

# Galaxy data fusion

Pauline Barmby

Physics & Astronomy,

Institute for Earth & Space Exploration,  
Rotman Institute of Philosophy

Western University,  
London Canada



CFHT/Coelum/Cuillandre



Optical



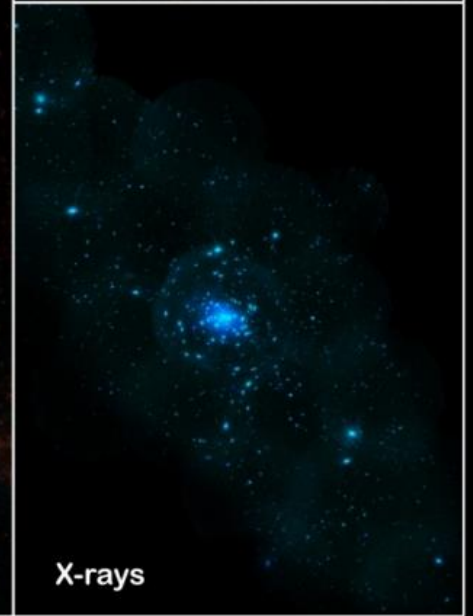
Composite



Infrared & X-rays



Infrared



X-rays

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



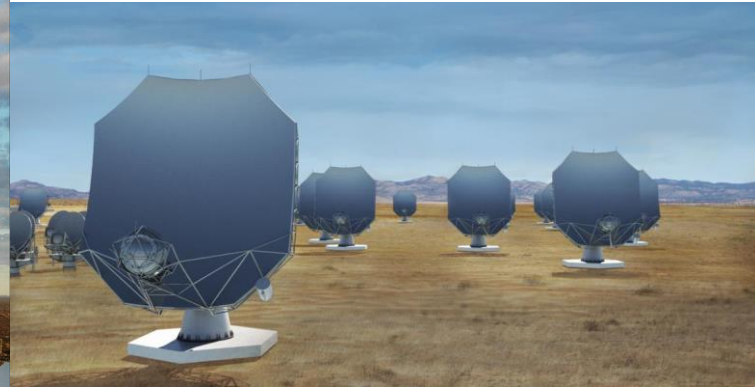


# What do we want to know about galaxies?

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

Daily raw data generated by the  
Square Kilometer Array:

**15 million**



64GB iPods



**2x**

daily global internet traffic

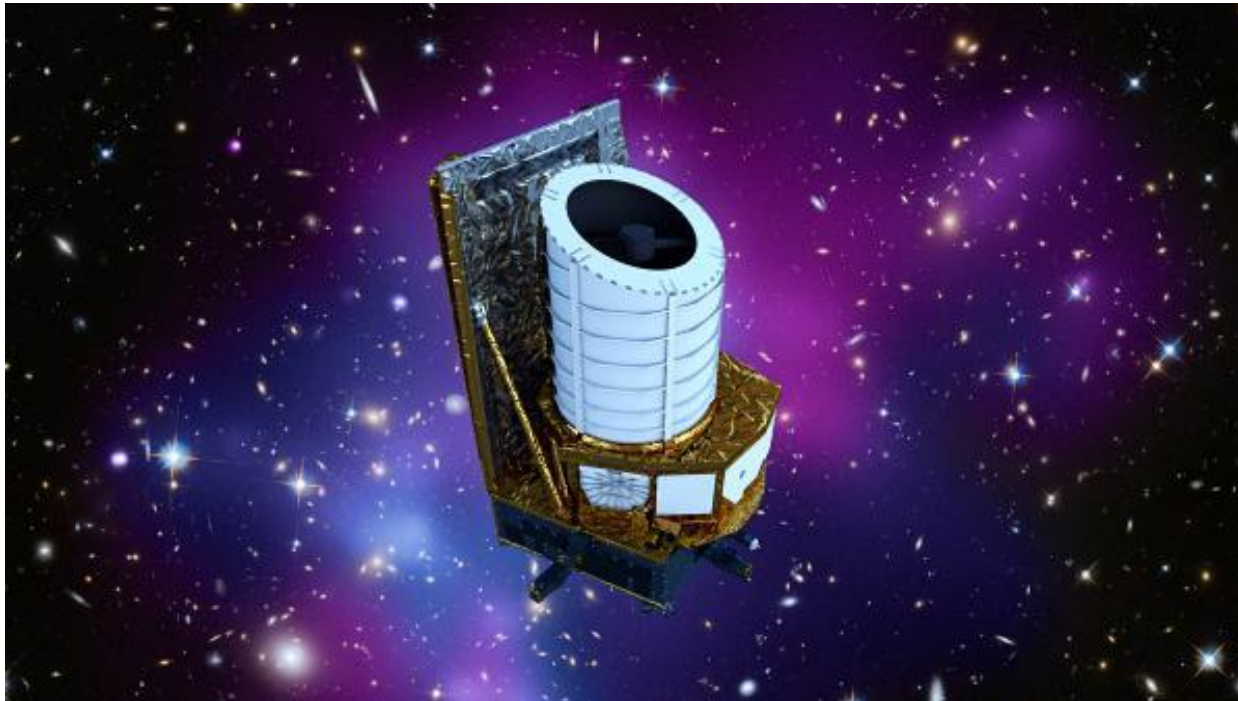


SKAO, NAOJ

Western Science



# *Euclid* mission (ESA)



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

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

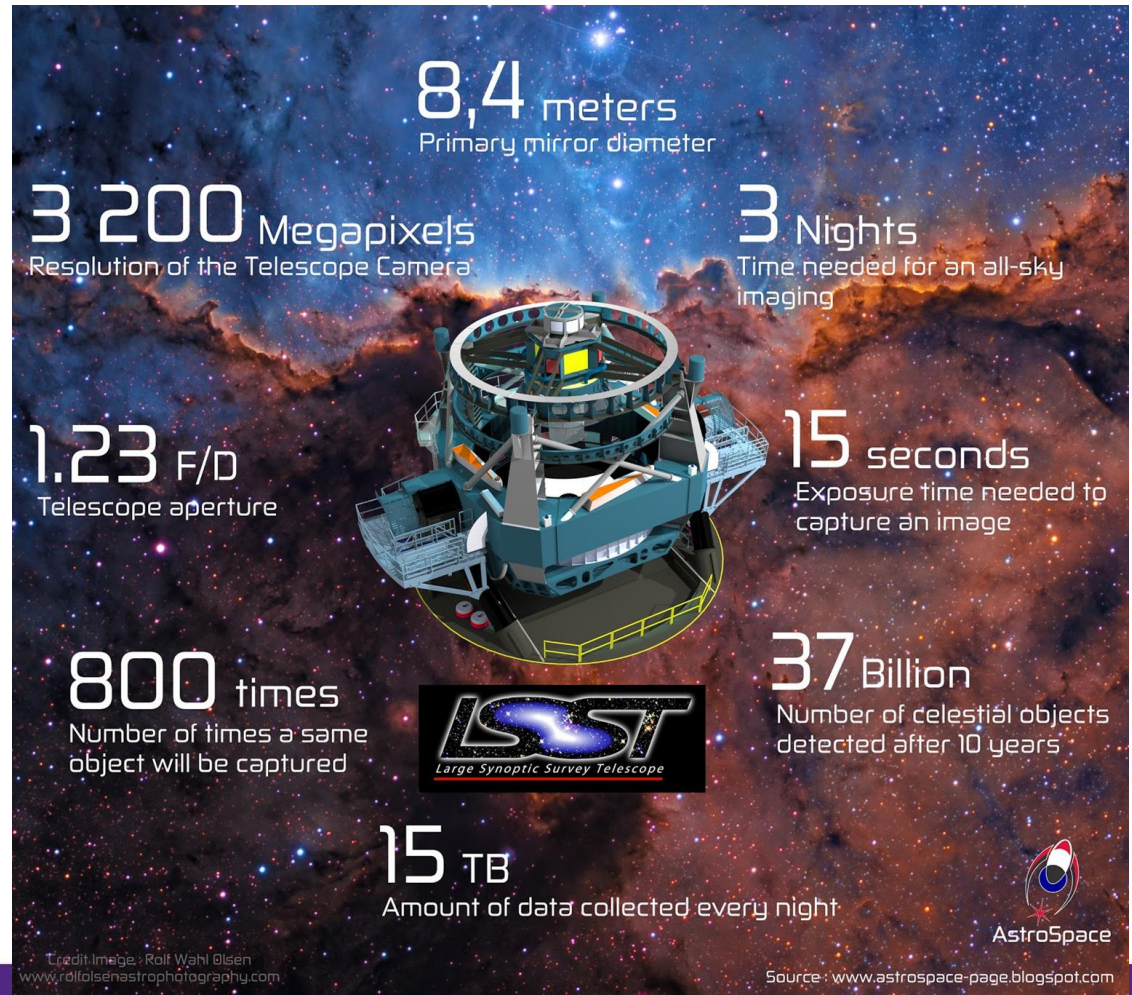
# Rubin/Legacy Survey of Space & Time (Chile)

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

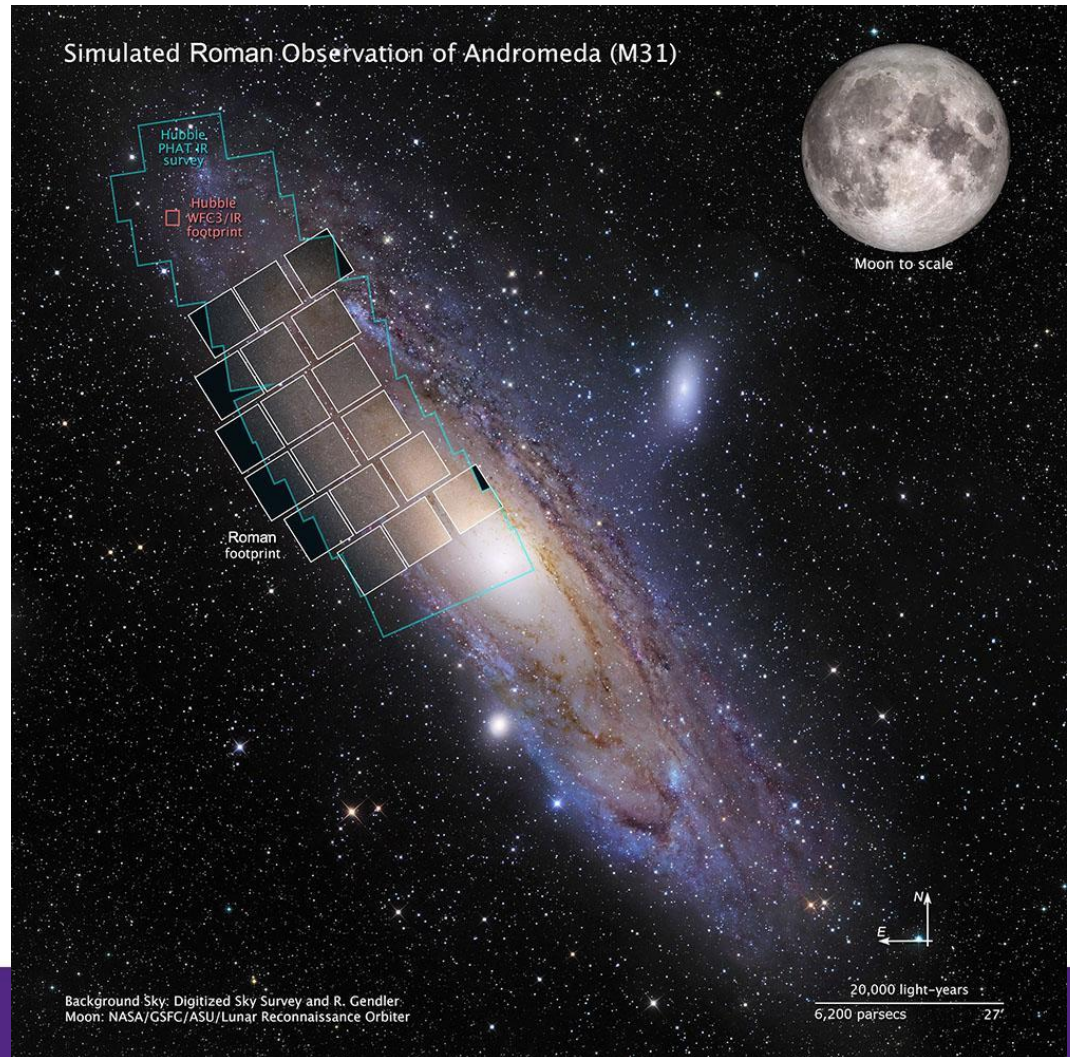


# Nancy Grace Roman Space Telescope (NASA)

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

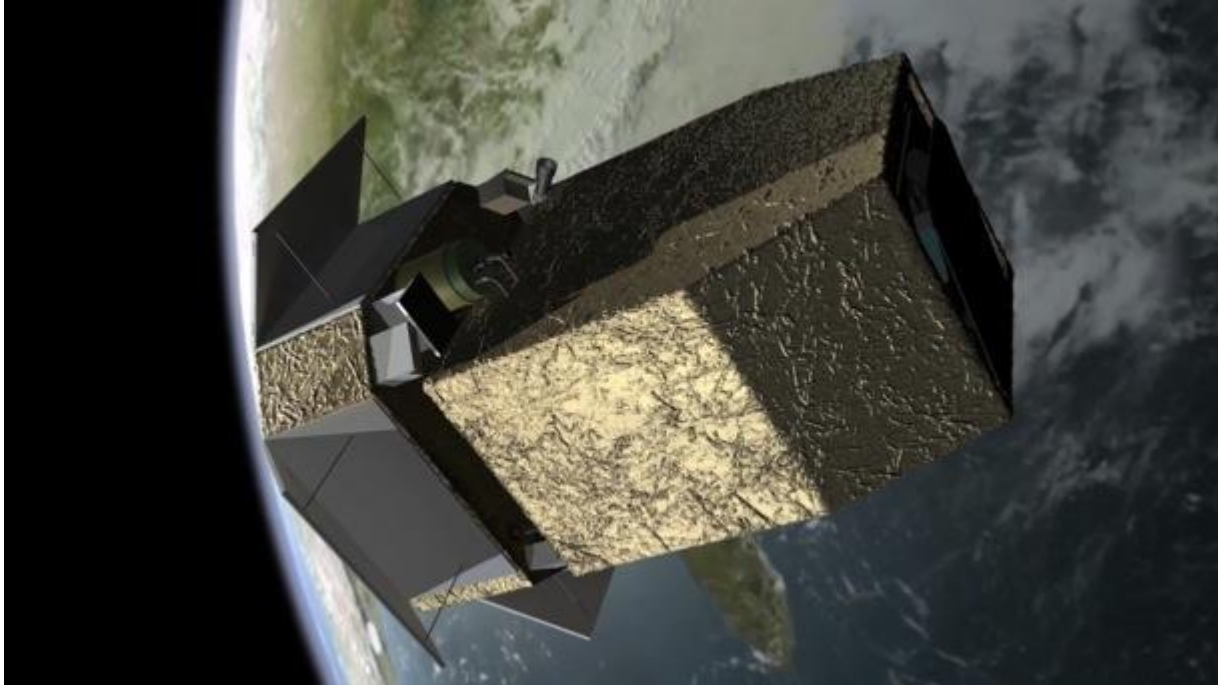
Visible & infrared imaging surveys; coronagraph

Launch 202?





# 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

# What are the challenges?

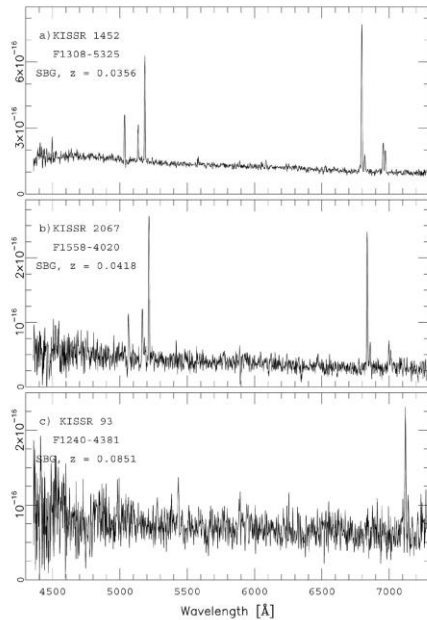
Well, everything.

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

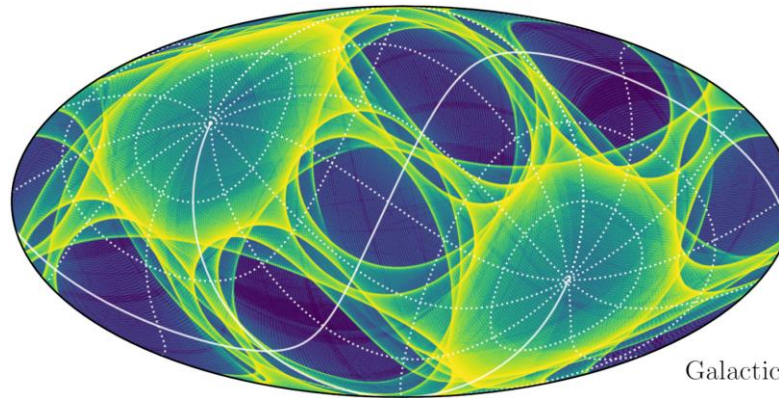


# Astro-data can differ from other big data

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 <i>new</i>
Standard S4 <i>new</i>	200	250 GB	1 TB <i>new</i>
Standard S6 <i>new</i>	400	250 GB	1 TB <i>new</i>
Standard S7 <i>new</i>	800	250 GB	1 TB <i>new</i>
Standard S9 <i>new</i>	1600	250 GB	1 TB <i>new</i>
Standard S12 <i>new</i>	3000	250 GB	1 TB <i>new</i>
Premium P1	125	500 GB	1 TB <i>new</i>
Premium P2	250	500 GB	1 TB <i>new</i>
Premium P4	500	500 GB	1 TB <i>new</i>
Premium P6	1000	500 GB	1 TB <i>new</i>
Premium P11 – P15	1750 – 4000	4 TB	4 TB
Premium RS PRS1	125	500 GB	1 TB <i>new</i>
Premium RS PRS2	250	500 GB	1 TB <i>new</i>
Premium RS PRS4	500	500 GB	1 TB <i>new</i>
Premium RS PRS6	1000	500 GB	1 TB <i>new</i>

# Astro-data is (very) widely distributed

Canada

Canadian Astronomy Data Centre

CADC Home

Search for data by target

Search

[Advanced Search](#)

## Telescope Data Products



[Gemini](#)



[CFHT](#)



[JCMT](#)



[HST](#)



[BLAST](#)

## Advanced Data Products



[MegaPipe](#)



[HLA](#)



[IRIS](#)

## Services



[Meetings](#)



[Community](#)



Select a collection...

All MAST Observations

[About Collections...](#)

[Upload Target List](#)

and enter target:

Enter object name or RA and Dec

[Show Examples...](#) [Random Search](#)

[User Manual/Help](#) [Leave Feedback](#) [About This Site](#)

[Home Page](#)

## MAST: Barbara A. Mikulski Archive for Space Telescopes

The MAST Portal lets you search multiple collections of astronomical datasets all in one place. Use this tool to find astronomical data, publications, and images.

### Quick Start:

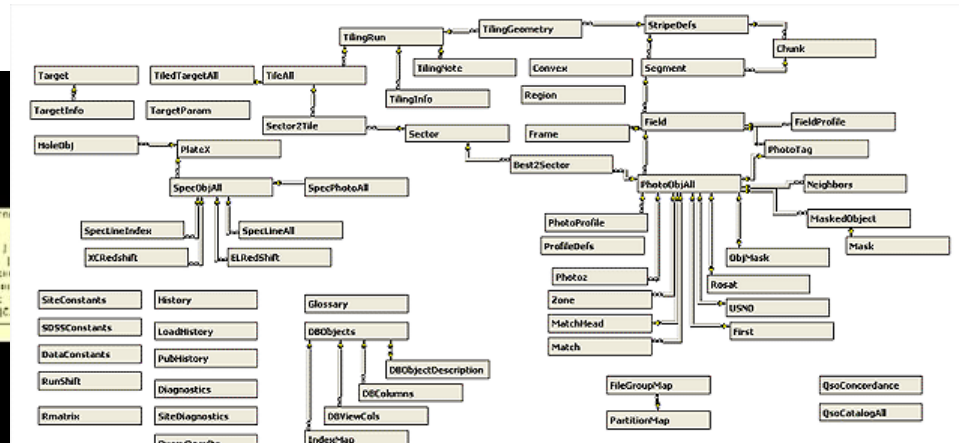
1. Select a collection and enter a new search target OR upload an existing list

### Currently available data collections:

- MAST Observations: Millions of observations from Hubble, Kepler, G. IUE, FUSE, and more.
- Virtual Observatory: Search thousands of astronomical data archives around the world for images, spectra, and catalogs.
- Hubble Source Catalog: A master catalog with a hundred million measurements of objects in Hubble images.



CADC SDSS. [cnesuvsqaastroide.github.io/](https://cnesuvsqaastroide.github.io/)

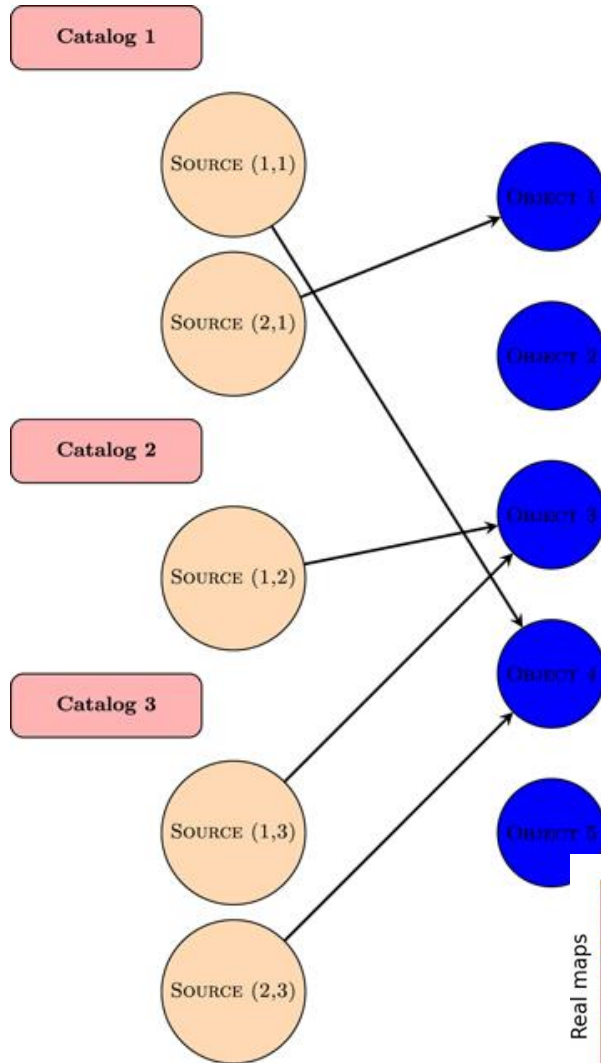


```
SELECT * FROM gaia
JOIN tycho2
ON 1=CONTAINS (POINT('ICRS',gaia.ra, gaia.dec),
CIRCLE('ICRS', tycho2.ra, tycho2.dec, 2/3600))
```

# Western Science

# Possible solutions 1

Salvato et al 2018



Sophisticated and statistically sound cross-match & deblending algorithms

Lam et al 2025

Input:

Primary Catalogue
x1
x2
...

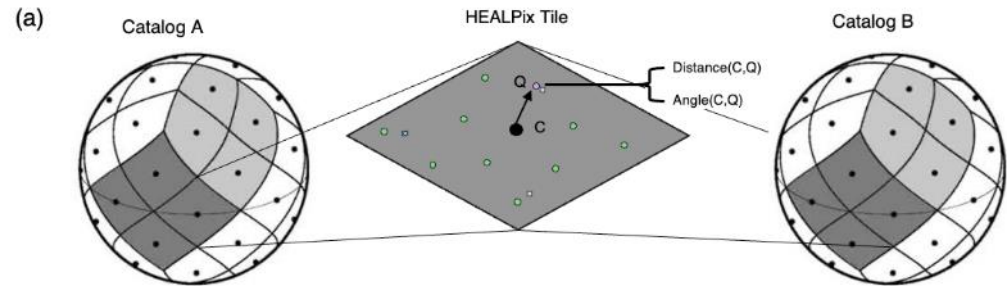
2nd Catalogue
b1
b2
...

3rd Catalogue
c1
c2
...

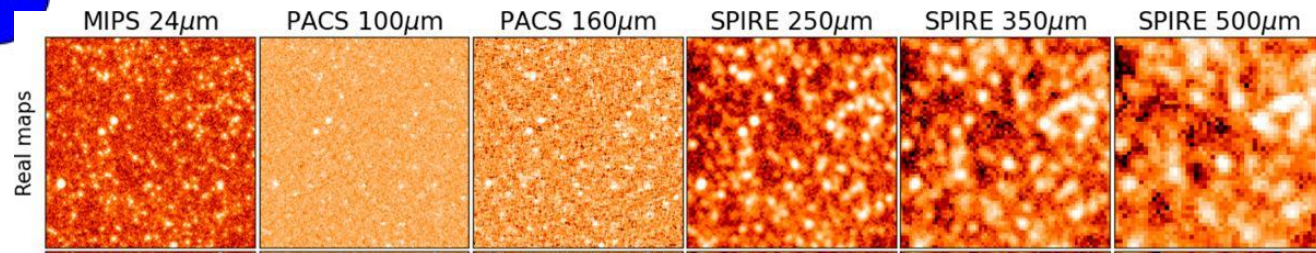
Output:

Primary Cat. Entry	2nd Cat. Entry	3rd Cat. Entry	Probability
x1	b1	c1	...
x1	b1	c2	...
x1	b1	(none)	...
x1	b2	c1	...
x1	b2	c2	...
x1	b2	(none)	...
x1	(none)	c1	...
x1	(none)	c2	...
x1	(none)	(none)	...

source x1



Wang et al 2024

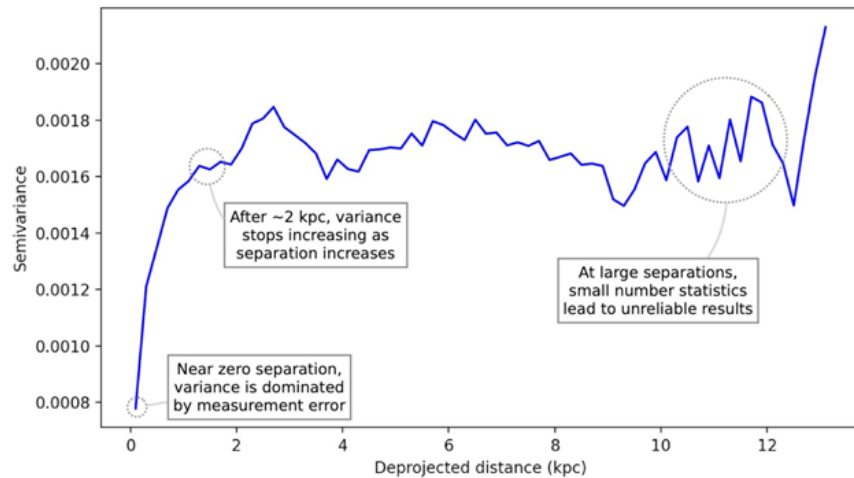


Nguyen, Basu & Budavari 2022

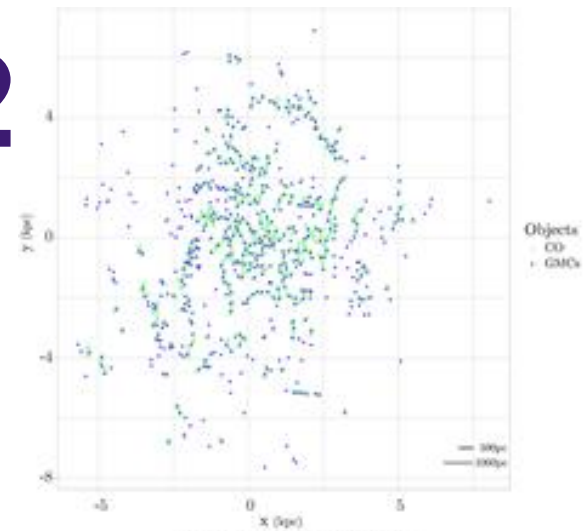


# Possible solutions 2

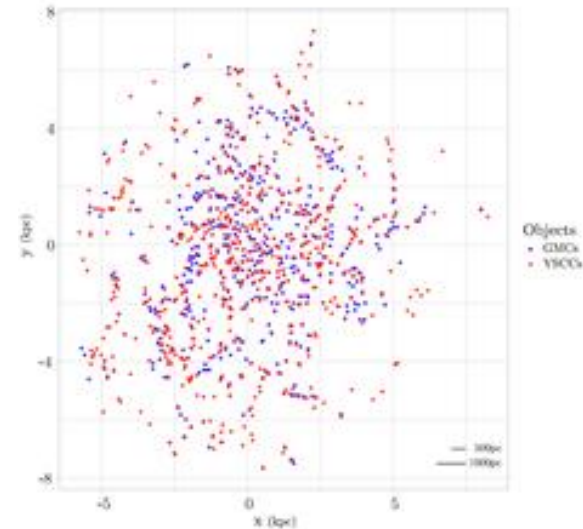
## spatial statistics approaches



Metha et al 2021



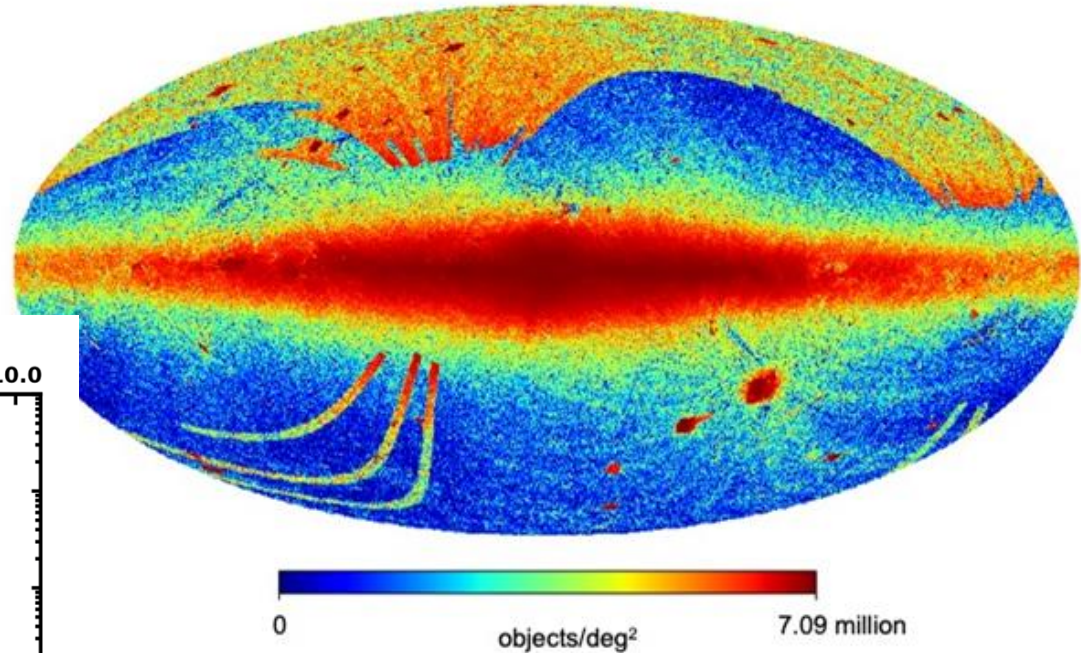
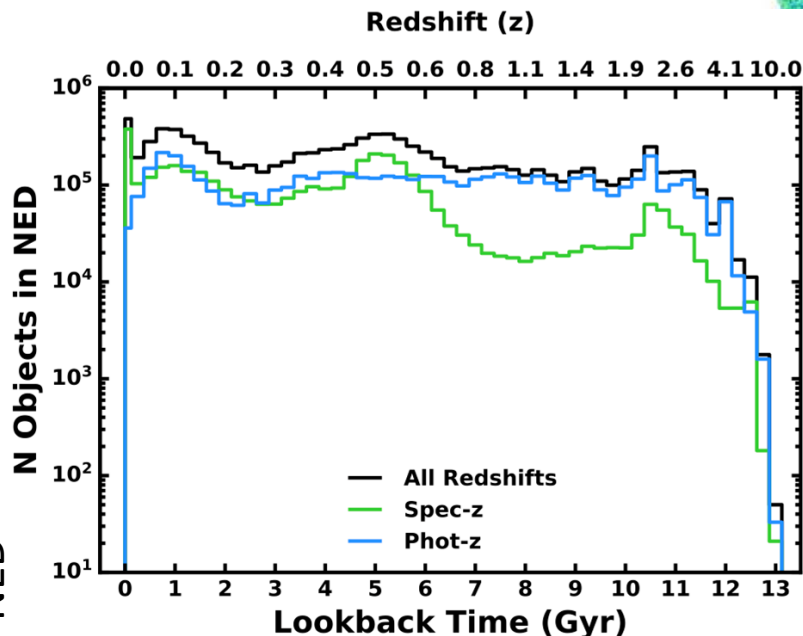
(a) CO filament and GMCs



(b) GMCs and YSCCs

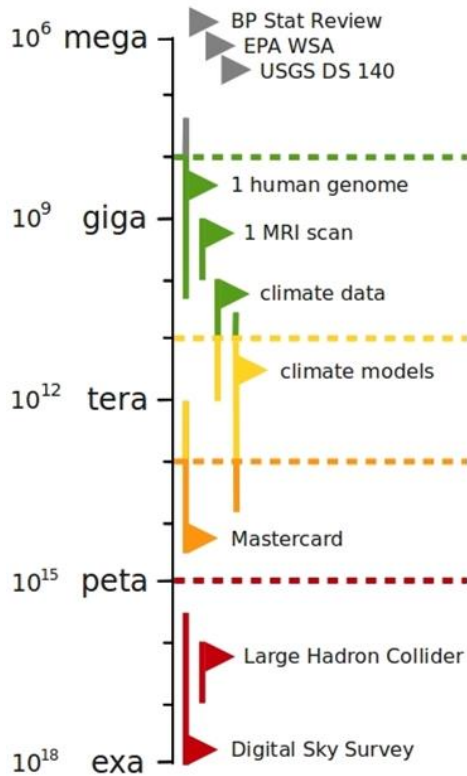
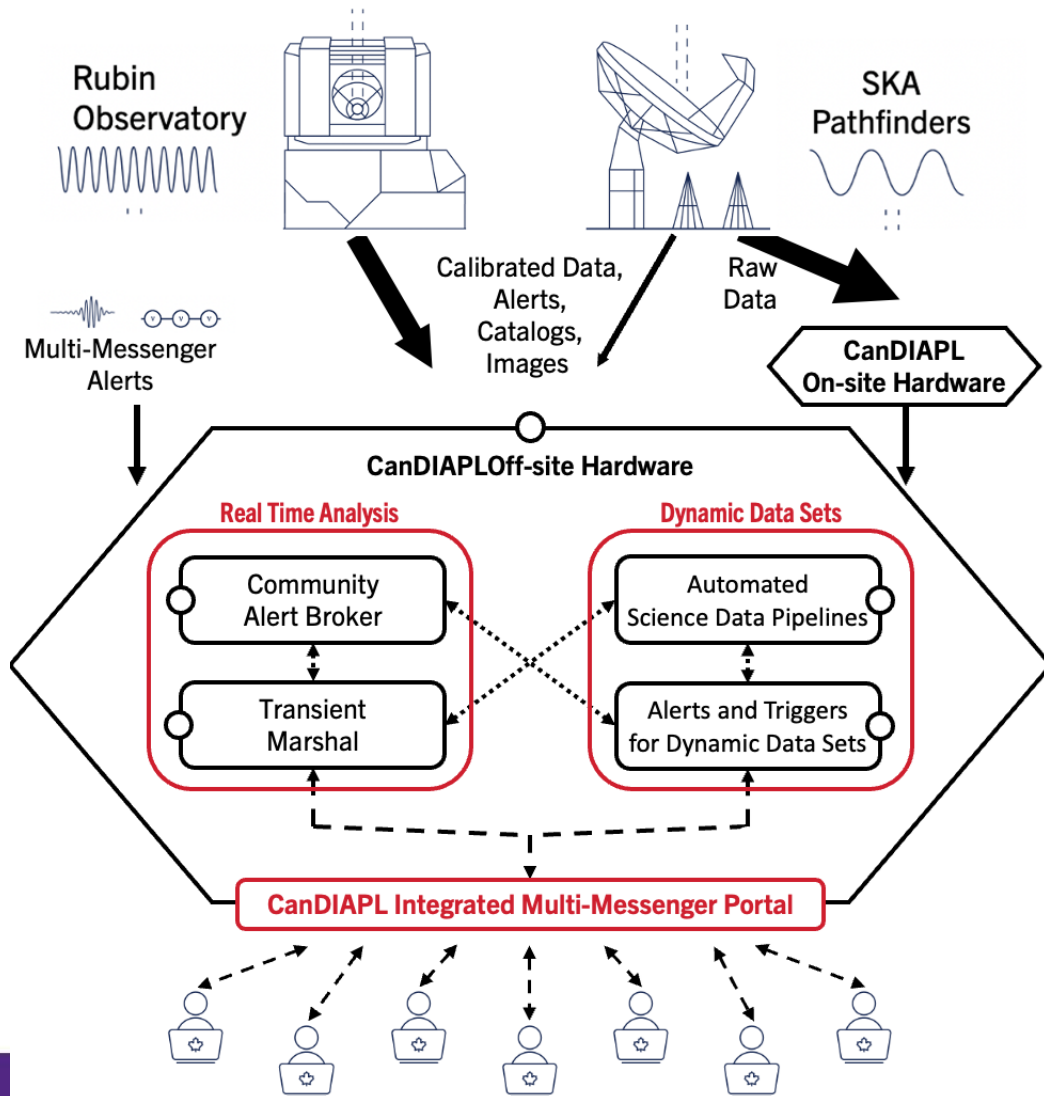
Li & Barmby 2021

# Curated databases: SIMBAD, NED, HyperLEDA





# 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](#)