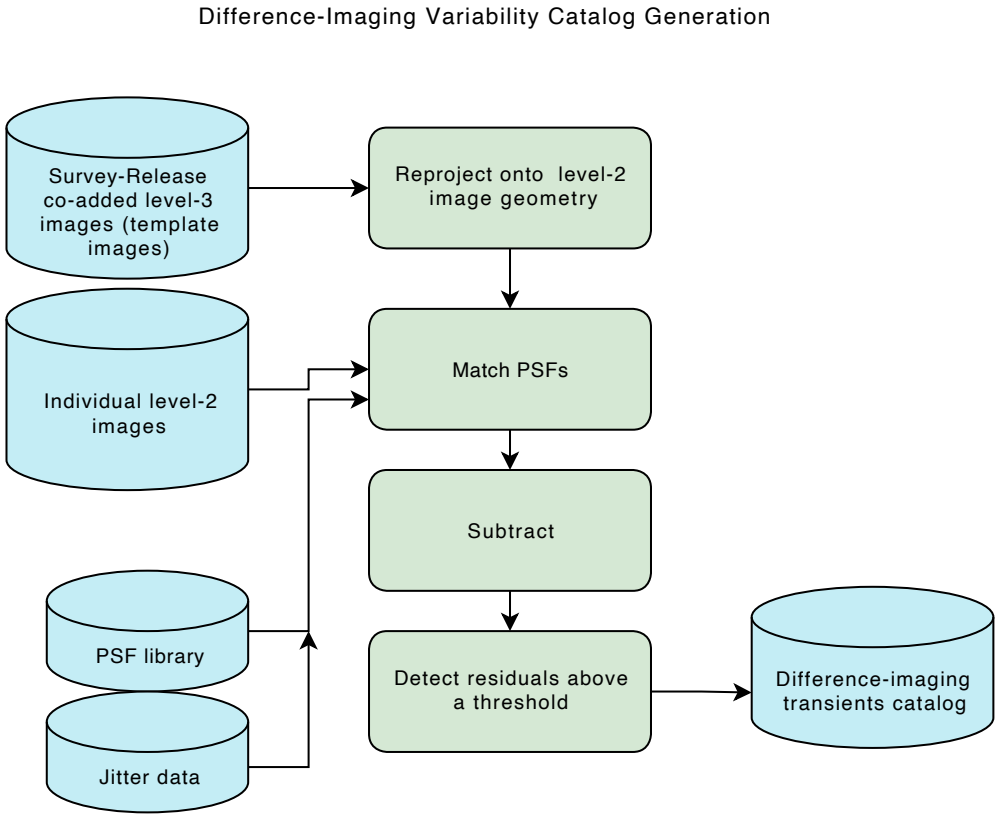


Possible strategy for Roman



Difference imaging data-processing strategy

- **Better for finding embedded variable sources**
  - Superimposed on a host galaxy
  - In crowded regions
- **Implementation details very much TBD**
  - E.g. Are template images coming from survey releases or other kinds of stacks?
  - How (or even whether) to do the PSF matching?
  - Will astrometric corrections be good enough without iteration?
  - Will jitter be small enough to ignore?



Concluding Remarks

- With a stable PSF and wide field of view, Roman will be a superb instrument for time-domain discoveries.
- The SOC at STScI is currently designing the pipeline for time-domain catalogs.
- Now is a good time to consider implementation strategies and data-processing trades.