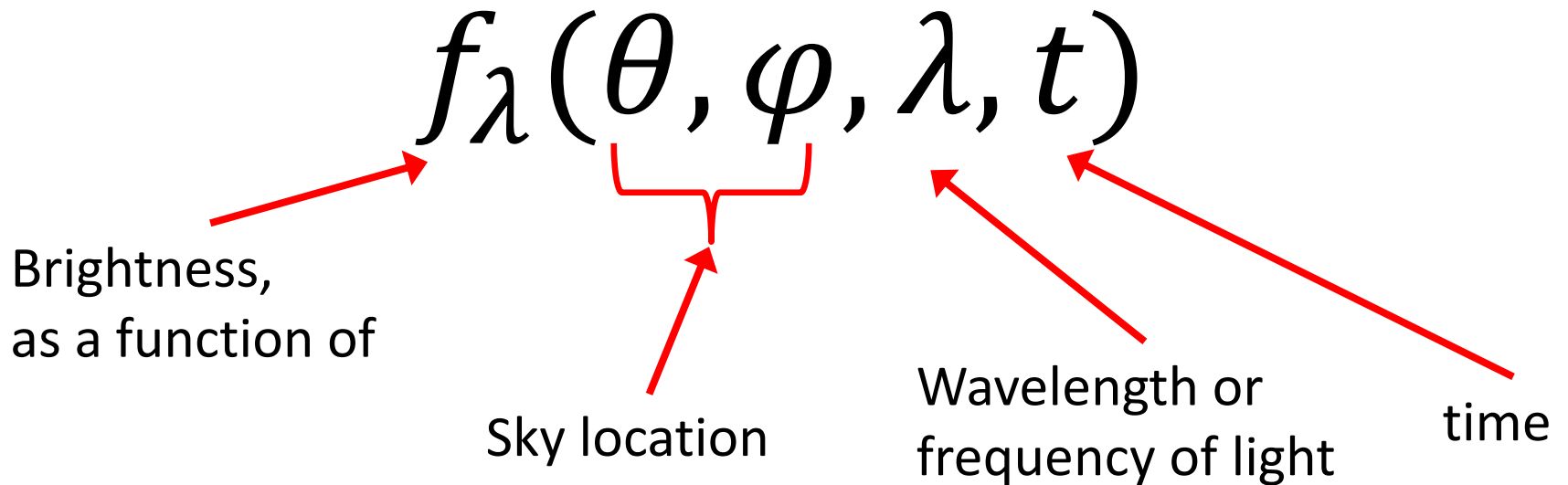# Big datasets in astronomy

Pauline Barmby
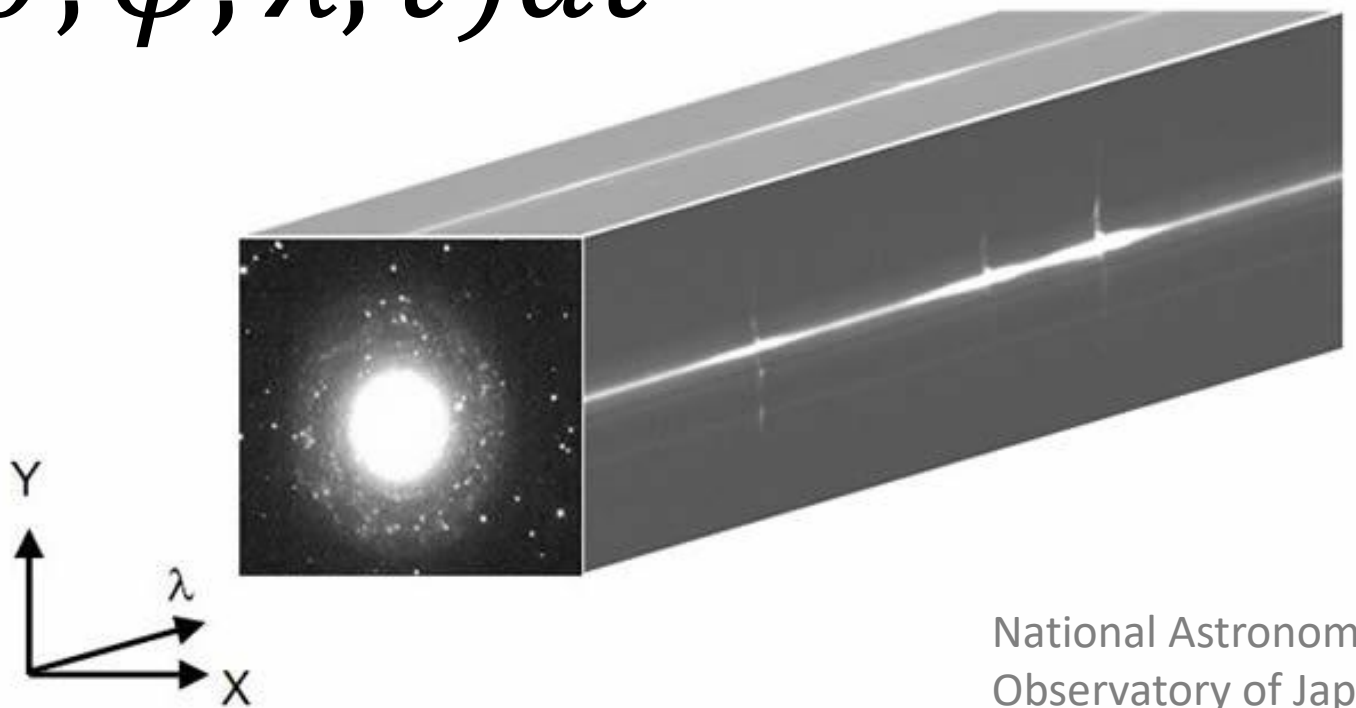Physics & Astronomy, Institute for Earth & Space
Exploration, Rotman Institute of Philosophy
Western University
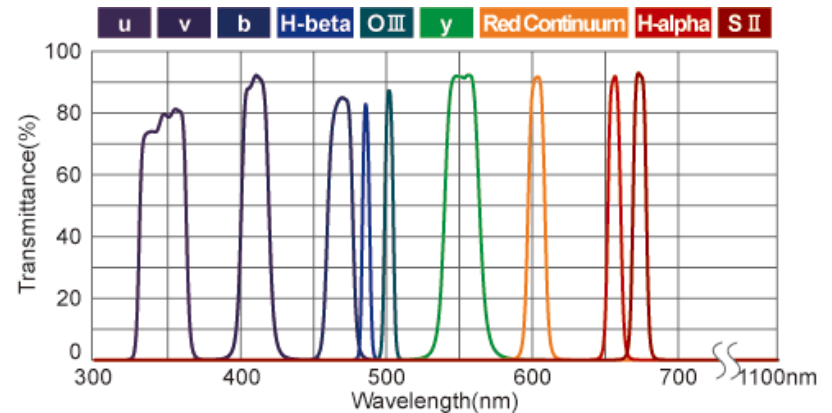
What do astronomers measure?

$$f_\lambda(\theta, \varphi, \lambda, t)$$
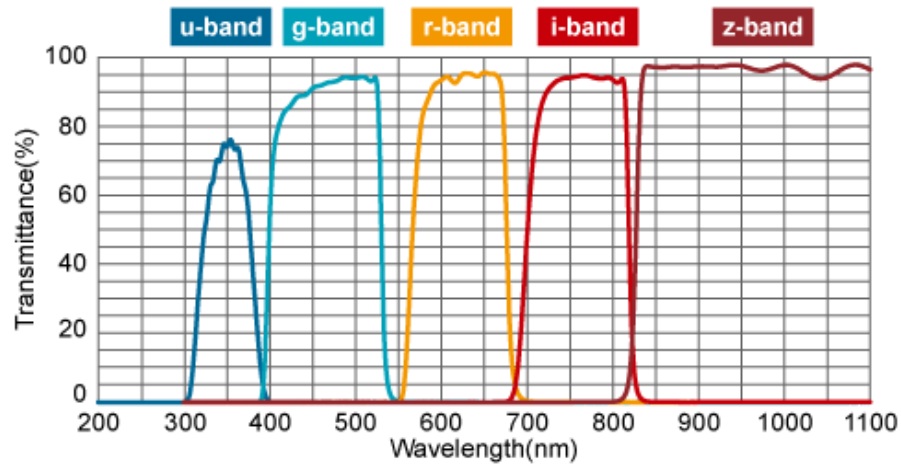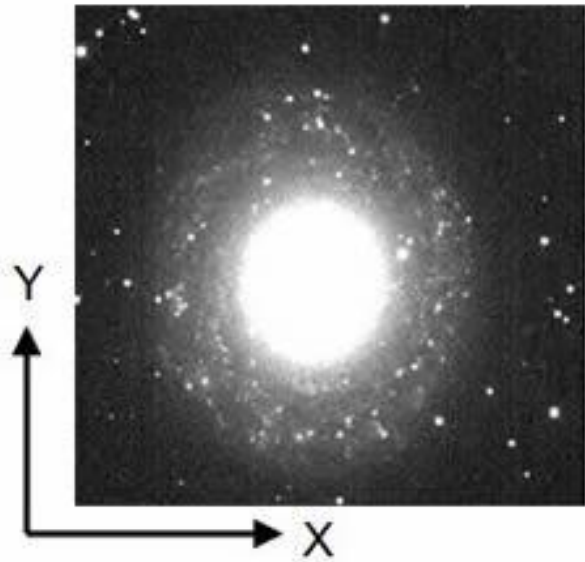
Brightness,
as a function of

Sky location

Wavelength or
frequency of light

time

Western Science

Data cube

$$\int f_\lambda (\theta, \varphi, \lambda, t)\, dt$$

National Astronomical
Observatory of Japan

Western Science

Image
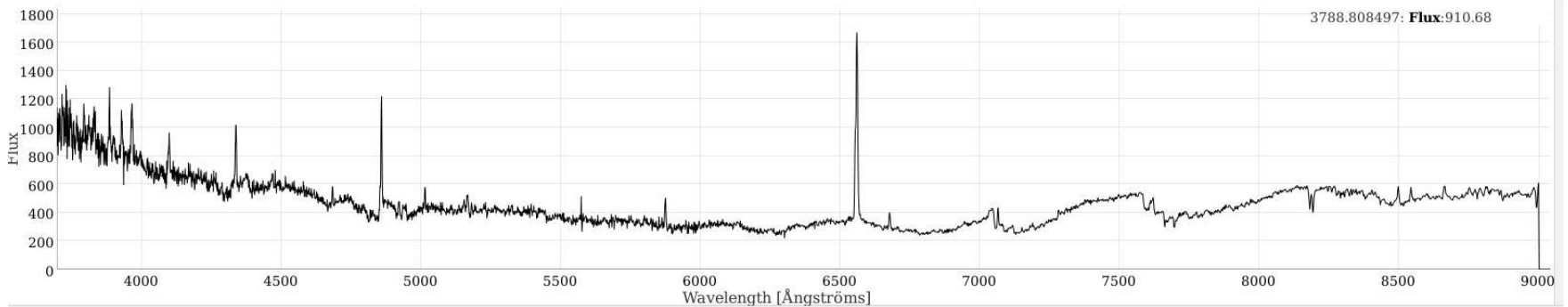
$$\int f_\lambda\left(\theta, \varphi, \lambda, t\right) dt d\lambda$$

National Astronomical
Observatory of Japan

KOS, Inc.

Western Science

Spectrum

$$\int f_\lambda (\theta, \varphi, \lambda, t) dt d\Omega$$

NAOJ

Skoda et al 2016

3788.808497: **Flux**:910.68
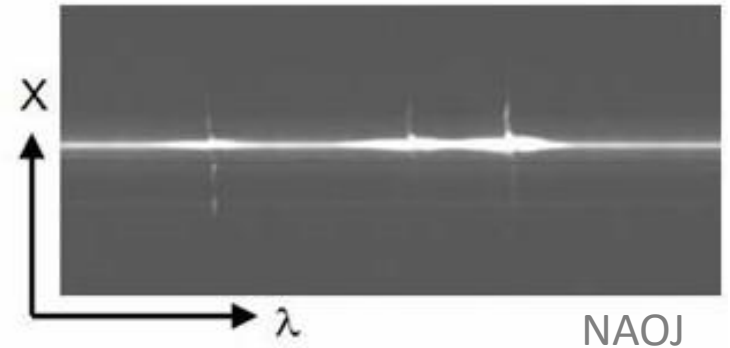
Time series photometry

$$\int f_\lambda \left(\theta, \varphi, \lambda, t \right) d\lambda d\Omega$$
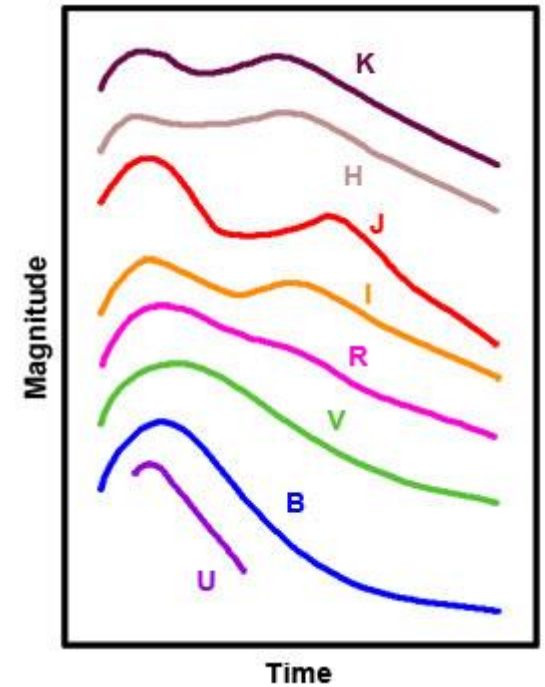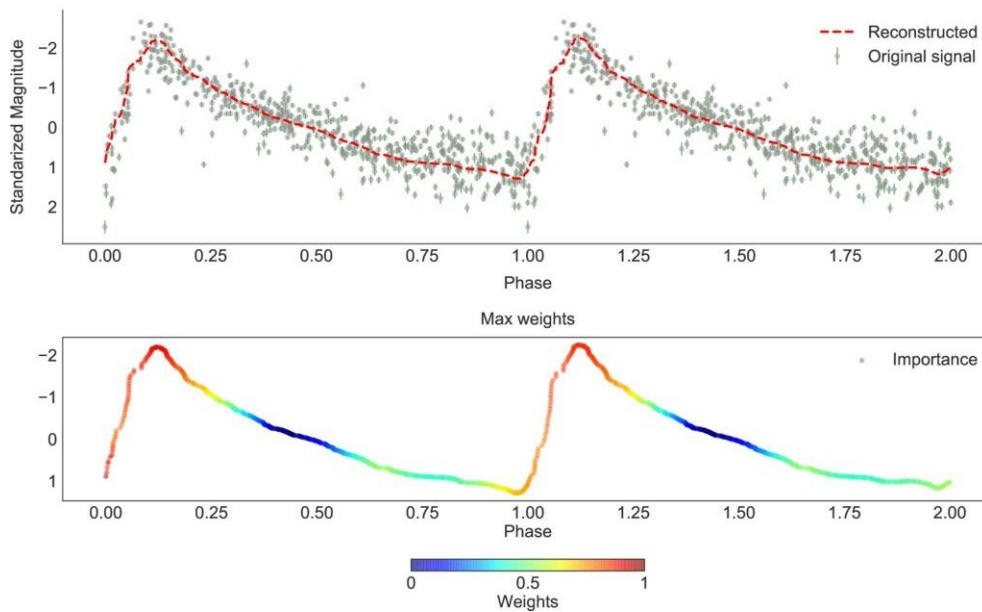
Pieringer et al 2019

Swinburne Astronomy Online

Western ⚜ Science

Capturing the light

LGSS

Cable De-Rotator

Focal Plate

Optical Relay System

NIR Spectrograph

VIS Spectrograph

Rotating Structure

Volume For Static
Support Structure

Electronic & Control Cabinets

MOSAIC

ESO
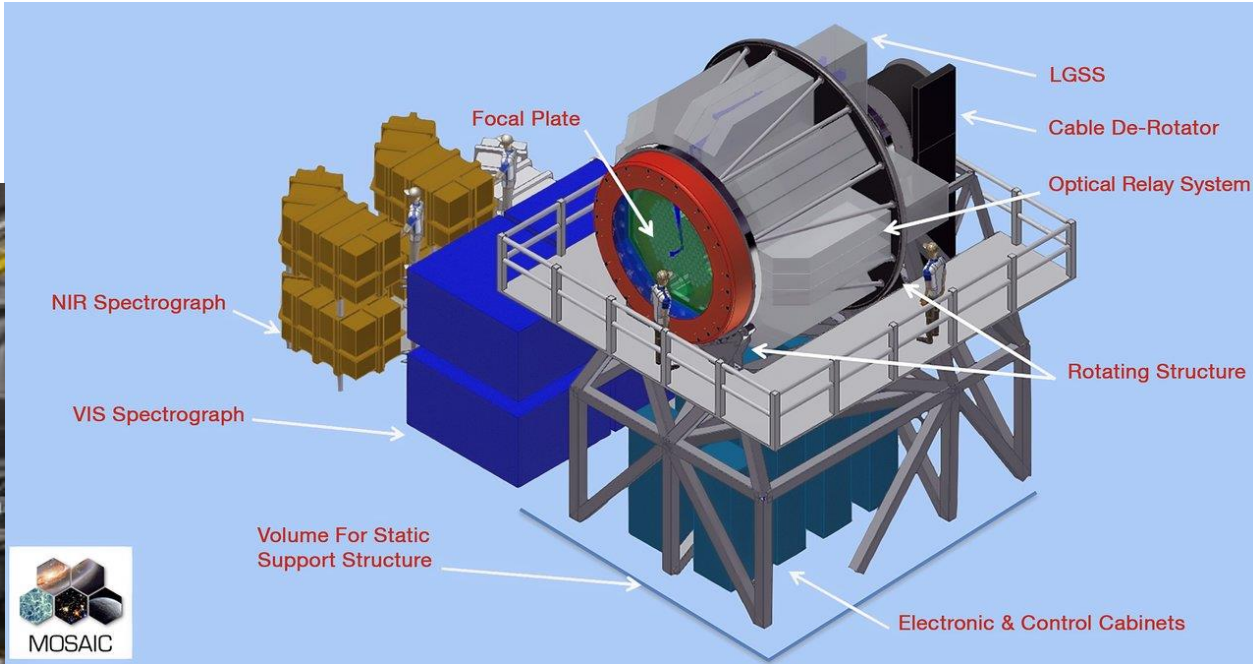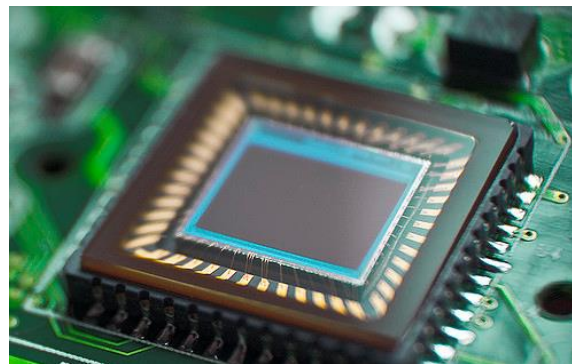
EAO

M. Laskowski via Flickr

Western Science

Measurements (and uncertainties!) from data

Location, structure, luminosity

Motion, composition, physical conditions

Measurements are compiled into catalogues

MESSIER CATALOG

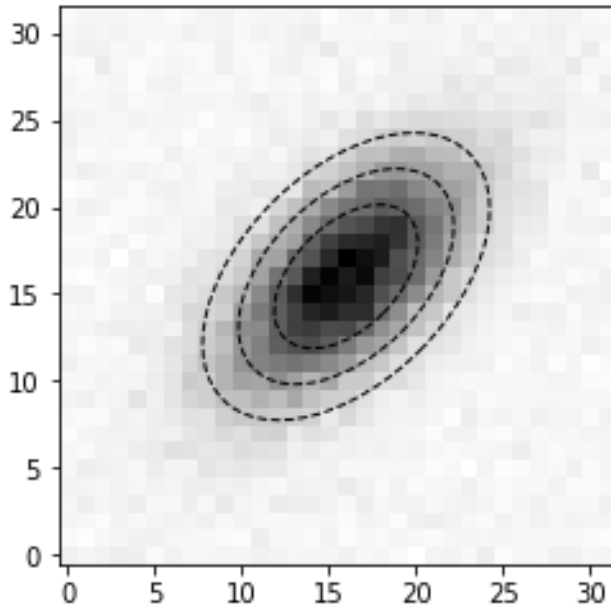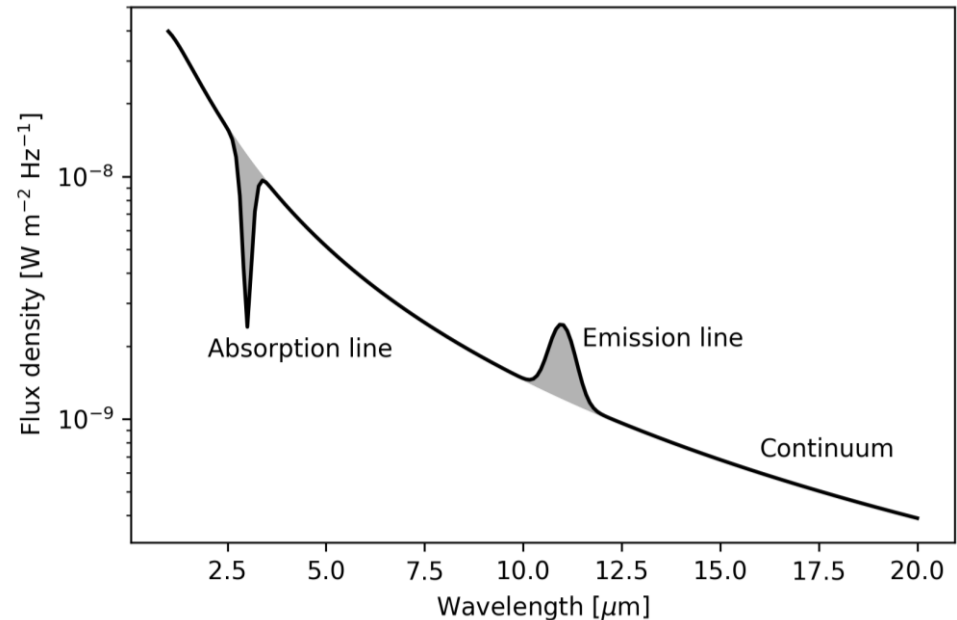| MESSIER NUMBER	OTHER DESIGNATION	RIGHT ASCENSION	DECLINATION	CONSTELLATION	TYPE	MAGNITUDE	SIZE
M1	NGC 1952	5h35m	22°01'	Taurus	Supernova remnant	8.0	6' by 4'
M2	NGC 7089	21h33m	-0°49'	Aquarius	Globular cluster	6.3	16'
M3	NGC 5272	13h42m	28°23'	Canes Venatici	Globular cluster	5.9	18'
M4	NGC 6121	16h24m	-26°31'	Scorpius	Globular cluster	5.4	36'
M5	NGC 5904	15h19m	2°05'	Serpens	Globular cluster	5.7	23'
M6	NGC 6405	17h40m	-32°16'	Scorpius	Open cluster	4.2	25'
M7	NGC 6475	17h54m	-34°47'	Scorpius	Open cluster	2.8	75'
M8	NGC 6523	18h04m	-24°23'	Sagittarius	Nebula	6.0	90' by 40'
M9	NGC 6333	17h19m	-18°31'	Ophiuchus	Globular cluster	7.8	12'
M10	NGC 6254	16h57m	-4°06'	Ophiuchus	Globular cluster	6.6	20'
M11	NGC 6705	18h51m	-6°16'	Scutum	Open cluster	5.3	14'
M12	NGC 6218	16h47m	-1°57'	Ophiuchus	Globular cluster	6.8	16'
M13	NGC 6205	16h42m	36°28'	Hercules	Globular cluster	5.3	20'
M14	NGC 6402	17h38m	-3°15'	Ophiuchus	Globular cluster	7.6	11'
M15	NGC 7078	21h30m	12°10'	Pegasus	Globular cluster	6.0	18'
M16	NGC 6611	18h19m	-13°47'	Serpens	Open cluster	6.0	7'
M17	NGC 6618	18h21m	-16°11'	Sagittarius	Nebula	7.0	11'
M18	NGC 6613	18h20m	-17°08'	Sagittarius	Open cluster	6.9	9'
M19	NGC 6273	17h03m	-26°16'	Ophiuchus	Globular cluster	6.8	17'
M20	NGC 6514	18h03m	-23°02'	Sagittarius	Nebula	9.0	28'
M21	NGC 6531	18h05m	-22°30'	Sagittarius	Open cluster	5.9	13'
M22	NGC 6656	18h36m	-23°54'	Sagittarius	Globular cluster	5.2	32'
M23	NGC 6494	17h57m	-19°01'	Sagittarius	Open cluster	5.5	27'
M24	NGC 6603	18h17m	-18°29'	Sagittarius	Star cloud	2.5	90'
M25	IC 4725	18h32m	-19°15'	Sagittarius	Open cluster	4.6	32'
M26	NGC 6694	18h45m	-9°24'	Scutum	Open cluster	8.0	15'
M27	NGC 6853	20h00m	22°43'	Vulpecula	Planetary nebula	7.3	8' by 5.7'
M28	NGC 6626	18h25m	-24°52'	Sagittarius	Globular cluster	6.9	11'
M29	NGC 6913	20h24m	38°32'	Cygnus	Open cluster	6.6	7'
M30	NGC 7099	21h40m	-23°11'	Capricornus	Globular cluster	6.9	12'
M31	NGC 224	0h43m	41°16'	Andromeda	Galaxy	3.4	178' by 63'
M32	NGC 221	0h43m	40°52'	Andromeda	Galaxy	8.2	8' by 6
M33	NGC 598	1h34m	30°39'	Triangulum	Galaxy	5.7	73' by 45'
M34	NGC 1039	2h42m	42°47'	Perseus	Open cluster	5.2	35'
M35	NGC 2168	6h09m	24°20'	Gemini	Open cluster	5.1	28'
M36	NGC 1960	5h36m	34°08'	Auriga	Open cluster	6.0	12'
M37	NGC 2099	5h52m	32°33'	Auriga	Open cluster	5.6	24'
M38	NGC 1922	5h28m	35°50'	Auriga	Open cluster	6.4	21'
M39	NGC 7092	21h32m	48°26'	Cygnus	Open cluster	4.6	32'
M40	Winnecke 4	12h22m	58°05'	Ursa Major	Double star	9.0/9.6	0.8'
M41	NGC 2287	6h46m	-20°44'	Canis Major	Open cluster	4.5	38'
M42	NGC 1976	5h35m	-5°27'	Orion	Nebula	3.7	85' by 60'
M43	NGC 1982	5h36m	-5°16'	Orion	Nebula	6.8	20' by 15'
M44	NGC 2632	8h40m	19°59'	Cancer	Open cluster	3.1	95'
M45	The Pleiades	3h47m	24°07'	Taurus	Open cluster	1.5	110'
M46	NGC 2437	7h42m	-14°49'	Puppis	Open cluster	6.1	27'
M47	NGC 2422	7h37m	-14°30'	Puppis	Open cluster	4.4	30'
M48	NGC 2548	8h14m	-5°48'	Hydra	Open cluster	5.8	54'
M49	NGC 4472	12h30m	8°00'	Virgo	Galaxy	8.4	9' by 7.5'
M50	NGC 2323	7h03m	-8°20'	Monoceros	Open cluster	5.9	16'
M51	NGC 5194	13h30m	47°12'	Canes Venatici	Galaxy	8.4	11' by 7'
M52	NGC 7654	23h24m	61°35'	Cassiopeia	Open cluster	6.9	13'
M53	NGC 5024	13h13m	18°10'	Coma Berenices	Globular cluster	7.7	13'
M54	NGC 6715	18h55m	-30°29'	Sagittarius	Globular cluster	7.2	12'
M55	NGC 6809	19h40m	-30°58'	Sagittarius	Globular cluster	6.3	19'
M56	NGC 6779	19h17m	30°11'	Lyra	Globular cluster	8.4	8.8'
M57	NGC 6720	18h54m	33°02'	Lyra	Planetary nebula	8.8	1.4' by 1'
M58	NGC 4579	12h38m	11°49'	Virgo	Galaxy	9.6	5.5' by 4.5'
M59	NGC 4621	12h42m	11°39'	Virgo	Galaxy	9.6	5' by 3.5'
M60	NGC 4649	12h44m	11°33'	Virgo	Galaxy	8.8	7' by 6'
M61	NGC 4303	12h22m	4°28'	Virgo	Galaxy	9.6	6' by 5.5'
M62	NGC 6266	17h01m	-30°07'	Ophiuchus	Globular cluster	6.7	15'
M63	NGC 5055	13h16m	42°02'	Canes Venatici	Galaxy	8.6	10' by 6'
M64	NGC 4826	12h57m	21°41'	Coma Berenices	Galaxy	8.5	9.3' by 5.4'
M65	NGC 3623	11h19m	13°05'	Leo	Galaxy	8.8	8' by 1.5'
M66	NGC 3627	11h20m	12°59'	Leo	Galaxy	9.0	8' by 2.5'
M67	NGC 2682	8h50m	11°49'	Cancer	Open cluster	6.0	30'
M68	NGC 4590	12h40m	-26°45'	Hydra	Globular cluster	7.6	11'
M69	NGC 6637	18h31m	-32°21'	Sagittarius	Globular cluster	7.4	9.8'
M70	NGC 6681	18h43m	-32°18'	Sagittarius	Globular cluster	7.8	8'
M71	NGC 6838	19h54m	18°47'	Sagitta	Globular cluster	8.0	7.2'
M72	NGC 6981	20h54m	-12°32'	Aquarius	Globular cluster	9.2	6.6'
M73	NGC 6994	20h59m	-12°38'	Aquarius	Open cluster	8.9	2.8'
M74	NGC 628	1h37m	15°47'	Pisces	Galaxy	8.5	10.2' by 9.5'
M75	NGC 6864	20h06m	-21°55'	Sagittarius	Globular cluster	8.6	6.8'
M76	NGC 650	1h42m	51°34'	Perseus	Planetary nebula	10.1	2.7' by 1.8'
M77	NGC 1068	2h43m	-0°01'	Cetus	Galaxy	8.9	7' by 6'
M78	NGC 2068	5h47m	0°03'	Orion	Nebula	8.0	8' by 6'
M79	NGC 1904	5h25m	-24°33'	Lepus	Globular cluster	7.7	9.6'
M80	NGC 6093	16h17m	-22°59'	Scorpius	Globular cluster	7.3	10'
M81	NGC 3031	9h56m	69°04'	Ursa Major	Galaxy	6.9	21' by 10'
M82	NGC 3034	9h56m	69°41'	Ursa Major	Galaxy	8.4	9' by 4'
M83	NGC 5236	13h37m	-29°52'	Hydra	Galaxy	7.5	11' by 10'
M84	NGC 4374	12h25m	12°53'	Virgo	Galaxy	9.1	5'
M85	NGC 4382	12h25m	18°11'	Coma Berenices	Galaxy	9.1	7.1' by 5.2'
M86	NGC 4406	12h26m	12°57'	Virgo	Galaxy	8.9	7.5' by 5.5'
M87	NGC 4486	12h31m	12°24'	Virgo	Galaxy	8.6	7'
M88	NGC 4501	12h32m	14°25'	Coma Berenices	Galaxy	9.6	7' by 4'
M89	NGC 4552	12h36m	12°33'	Virgo	Galaxy	9.7	4'
M90	NGC 4569	12h37m	13°10'	Virgo	Galaxy	9.5	9.5' by 4.5'
M91	NGC 4548	12h35m	14°30'	Coma Berenices	Galaxy	10.1	5.4' by 4.4'
M92	NGC 6341	17h17m	43°08'	Hercules	Globular cluster	6.5	14'
M93	NGC 2447	7h45m	-23°52'	Puppis	Open cluster	6.2	22'
M94	NGC 4736	12h51m	41°07'	Canes Venatici	Galaxy	8.2	7' by 3'
M95	NGC 3351	10h44m	11°42'	Leo	Galaxy	9.7	4.4' by 3.3'
M96	NGC 3368	10h47m	11°49'	Leo	Galaxy	9.2	6' by 4'
M97	NGC 3587	11h15m	55°01'	Ursa Major	Planetary nebula	9.9	3.4' by 3.3'
M98	NGC 4192	12h14m	14°54'	Coma Berenices	Galaxy	10.1	9.5' by 3.2'
M99	NGC 4254	12h19m	14°25'	Coma Berenices	Galaxy	9.9	5.4' by 4.8'
M100	NGC 4321	12h23m	15°49'	Coma Berenices	Galaxy	9.3	7' by 6'
M101	NGC 5457	14h03m	54°21'	Ursa Major	Galaxy	7.9	22'
M102	NGC 5866	15h07m	55°46'	Draco	Galaxy	10.0	5.2' by 2.3'
M103	NGC 581	1h33m	60°42'	Cassiopeia	Open cluster	7.4	6'
M104	NGC 4594	12h40m	-11°37'	Virgo	Galaxy	8.0	9' by 4'
M105	NGC 3379	10h48m	12°35'	Leo	Galaxy	9.3	2'
M106	NGC 4258	12h19m	47°18'	Canes Venatici	Galaxy	8.3	19' by 8'
M107	NGC 6171	16h33m	-13°03'	Ophiuchus	Globular cluster	7.8	13'
M108	NGC 3556	11h12m	55°40'	Ursa Major	Galaxy	10.0	8' by 1'
M109	NGC 3992	11h58m	53°23'	Ursa Major	Galaxy	9.8	7' by 4'

M28   M29   M30   M31

M33   M34   M35   M36

Astro-data can differ from other big data

Astrostatistics.psu.edu
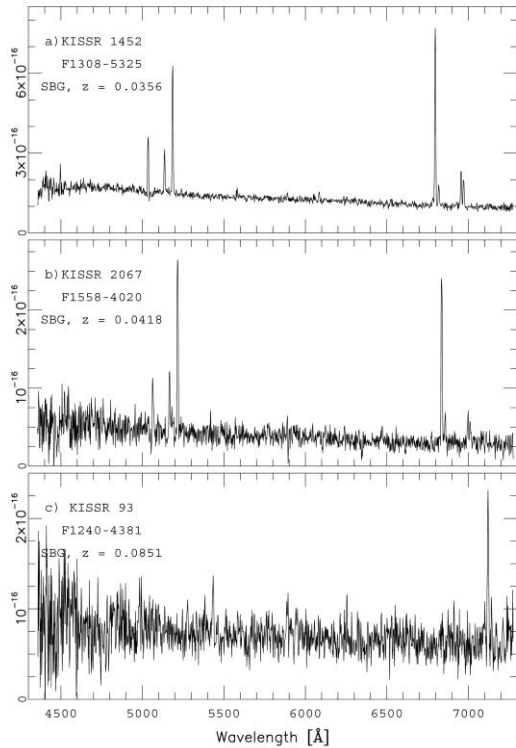
Noisy

B. Campbell/NRAO
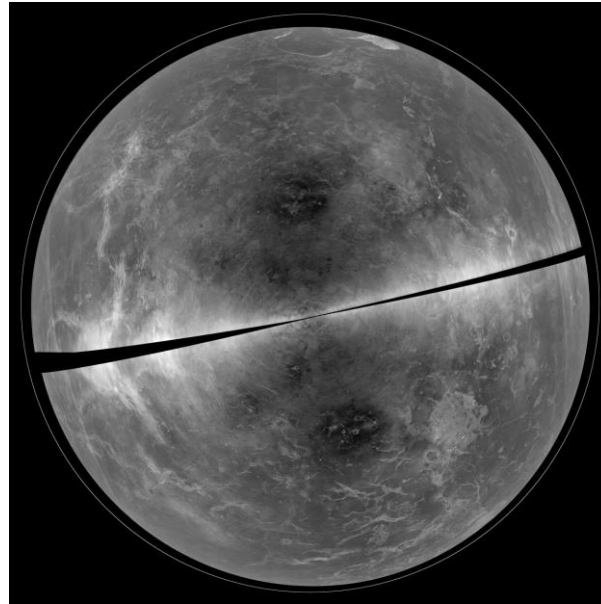
Incomplete/truncated

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

Doesn't always fit into a database model

Astro-data is very often public!

Data(base) practices are still evolving

CADC, SDSS, cnesuvsqastroide.github.io/
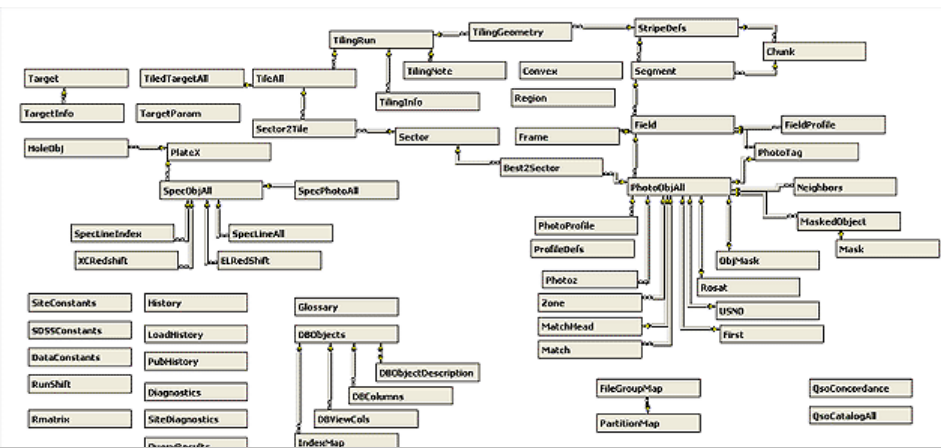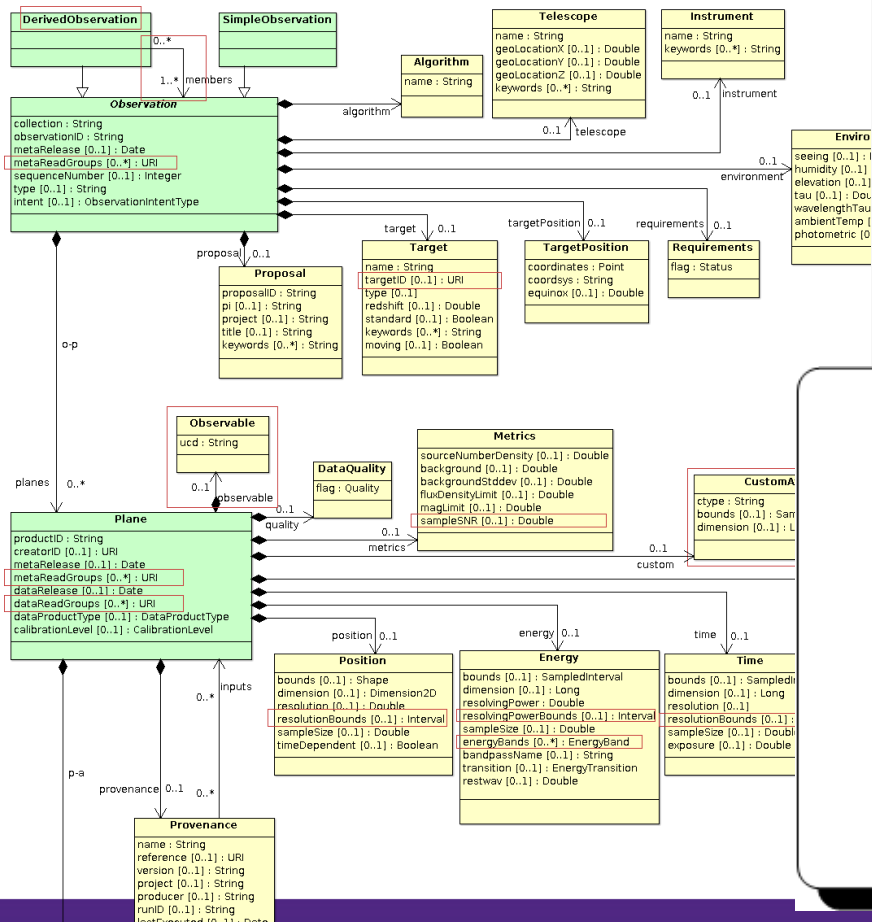
Astronomical Data Query Language

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

IVOA

Western✻Science

Meta-databases: SIMBAD, NED, etc

NED
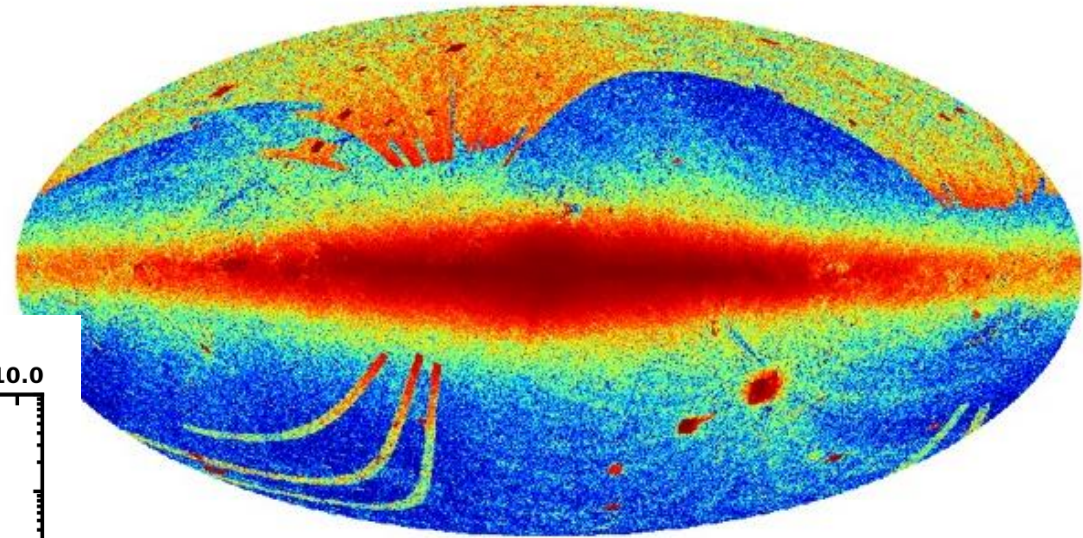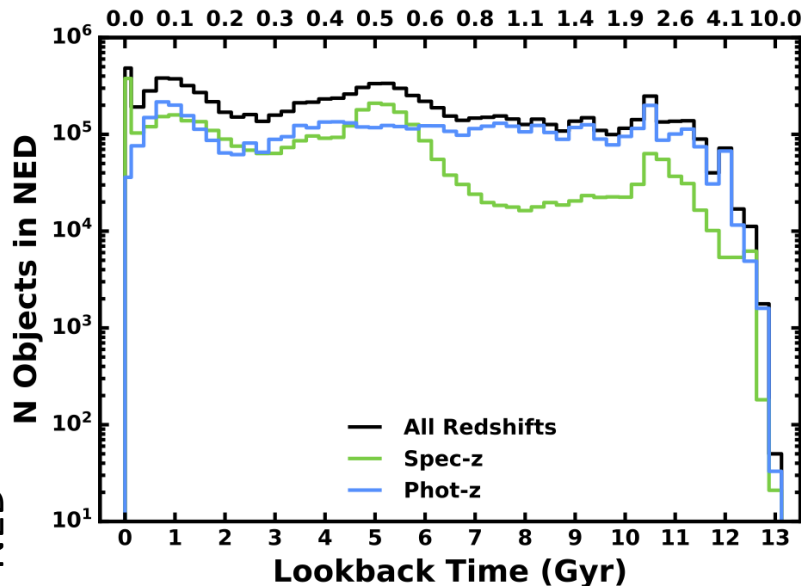
Gaia: a billion stars, many database columns

ESA, M. Richmond

Gaia mission: numbers

- 2×10^9 objects in Milky Way & beyond, first distance for 99.9%

- About 1% of Milky Way stars

- Faintest stars: 10^{-6}x human eye detection limit

- Positional accuracy: 24 µarcsec = 0.11 nanoradian

- Nearest stars: distances measured to 0.001%

- Stars near Galactic Centre (30k lt-yr): distances to 20%

- 400+ people in Gaia Data Processing and Analysis

Consortium

Western Science

Digital telescopes: CHIME & CHORD (Canada)

13 Tb/sec

phys.org

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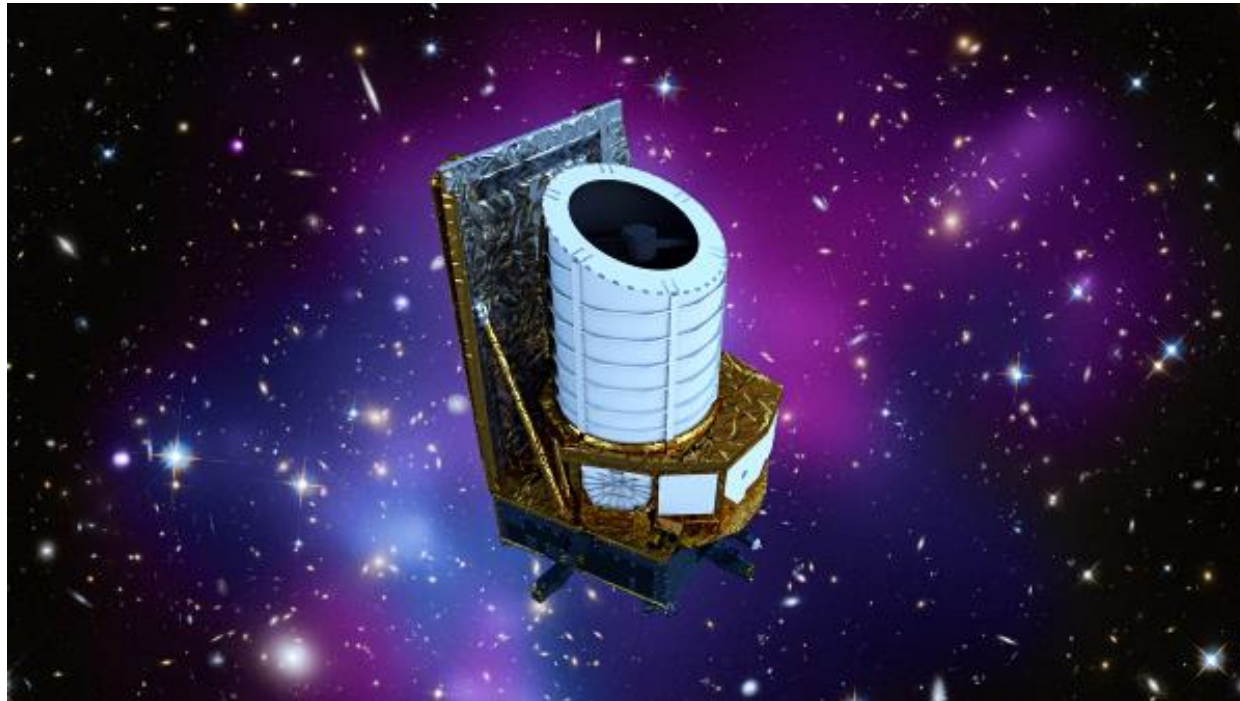
- 1.2-m diameter visible/near-infrared telescope designed for surveys; imaging and spectroscopy
- launched July 2023, first public data 2025

Western ⚜ Science

Rubin/Legacy Survey of Space & Time (Chile)

Under construction in Chile, operations to begin 2025

Will survey entire southern sky once/week over 10 years
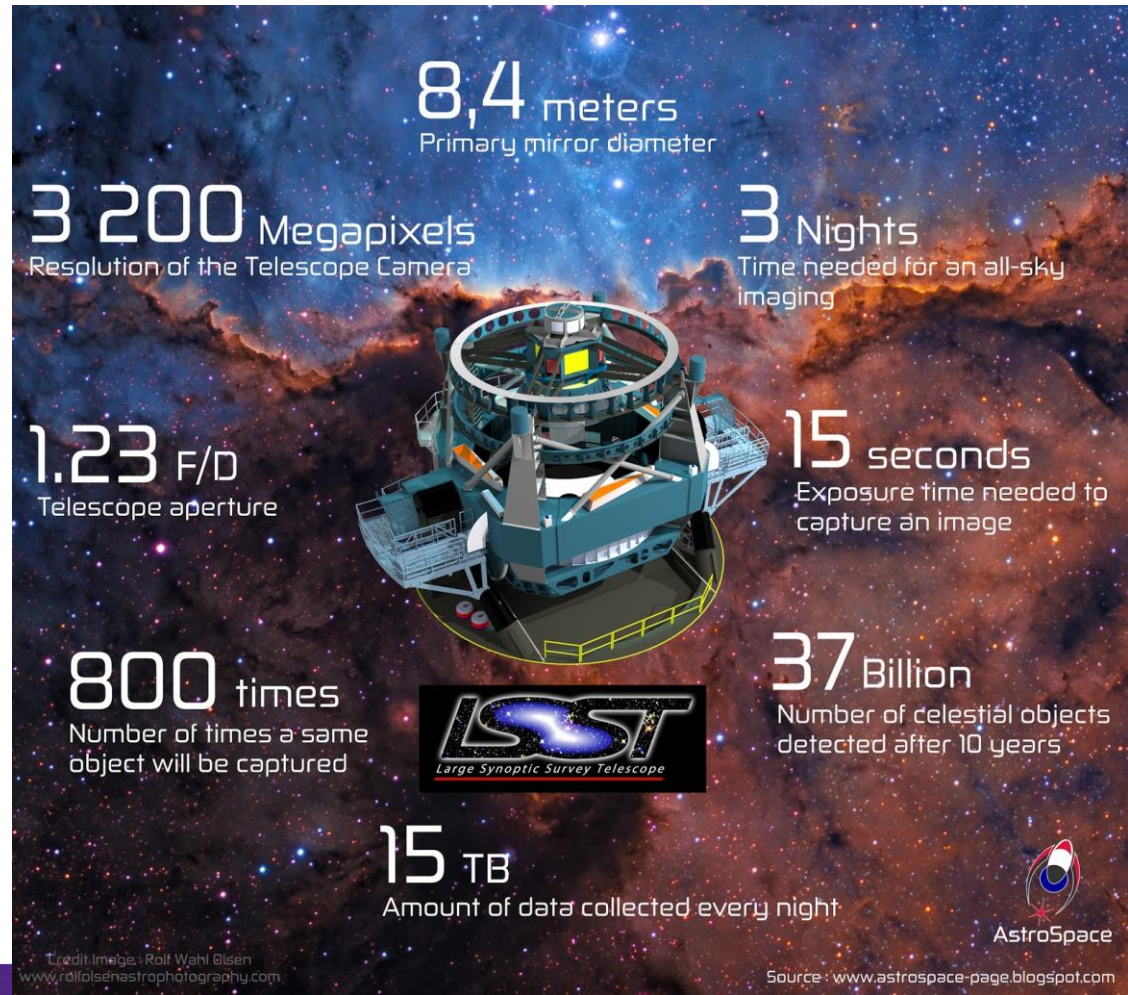
Astrospace-page.blogspot.com

Nancy Grace Roman Space Telescope (NASA)

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

Visible and infrared imaging surveys and coronagraph

Launch 202?

NASA

Simulated Roman Observation of Andromeda (M31)

Hubble PHAT IR survey

Hubble WFC3/IR footprint

Moon to scale

Roman footprint

N

E

Background Sky: Digitized Sky Survey and R. Gendler
Moon: NASA/GSFC/ASU/Lunar Reconnaissance Orbiter

20,000 light-years
6,200 parsecs   27'
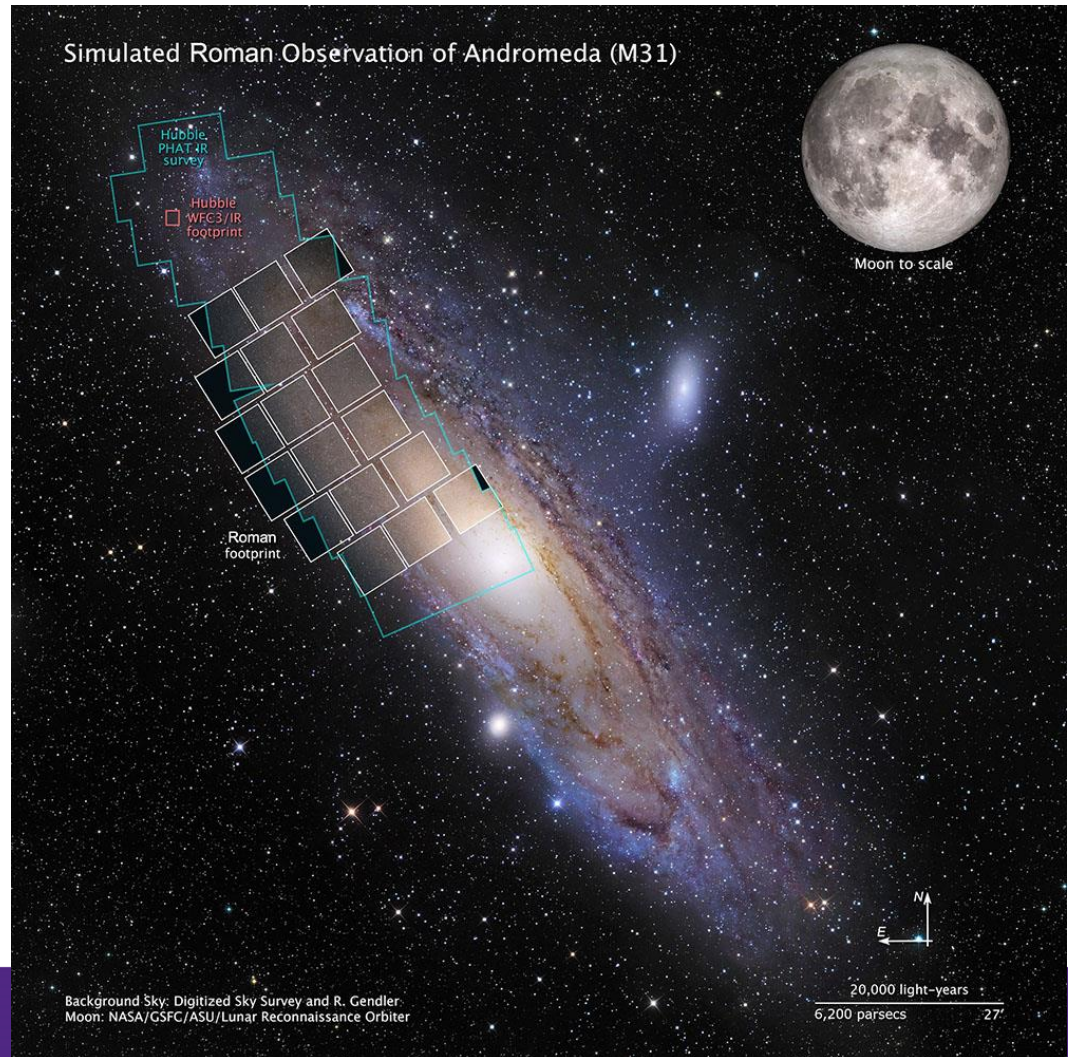
SKA Observatory (S. Africa, Australia, NZ)
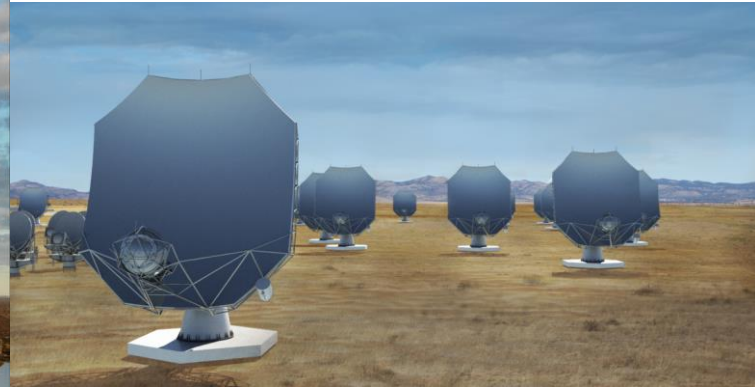
300 PB/year,
operations begin 2027

Daily raw data generated by the Square Kilometer Array:

15 million 64GB iPods

2x daily global internet traffic

ibm.com

SKAO, NAOJ

Western Science

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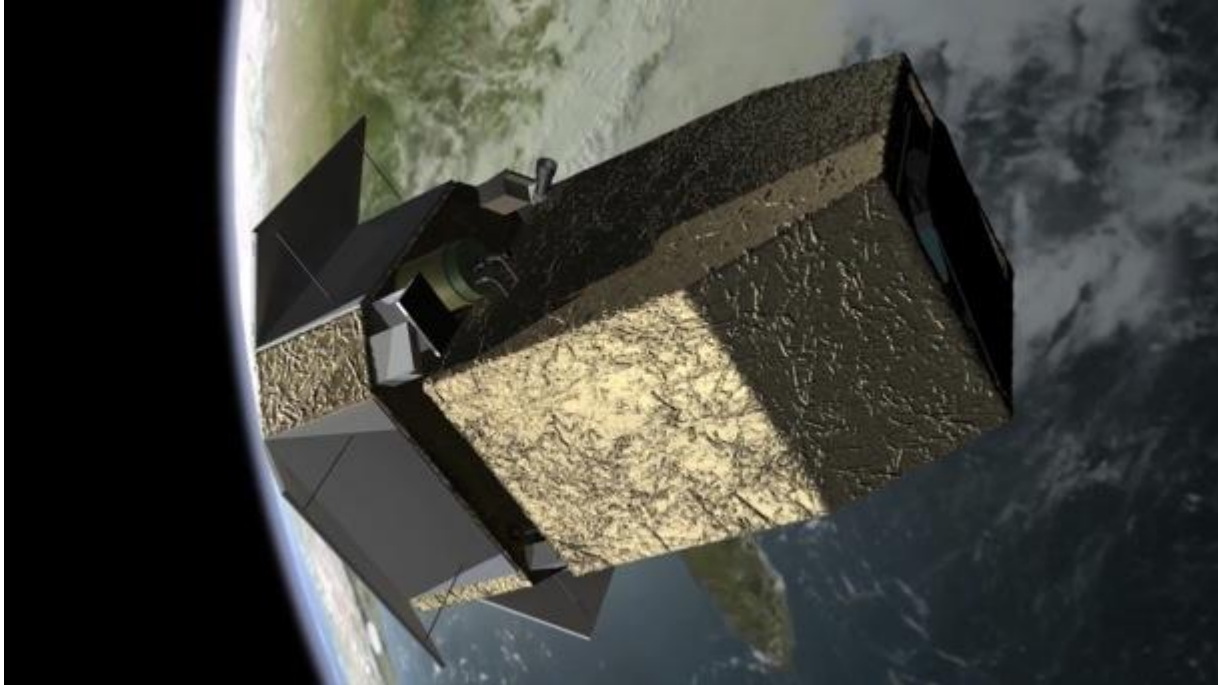
- 1-m diameter visible/ultraviolet telescope with large field of view; imaging and spectroscopy
- CSA-led with Japan, India, US: funding advocacy in progress

Western Science

@PBarmby

Putting it all together: CanDIAPL
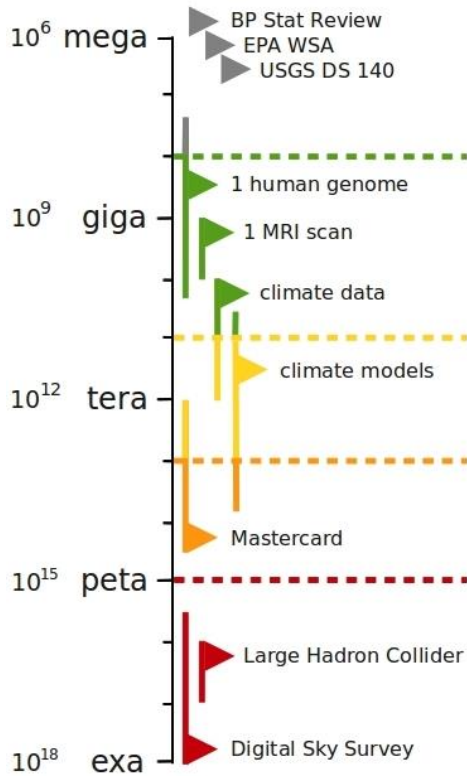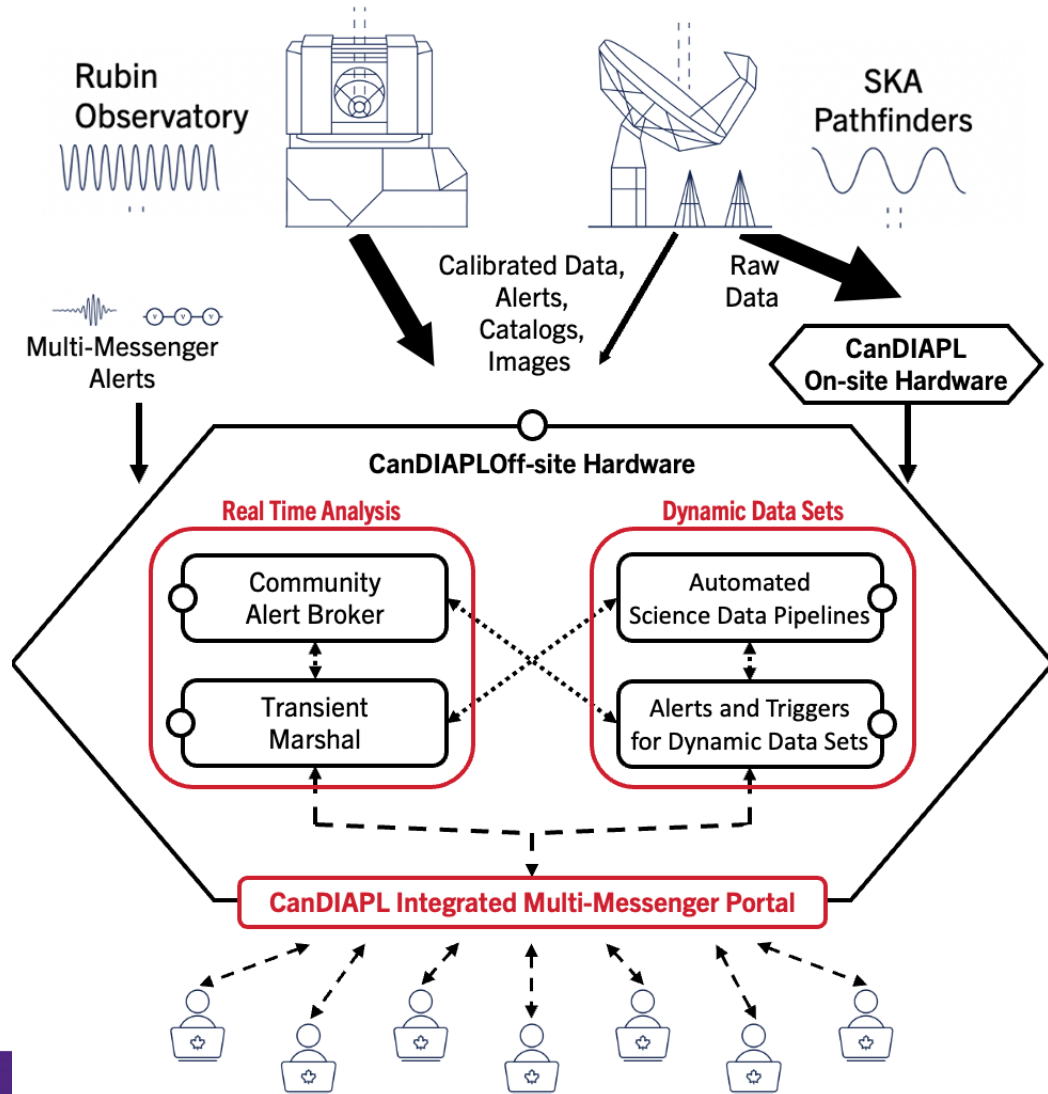
Gaia science: what did *Gaia* miss?
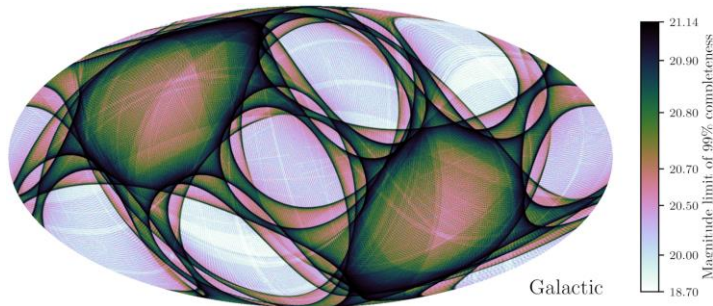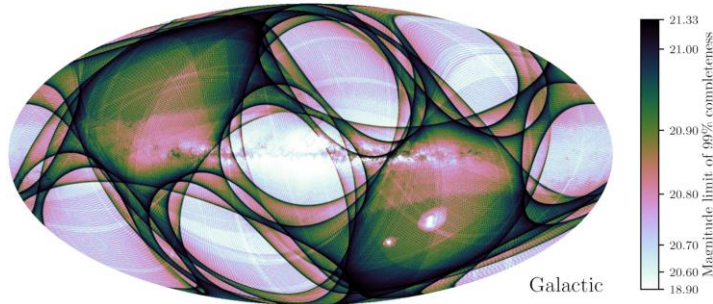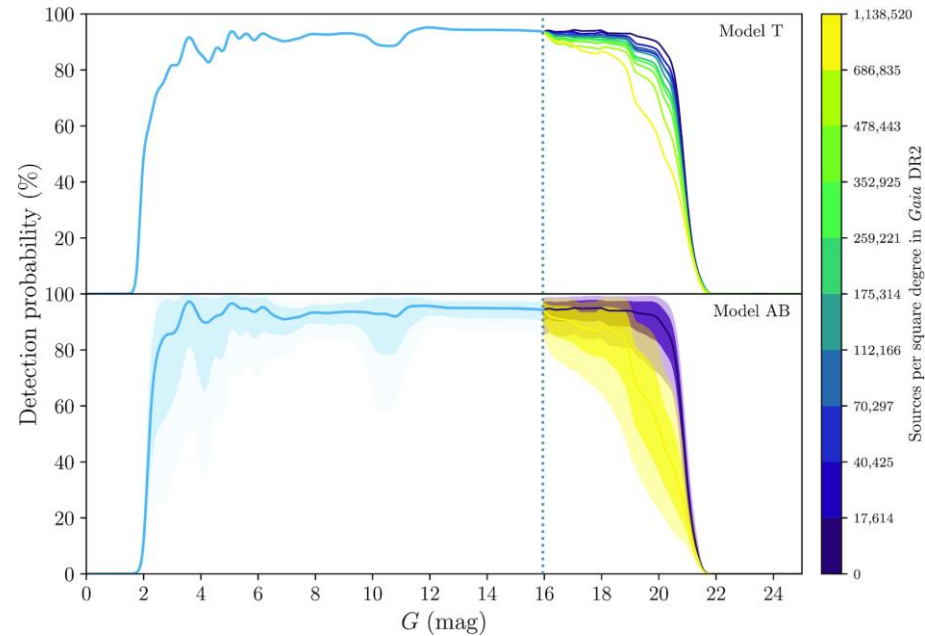
Boubert & Everall et al 2020, MNRAS, 497, 4246–426 **DOI** 10.1093/mnras/staa2305

Gaia science: nearby galaxy rotation
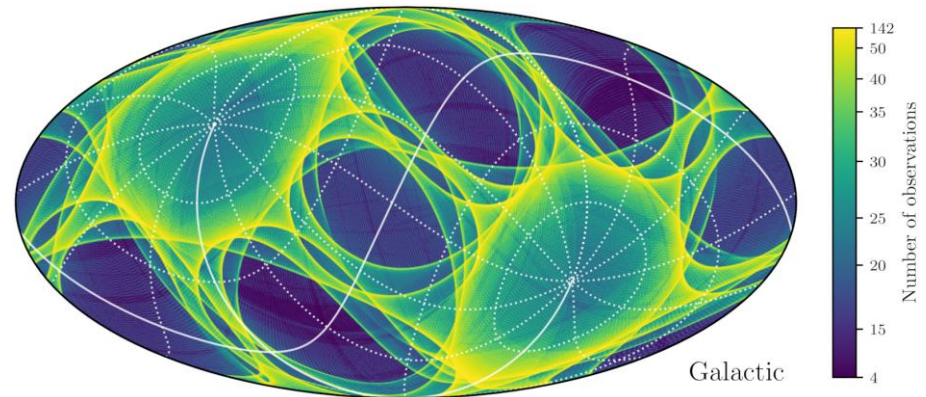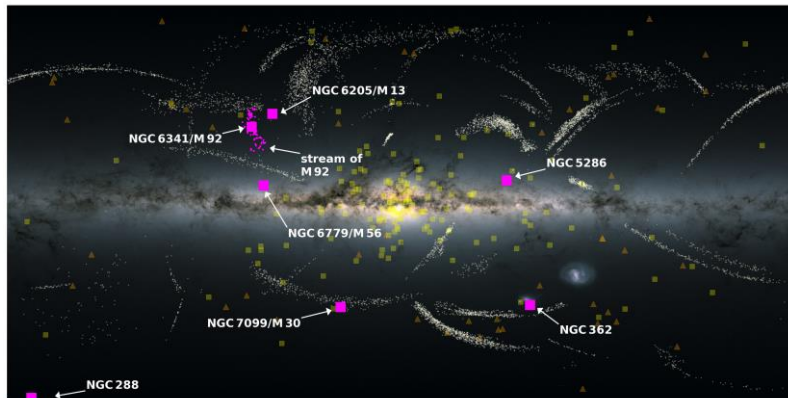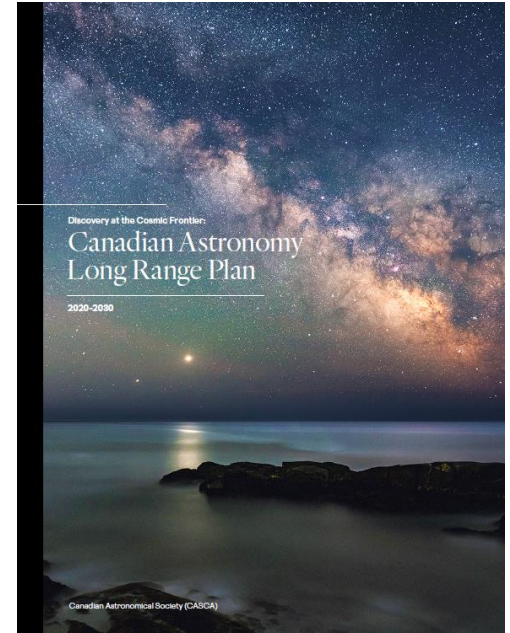
Western ⚜ Science
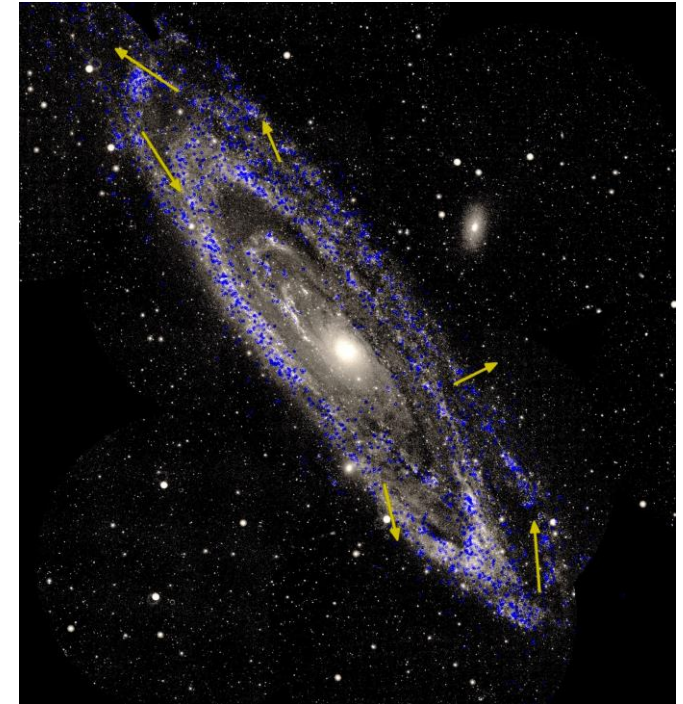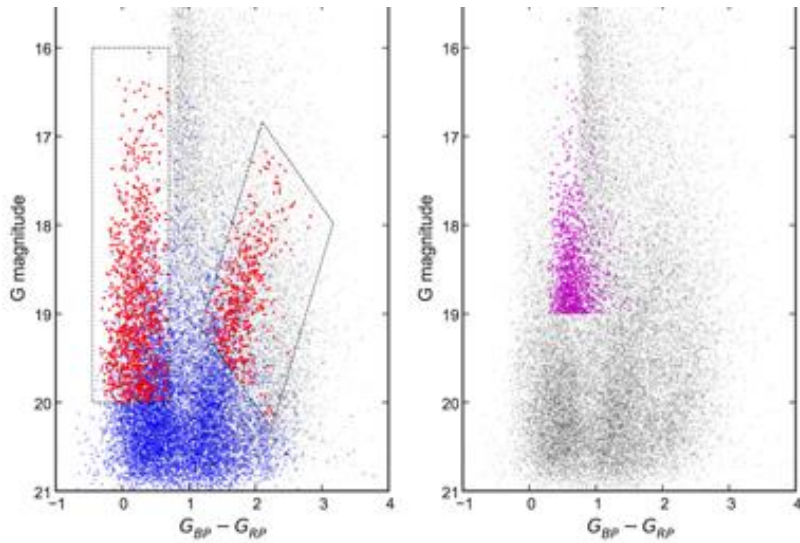
Gaia science: classifying nearby galaxy contents
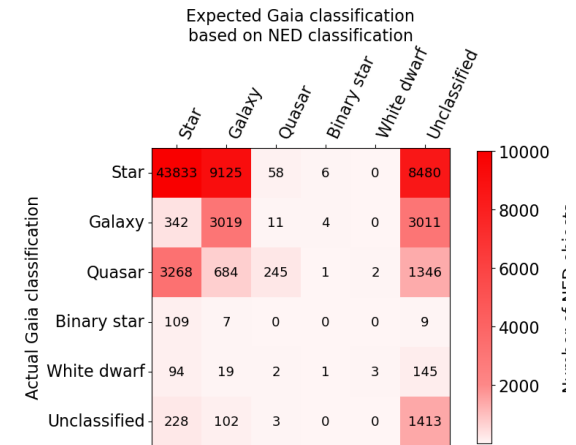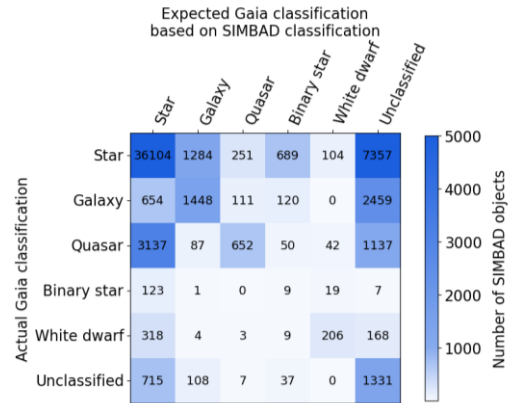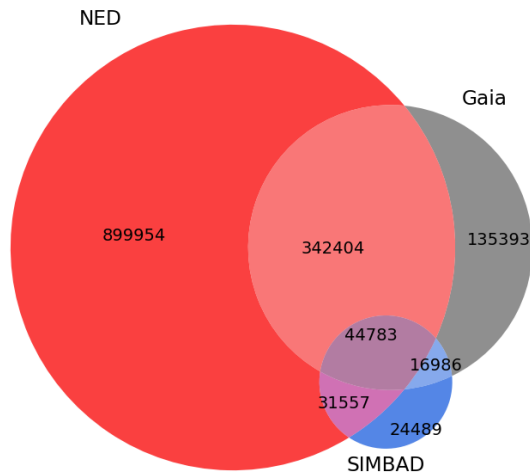
Hales & Barmby 2023 MNRAS, submitted

To learn more..

- [Astronomical observations: a guide for allied researchers](#)
- [Data in observational astronomy](#)
- [Gaia archive](#)
- [casca.ca/lrp2020](#) & US counterpart [Astro2020](#)

Boubert & Everall 2020, ESA

Western ⚜ Science