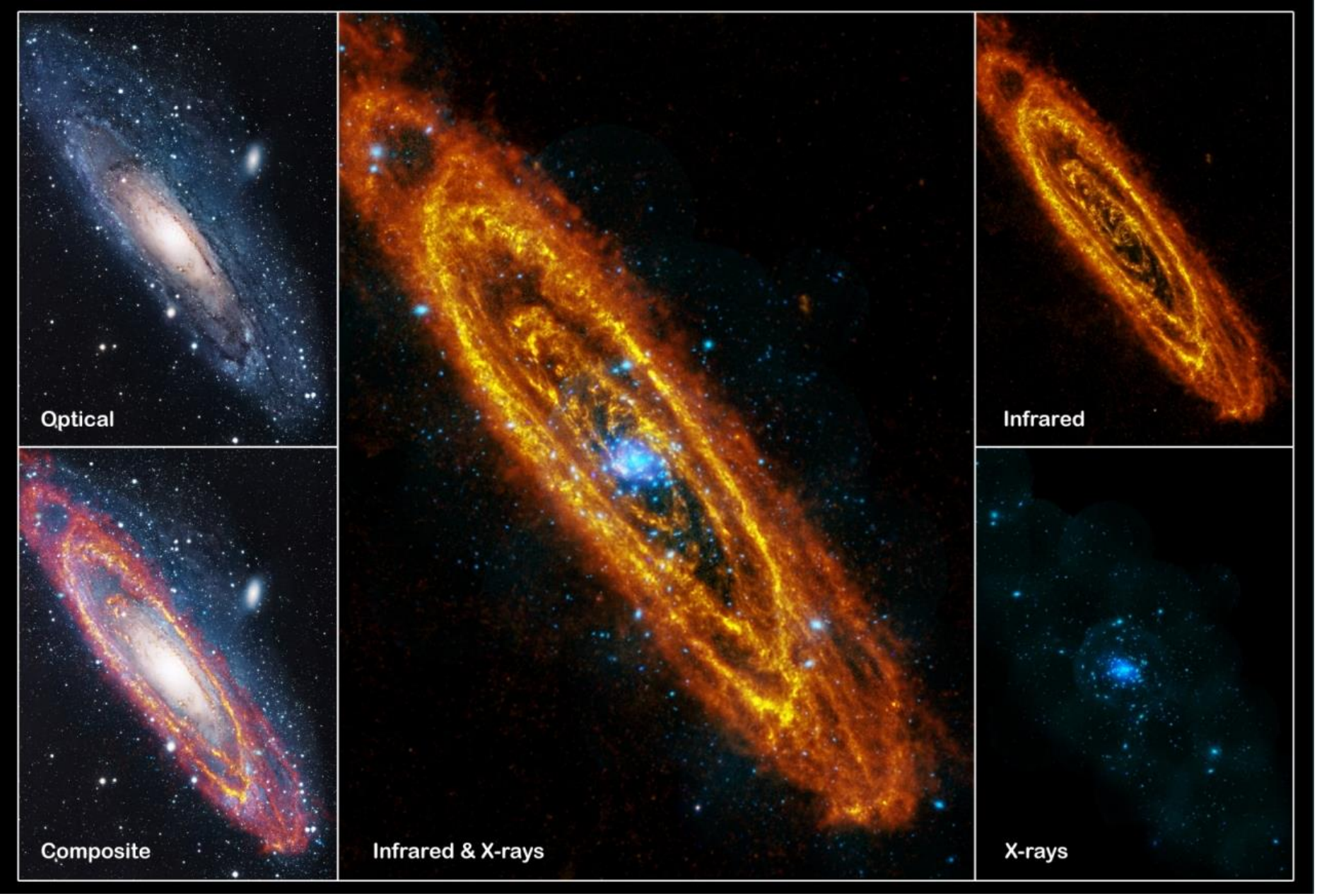
Galaxy data fusion

Pauline Barmby
Physics & Astronomy,
Institute for Earth & Space Exploration,
Rotman Institute of Philosophy

Western University,
London Canada

CFHT/Coelum/Cuillandre



ESA/Herschel/J. Fritz, U. Gent; ESA/XMM-Newton/W. Pietsch; R. Gendler

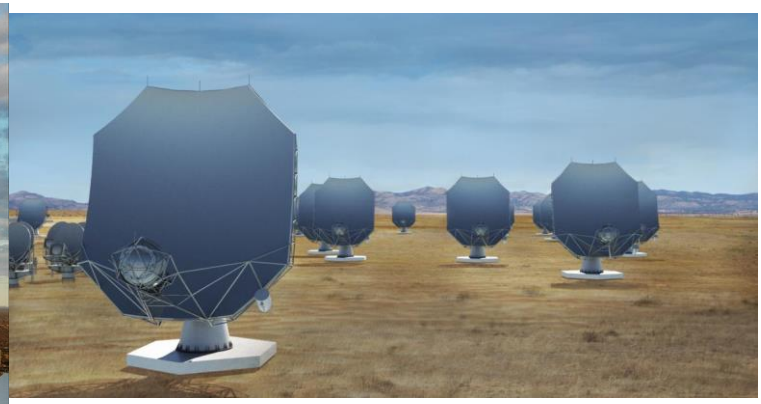

NASA/ESA/Holwerda

What do we want to know about galaxies?

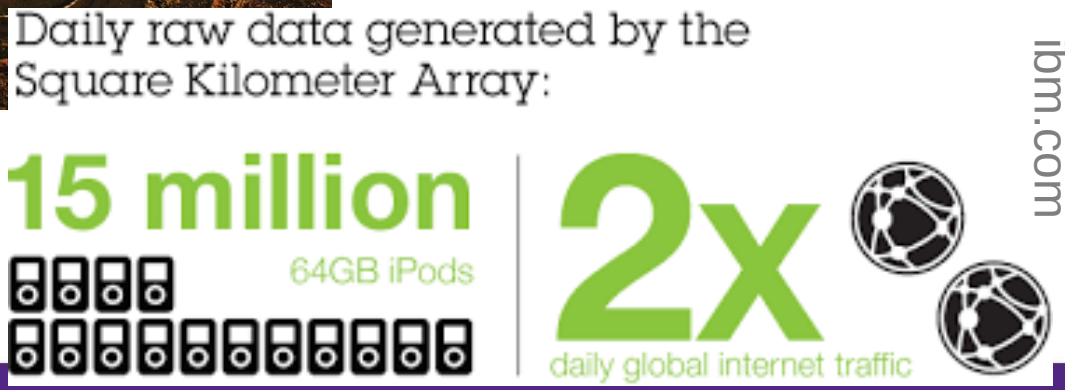
Well, everything.

- What accounts for the range of galaxy properties?
- How (when, where) do they make stars?
- How (when) do they grow supermassive black holes?
- What about their interstellar medium?
- How do the stars, ISM, and AGN interact?
- How does the local environment affect all of the above?

SKA Observatory (S. Africa, Australia, NZ)

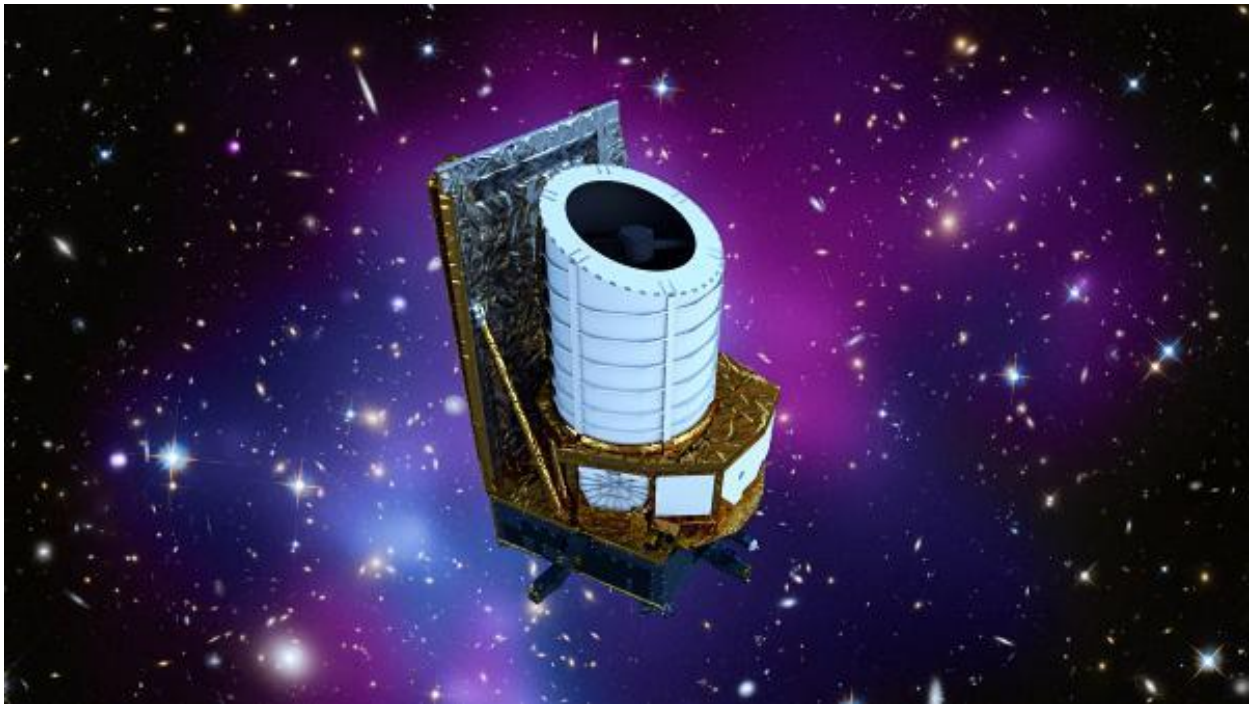


300 PB/year,
operations begin 2027



SKAO, NAOJ

Euclid mission (ESA)



ESA/ATG medialab; NASA, ESA, CXC, C. Ma, H. Ebeling and E. Barrett al., STScI

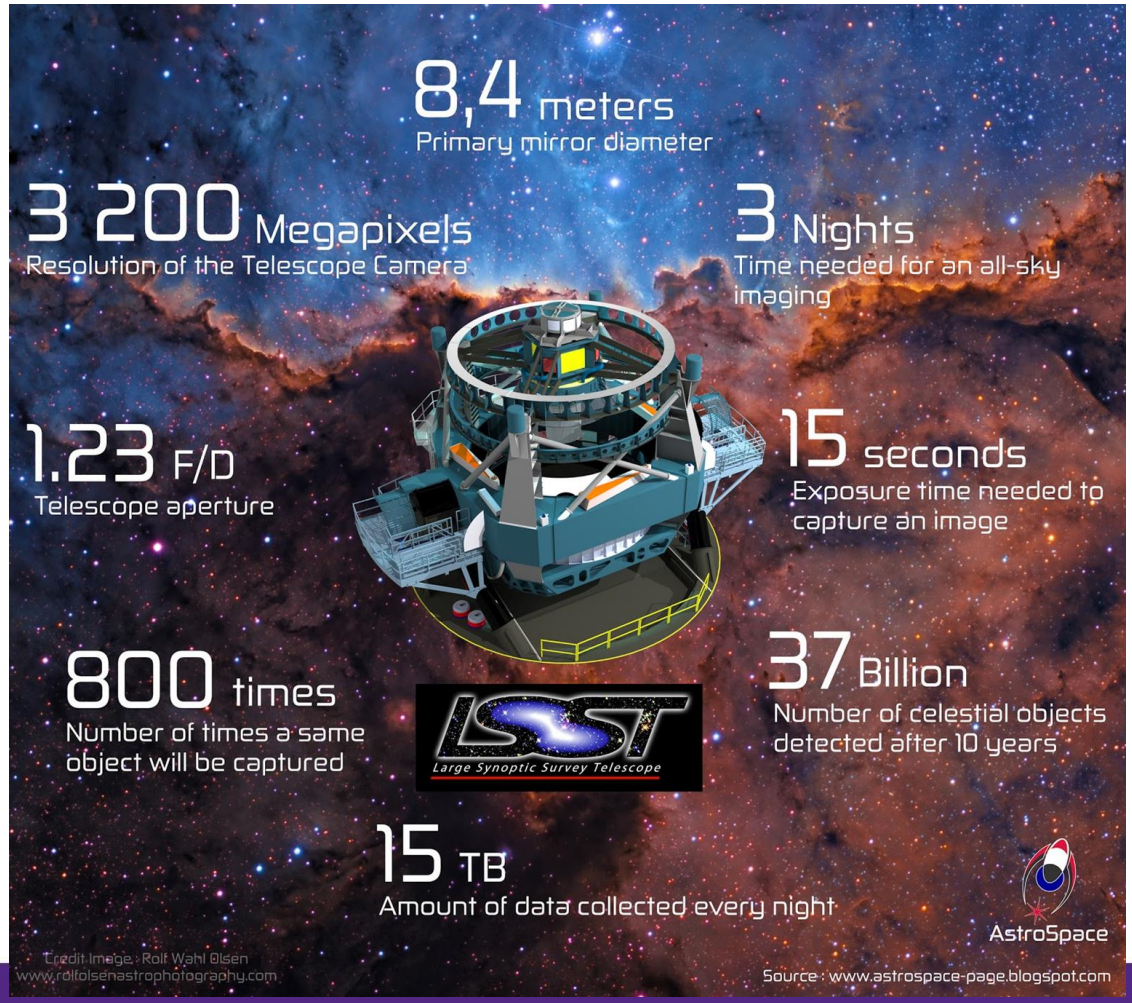
- 1.2-m diameter visible/near-infrared telescope designed for surveys; imaging and spectroscopy
- launched July 2023, first public data 2025

Rubin/Legacy Survey of Space & Time (Chile)

Under construction in Chile, operations to begin very soon

Will survey entire southern sky once/week over 10 years

Informatics and Statistics Science Collaboration!
https://issc.science.lsst.org/



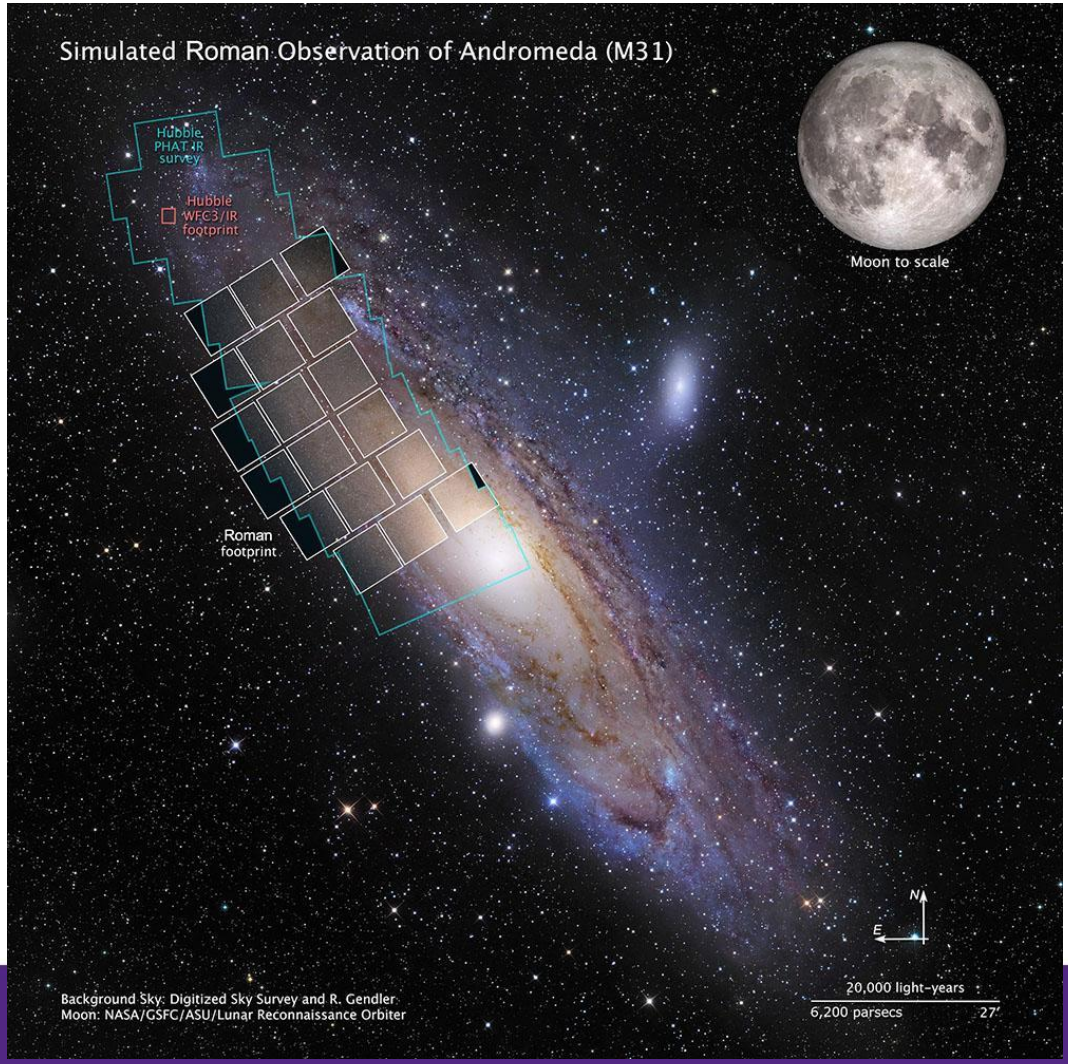
Astrospace-page.blogspot.com

Nancy Grace Roman Space Telescope (NASA)

Hubble-sized telescope, but with 100x larger field of view

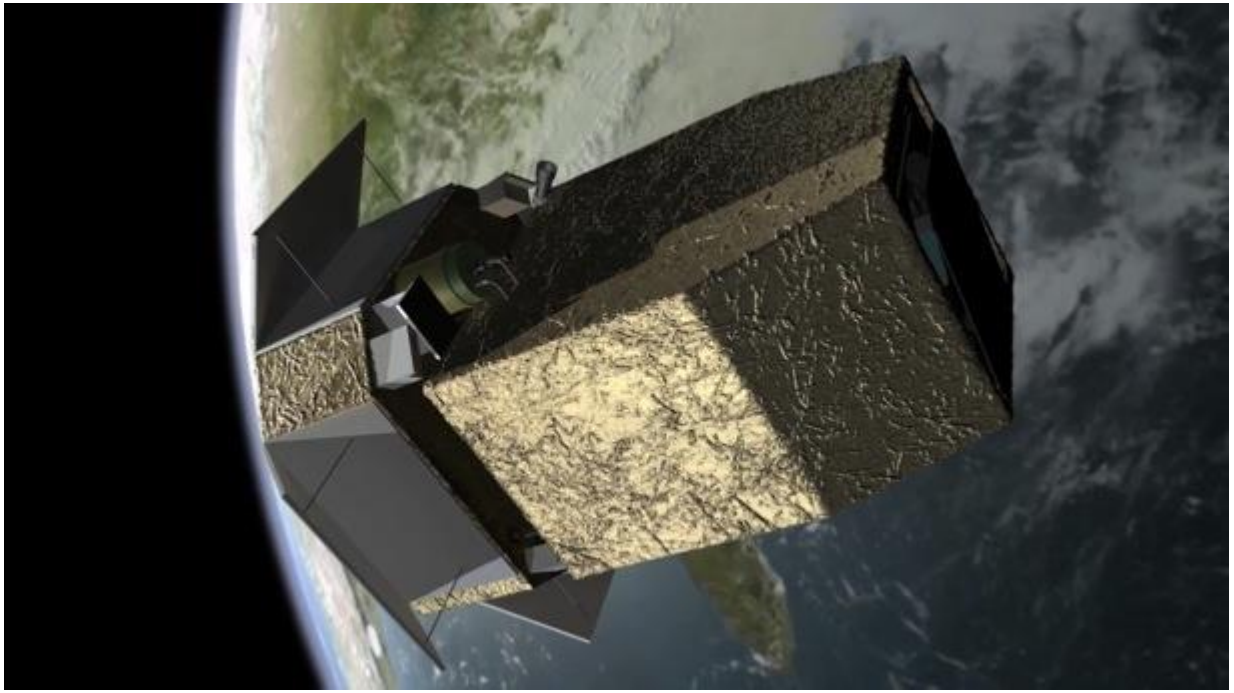
Visible & infrared imaging surveys; coronagraph

Launch 202?



NASA

CASTOR (CSA)



- 1-m diameter visible/ultraviolet telescope with large field of view; imaging and spectroscopy
- CSA-led with Japan, India, US: funding advocacy in progress

CSA & CASTOR team

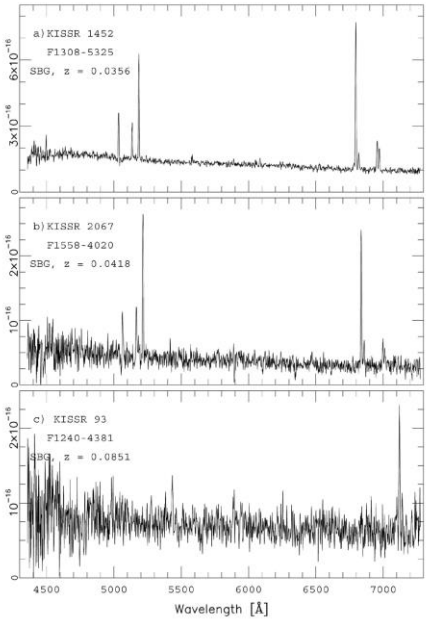
What are the challenges?

Well, everything.

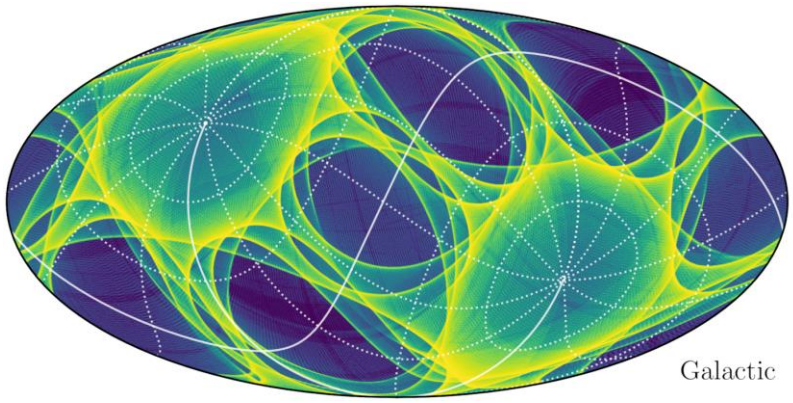
- Multiple physical scales
- Multiple observational wavelengths (& redshift!)
- Local versus distant galaxies
- Distances!
- Names and positions!
- Terminology of different sub-fields

Astro-data can differ from other big data

Noisy



Incomplete/truncated/ weird selection functions

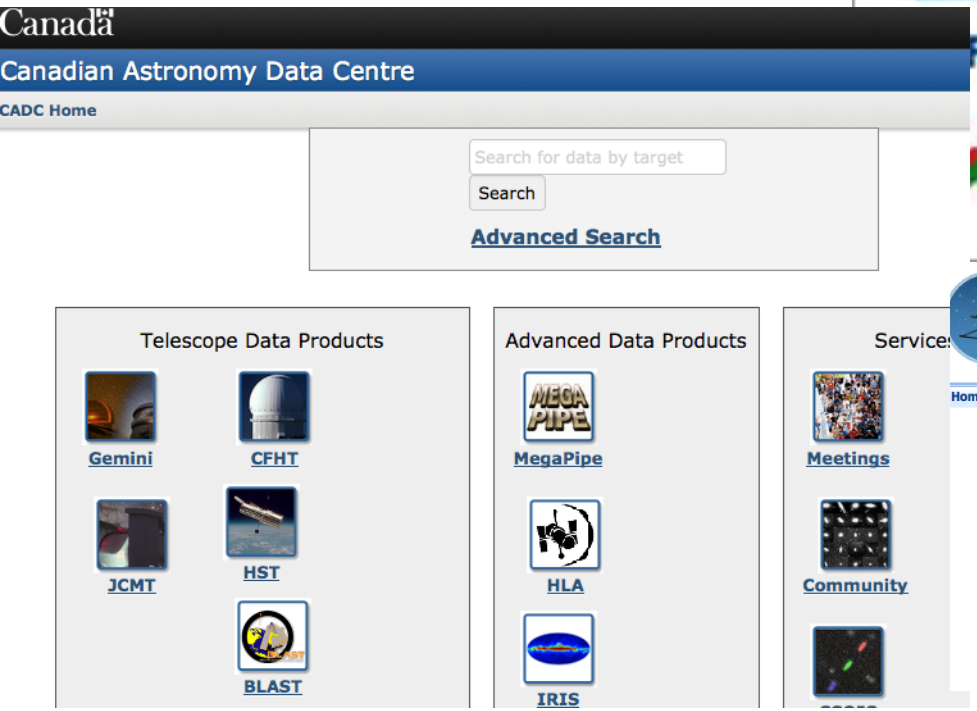


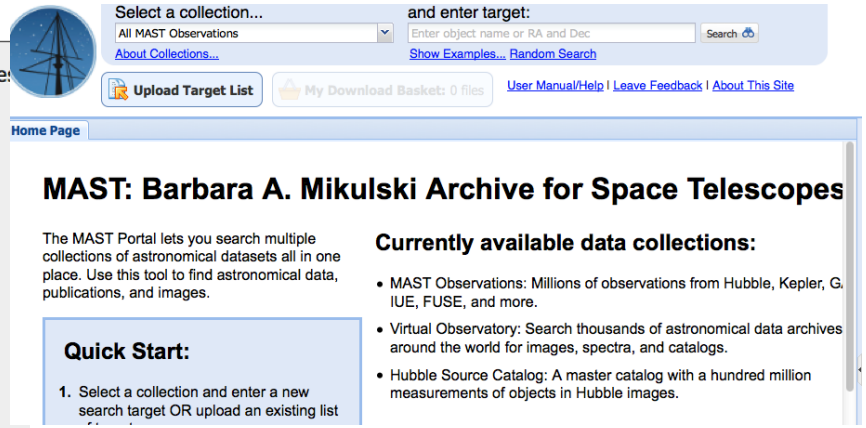
Boubert & Everall 2020

Doesn't always fit into a database model

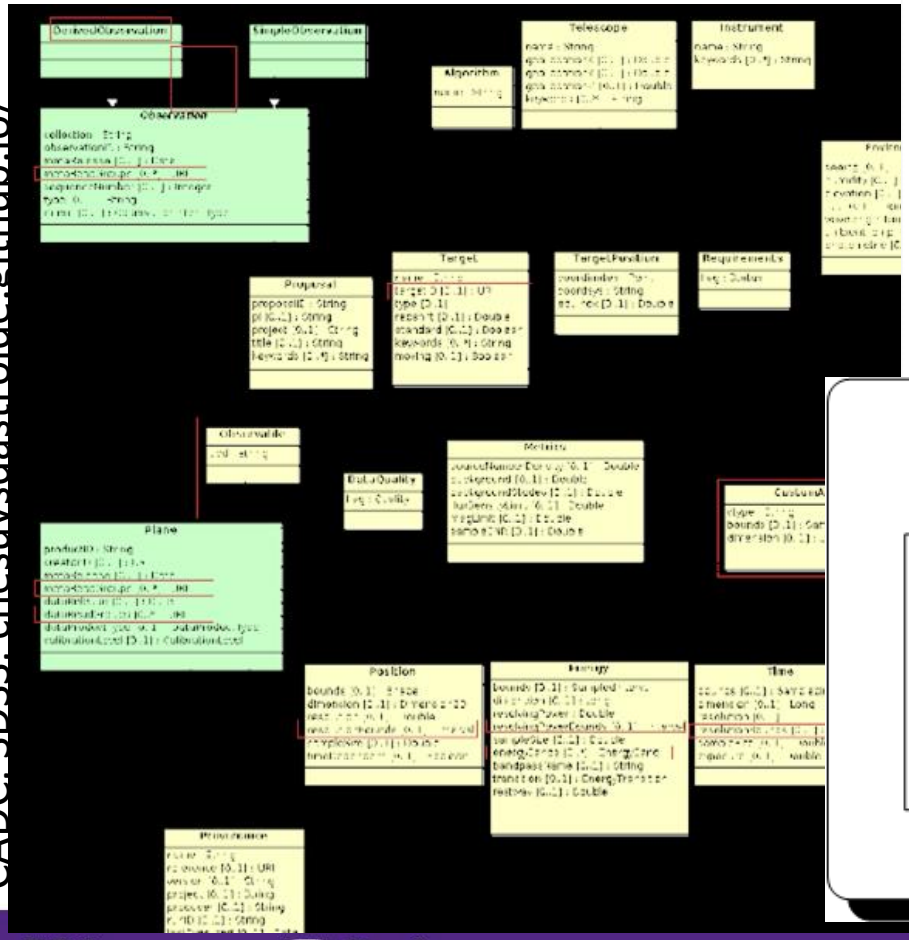
Performance level	Database DTUs	Storage included	Storage limit
Standard S0 – S2	10 – 50	250 GB	250 GB
Standard S3	100	250 GB	1 TB *new*
Standard S4 *new*	200	250 GB	1 TB *new*
Standard S6 *new*	400	250 GB	1 TB *new*
Standard S7 *new*	800	250 GB	1 TB *new*
Standard S9 *new*	1600	250 GB	1 TB *new*
Standard S12 *new*	3000	250 GB	1 TB *new*
Premium P1	125	500 GB	1 TB *new*
Premium P2	250	500 GB	1 TB *new*
Premium P4	500	500 GB	1 TB *new*
Premium P6	1000	500 GB	1 TB *new*
Premium P11 – P15	1750 – 4000	4 TB	4 TB
Premium RS PRS1	125	500 GB	1 TB *new*
Premium RS PRS2	250	500 GB	1 TB *new*
Premium RS PRS4	500	500 GB	1 TB *new*
Premium RS PRS6	1000	500 GB	1 TB *new*

Astro-data is (very) widely distributed

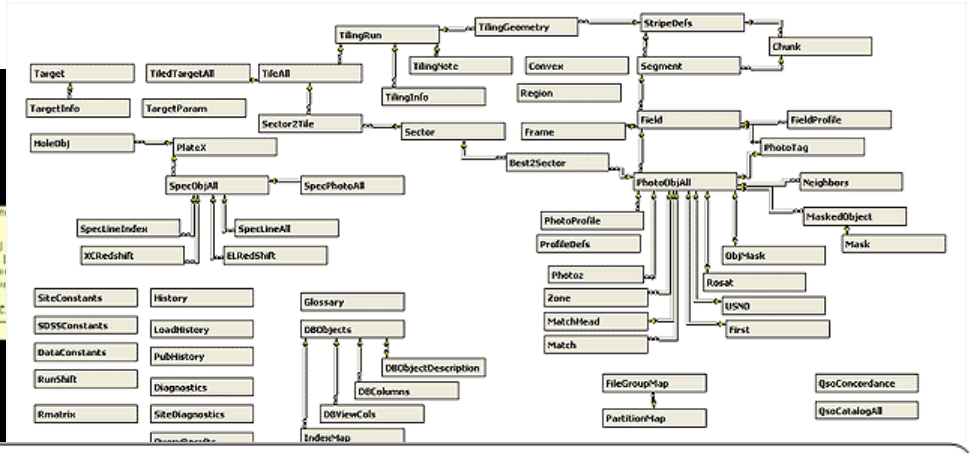




Data(base) practices are still evolving



CADC, SDSS, cnesuvsgastroide.github.io/



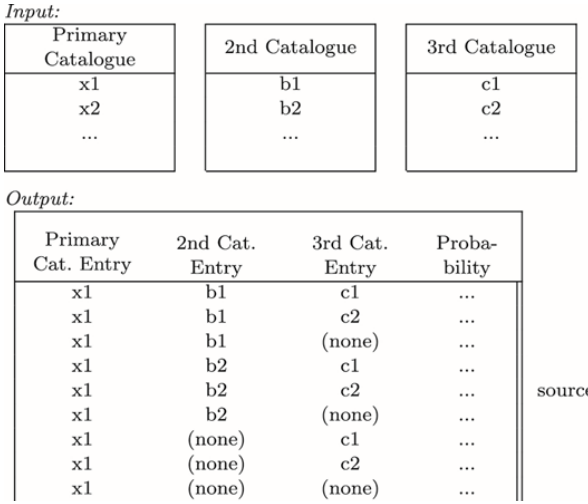
Astronomical Data Query Language

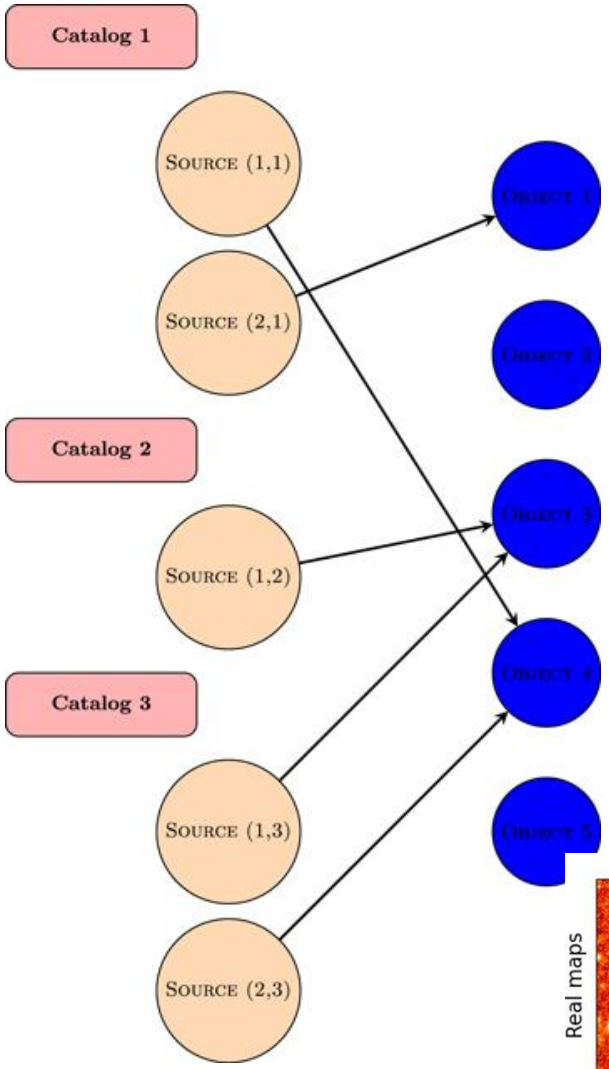
```
SELECT * FROM gaia
JOIN tycho2
ON 1=CONTAINS (POINT('ICRS',gaia.ra, gaia.dec),
CIRCLE('ICRS', tycho2.ra, tycho2.dec, 2/3600))
```

IVOA

Possible solutions 1

Salvato et al 2018

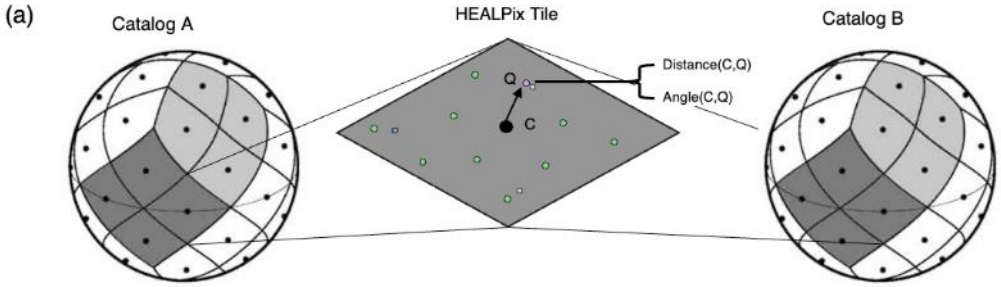




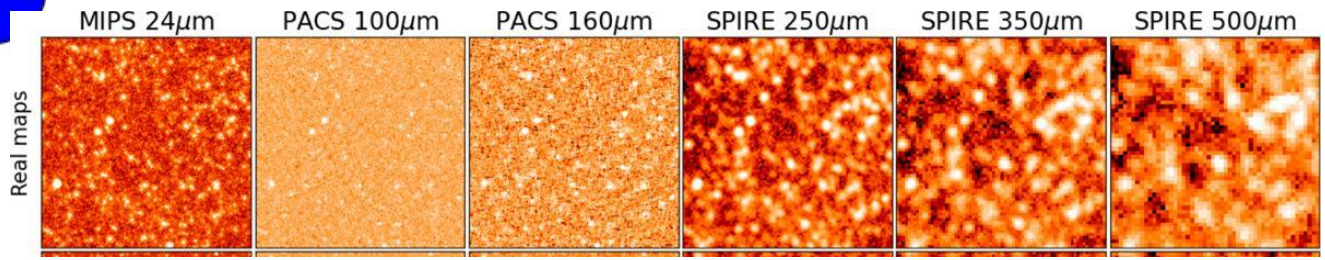
Nguyen, Basu & Budavari 2022

Sophisticated and statistically sound cross-match & deblending algorithms

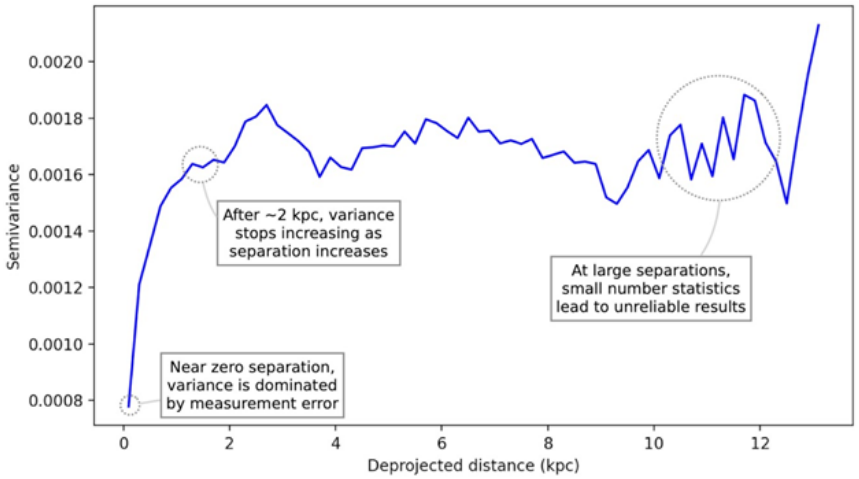
Lam et al 2025



Wang et al 2024

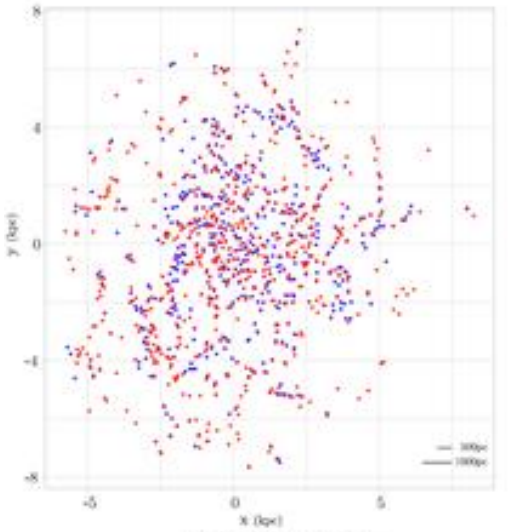


Possible solutions 2 spatial statistics approaches



Metha et al 2021

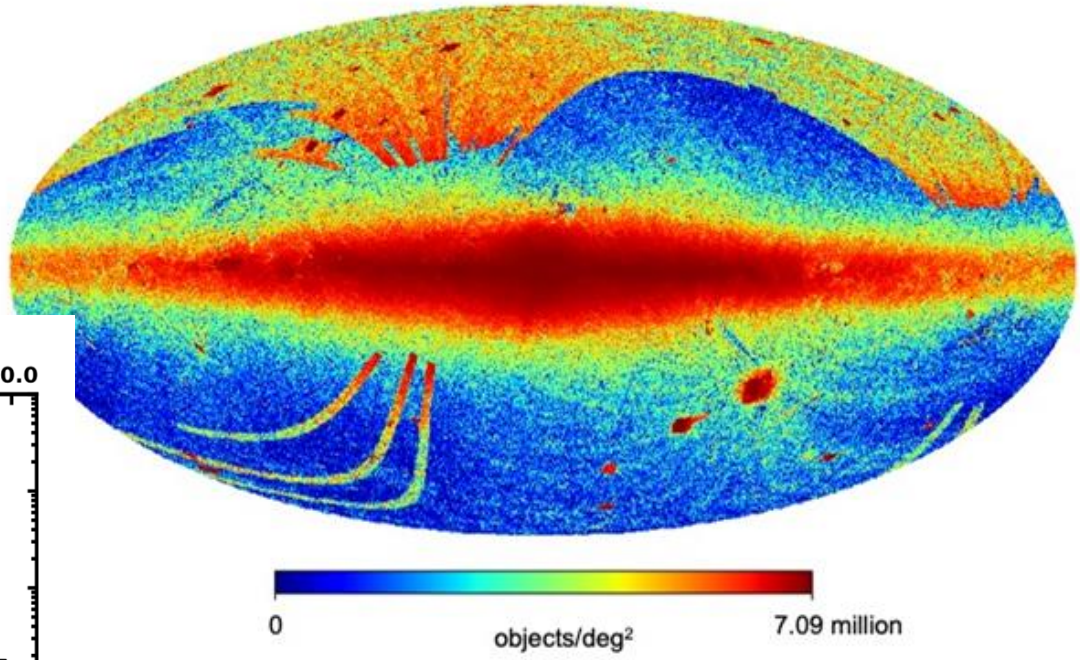
(a) CO filament and GMCs

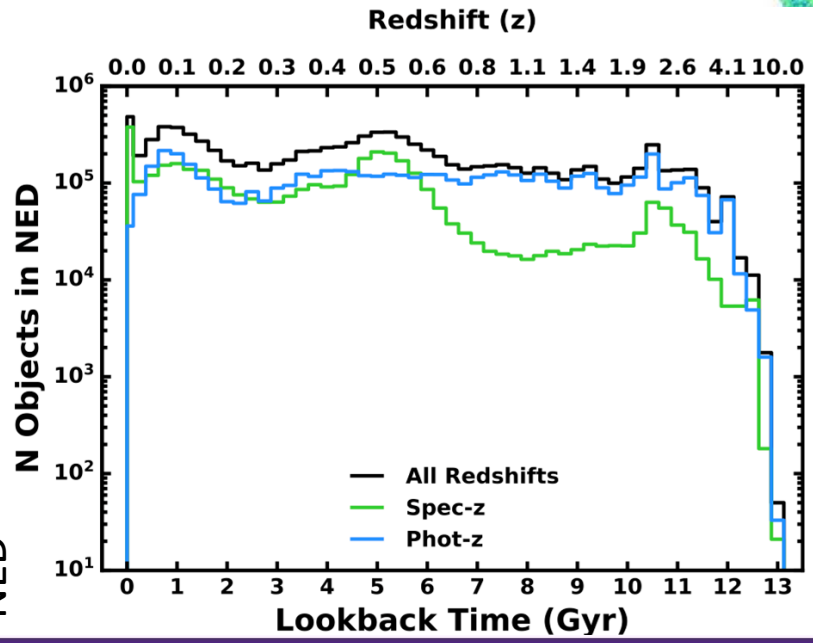


(b) GMCs and YSCCs

Li & Barmby 2021

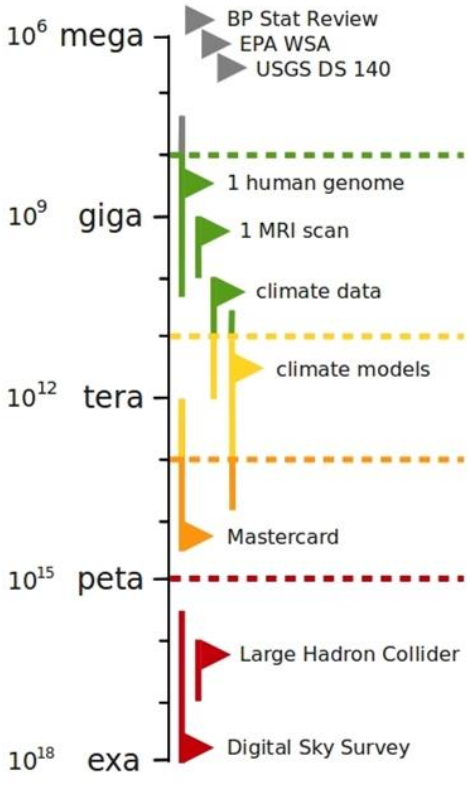
Curated databases: SIMBAD, NED, HyperLEDA



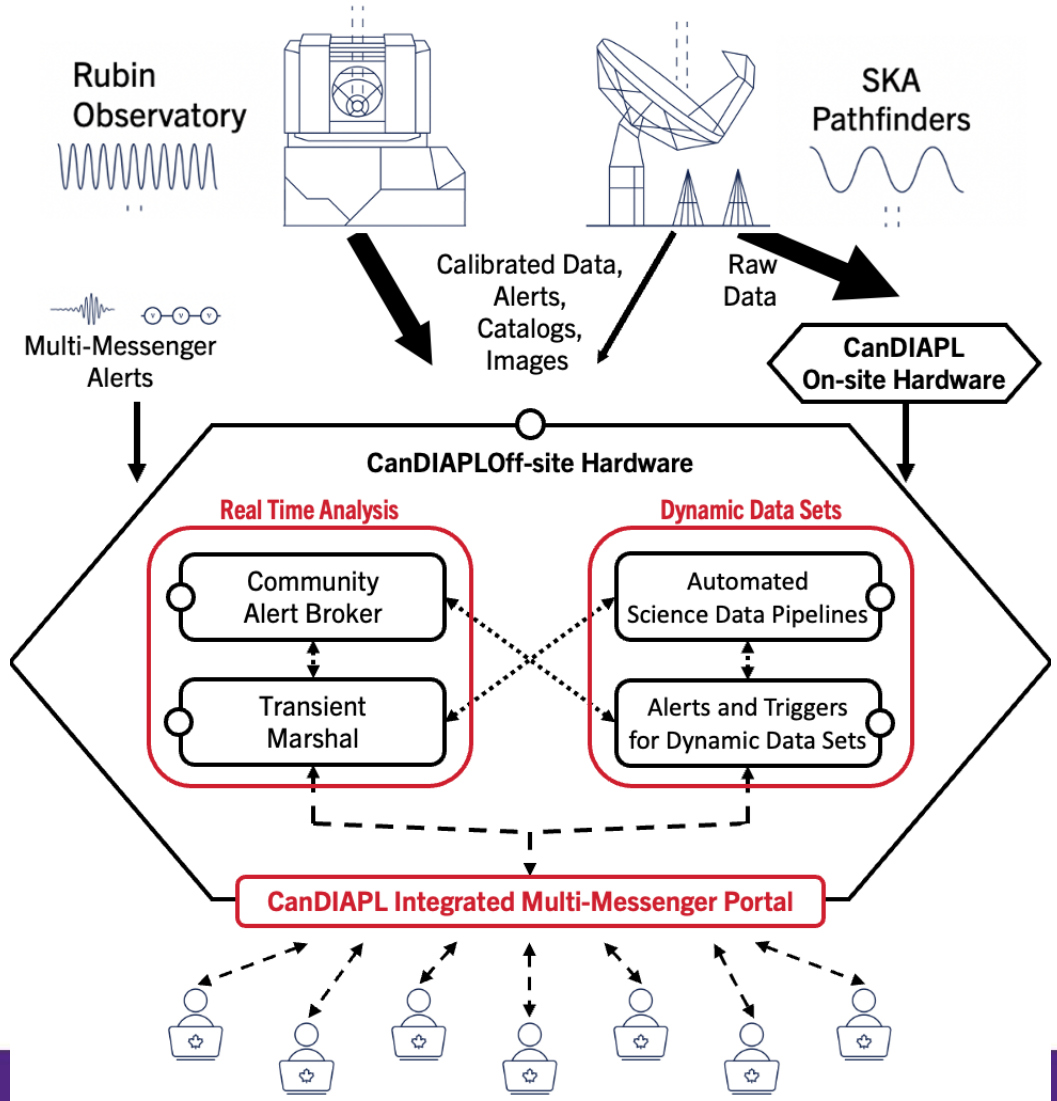


NED

Putting it all together: CanDIAPL



Mazamascience.com



More info

- “Astronomical data fusion: recent progress and future prospects — a survey,” Yu et al 2019 ExA, 47, 359

- "A ‘Rosetta Stone’ for Studies of Spatial Variation in Astrophysical Data: Power Spectra, Semivariograms, and Structure Functions”, Metha & Berger 2024

- Astronomical observations: a guide for allied researchers Barmby 2019