Exploring Roman Data in the Cloud

Integrating ipyaladin and Jdaviz in the Roman Research Nexus

Tom Donaldson, Sarah Betti, Jackie Champagne, Brian Charlow,
Patrick Custer, Hatice Karatay, Brett Morris, Celia Parts
Space Telescope Science Institute

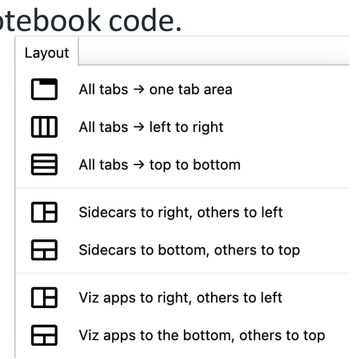
Cloud analysis for Roman's data volume

Roman's will produce 500 times more data per day than HST, and 23 times more than JWST, making local downloads impractical for many workflows. Analysis near the data is essential.
The Roman Research Nexus (RRN) is STScI's cloud science platform on AWS, combining JupyterLab, Roman software, shared workspaces and scalable compute.
The tools described here will be included automatically in the RRN, accessible via both UI and API.

One notebook, connected views

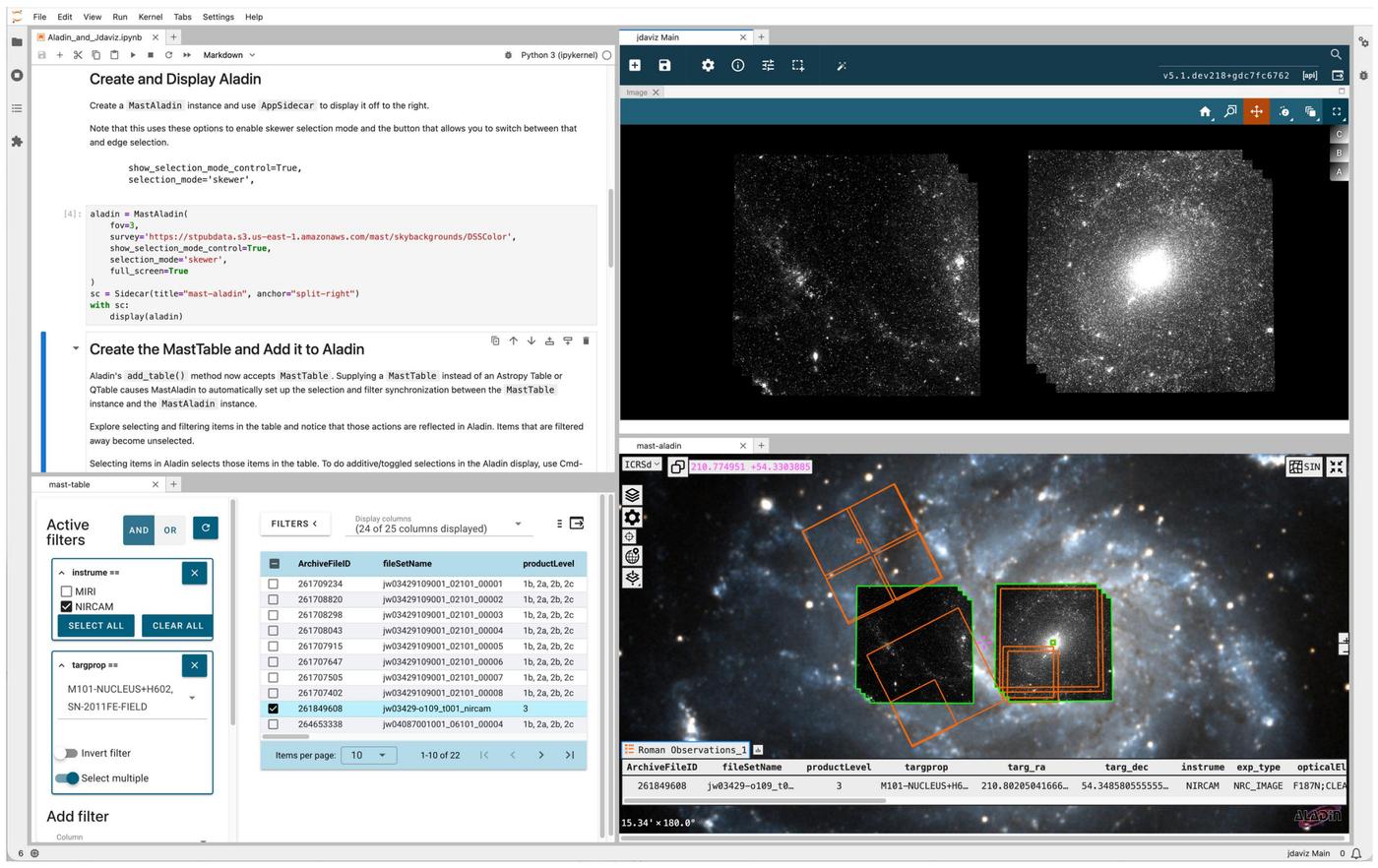
motorcycle

Layout menu arranges sidecar widgets beside notebook code.



mast-table

Package built with Solara/Vue3 to visualize, filter, and select table rows; see matching sources and footprints highlighted in Aladin.



Jdaviz

Interactive visualization and analysis of images, spectra, and catalogs. Allows for any mix of interactive point-and-click operations and advanced scripting via a Python API for bulk and reproducible analysis.

mast-aladin

ipyaladin / Aladin Lite

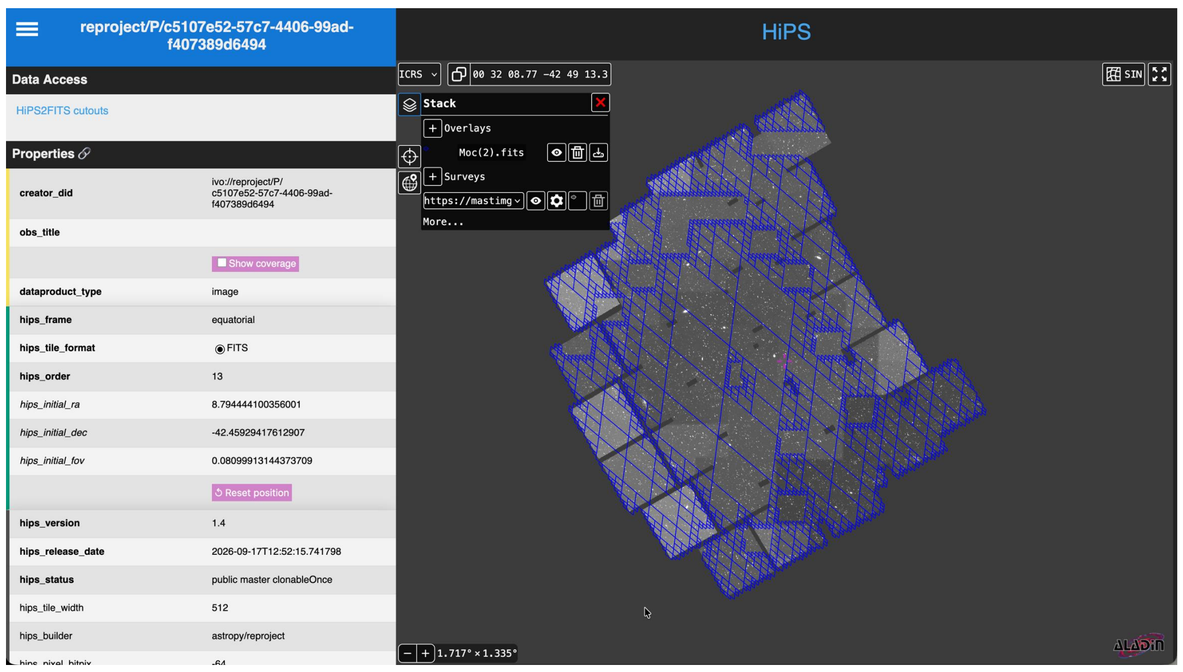
Browse the sky with HiPS backgrounds overlaid with catalogs and observation footprints. Search MAST at the displayed sky position. Load Parquet catalogs; link sky and table selections.

MastManager

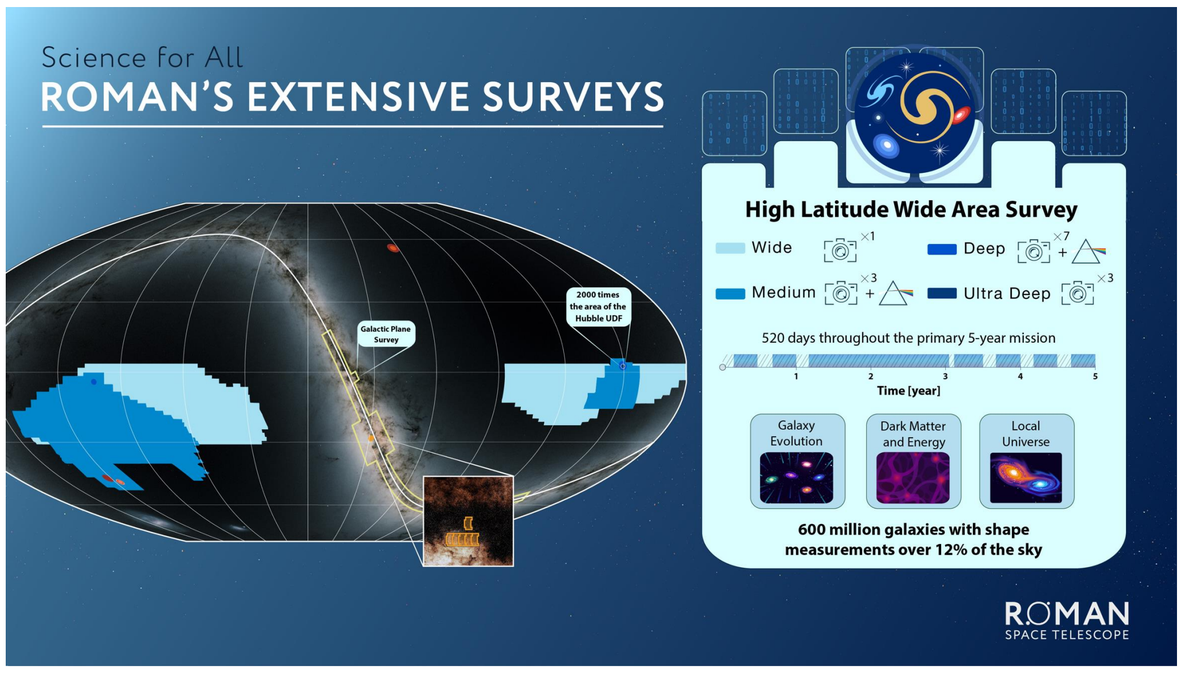
Automatically registers instances of mast-aladin and Jdaviz to enable synchronization of selected displays.

From Roman skycells to a zoomable sky

Hierarchical Progressive Surveys (HiPS) organize sky images into a multi-resolution HEALPix tile pyramid.
Aladin and other compatible viewers retrieve only the tiles needed for the current view.



HiPS generated from simulated Roman imaging.
HiPS data and screenshot generation: P. Custer (STScI).
Display tool: Aladin Lite



HLWAS Wide, Medium and Deep: Roughly a tenth of the sky.
Image Composition: J. Olmsted, J. Kang, and C. Nieves (STScI).
Background sky image: ESA/Gaia/DPAC. Acknowledgement: Javier Sanchez (STScI).

Submit L3 skycell references → **Amazon SQS** Queue / retries / DLQ → **AWS Lambda** Read skycell; run reproject ↔ **Amazon S3** Input + HiPS + overlaps

HEALPix-derived sky tiles; L3 ASDF skycells: 5,000 x 5,000 pixels (4,800-pixel core + 100-pixel border).
Astropy's reproject package is the core of our HiPS processing.
Memory-aware processing writes tiles to S3 and incrementally merges overlap tiles.
Same-region S3 / Lambda: no transfer charge. Default concurrency: Up to 1,000.

Preliminary in-house benchmarks on simulated data

~1 square degree / order 13 / 335 skycells

~3 min wall time | ~$1 run cost | ~21 GB output

Five-year sky coverage alone does not set production cost.
Overlapping surveys by waveband and epoch remain under study; compute, storage and serving costs will inform those choices.

STScI code: github.com/spacetelescope/ { mast-aladin | mast-table | motorcycle | jdaviz }
ipyaladin / Aladin Lite: CDS, Strasbourg | doi:10.26093/kpaw-kb74
Roman Research Nexus
Jdaviz | doi:10.5281/zenodo.5513927
Roman mission / NASA
reproject docs