Joint Survey Processing of Roman/Euclid data with ground-based (Rubin/LSST) data

"Whole>>Sum of Parts"

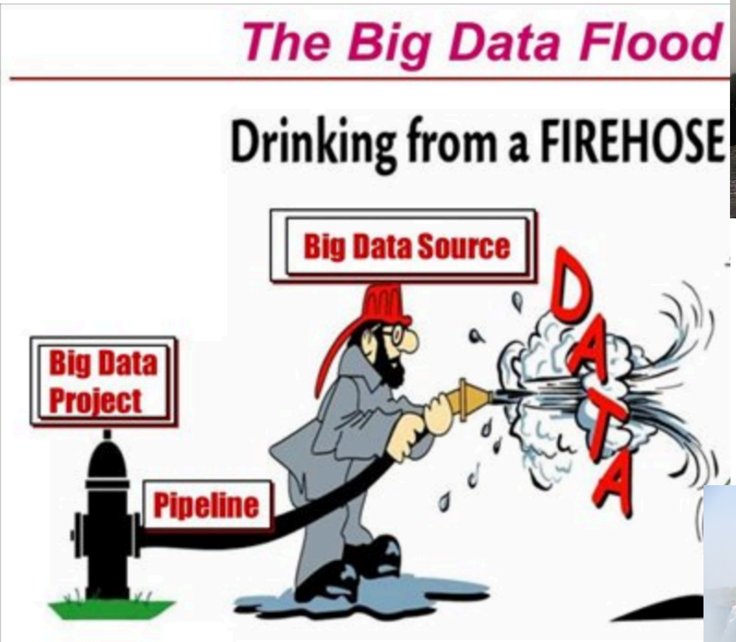
Ranga-Ram Chary (Caltech/IPAC)
for the joint survey processing (JSP) working group

~40 scientists/engineers from 10 different institutions did the groundwork for the final JSP report

rchary@caltech.edu

21 June 2021, arXiv:2008.10663

Faisst et al. 2021, arXiv:2103.09836



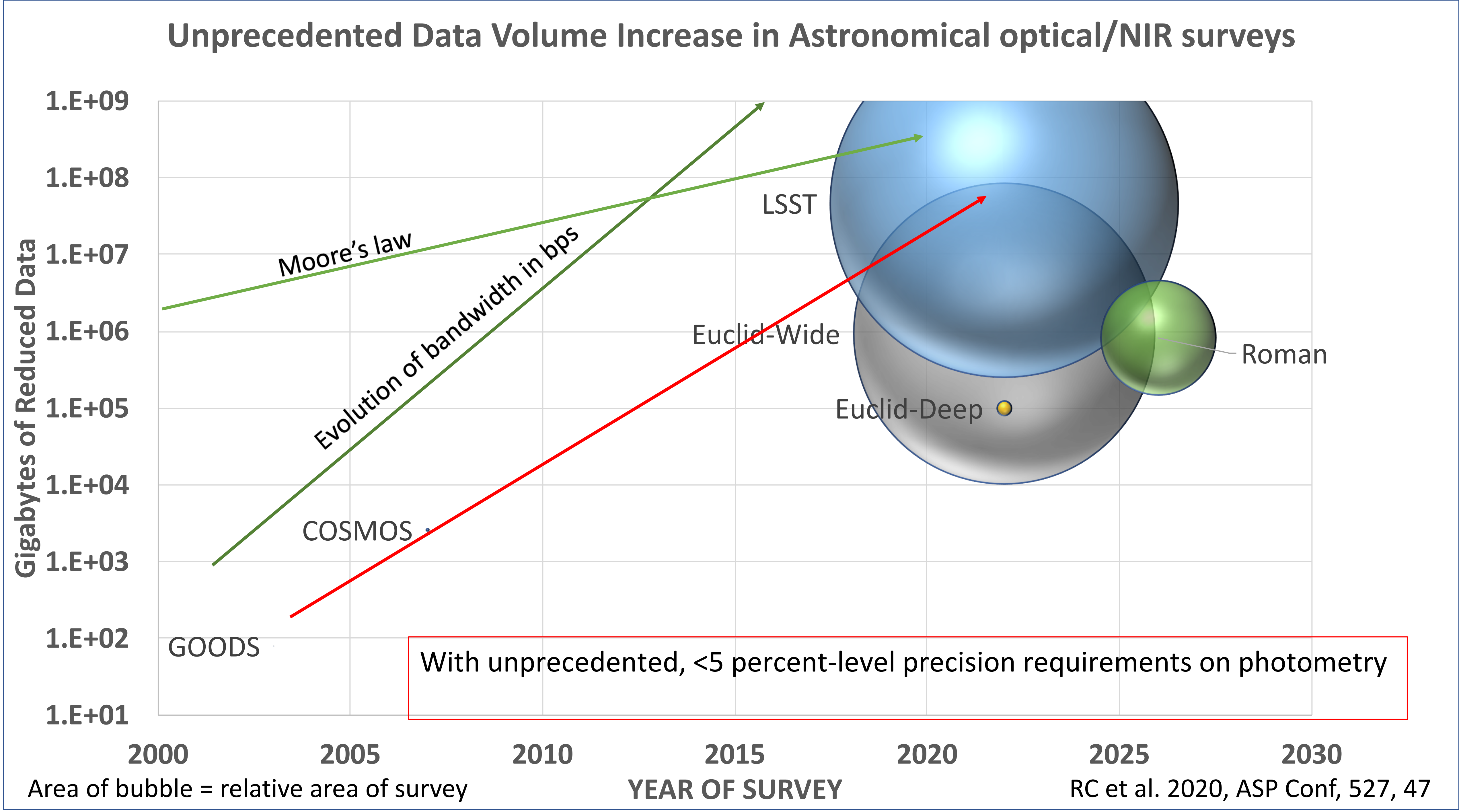
Euclid/Firehose (2022+)

LSST/Iguazu (2023+)

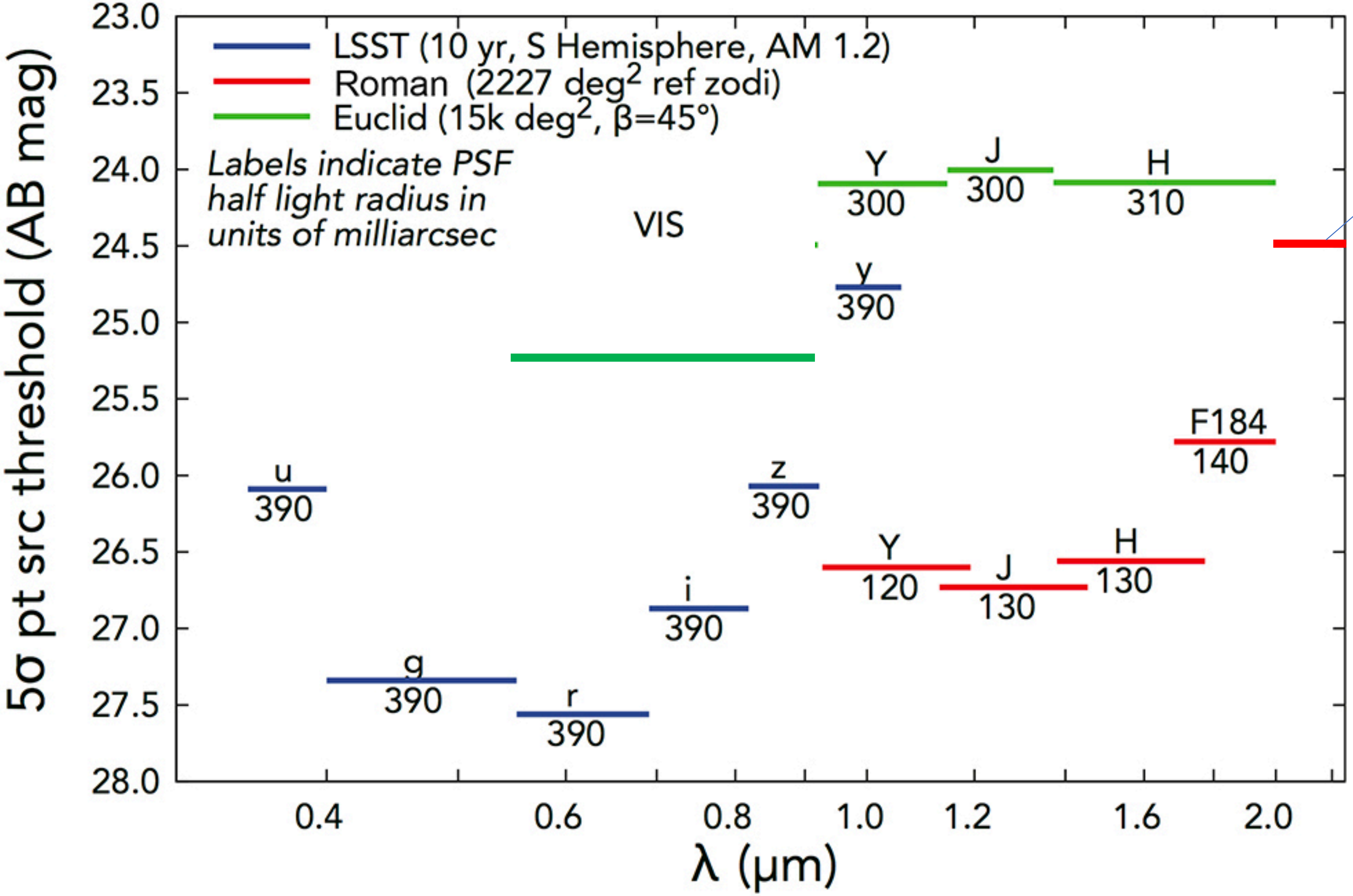
Roman/Victoria Falls (2026+)

Option 1: Fend for Yourself

Option 2: Channelize and reap the fruit
i.e. Joint Survey Processing


Unprecedented Data Volume Increase in Astronomical optical/NIR surveys
Gigabytes of Reduced Data
1.E+09
1.E+08
1.E+07
1.E+06
1.E+05
1.E+04
1.E+03
1.E+02
1.E+01
Moore's law
Evolution of bandwidth in bps
LSST
Euclid-Wide
Roman
Euclid-Deep
COSMOS
GOODS
With unprecedented, <5 percent-level precision requirements on photometry
2000
2005
2010
2015
2020
2025
2030
Area of bubble = relative area of survey
YEAR OF SURVEY
RC et al. 2020, ASP Conf, 527, 47

Sensitivities of LSST, Roman, and Euclid



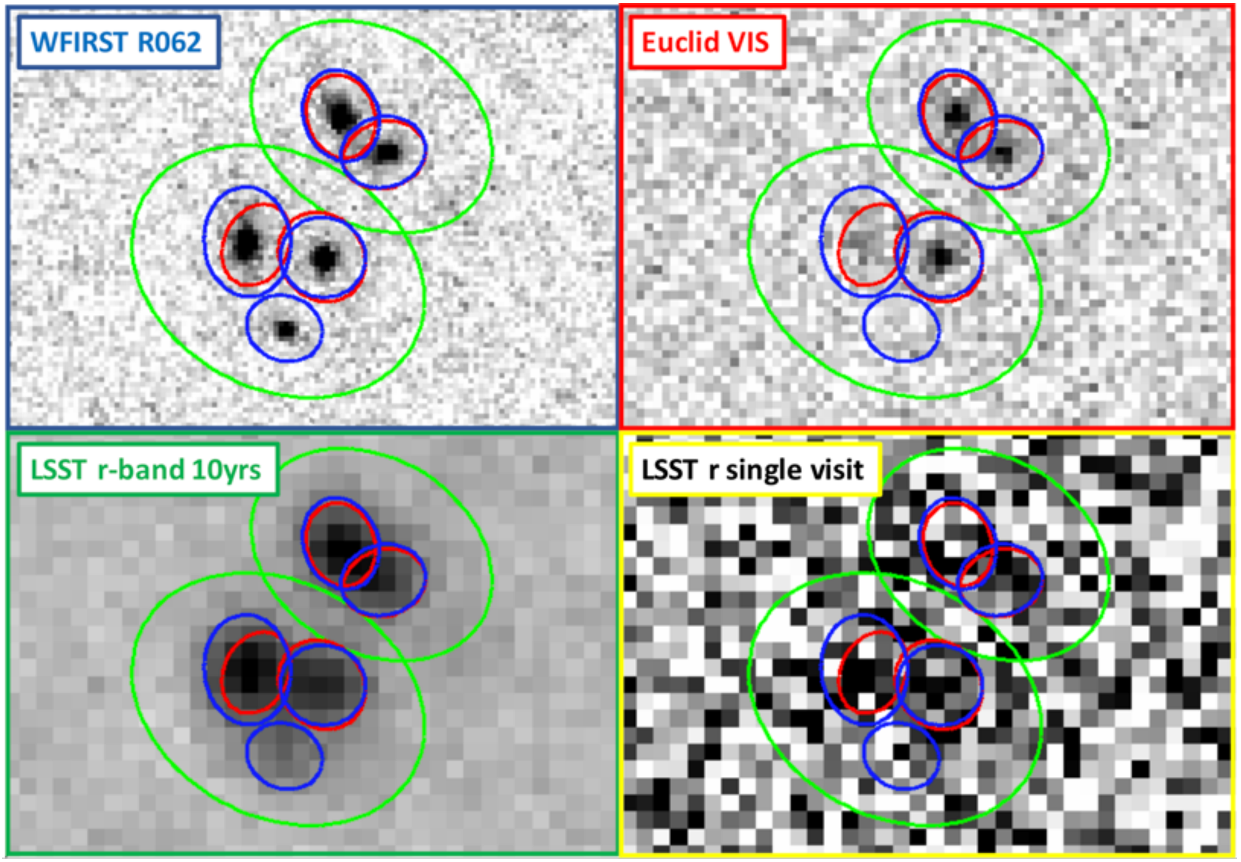
Roman Kp
A new WFI addition (likely not used for HLS)

Matched isophotes, deconfusion and precision

0.1” FWHM

0.2”

0.8”



This is in a typical extragalactic field. Confusion is higher in Galactic fields or cluster fields.

Lee & RC, 2019, ApJ

Matched isophote photometry at the <5% level is challenging

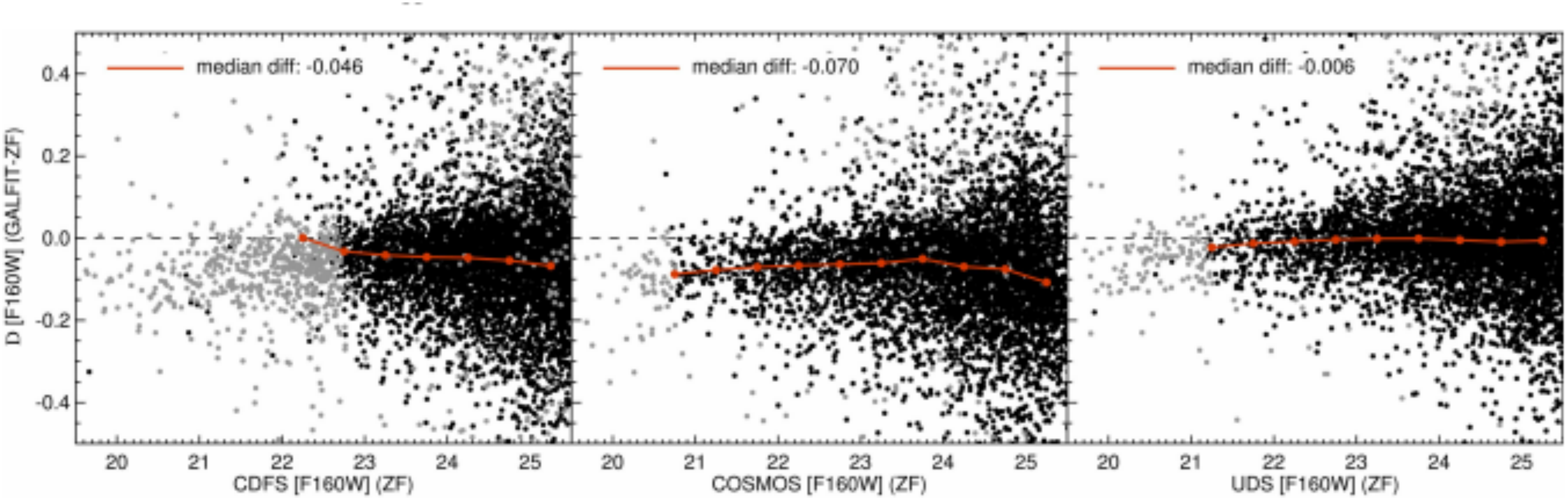
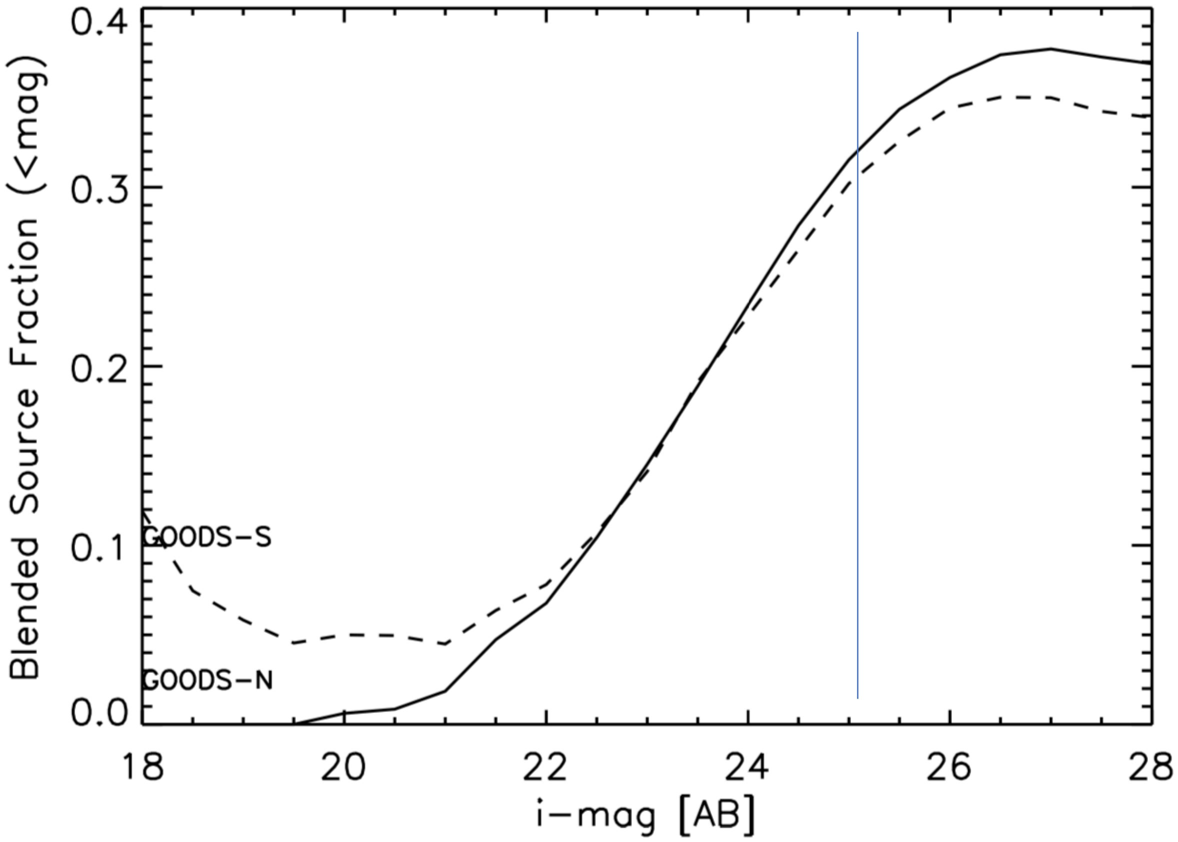


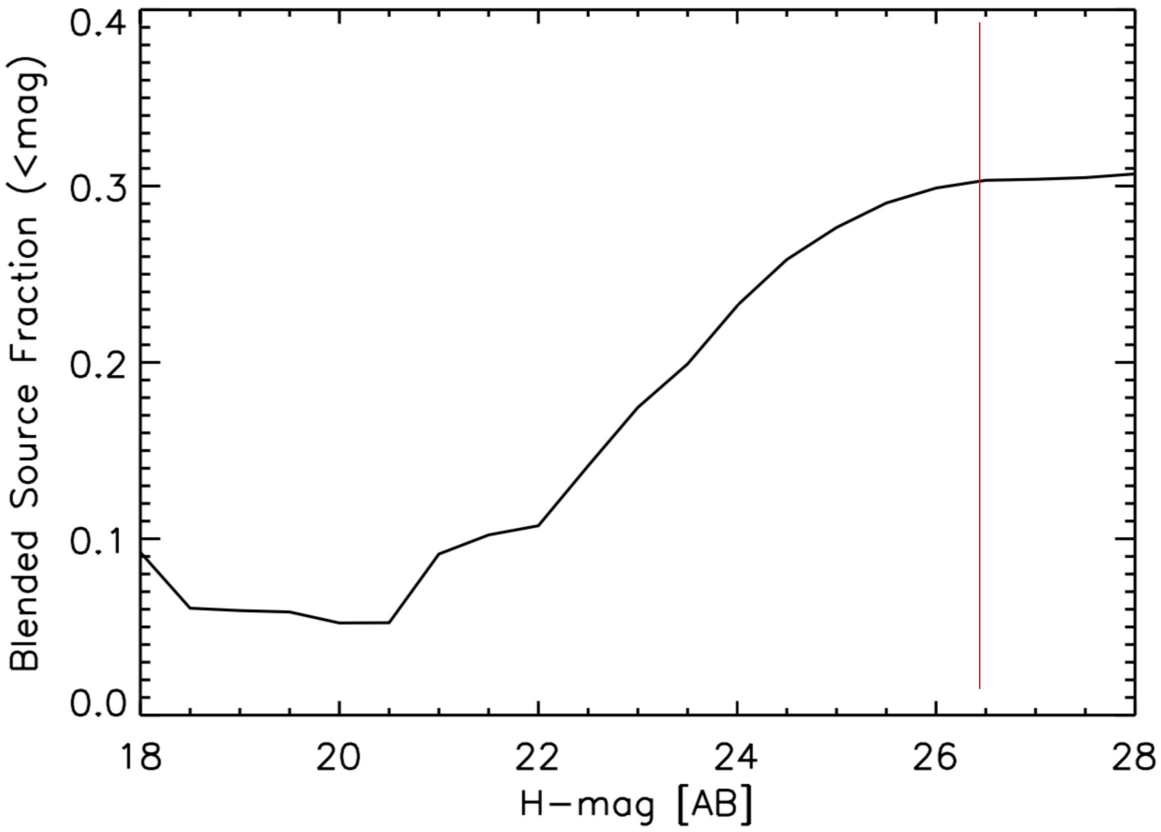
Figure 11. The difference between ZFOURGE and GALFIT F160W magnitudes plotted as a function of ZFOURGE magnitude, for galaxies with use = 1 and SEflag = 0 (excluding blended or contaminated sources). We show sources with GALFIT flag = 1 (a suspicious fit) in gray and sources with GALFIT flag = 0 (a good fit) with black datapoints. Bad fits (GALFIT flag >1) were ignored. The median magnitude difference for galaxies with GALFIT flag = 0 is shown by the red solid line and filled bulletpoints in bins of 0.5 mag. We also indicate the median offset in the legend. We find slightly brighter magnitudes with GALFIT, of 0.006–0.070 mag on average, with the difference presumably attributable to different techniques to derive total magnitudes and potential color gradients.

Straatman et al, 2016, Skelton et al. 2014

Blended source fraction in ground-based data (0.8” FWHM seeing)



Optical bands – Euclid/VIS relevant



NIR bands – Roman relevant

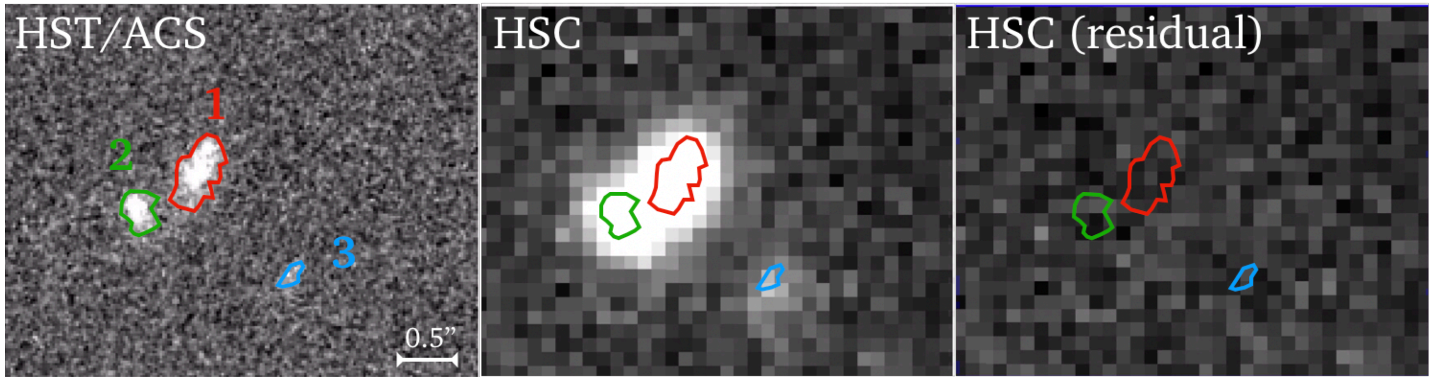
These are lower limits!

Ways of Combining Datasets

- Catalog matching [NOT joint processing]
  - Easy to handle, will be done by the projects at basic level
  - Photometric techniques from each project is different
  - Confusion and mis-association will plague the matching
- Joint fits in coadded pixel level data to photometry/positions/sizes of sources
  - Space based data can be used as a template for deconfusion, bulge/disk ratios
  - Consistent analysis across wavelengths, including extraction of limits
  - Uses robust color and location dependent PSFs
- Joint fits in single-epoch pixel level data to photometry/positions of sources
  - Astrometry from each epoch will be properly handled
  - Seeing variations will be factored into account during photometry
  - Transient sources will be properly handled, enabling selective stacks
  - Allows combining spectral grism data and imaging data
  - Most voluminous and challenging for infrastructure

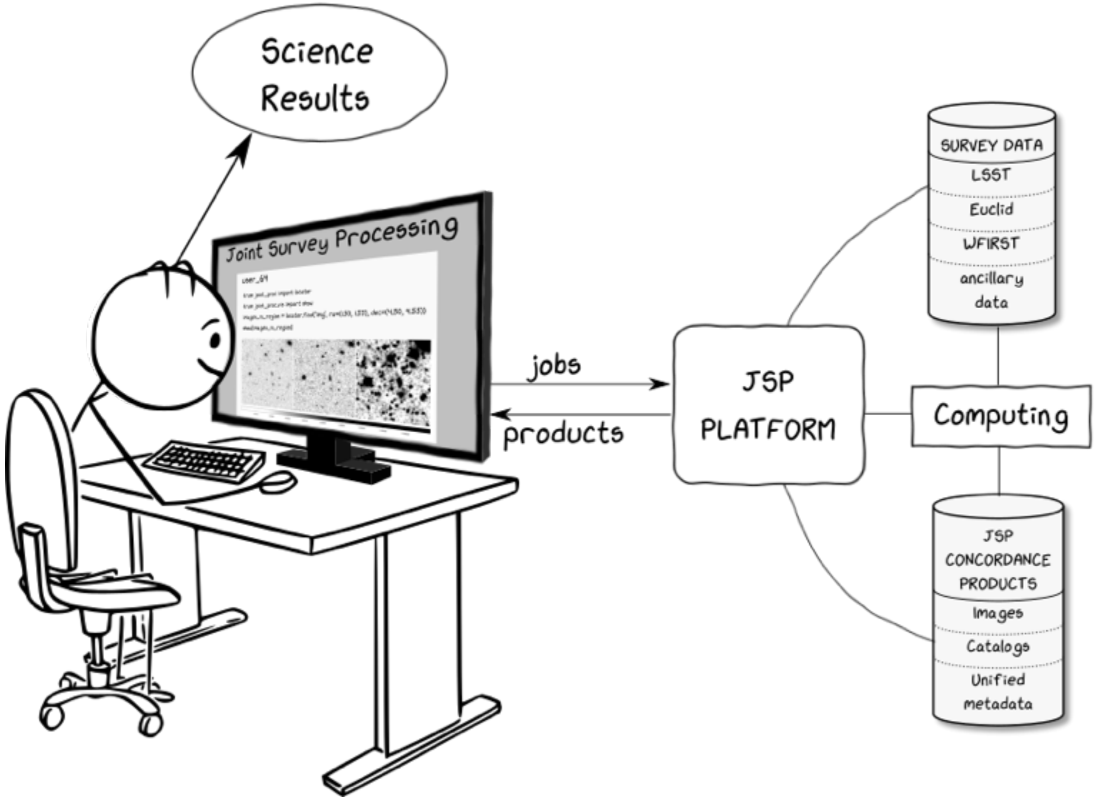
Joint Survey Processing for Euclid-Rubin-Roman

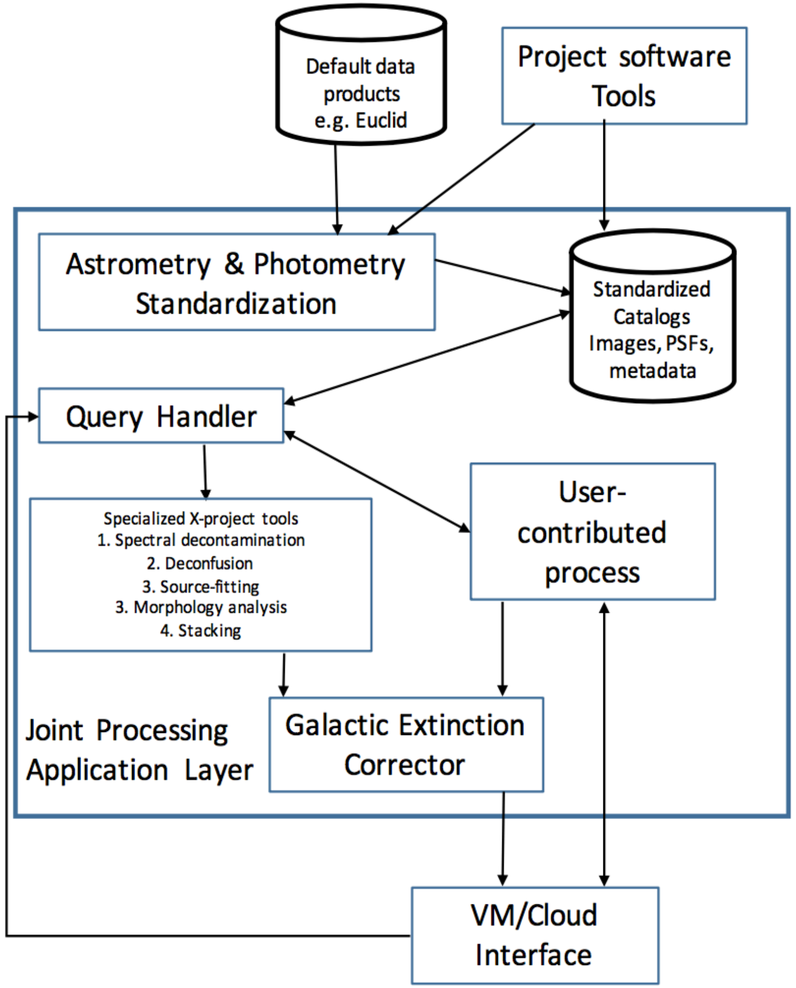
- Catalog matching is dead. Data are too deep, source density is too high and resolutions are too varied for precise cross-project photometry.



- Joint survey processing is pixel-level joint photometric analysis after matching astrometry, accounting for source surface brightness profiles, accounting for difference in PSFs and deblending.
- We have been doing prior-based photometry since 2003, first with Spitzer, and then Herschel – this is more challenging because ground-based seeing PSFs vary, and photometric precision requirements are more stringent (particularly for cosmology).
- JSP is needed not just on the image stacks – there are science cases for which JSP needs to be done on single epoch data.

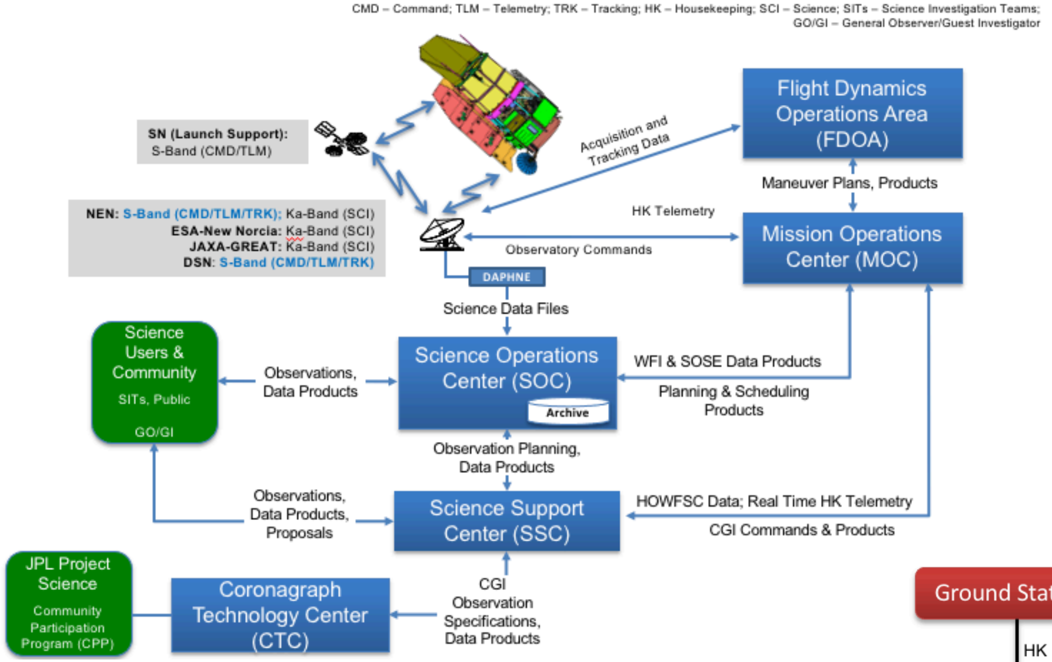
Basic JSP Architecture: Leveraging both existing HPC resources and cloud environment

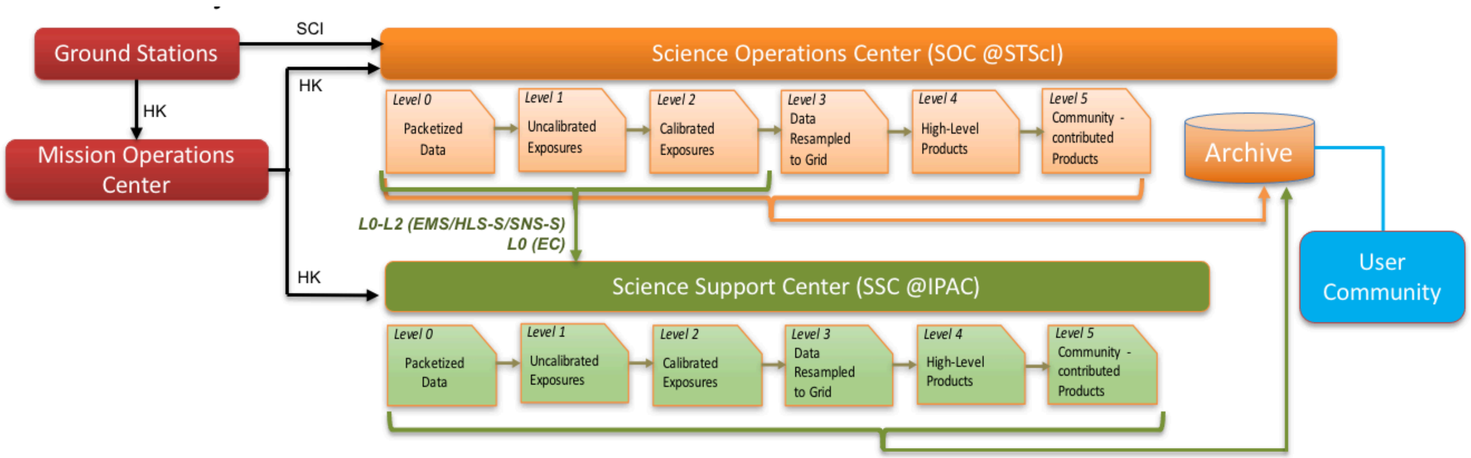




We start with reduced data products from each projects pipeline

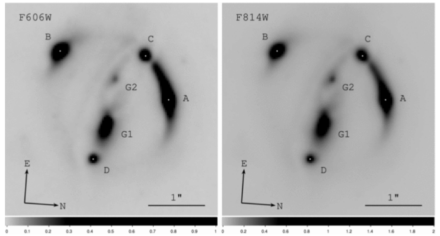
The NGRST (Roman) Data System





Slides from R. Akeson, L. Armus

JSP on Coadds vs Single-epoch images – Use cases



- Cosmology in the time domain
  - Strong lensing time delays
  - Z>1 Type Ia Supernovae identification

- Cosmology with coadds
  - Better shapes: Bulge/disk ratios for joint weak lensing shear estimates
  - Precise and accurate photometry for 3D WL and galaxy clustering
  - Reliable scene modeling for transient detection e.g. Type Ia SNe host galaxy
  - Accurate resolution-matched galaxy color selection for EoR studies (Faisst et al. 2021, arXiv:2103.09836)

- Astrophysics in the time domain
  - Supernovae identification
  - stellar motions for dark matter and tidal streams (currently obtaining 2 mas/year proper motion)
  - AGN, stellar and asteroidal variability
  - microlensing in the shot noise limit

Summary

- We should be concerned by the upcoming data volume and precision requirements.
- Catalog matching is dead.
- For the Whole>>Sum of Parts, requires capabilities for joint pixel level analysis and rapid visualization – the goal of JSP (arXiv:2008.10663).
  - Required for precision cosmology
  - Benefits all areas of astrophysics
- All projects (with datasets) should work together with the bigger Legacy projects to have a uniform and scalable data manipulation interface with smart software for easy integration.
- Build up >10Gbps networking capabilities at archives.
- Will need >Billion CPU hours for basic JSP, around 10 times that for all science projects with these datasets.
- HSC-HST JSP is a great prototype for Roman-Rubin (Faisst et al. 2021, arXiv:2103.09836). Euclid-Rubin JSP will provide a better understanding of limitations and systematics that need to be addressed.

Overview of Tasks

- Standardize the astrometry and photometry of the image-level data products and associated metadata available from each of these projects;
- Leverage these standardized products to generate the ultimately precise, deconfused, extinction-corrected, photometric catalogs over the entire overlapping sky area;
- Support the generation of such catalogs with other ancillary data sets that will be taken in support of Euclid and Roman e.g. with CFHT, Hyper-SuprimeCAM;
- Provide an interface for astronomical manipulation (e.g. shape fitting, diffuse emission extraction) of the standardized, calibrated frames from Euclid/LSST/Roman;
- Provide an interface for integration of the standardized Euclid/LSST/Roman catalogs to ancillary, lower-resolution, all-sky data products such as GALEX and WISE;
- Develop a high performance computing and networking environment which will enable manipulation of these data and allow monte-carlo simulations to be performed using the data sets from these projects.

Paradigm (Whole>$\boldsymbol{\Sigma}$parts)

- Well-documented set of standardized joint products
- Clear division of tasks between project, joint-processing and community and avoiding duplication
- Enables new multi-wavelength science while doing "old-science" better, reduces systematics
- Analysis and simulations are run on single copies of data located at the data processing institutions rather than moving (vast amounts) data between multiple institutions
- Provides a way for intellectual resources of projects to support community analysis – crucial when data-taking and processing is out of hands of end-user