

Problems of Astronomical Proportions

Research Challenges in the Era of Data-Intensive Astronomy

Professor Richard McDermid

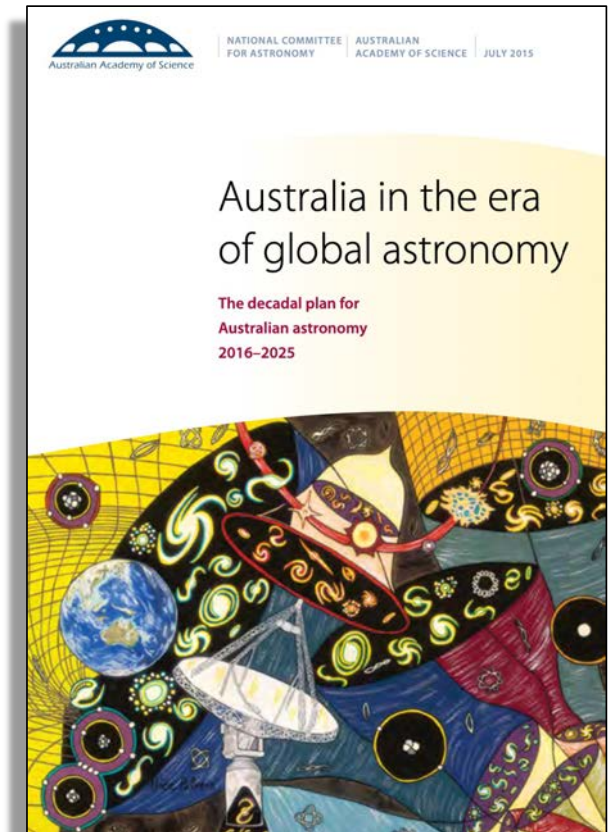
Macquarie University Astrophysics and Space Technologies Research Centre

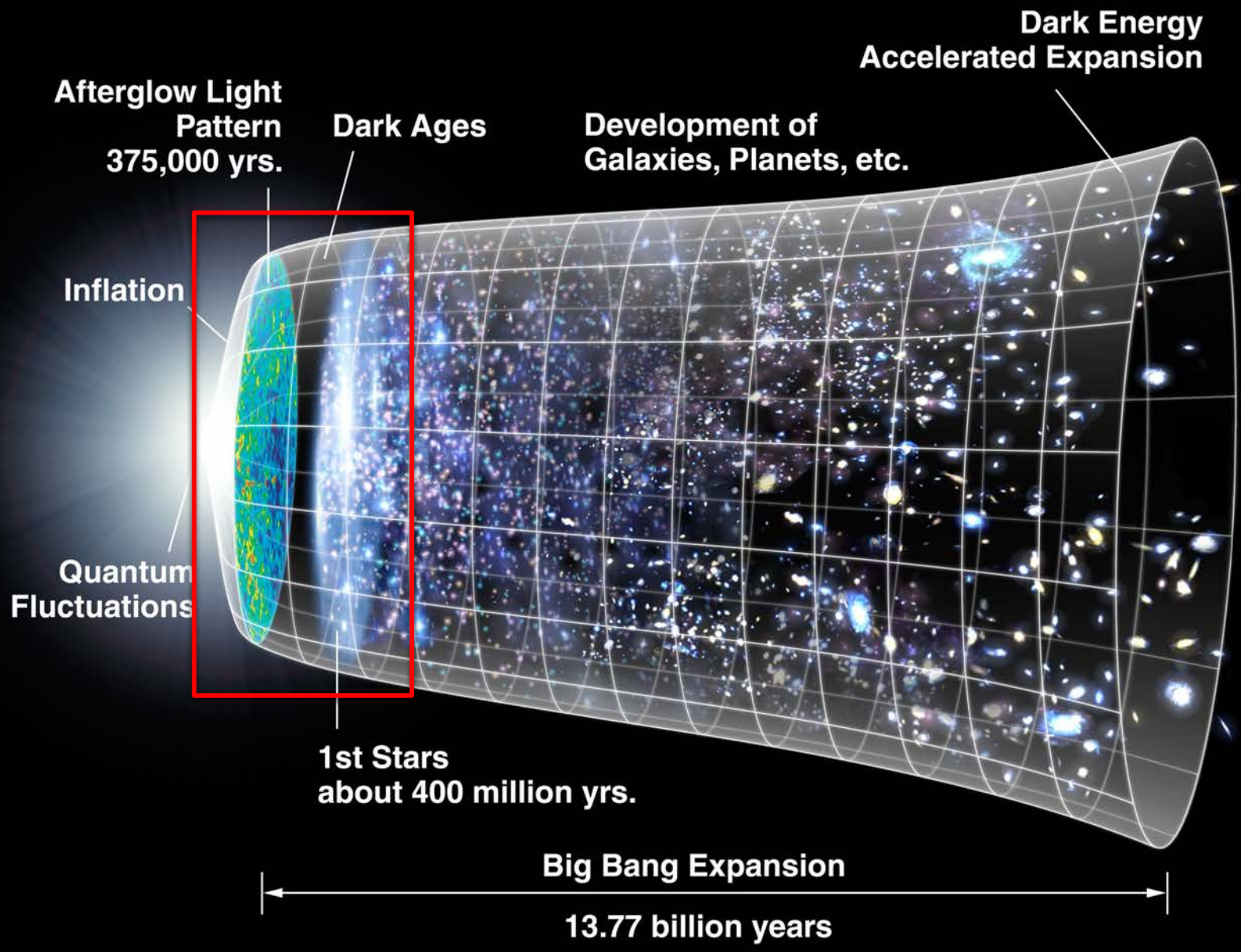
ARC Centre of Excellence for Astronomy in 3 Dimensions (ASTRO 3D)

Driving Questions of Modern Astrophysics

The Decadal Plan for Australian Astronomy

- How did the first stars and galaxies transform the Universe?
- What is the nature of dark matter and dark energy?
- How do galaxies form and evolve across cosmic time?
- How do stars and planets form?
- How are elements produced by stars and recycled through galaxies?
- What is the nature of matter and gravity at extreme densities?





Billions of years ago

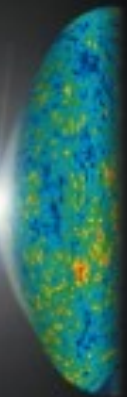
13.8

13.5

13.4

13

Big Bang



Cosmic
'dark ages'

Observed
period of
'cosmic dawn'

Oldest observed
galaxy

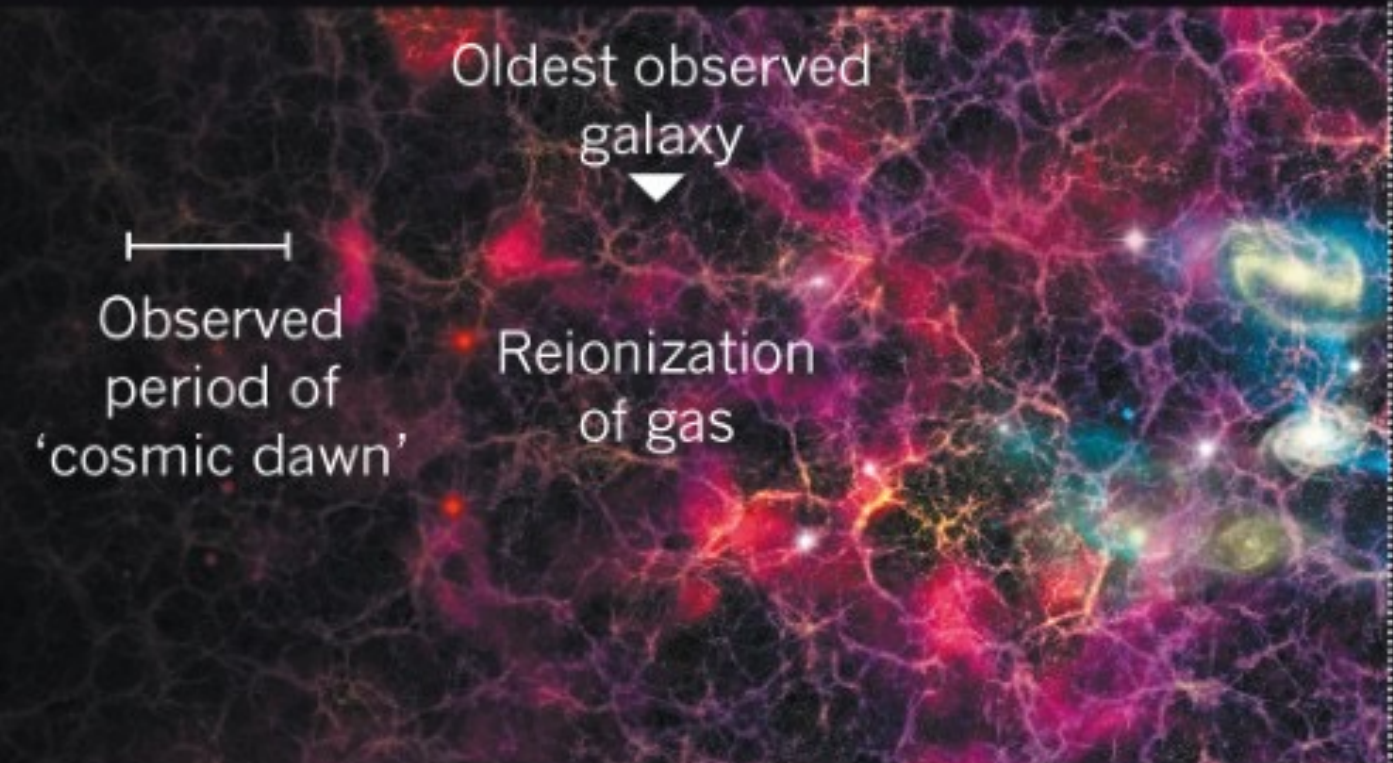
Reionization
of gas

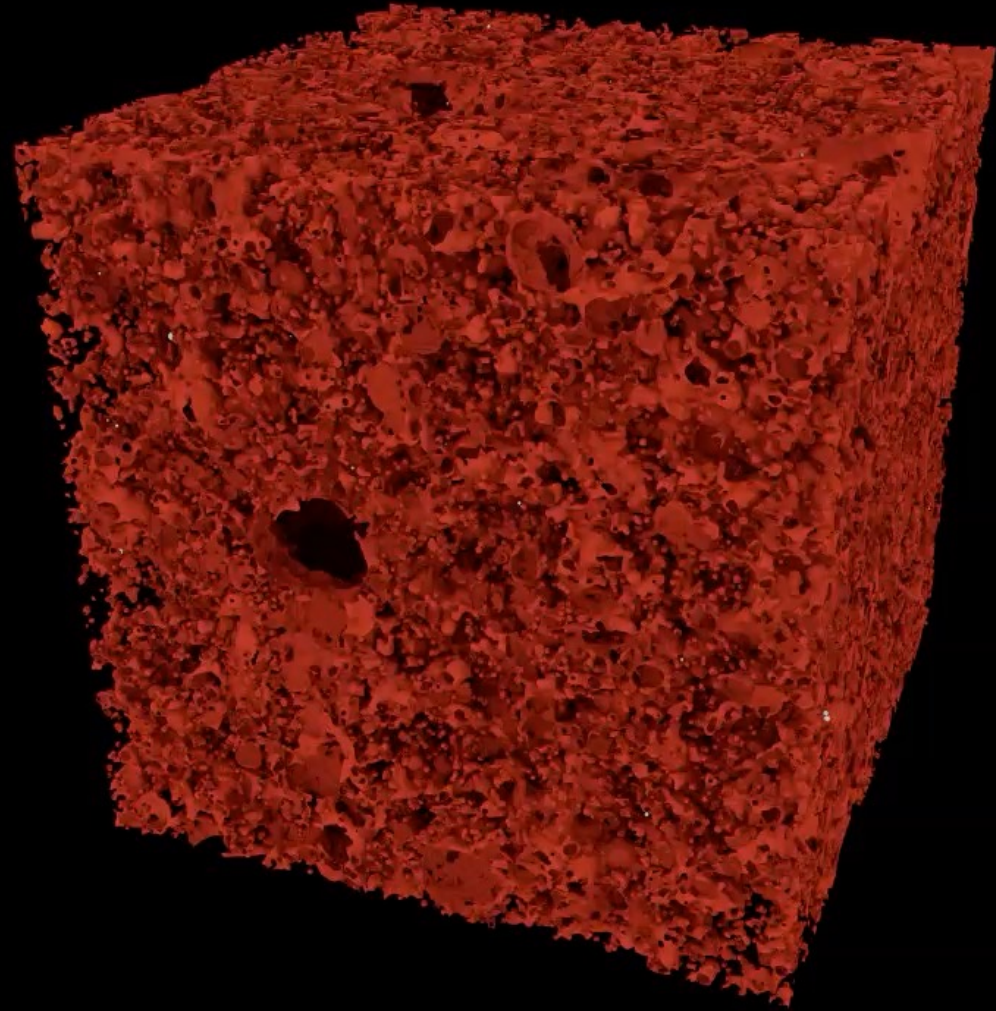
Recombination

First
stars

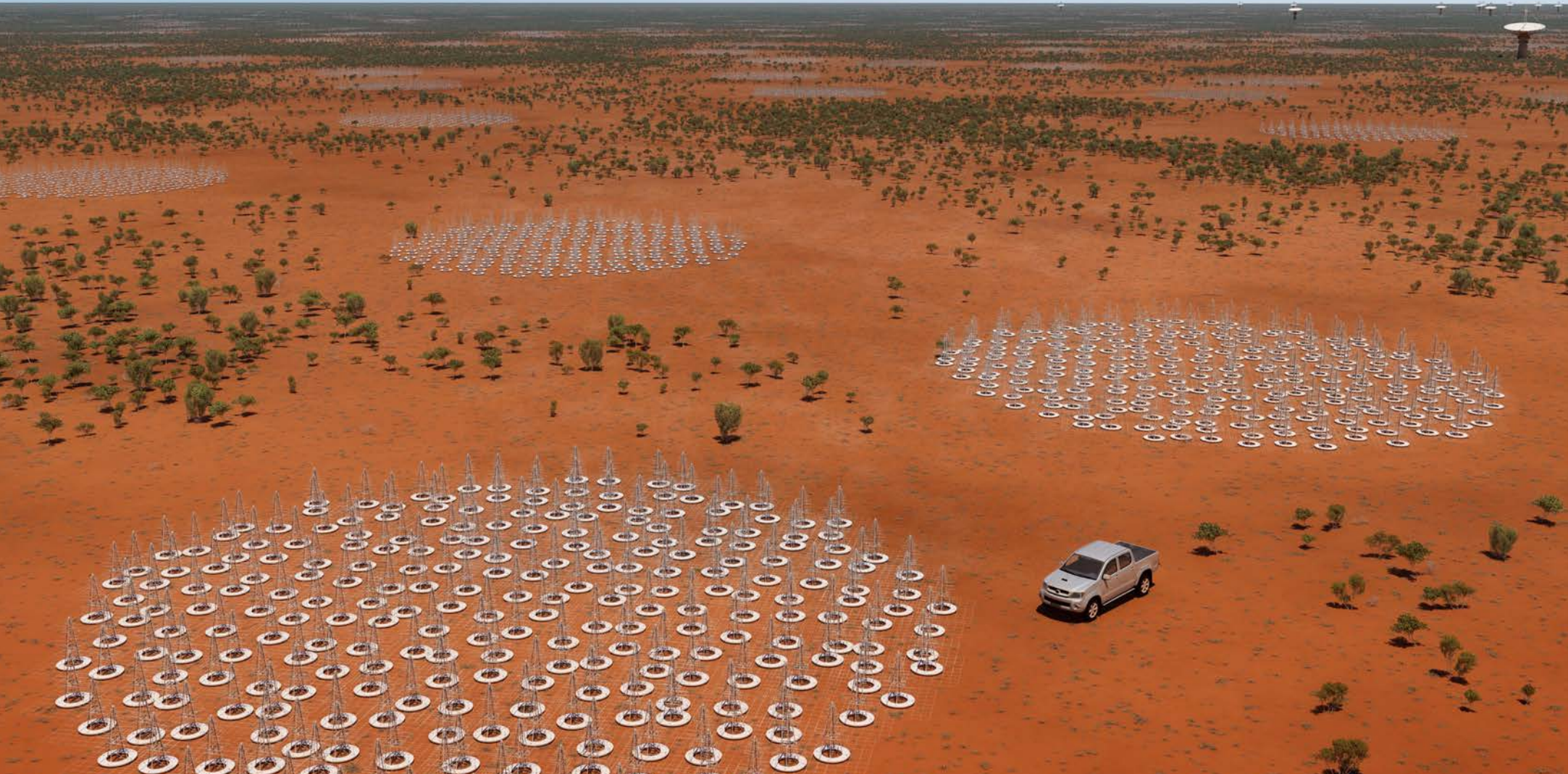
First
galaxies

©nature





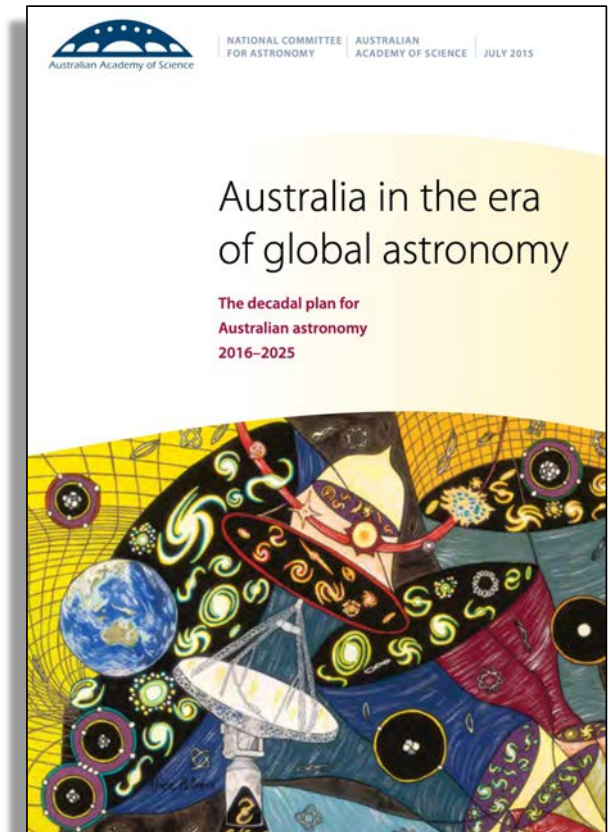
Square Kilometer Array – Low Frequency Facility
Western Australia (artist impression)

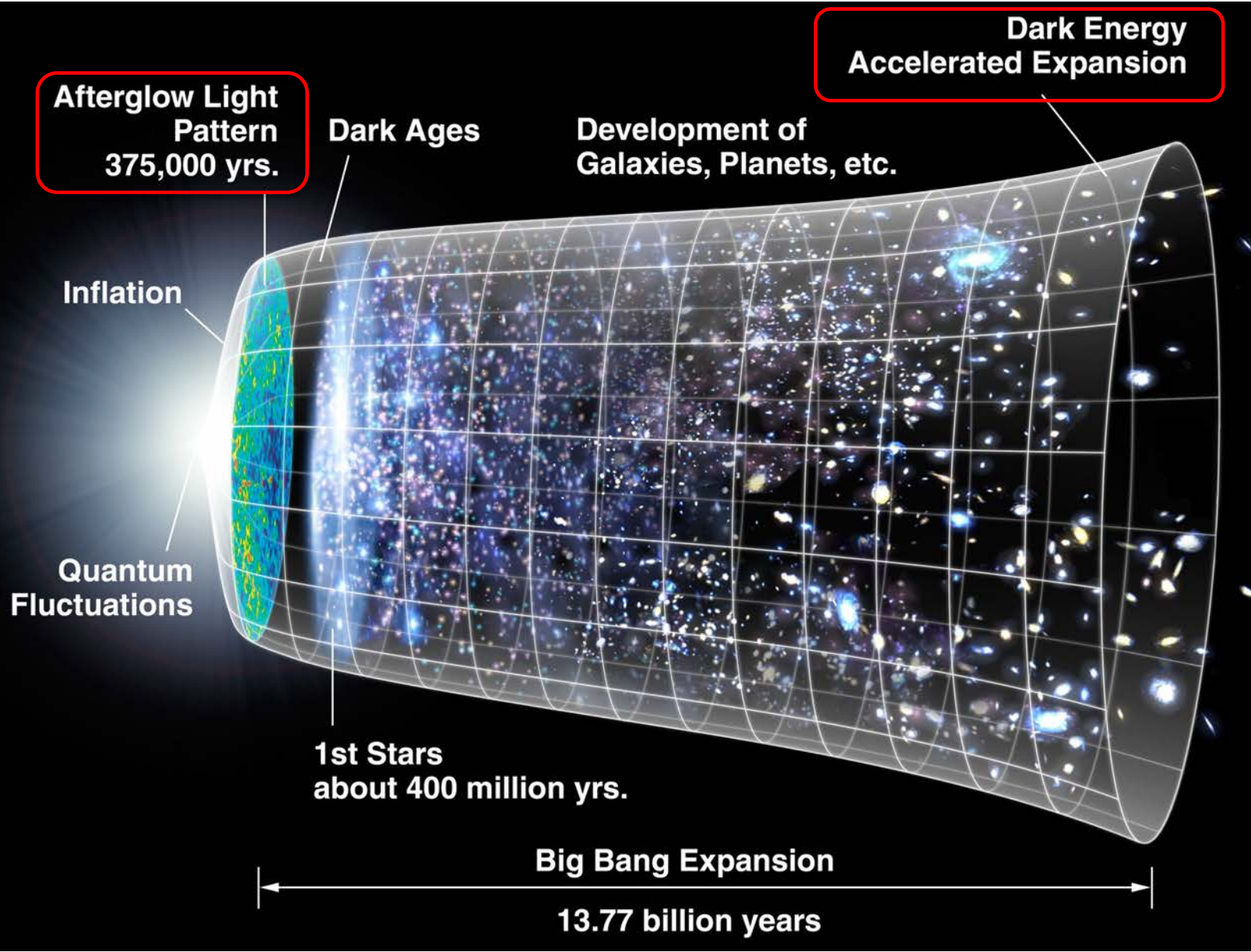


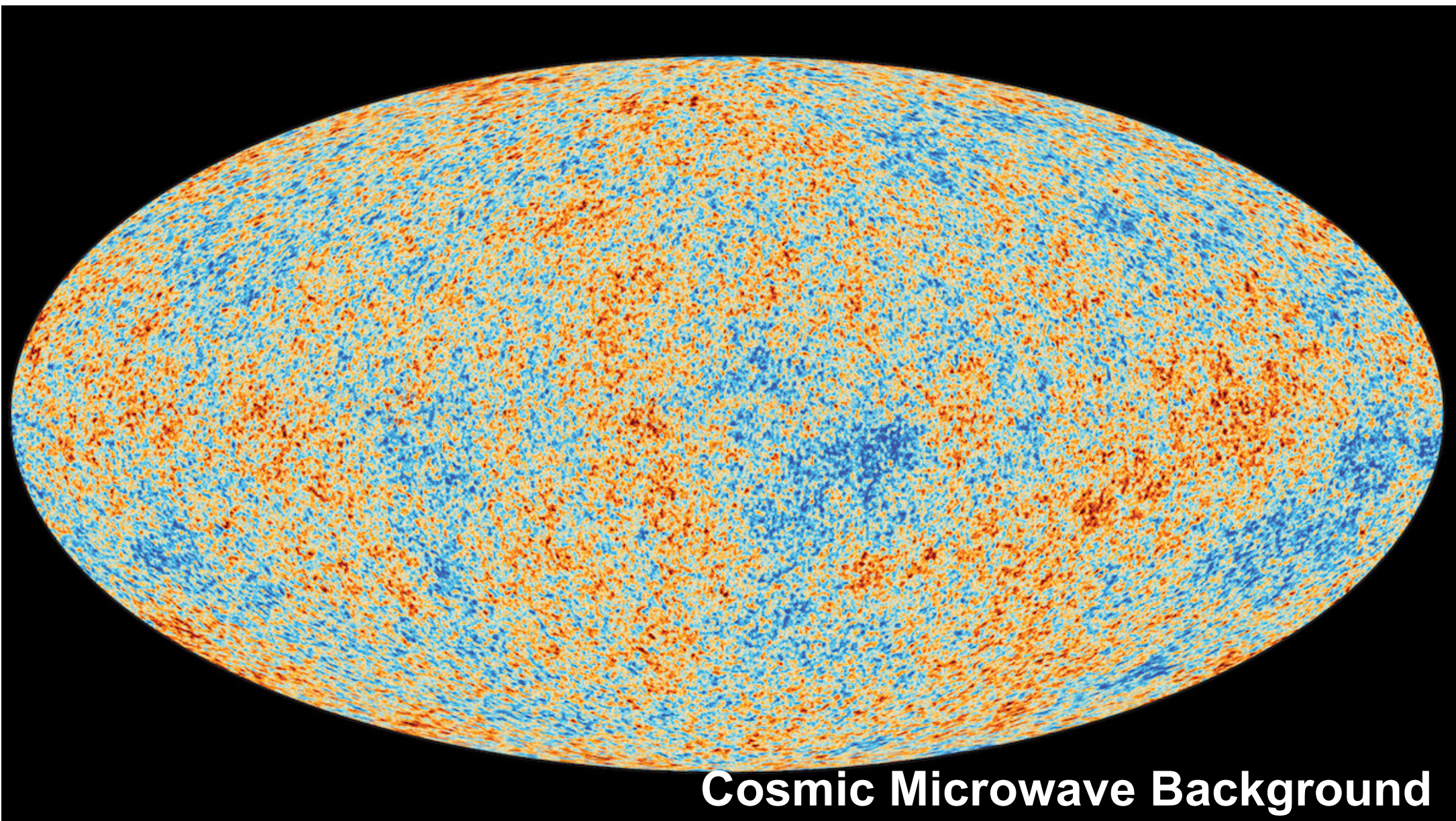
Driving Questions of Modern Astrophysics

The Decadal Plan for Australian Astronomy

- How did the first stars and galaxies transform the Universe?
- **What is the nature of dark matter and dark energy?**
- How do galaxies form and evolve across cosmic time?
- How do stars and planets form?
- How are elements produced by stars and recycled through galaxies?
- What is the nature of matter and gravity at extreme densities?

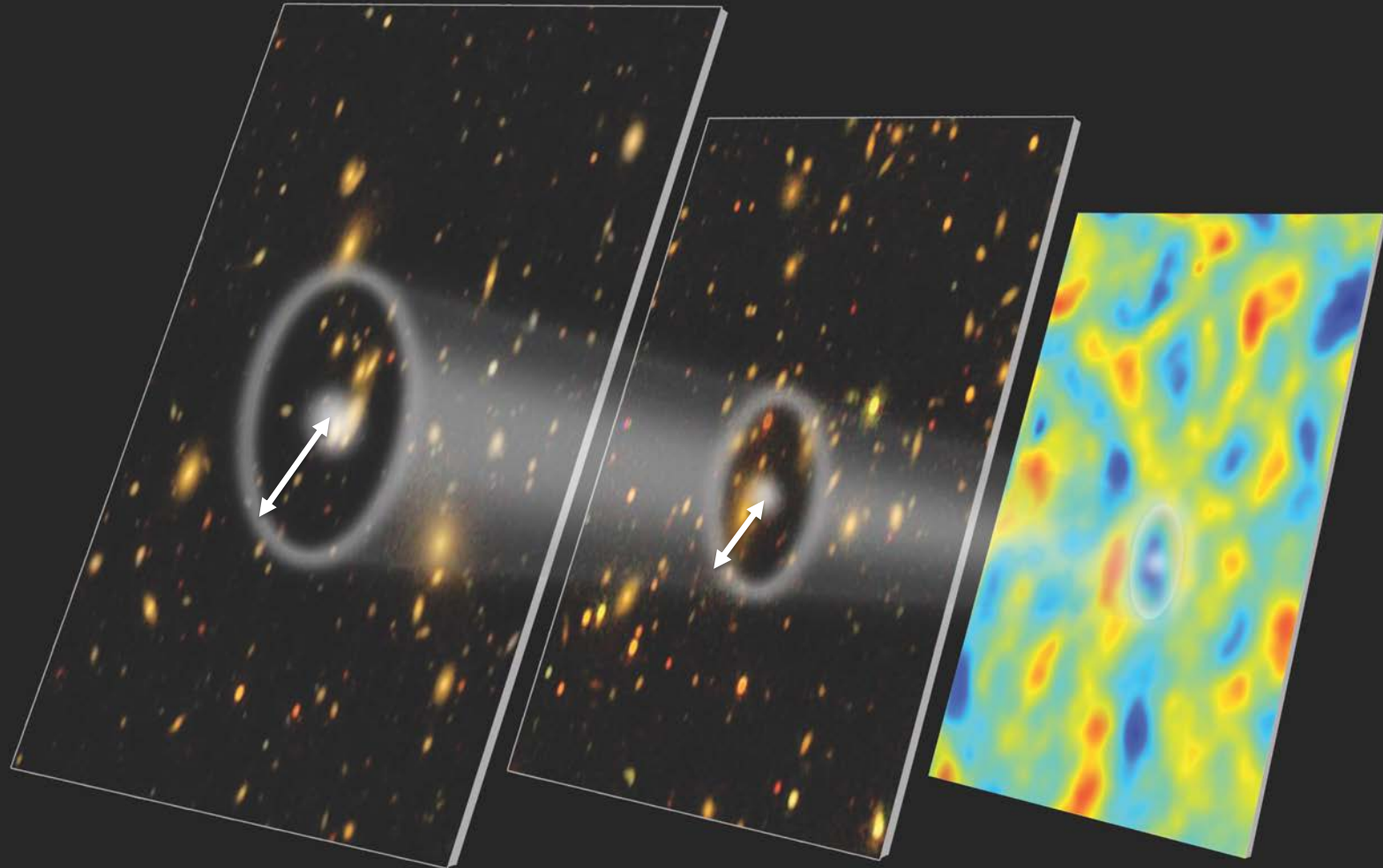






Cosmic Microwave Background

Baryonic Acoustic Oscillations



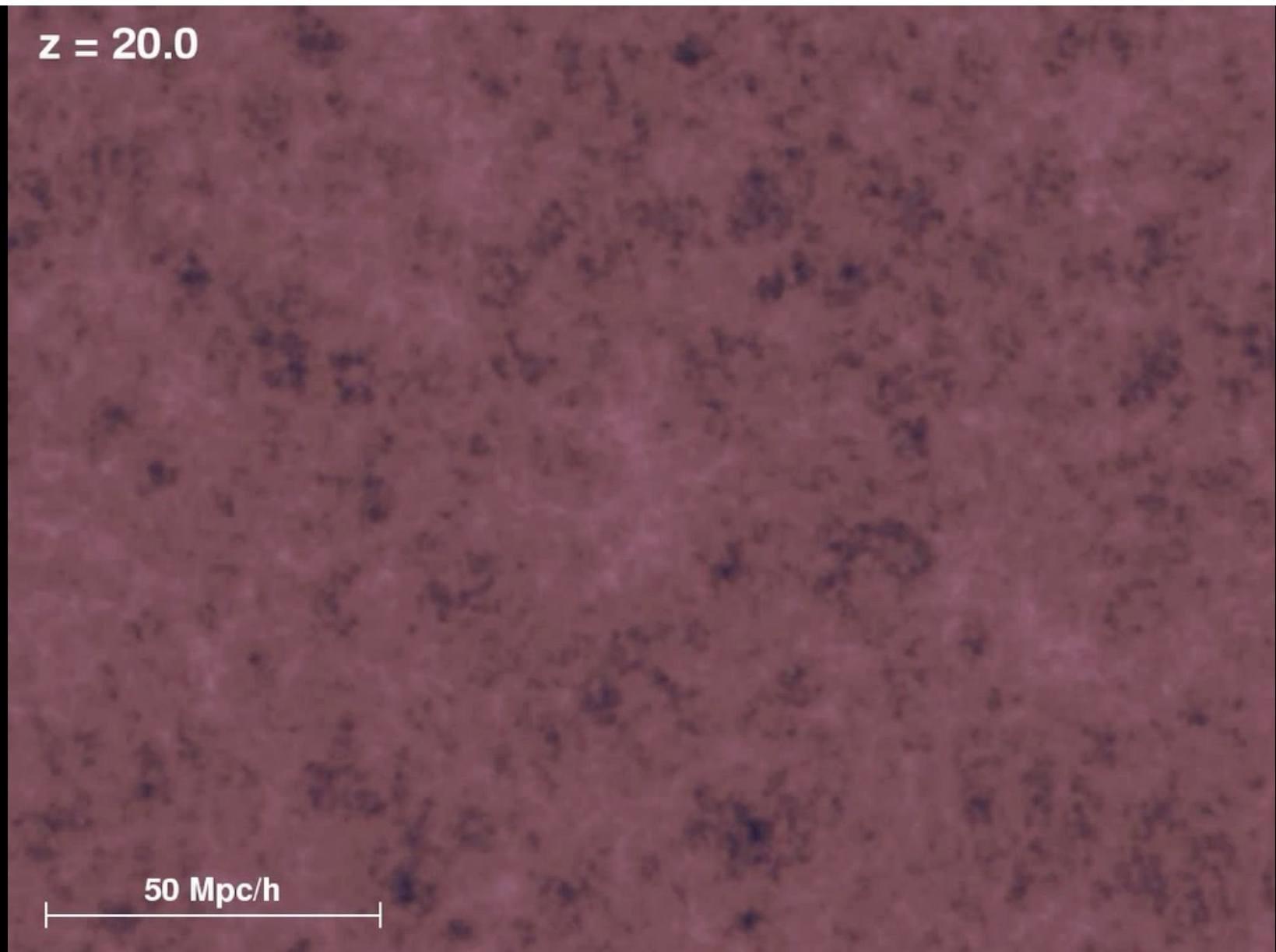
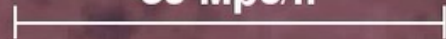
Galaxy map 3.8 billion years ago

Galaxy map 5.5 billion years ago

CMB 13.7 billion years ago

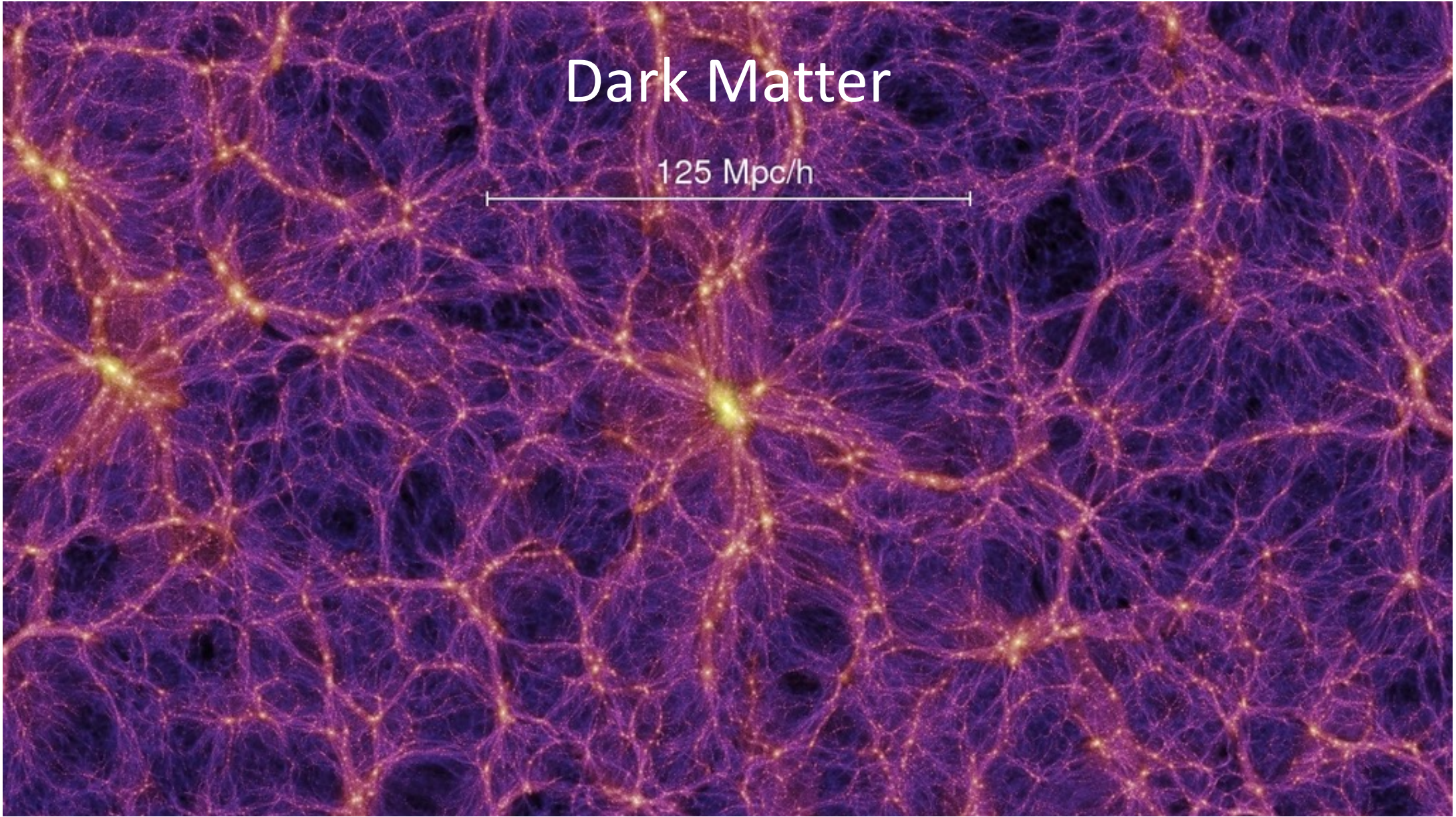
$z = 20.0$

50 Mpc/h

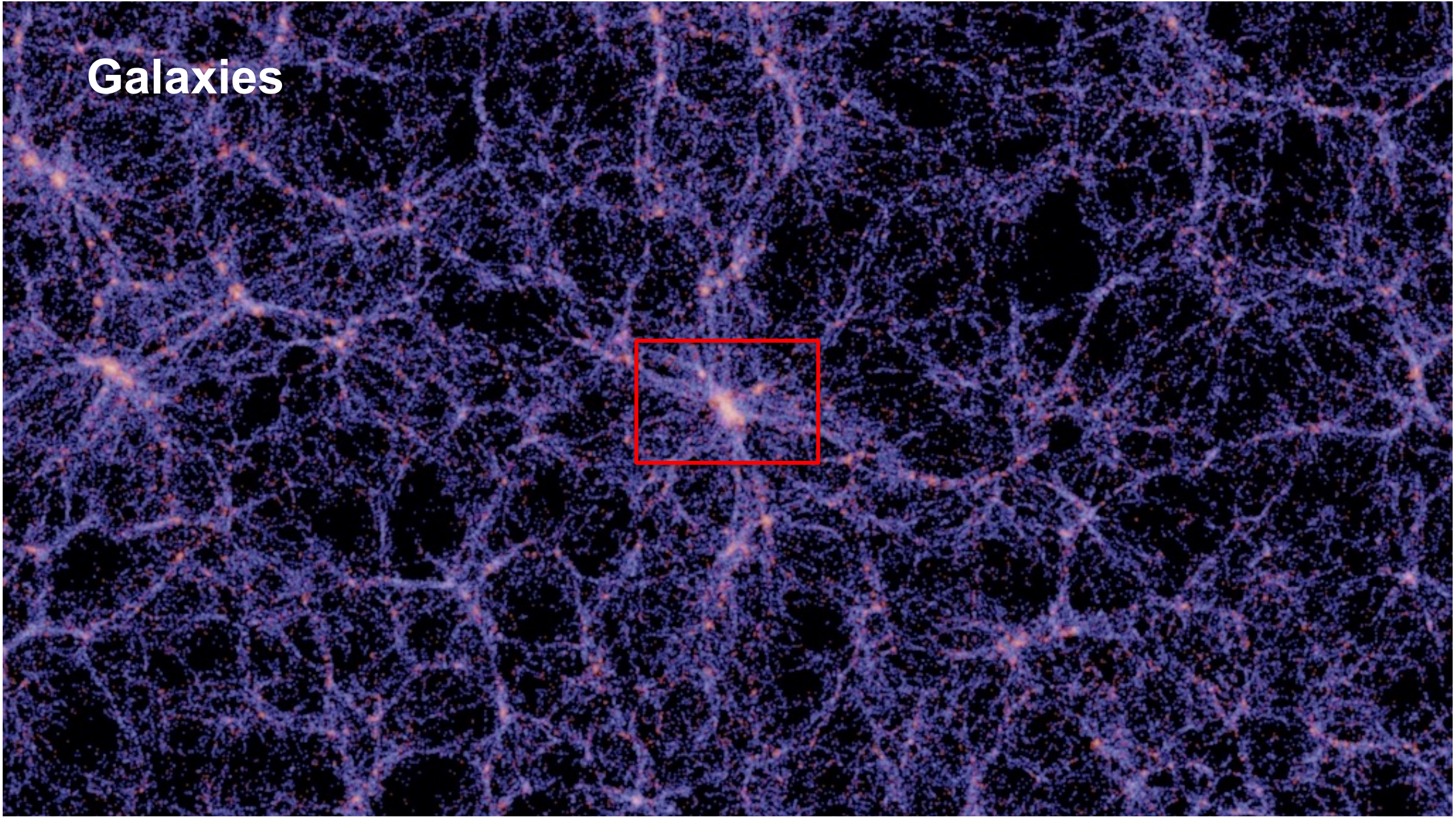


Dark Matter

125 Mpc/h



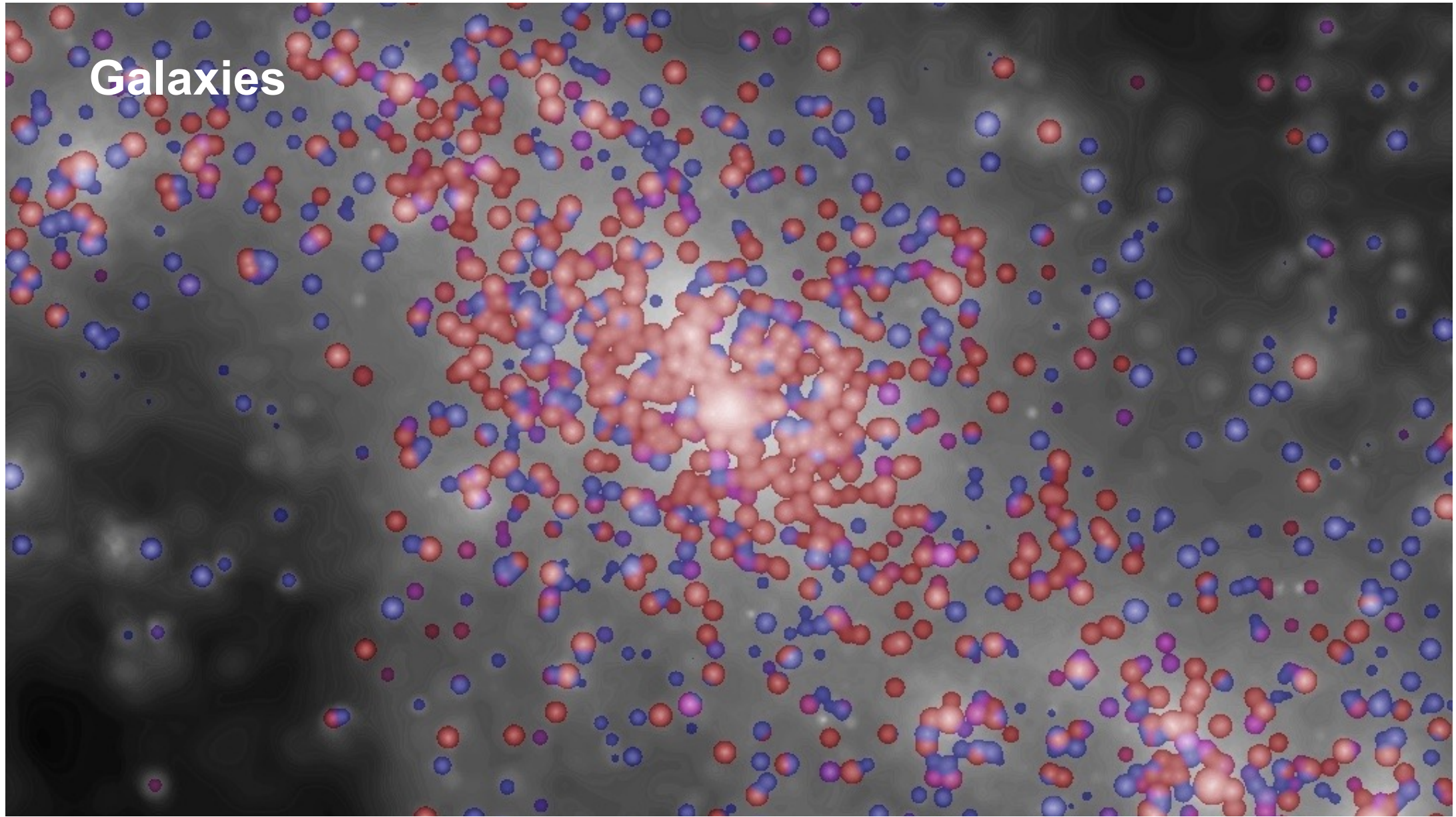
Galaxies

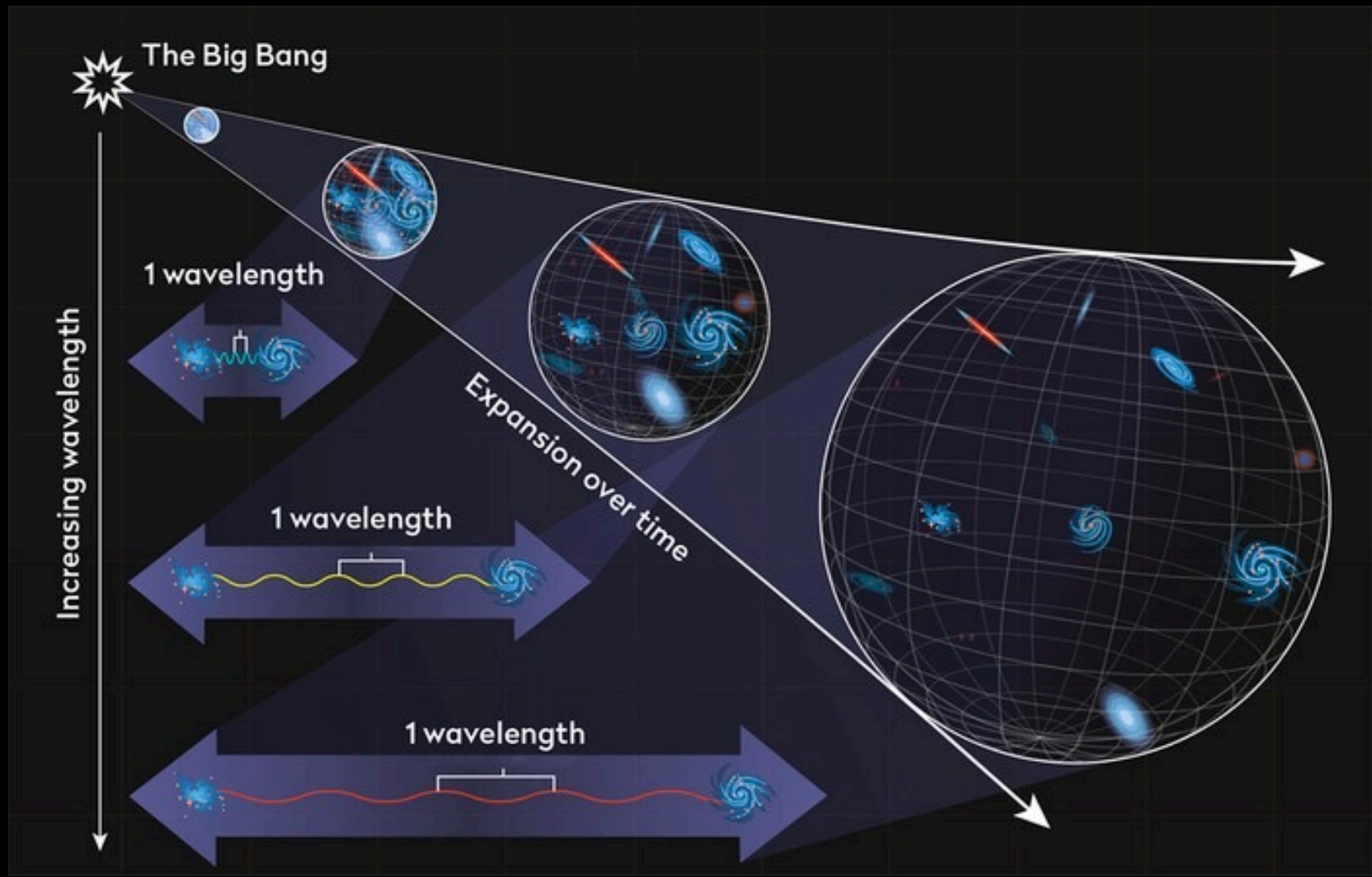


Dark Matter

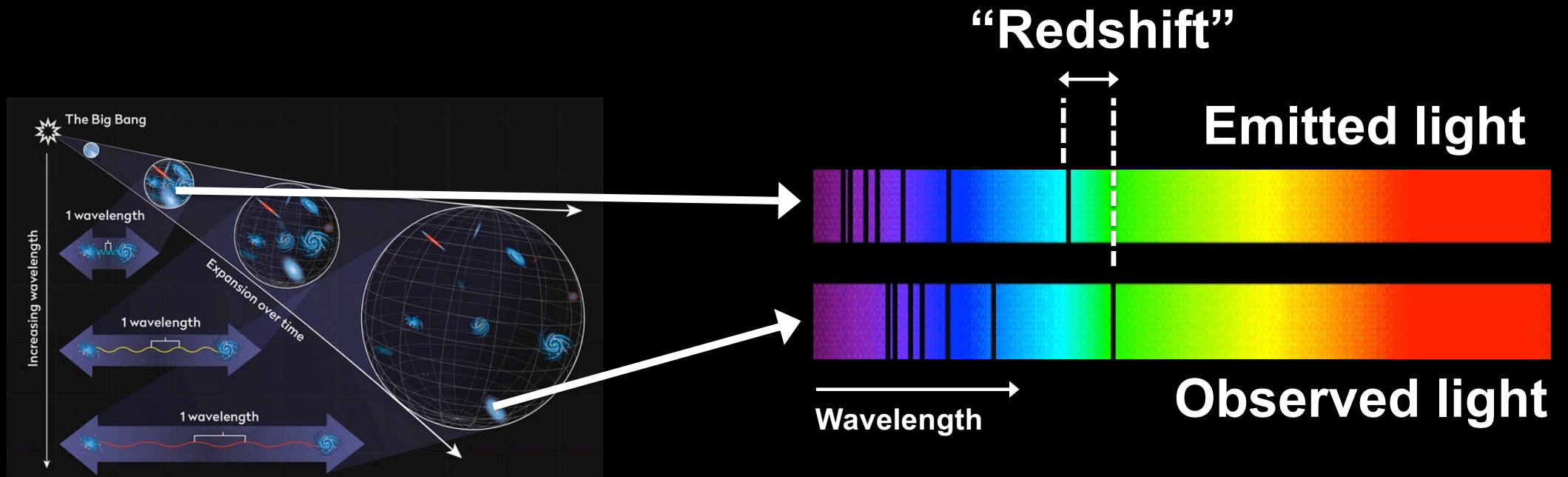


Galaxies



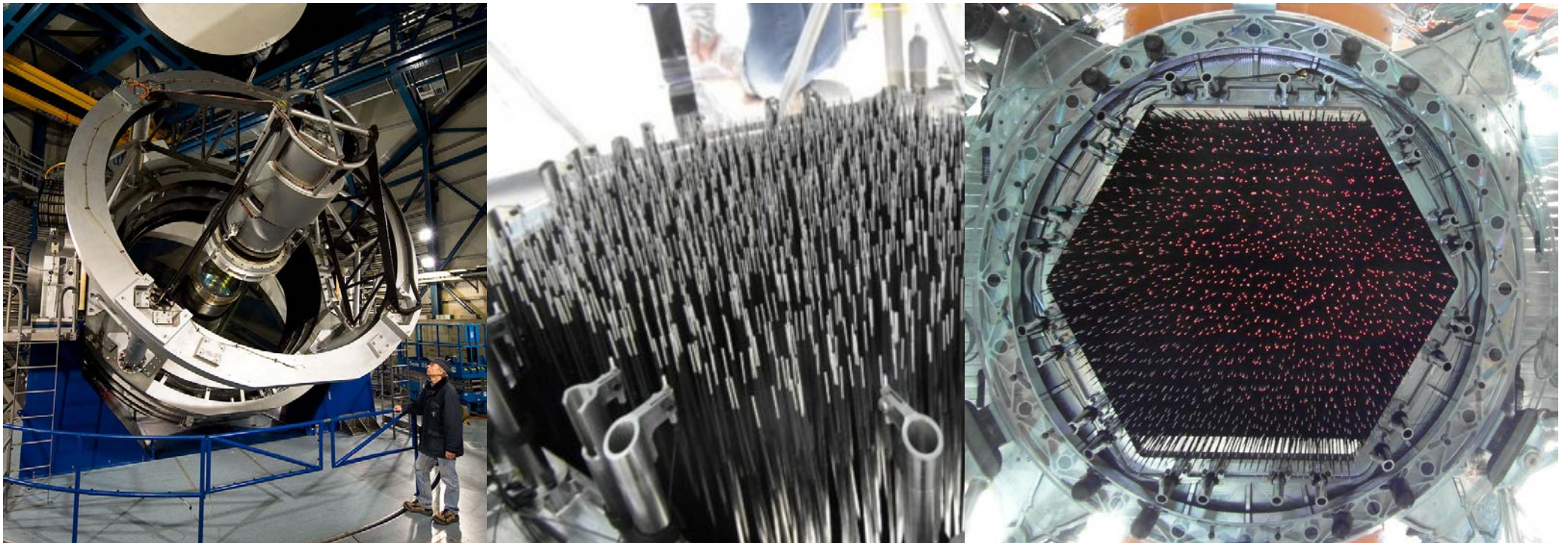


Use “redshift” of galaxy light to measure distance scale with time



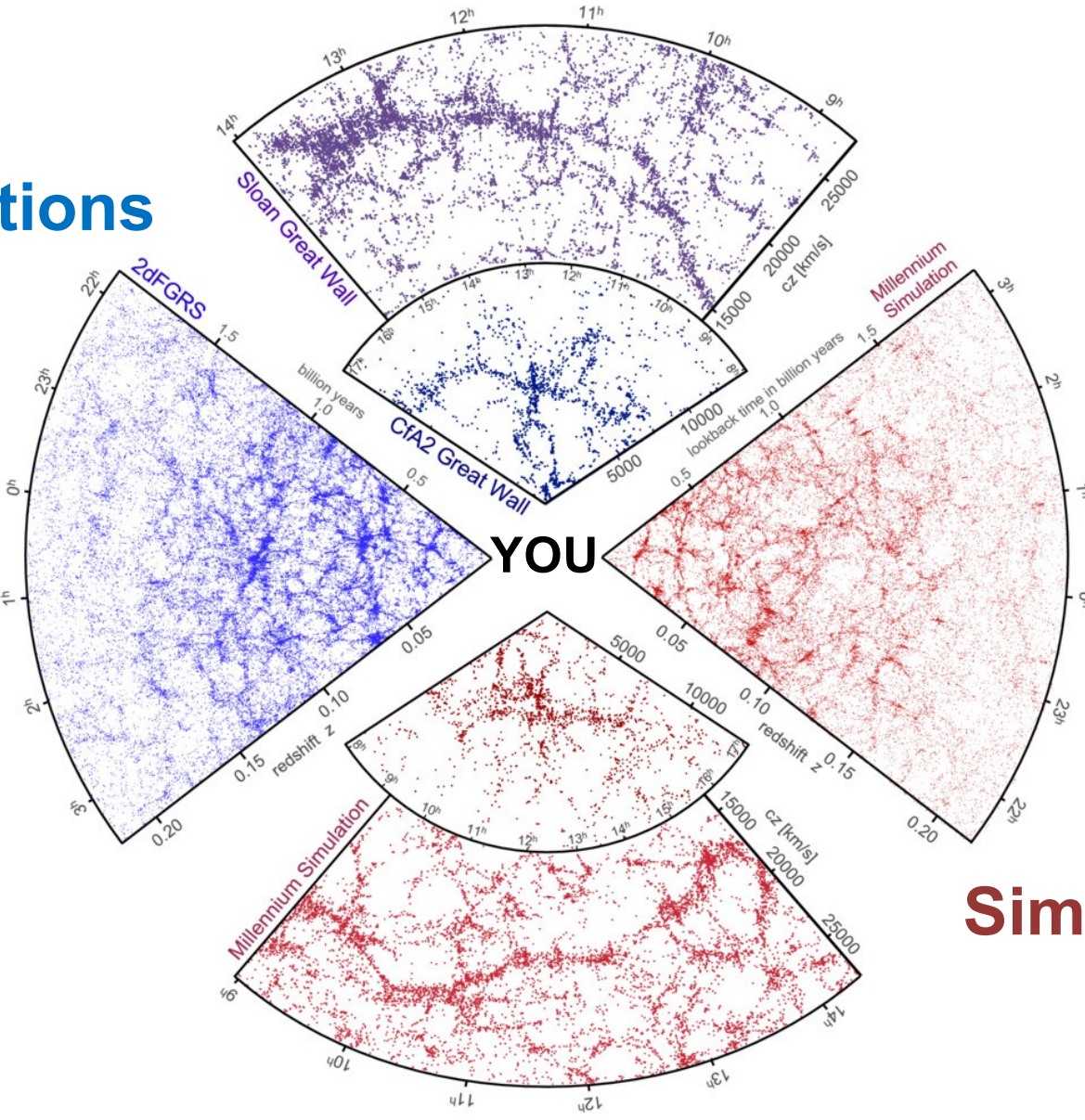
Use "redshift" of galaxy light to measure distance scale with time

4MOST – Multi-Object Spectroscopic Telescope



- Dedicated 4m telescope in Chile
- Captures 2400 spectra simultaneously with configurable field of fibers
- Fiber positioner designed and built by AAO

Observations

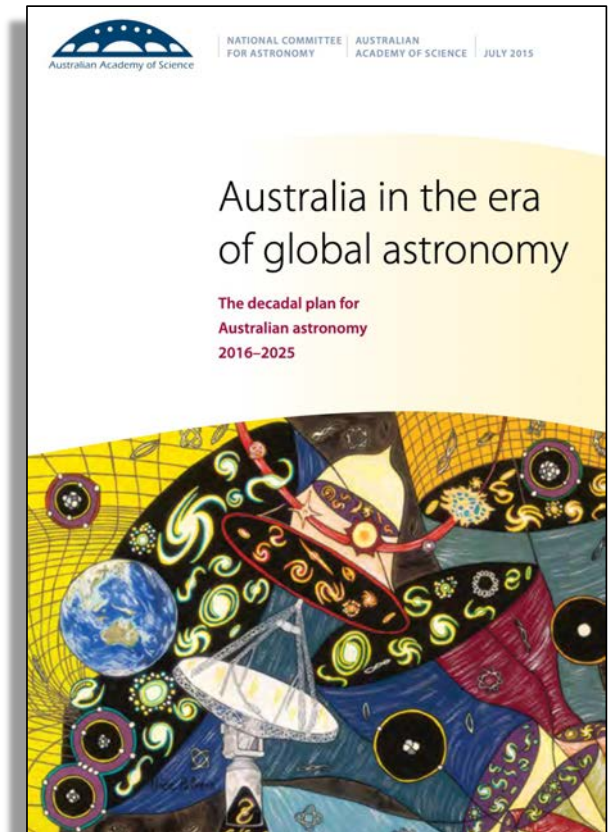


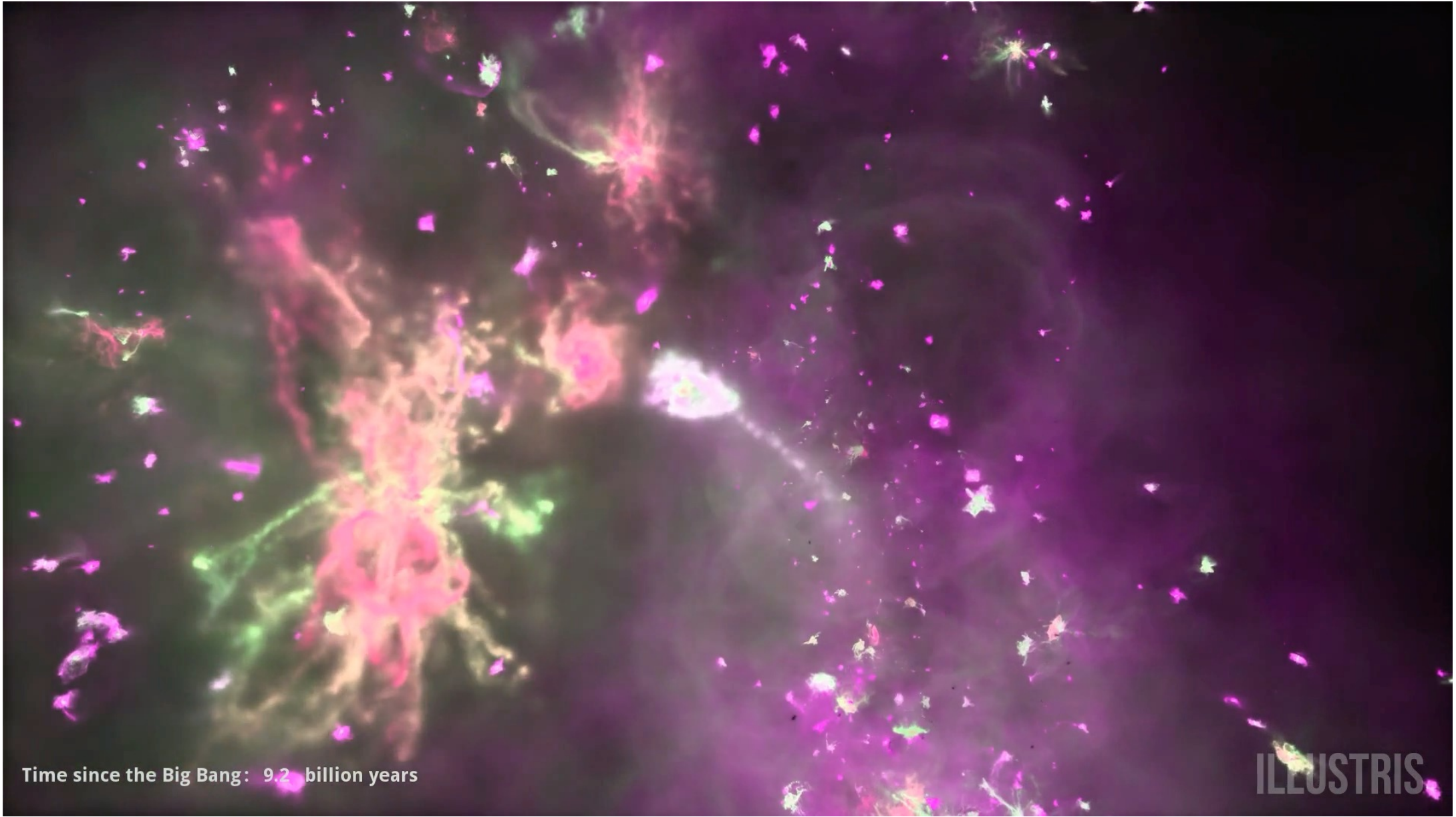
Simulations

Driving Questions of Modern Astrophysics

The Decadal Plan for Australian Astronomy

- How did the first stars and galaxies transform the Universe?
- What is the nature of dark matter and dark energy?
- How do galaxies form and evolve across cosmic time?
- How do stars and planets form?
- How are elements produced by stars and recycled through galaxies?
- What is the nature of matter and gravity at extreme densities?

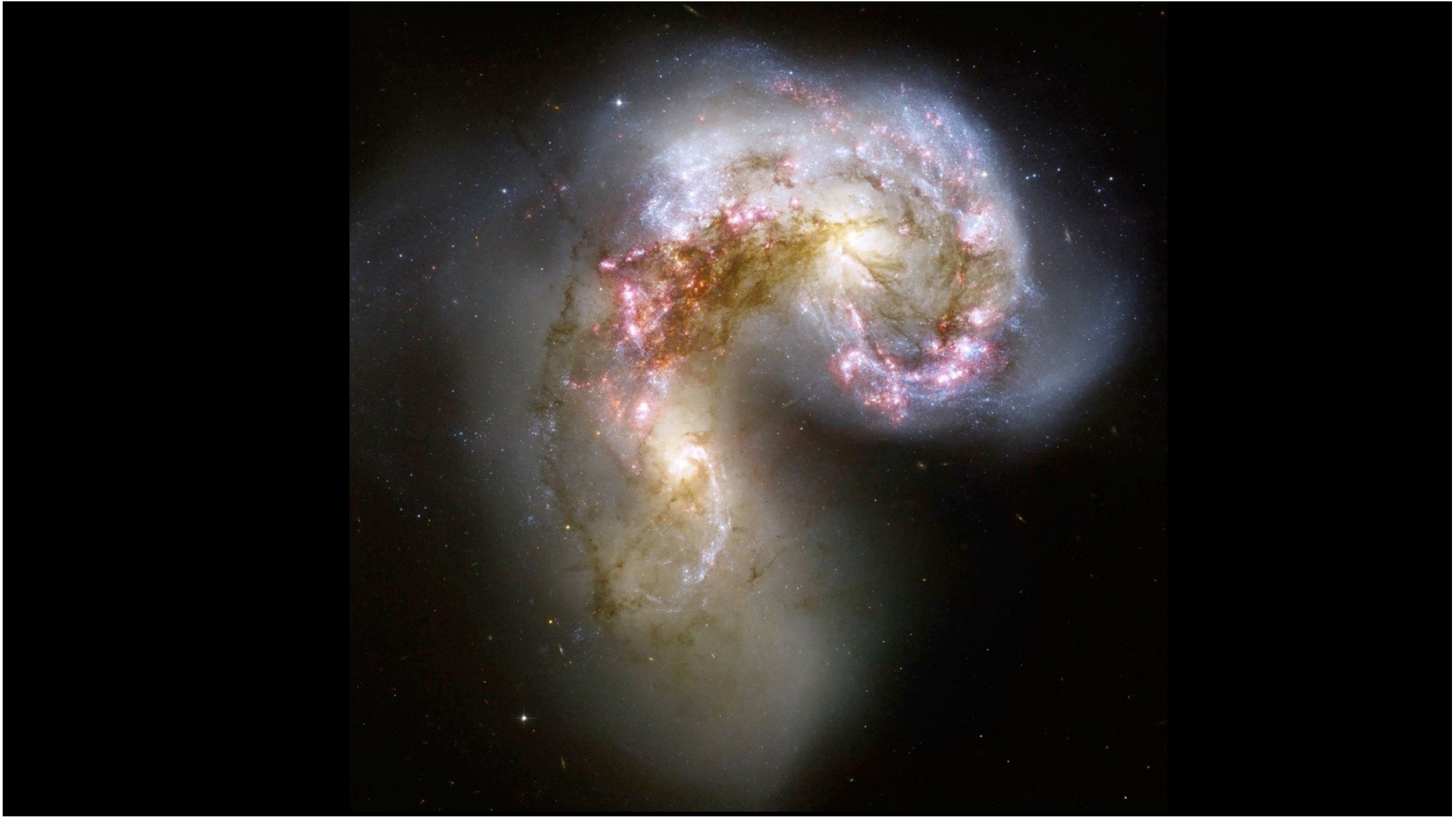




Time since the Big Bang: 9.2 billion years

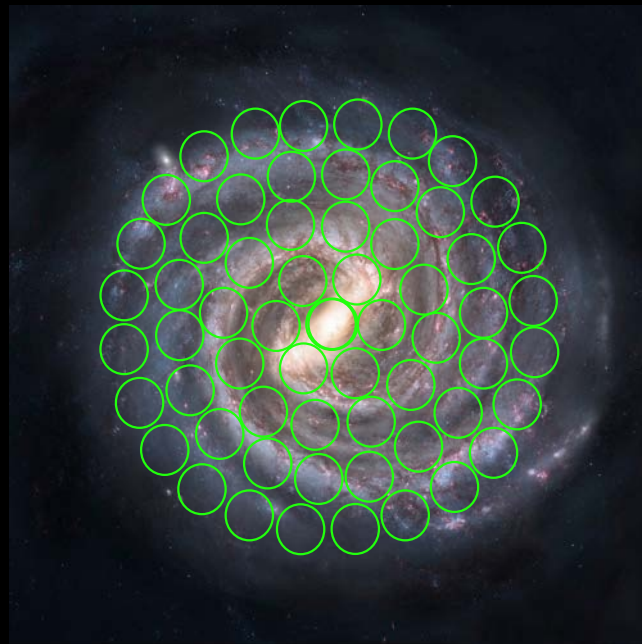
ILLUSTRIS



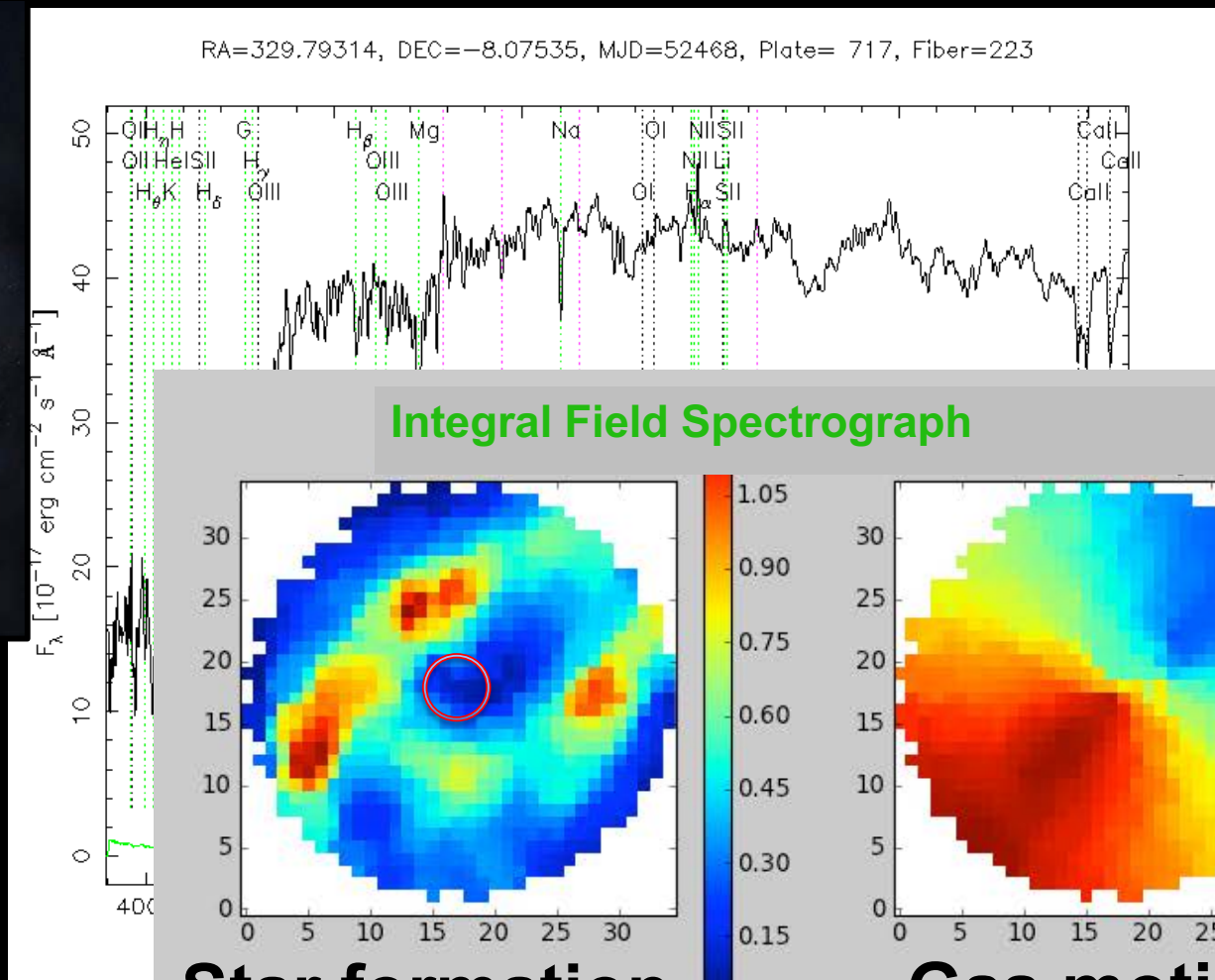




Integral Field Spectroscopy

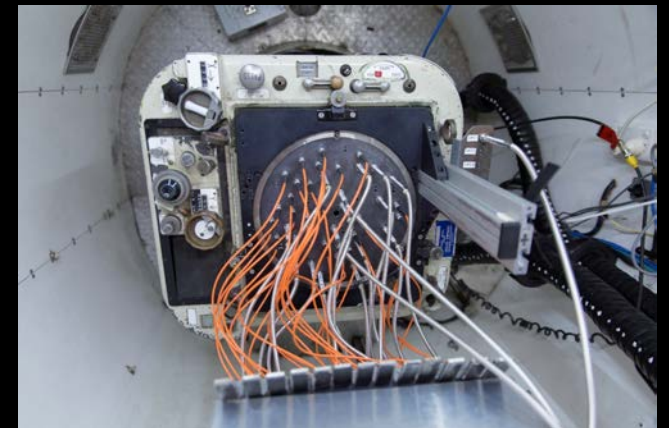
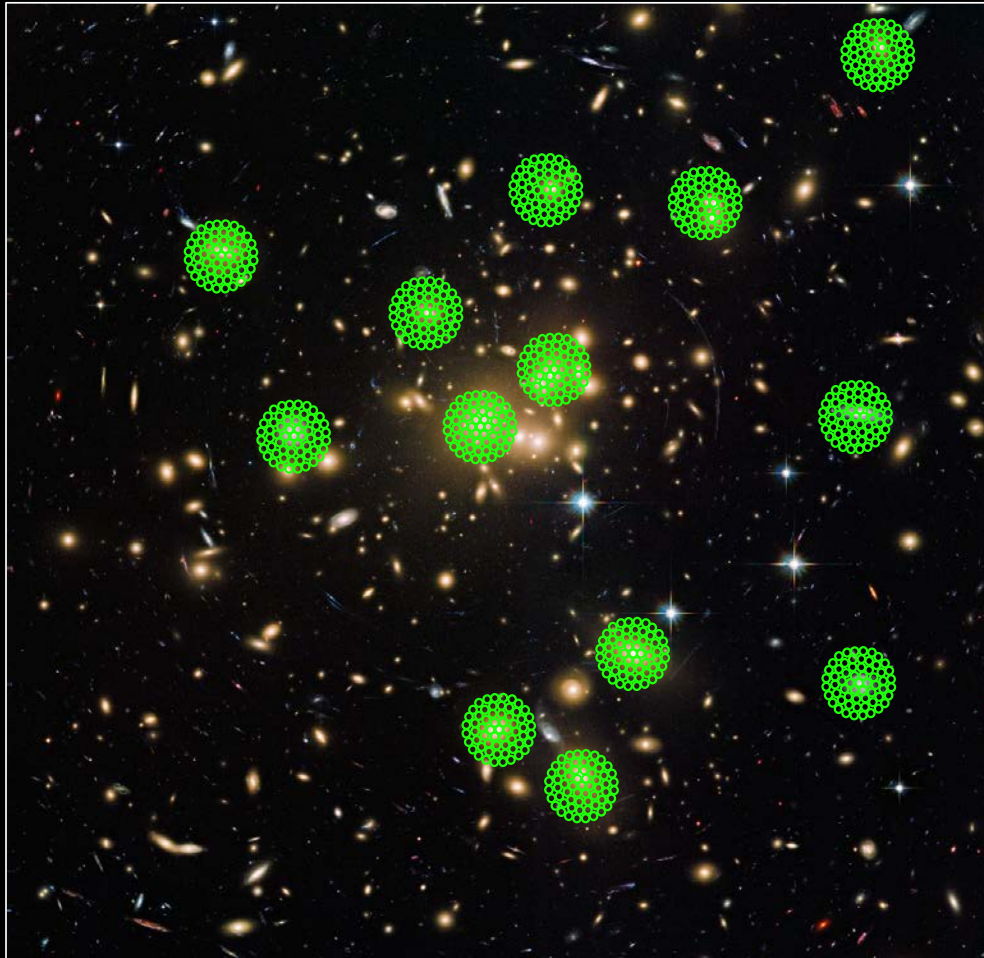


SDSS: image and spectrum



Multi - Integral Field Spectroscopy

Anglo-Australian Telescope, NSW



SAMI multi-IFU fiber feed

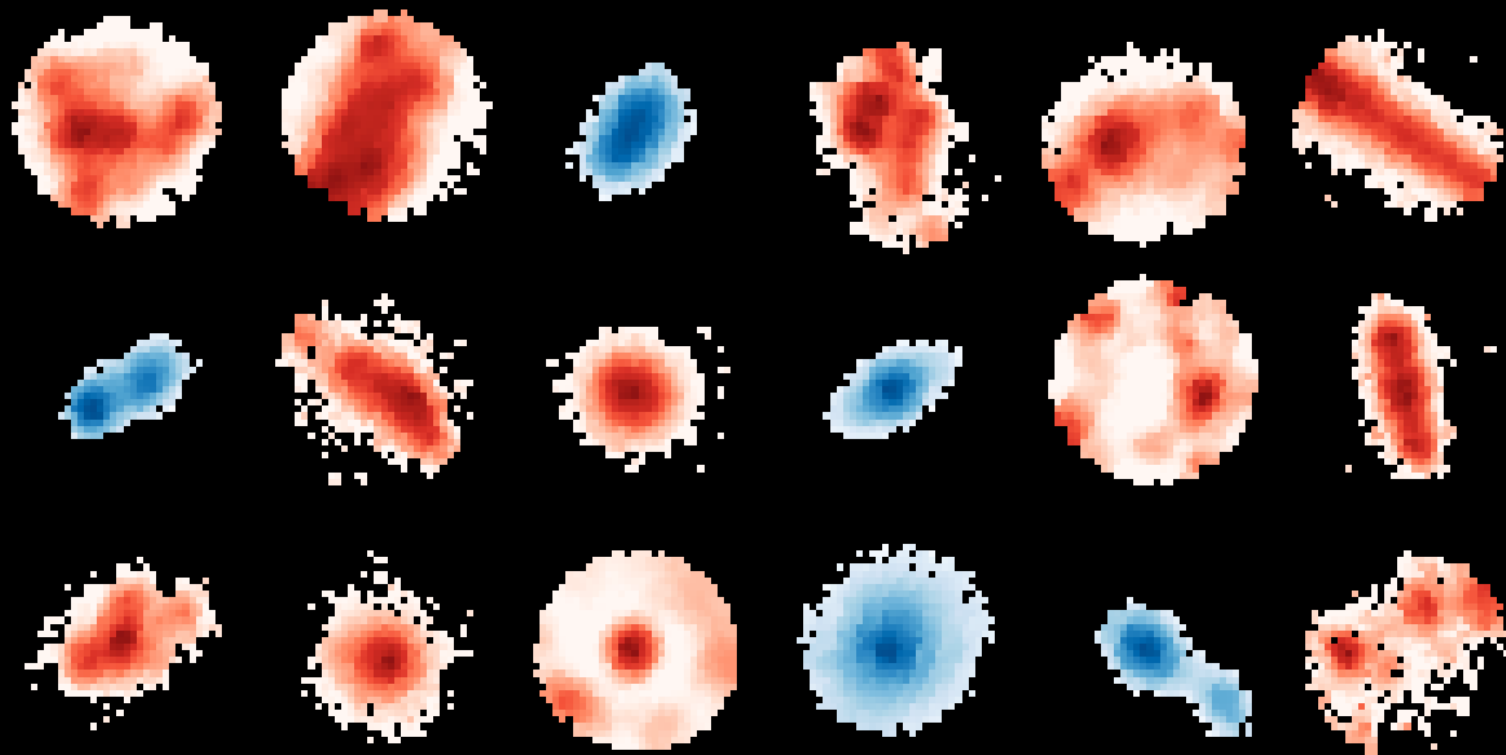
Galaxy Image



<http://sami-survey.org/edr>

SDSS

Stars and Gas

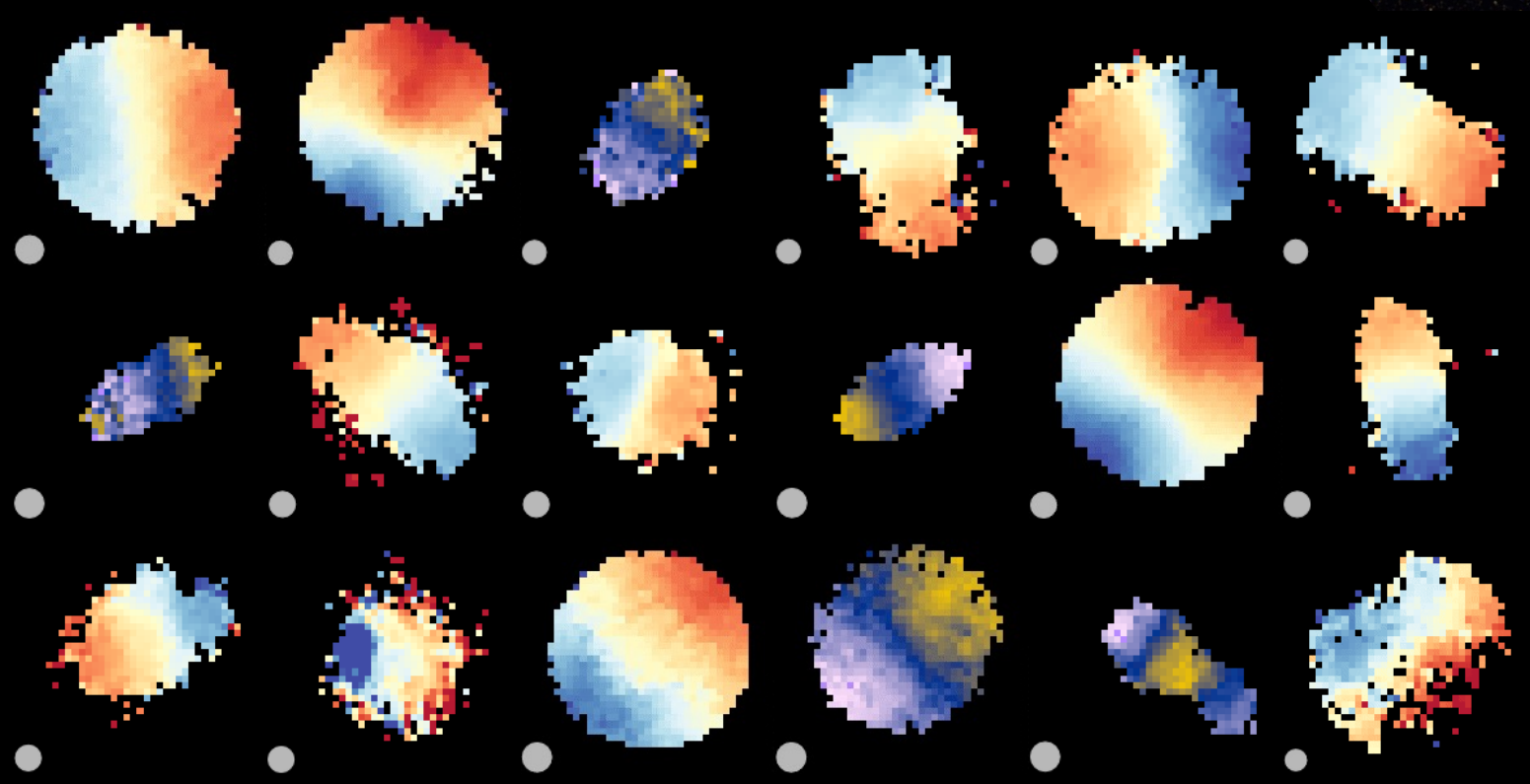
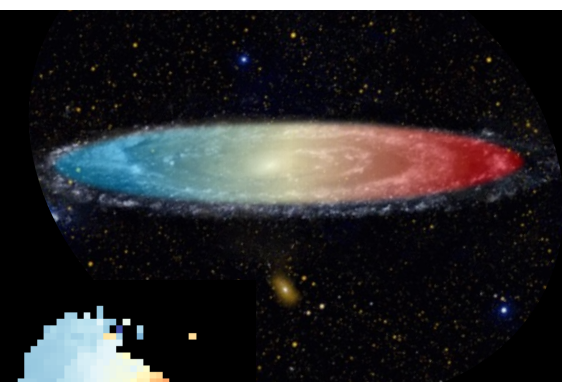


<http://sami-survey.org>

Stellar Flux

H α Flux

Orbital Motion of Stars and Gas



<http://sami-survey.org>

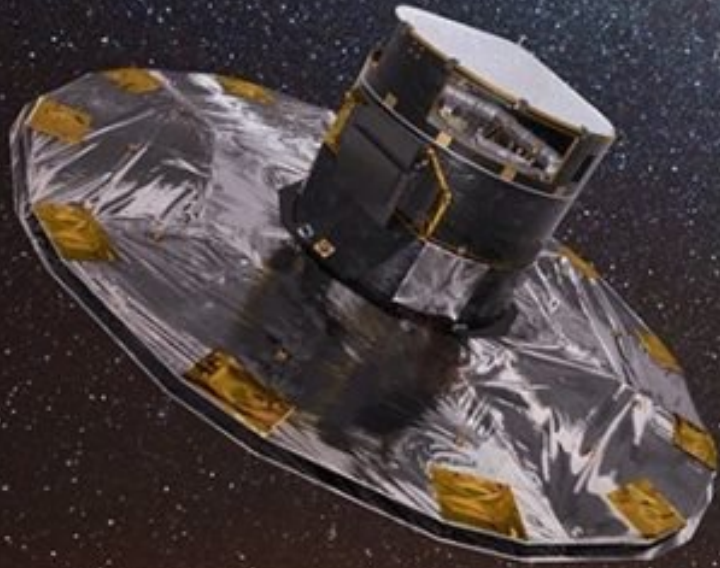
Stellar velocity

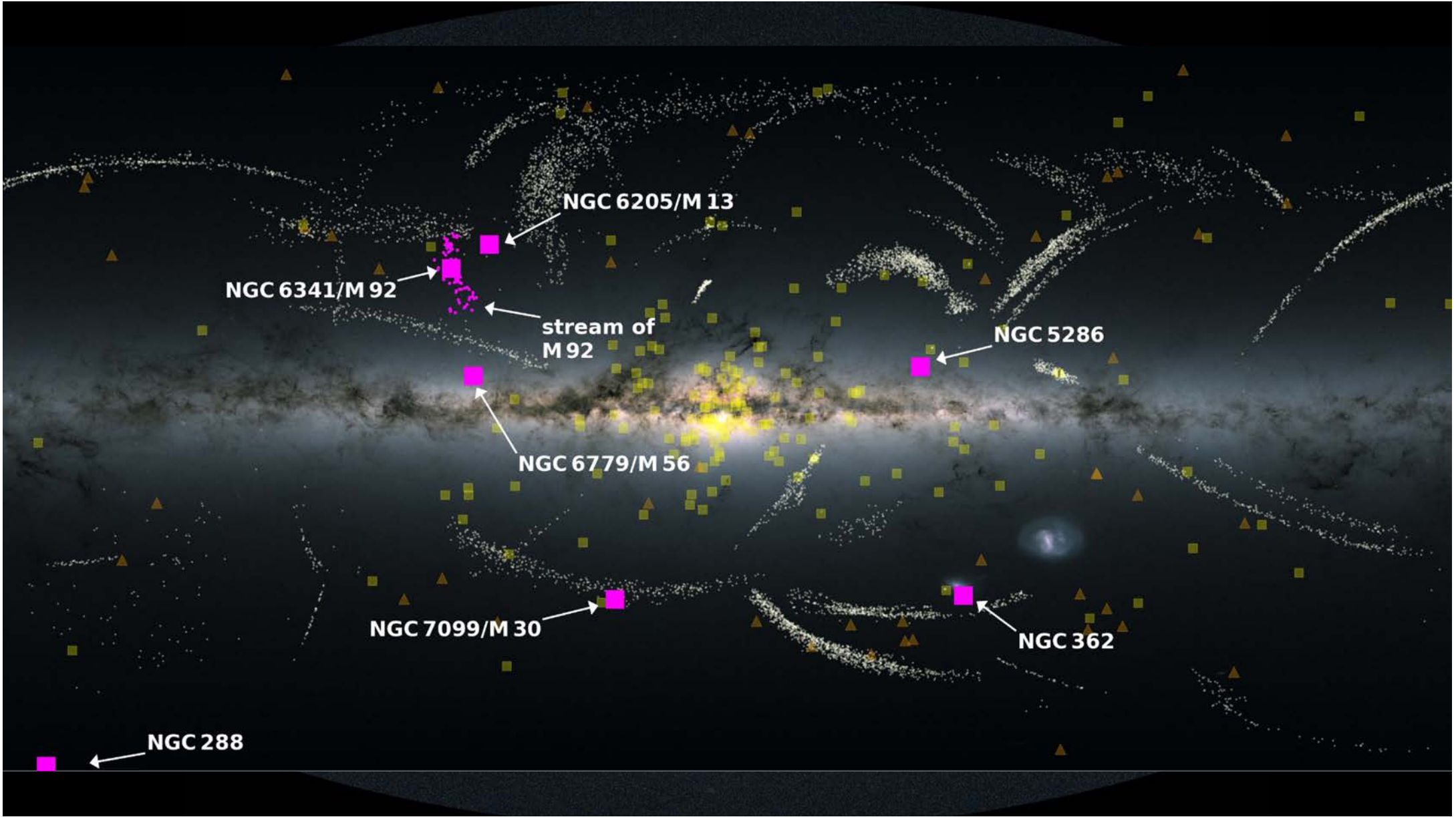
Gas velocity

$z=0.71$



GAIA satellite – launched 2013
Measure distances to 1 billion stars





NGC 6341/M 92

NGC 6205/M 13

stream of
M 92

NGC 5286

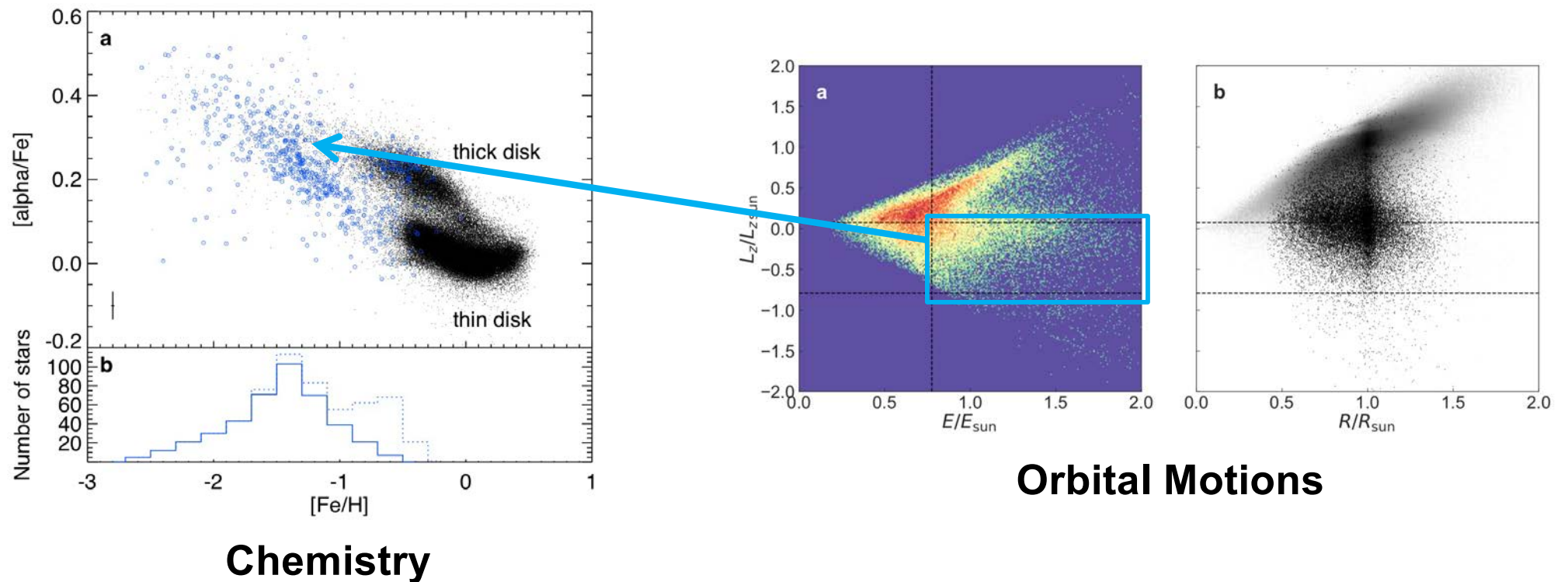
NGC 6779/M 56

NGC 7099/M 30

NGC 362

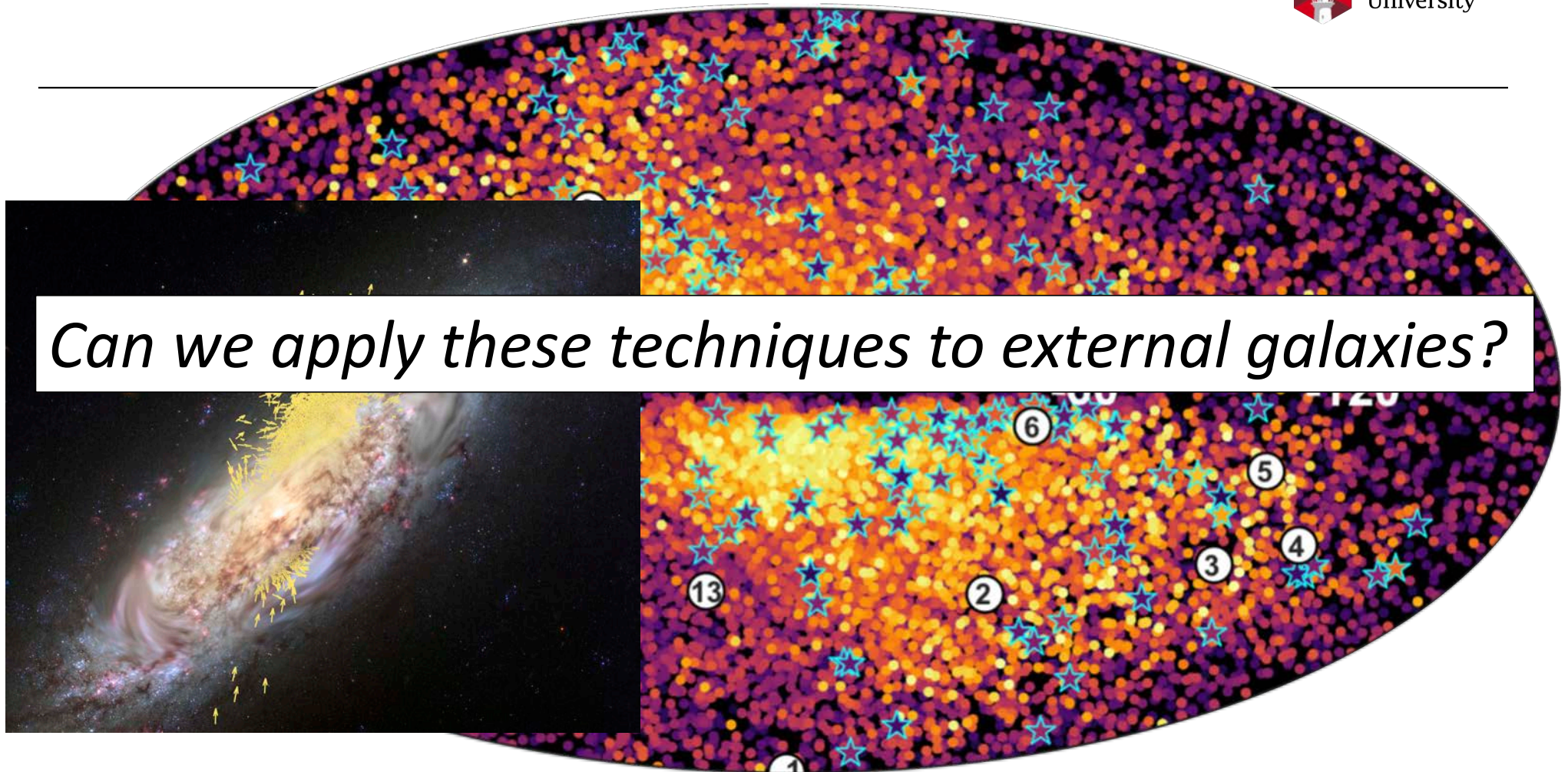
NGC 288

Finding Galactic Substructure in Multi-Label Data



Clustering in parameter space can be used to reveal the history of how our galaxy formed

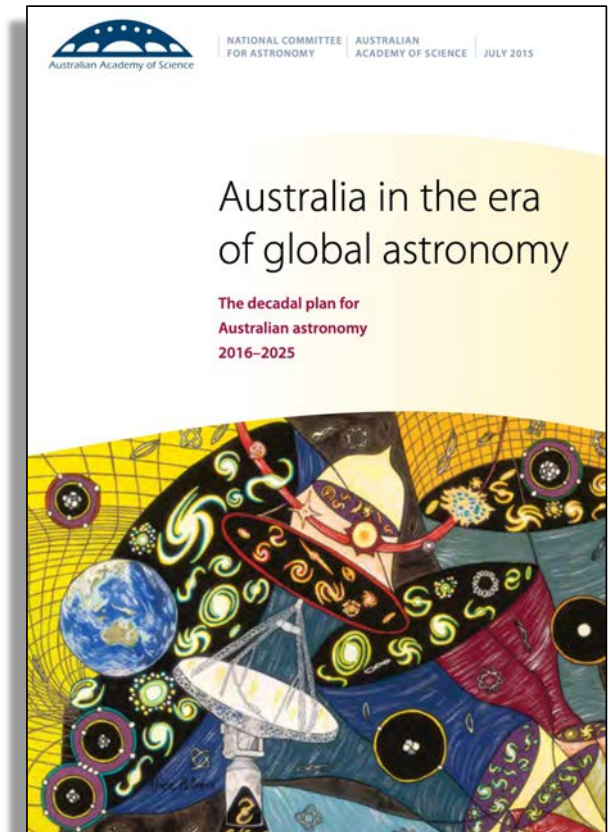
Finding Galactic Substructure in Multi-Label Data



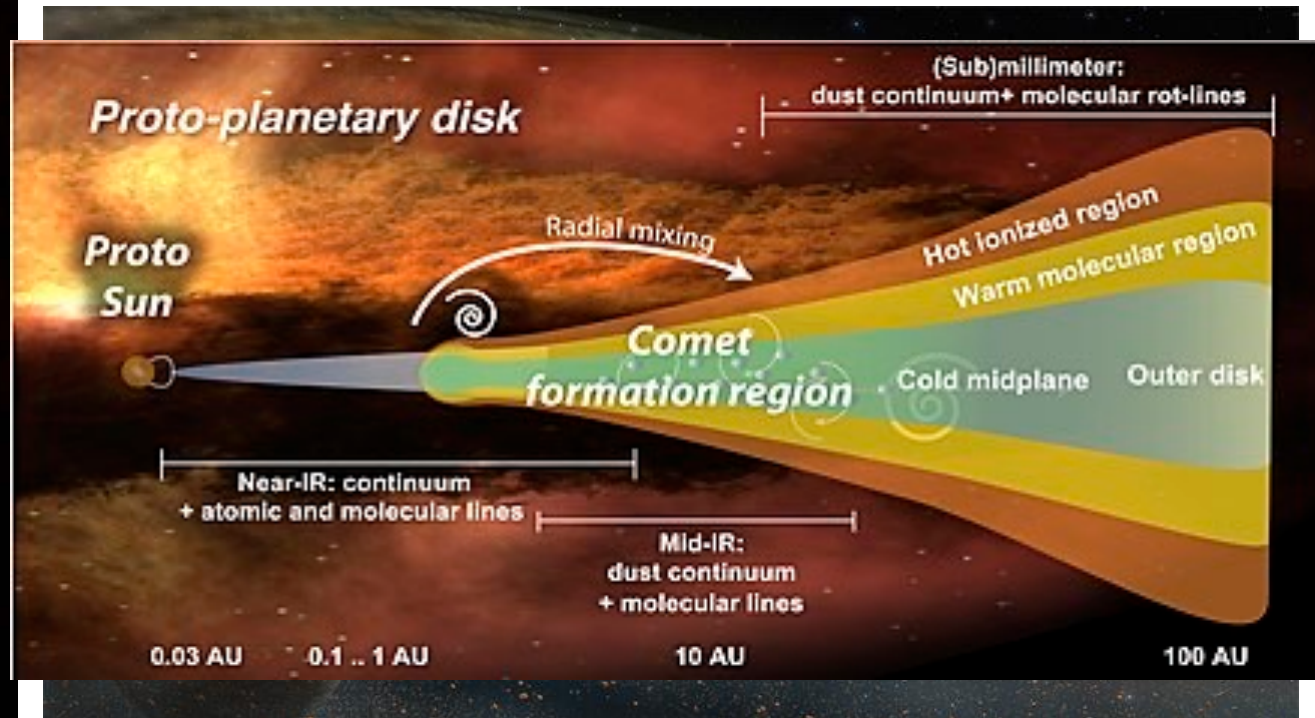
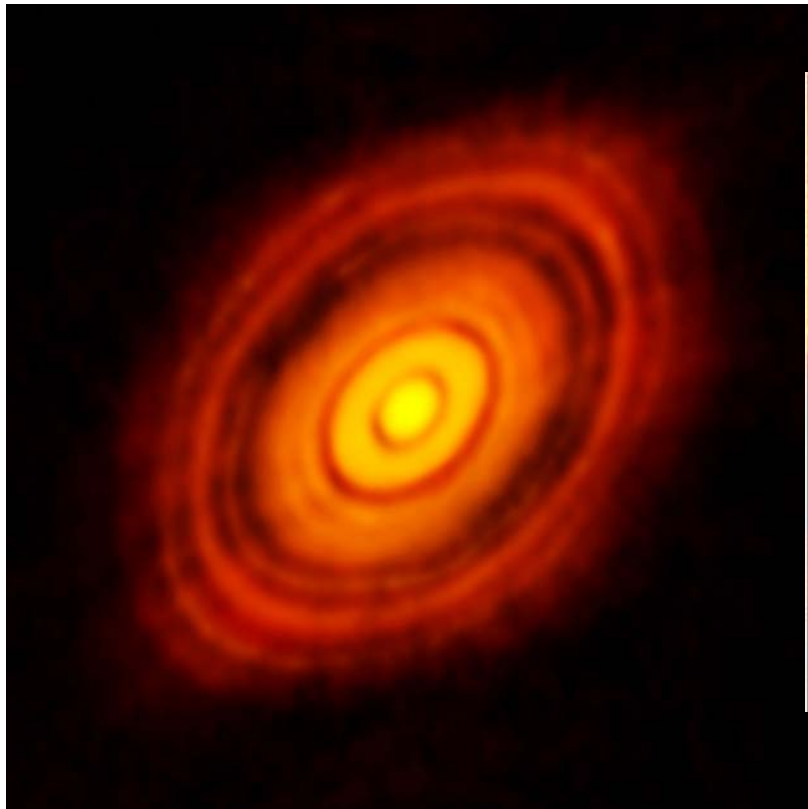
Driving Questions of Modern Astrophysics

The Decadal Plan for Australian Astronomy

- How did the first stars and galaxies transform the Universe?
- What is the nature of dark matter and dark energy?
- How do galaxies form and evolve across cosmic time?
- **How do stars and planets form?**
- How are elements produced by stars and recycled through galaxies?
- What is the nature of matter and gravity at extreme densities?



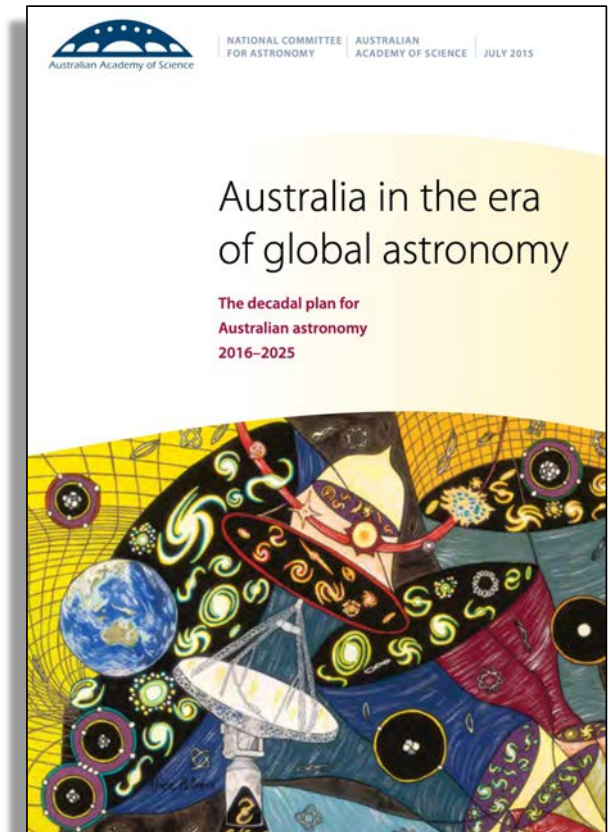
Forming Stars and Planets



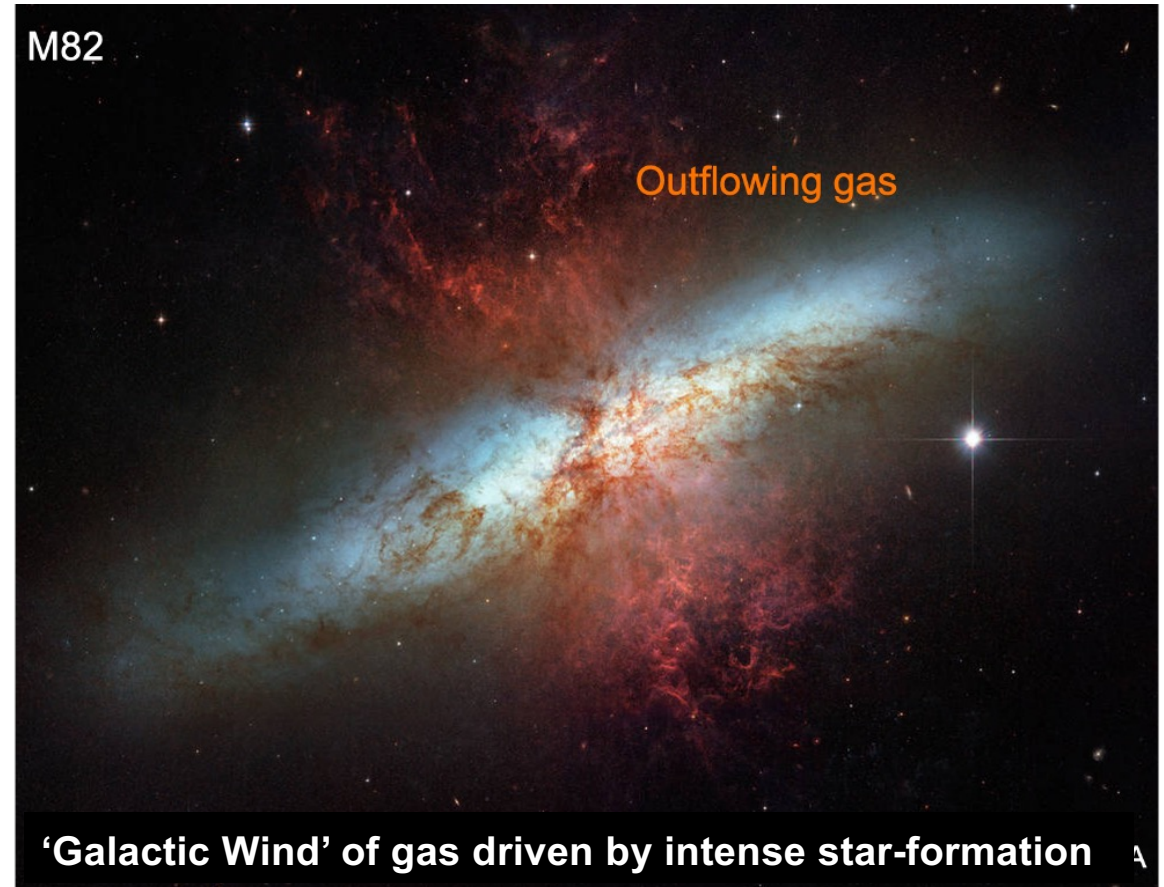
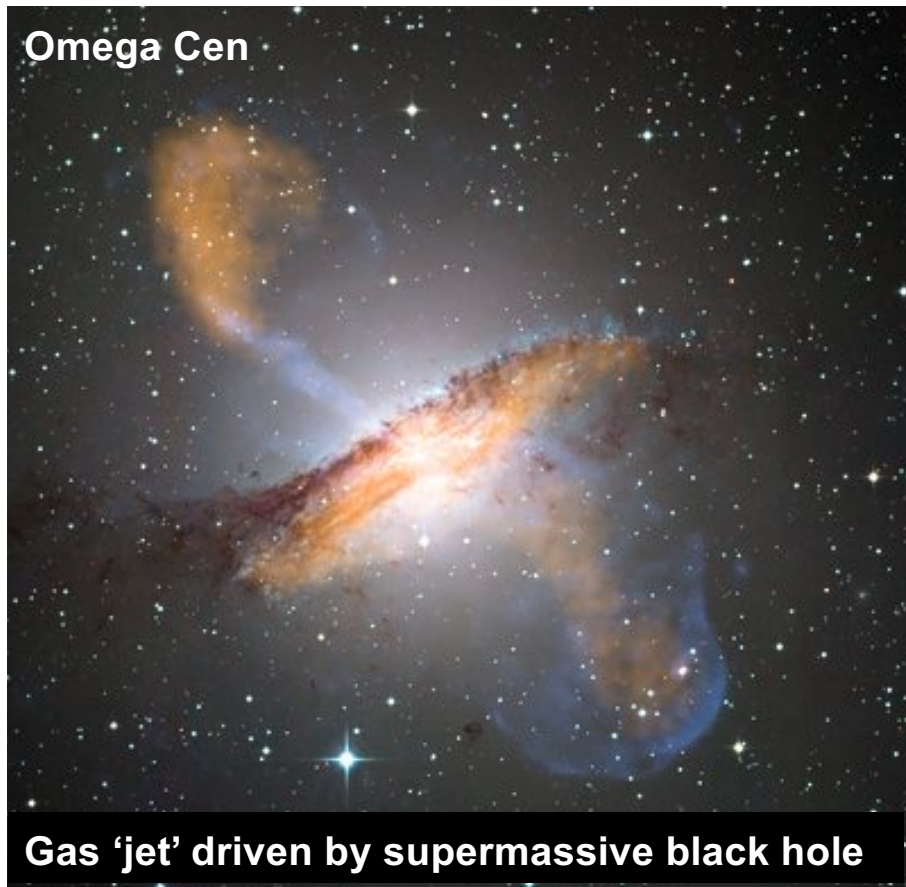
Driving Questions of Modern Astrophysics

The Decadal Plan for Australian Astronomy

- How did the first stars and galaxies transform the Universe?
- What is the nature of dark matter and dark energy?
- How do galaxies form and evolve across cosmic time?
- How do stars and planets form?
- How are elements produced by stars and recycled through galaxies?
- What is the nature of matter and gravity at extreme densities?



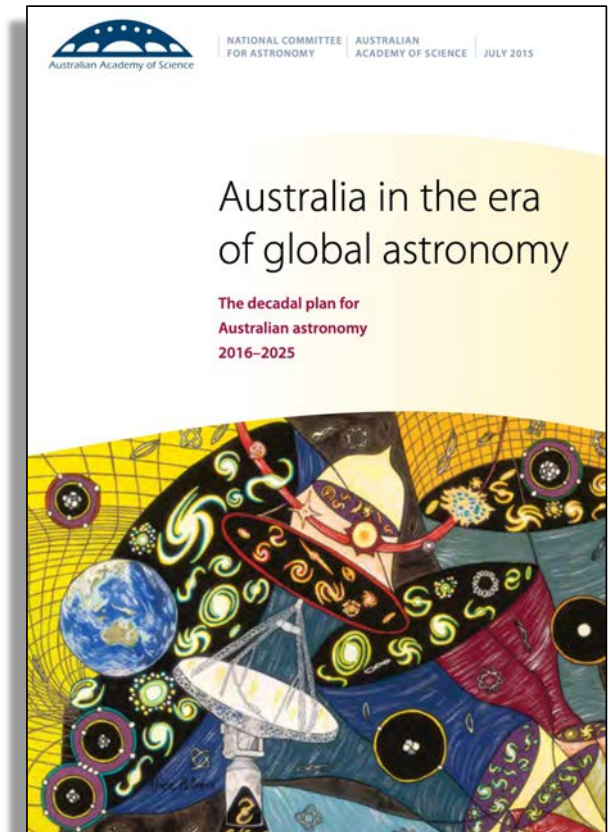
Recycling of elements in galaxies



Driving Questions of Modern Astrophysics

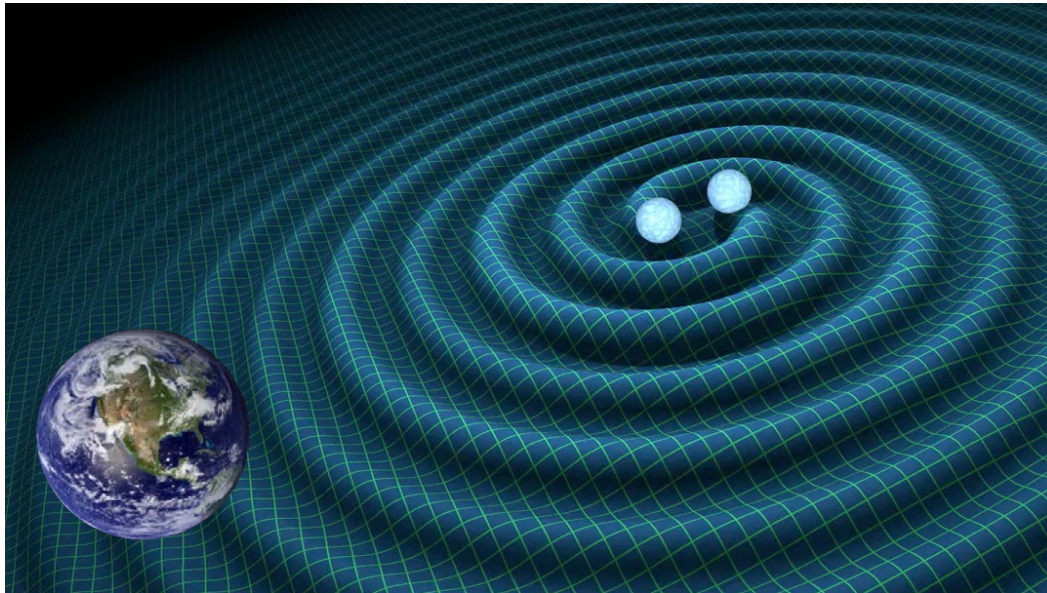
The Decadal Plan for Australian Astronomy

- How did the first stars and galaxies transform the Universe?
- What is the nature of dark matter and dark energy?
- How do galaxies form and evolve across cosmic time?
- How do stars and planets form?
- How are elements produced by stars and recycled through galaxies?
- What is the nature of matter and gravity at extreme densities?



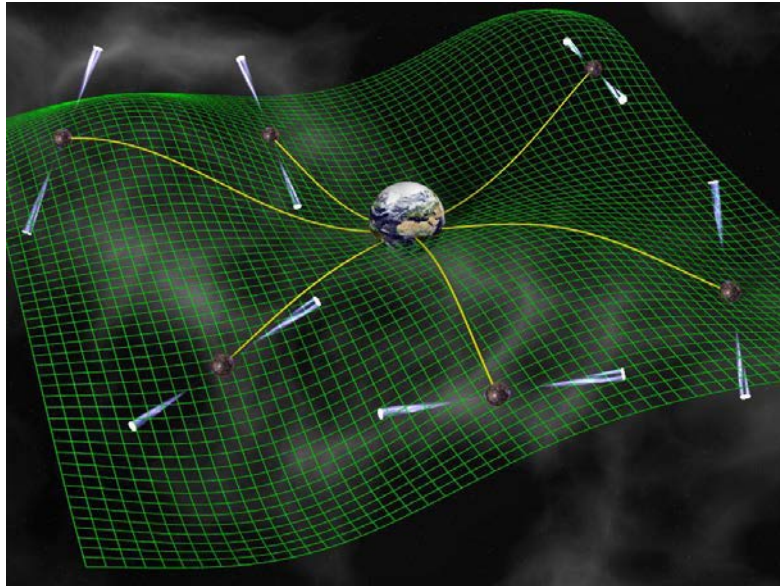
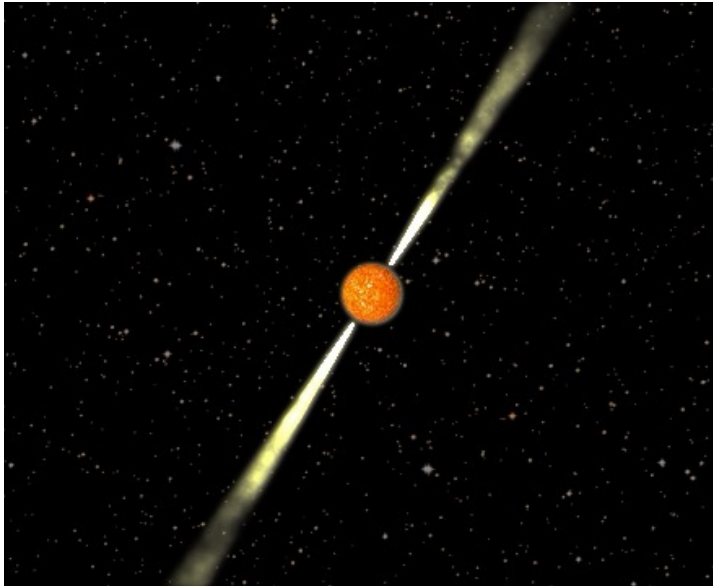
Gravitational Waves – A New Window on the Universe

- Nobel prize 2017 – detection of gravitational waves
- First detection was high frequency – 1-100x solar mass black holes

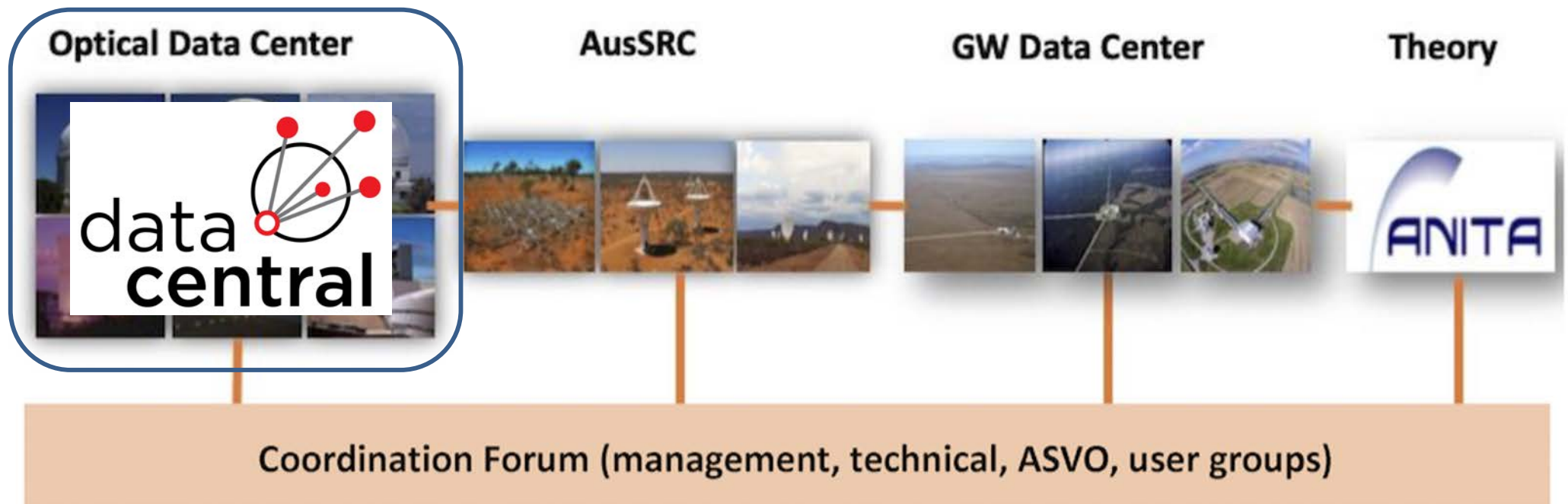


Gravitational Waves – A New Window on the Universe

- Nobel prize 2017 – detection of gravitational waves
- First detection was high frequency – 1-100x solar mass black holes
- Australia is leader in low-frequency detection using pulsars
- Potential to detect supermassive black hole mergers – 10^9 solar masses



















Australian Data Landscape





- > Part of NCRIS-funded national astronomy data capability, hosted at Macquarie / AAO
- > Supporting multiple national and international observational surveys
- > Applying international standards of FAIR data
- > Much more than an archive...

 <p>SAMI</p>	 <p>OzDES</p>	 <p>LoBES</p>	 <p>GALAH</p>	<p>RIE</p>
 <p>RAVE</p>	 <p>GAMA</p>	 <p>MAGPI</p>	 <p>Hector</p>	
 <p>Hi-KIDS</p>	 <p>Huntsman</p>	 <p>6dFGS</p>	 <p>Veloce</p>	
 <p>2dFGRS</p>	 <p>SkyMapper</p>	 <p>DEVILS</p>	 <p>TYPHOON</p>	

Data Central: Facilitating collaboration

Bundle exactly what you need



No two research teams are alike, and not all teams need all features. We offer a custom range of services that can be enabled when you register your team.

Team Management

Cloud Space

Private Wiki

Public Website

Mailing Lists

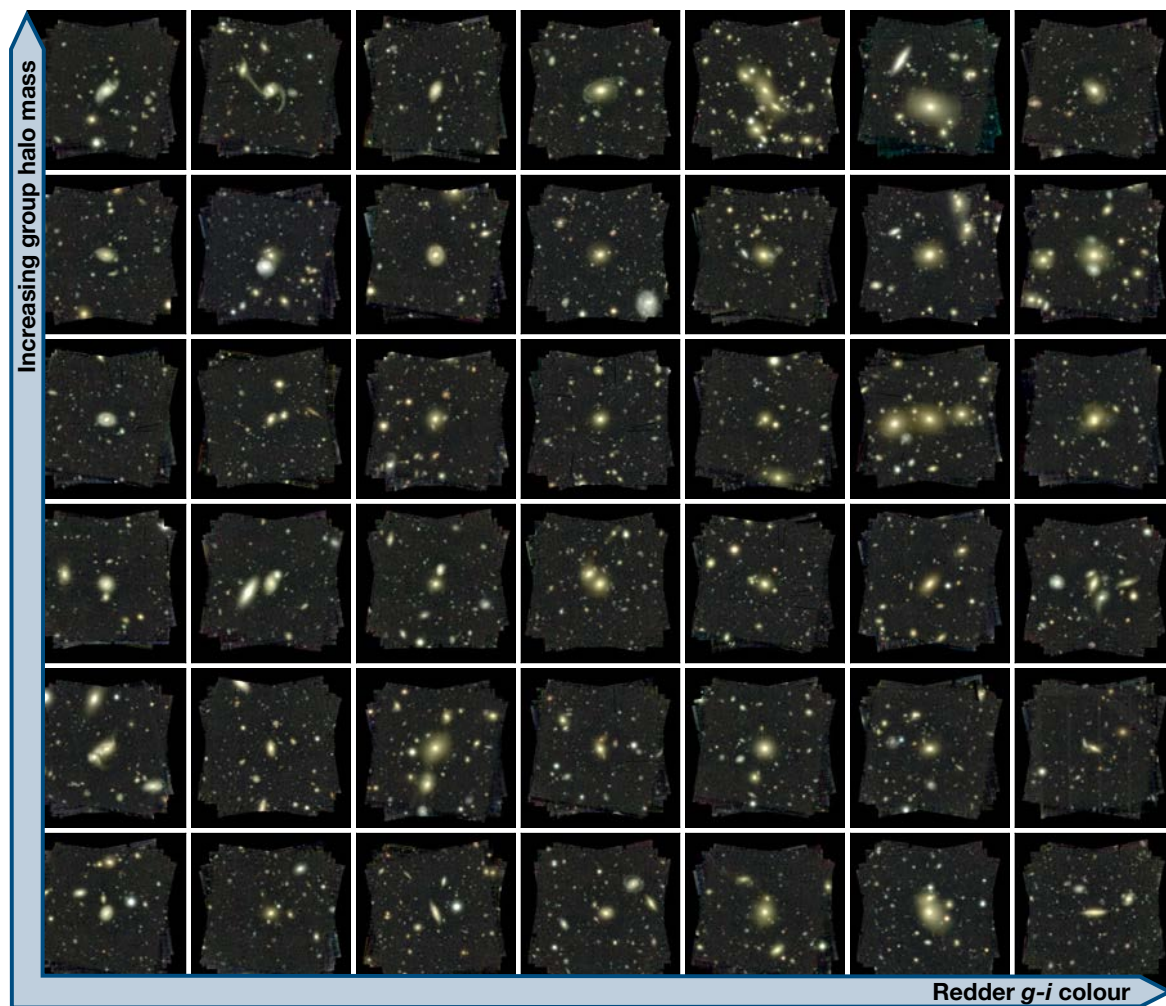
In Prep

Document Central

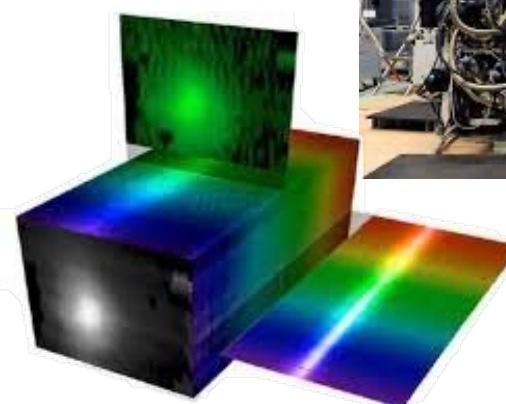
Desktops

Chat

Example: MAGPI Larger Program



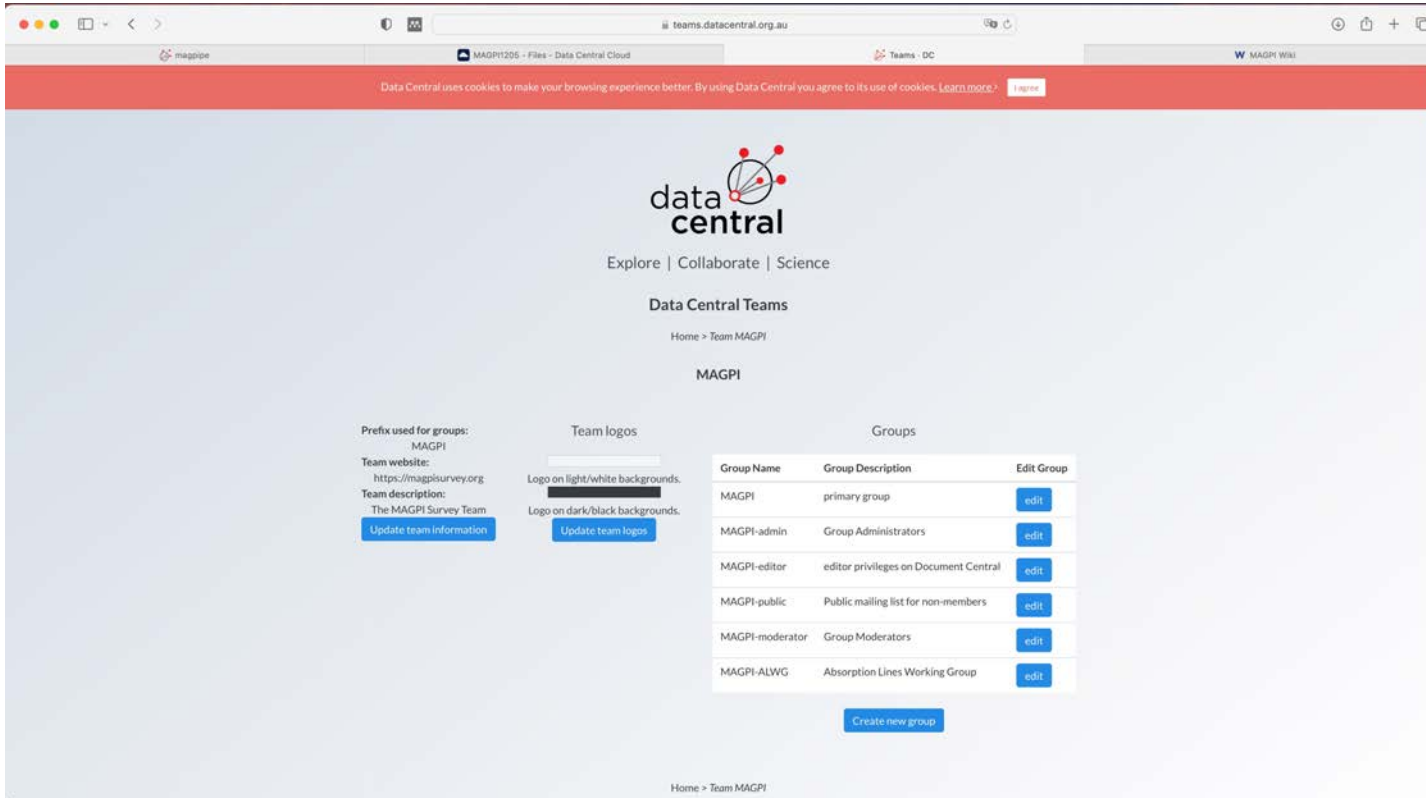
MUSE Spectrograph



Data Cube

Each field = 12Gb of raw data, plus 100s x more calibrations. 60-field survey > 100Tb

Example: MAGPI Larger Program



The screenshot shows the Data Central Teams management interface for the MAGPI team. The page title is "Data Central Teams" and the breadcrumb is "Home > Team MAGPI". The main heading is "MAGPI".

On the left, there are fields for team configuration:

- Prefix used for groups: MAGPI
- Team website: <https://magpisurvey.org>
- Team description: The MAGPI Survey Team

Below these fields are two buttons: "Update team information" and "Update team logos".

In the center, there are two input fields for "Team logos":

- Logo on light/white backgrounds.
- Logo on dark/black backgrounds.

Below these is a button: "Update team logos".

On the right, there is a table of "Groups":

Group Name	Group Description	Edit Group
MAGPI	primary group	edit
MAGPI-admin	Group Administrators	edit
MAGPI-editor	editor privileges on Document Central	edit
MAGPI-public	Public mailing list for non-members	edit
MAGPI-moderator	Group Moderators	edit
MAGPI-ALWG	Absorption Lines Working Group	edit

At the bottom of the table is a button: "Create new group".

Data Central enables:

- › Easy team roles management

Example: MAGPI Larger Program

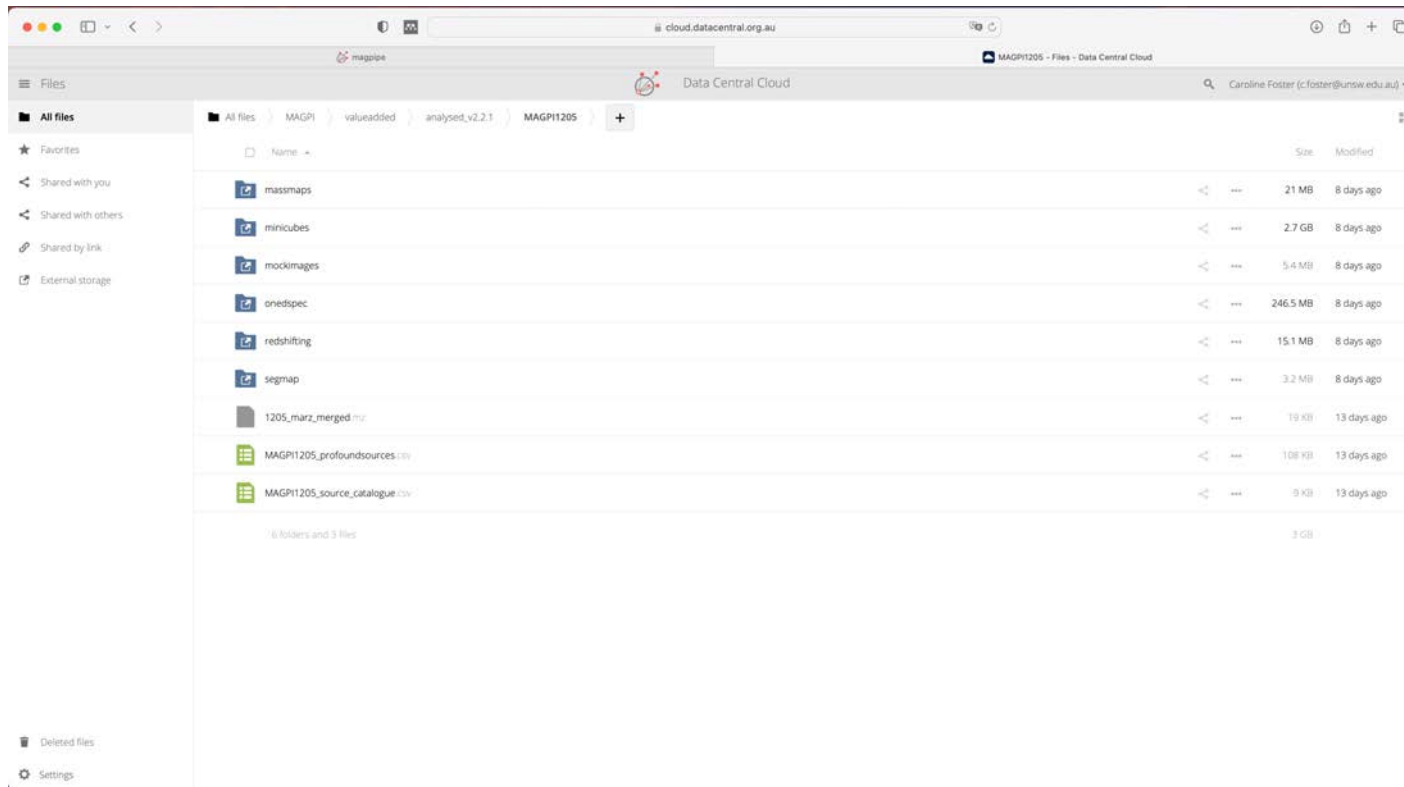


The screenshot shows a web browser displaying the MAGPI Wiki page. The page has a dark blue header with the text "MAGPI WIKI" and a search bar. Below the header is a navigation menu with options like "Home", "All Pages", and "Settings". The main content area features a large "MAGPI" logo with a bird perched on a swoosh. Below the logo is a "Page links" section with a table of links: Home, Survey Design, Science projects, Working groups, Resources, Team members, and Policies. The page also includes sections for "About MAGPI", "Staying in touch", and "Calendar". At the bottom, there is a "MAGPI Survey" calendar for November 2022.

Data Central enables:

- › Easy team roles management
- › Team wiki

Example: MAGPI Larger Program



Data Central enables:

- › Easy team roles management
- › Team wiki
- › Controlled cloud data storage

Example: MAGPI Larger Program



```

magpi@dccompute2: ~/magpipe/stellar_kinematics
File Edit View Search Terminal Tabs Help

magpi@dccompute2: ~/magpipe/stellar_kinematics
magpi@dccompute2: ~/magpipe/stellar_kinematics/01_kinematics

ts
[ RUNNING ] Loading annuli template weights from 01_kinematics/MAGPI1201/1201197199/results/1201197199_2MOM/1201197199_kin-annularbi
[ DONE ] Loading annuli template weights from 01_kinematics/MAGPI1201/1201197199/results/1201197199_2MOM/1201197199_kin-annularbi
ns-template-weights.fits
[ DONE ] Log-rebinning the individual spaxel spectra
[ DONE ] Log-rebinning the individual spaxel error spectra
[ DONE ] Writing: 1201197199_LogrebinSpectra.fits
[ RUNNING ] Preparing stellar templates for kinematics extraction
[ INFO ] 'indous' library provides air wavelengths.
[ DONE ] Preparing stellar templates for kinematics extraction
[ DONE ] Running PPXF in parallel mode
Running PPXF on 16368 spectra took 819.09s using 32 cores
There were no problems in the analysis.

[ DONE ] Writing: 1201197199_kinematics_ppxf-maps.fits
[ DONE ] Writing: 1201197199_kinematics_ppxf-output.fits
[ DONE ] Writing: 1201197199_kinematics_ppxf-linbestfit.fits
[ DONE ] Writing: 1201197199_kinematics_ppxf-polyweights.fits
Kinematic fits done.

[ WARNING ] Skipping CONTSUBKIN subtask of CONTSUB!

[ WARNING ] Skipping CONTSUBWHT subtask of CONTSUB!

[ WARNING ] Skipping CONTSUBSPX subtask of CONTSUB!

...
Running PPXF!
[ RUNNING ] Preparing the stellar population templates
[ INFO ] 'indous' library provides air wavelengths.
[ INFO ] Converting to vacuum wavelengths (resampling)
[ DONE ] Preparing the stellar population templates
[ RUNNING ] Running PPXF in parallel mode
| 29.52%
    
```

```

magpi@dccompute2: ~/magpipe/stellar_kinematics/01_kinematics
File Edit View Search Terminal Tabs Help

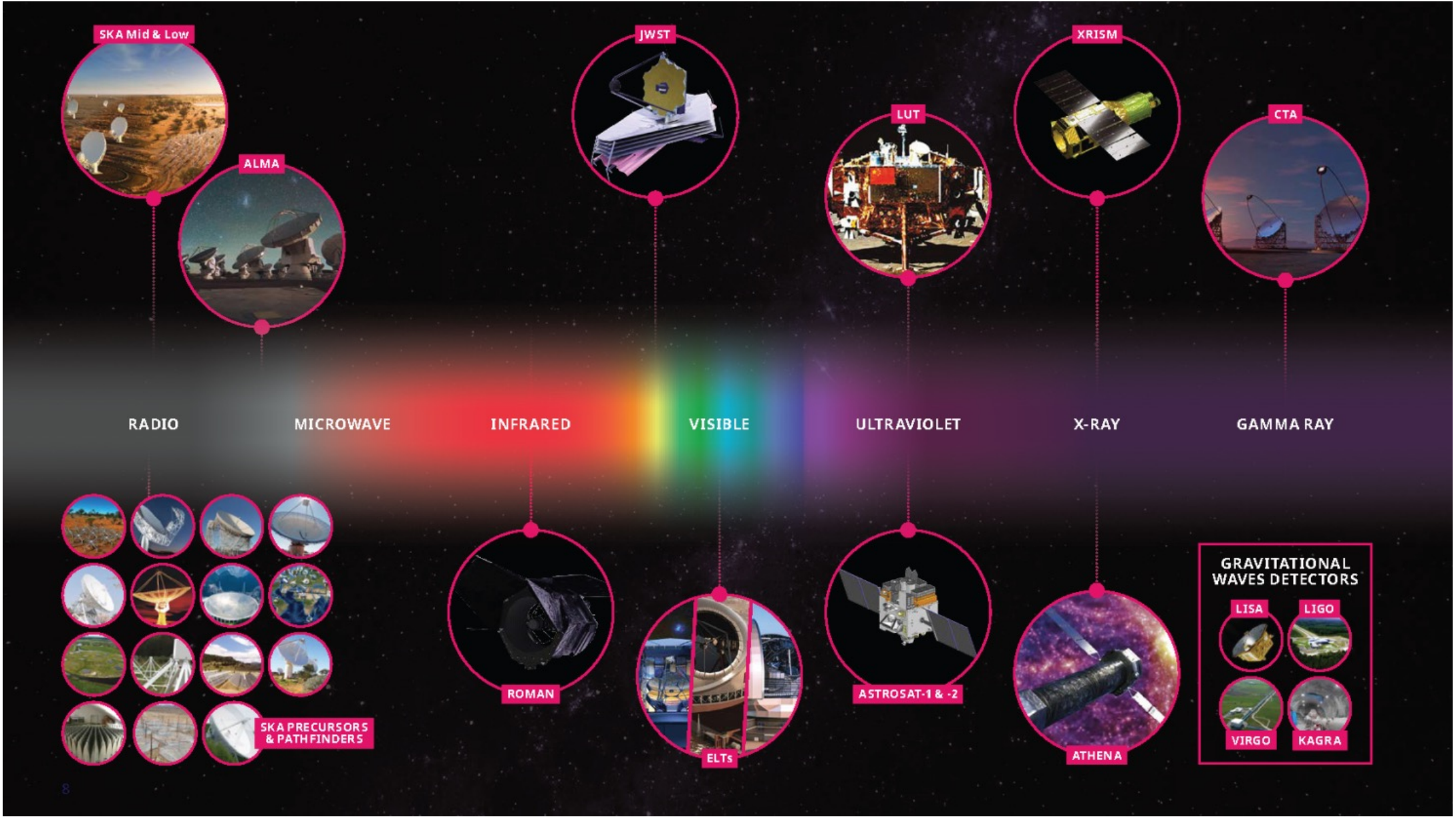
magpi@dccompute2: ~/magpipe/stellar_kinematics
magpi@dccompute2: ~/magpipe/stellar_kinematics/01_kinematics

top - 15:45:10 up 189 days, 6:34, 0 users, load average: 30.41, 28.07, 25.58
Tasks: 672 total, 35 running, 473 sleeping, 1 stopped, 0 zombie
%Cpu(s): 1.0 us, 9.0 sy, 89.3 ni, 0.5 id, 0.2 wa, 0.0 hi, 0.0 si, 0.0 st
KiB Mem : 52835766+total, 33852928 free, 13339912 used, 48116483+buff/cache
KiB Swap : 2097148 total, 1966332 free, 130816 used, 51121539+avail Mem

PID USER      PR  NI   VIRT   RES    SHR  S  %CPU  %MEM    TIME+  COMMAND
6212 dcadmin   30  10  52332  17208  4036  R   99.7   0.0   2:08.43  tessellImage
6418 magpi    35  15  2809820  1.775g  12952  R   91.5   0.4   1:29.50  gistPipeline
6389 magpi    35  15  2860340  1.823g  12780  R   89.3   0.4   1:33.53  gistPipeline
6403 magpi    35  15  2809820  1.775g  12948  R   85.3   0.4   1:22.18  gistPipeline
6404 magpi    35  15  2883508  1.845g  12948  R   85.0   0.4   1:17.10  gistPipeline
6421 magpi    35  15  2866580  1.829g  12952  R   82.4   0.4   1:26.12  gistPipeline
6391 magpi    35  15  2883508  1.845g  12948  R   76.2   0.4   1:14.22  gistPipeline
6416 magpi    35  15  2860340  1.823g  12780  R   74.9   0.4   1:11.07  gistPipeline
6419 magpi    35  15  2856152  1.775g  12712  R   73.6   0.4   1:08.71  gistPipeline
6423 magpi    35  15  2809820  1.775g  12948  R   72.3   0.4   1:14.07  gistPipeline
6401 magpi    35  15  2809820  1.770g  12948  R   71.7   0.4   1:15.01  gistPipeline
6407 magpi    35  15  2883508  1.845g  12948  R   71.7   0.4   1:10.89  gistPipeline
6415 magpi    35  15  2809820  1.767g  12712  R   71.0   0.4   1:10.57  gistPipeline
6396 magpi    35  15  2883508  1.845g  12948  R   70.7   0.4   1:13.39  gistPipeline
6424 magpi    35  15  2883508  1.845g  12780  R   70.7   0.4   1:10.05  gistPipeline
6422 magpi    35  15  2883508  1.845g  12948  R   70.4   0.4   1:09.86  gistPipeline
6408 magpi    35  15  2786652  1.753g  12784  R   68.1   0.3   1:06.59  gistPipeline
6411 magpi    35  15  2860340  1.823g  12948  R   66.8   0.4   1:08.18  gistPipeline
6414 magpi    35  15  2809820  1.775g  12712  R   64.8   0.4   1:08.56  gistPipeline
6409 magpi    35  15  2809820  1.775g  12884  R   64.5   0.4   1:08.60  gistPipeline
6399 magpi    35  15  2856152  1.811g  12952  R   63.8   0.4   1:10.04  gistPipeline
6400 magpi    35  15  2809820  1.775g  12784  R   63.8   0.4   1:10.27  gistPipeline
6397 magpi    35  15  2809820  1.775g  12780  R   61.9   0.4   1:11.19  gistPipeline
6412 magpi    35  15  2809820  1.775g  12948  R   61.9   0.4   1:13.16  gistPipeline
6398 magpi    35  15  2809820  1.775g  12948  R   60.6   0.4   1:12.32  gistPipeline
6402 magpi    35  15  2883508  1.845g  12784  R   60.3   0.4   1:11.62  gistPipeline
6417 magpi    35  15  2809820  1.775g  12780  R   60.3   0.4   1:05.75  gistPipeline
6394 magpi    35  15  2860472  1.775g  12952  R   58.6   0.4   1:12.85  gistPipeline
6410 magpi    35  15  2809820  1.775g  12948  R   58.3   0.4   1:11.57  gistPipeline
6393 magpi    35  15  2883508  1.845g  12948  R   57.7   0.4   1:12.82  gistPipeline
    
```

Data Central enables:

- › Easy team roles management
- › Team wiki
- › Controlled cloud data storage
- › Reduction + analysis on remote server-at-data



SKA Mid & Low



ALMA



JWST



LUT



XRISM



CTA



RADIO

MICROWAVE

INFRARED

VISIBLE

ULTRAVIOLET

X-RAY

GAMMA RAY



SKA PRECURSORS & PATHFINDERS

ROMAN



ELTs



ASTROSAT-1 & -2



ATHENA



GRAVITATIONAL WAVES DETECTORS

LISA



LIGO



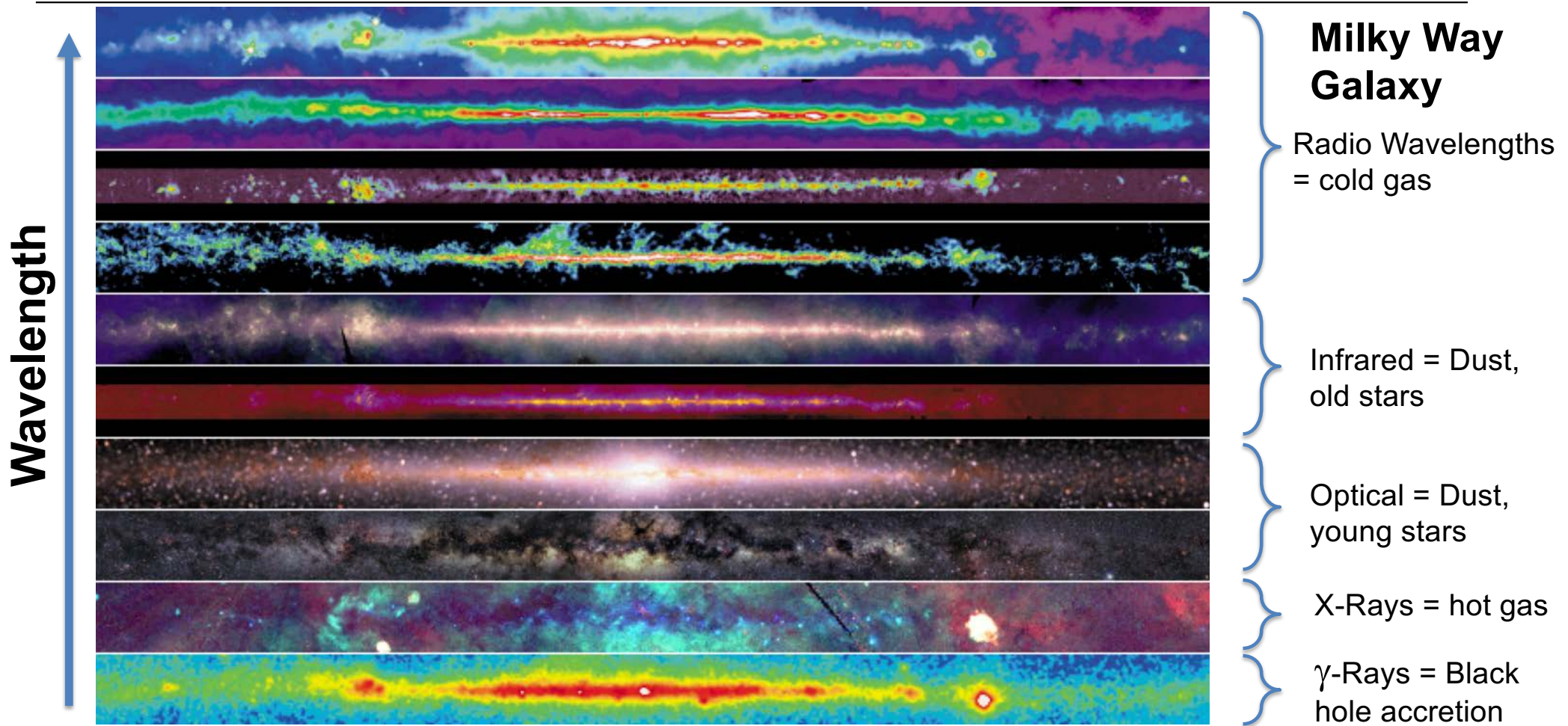
VIRGO



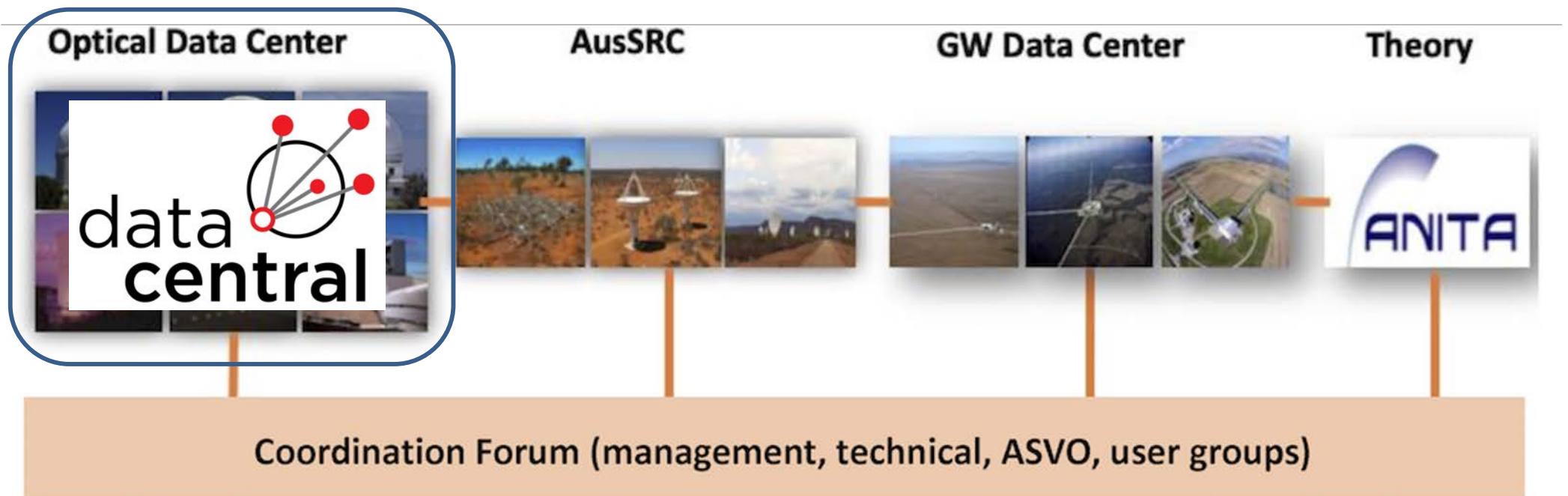
KAGRA



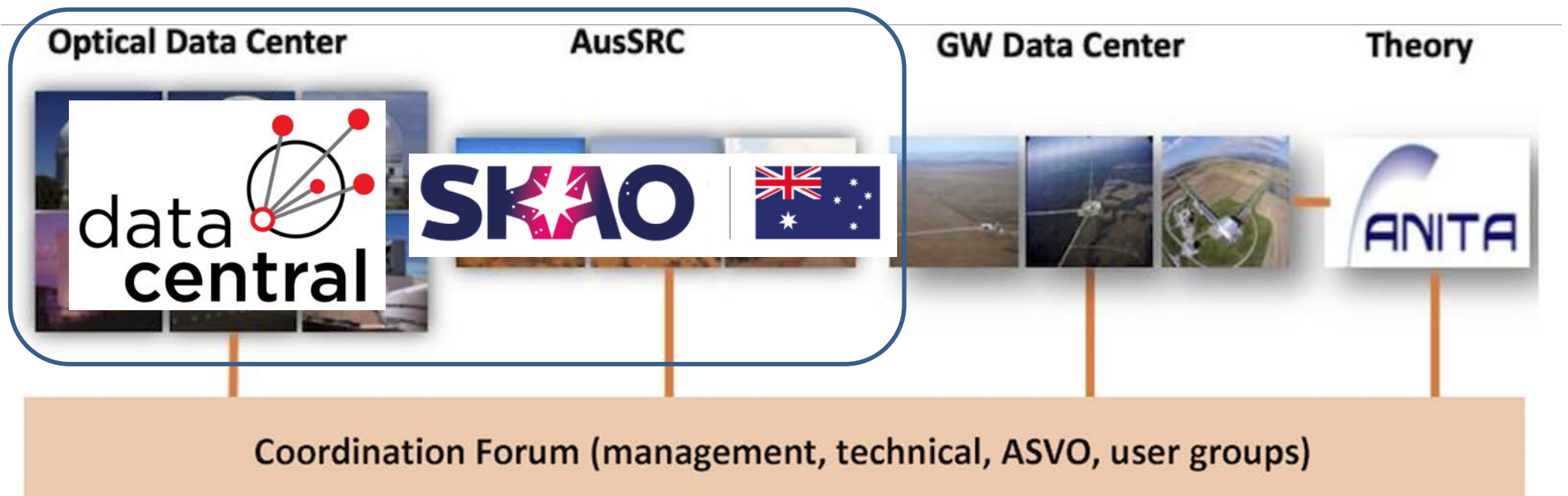
Multi-Wavelength Astronomy



Australian Data Landscape



Australian Data Landscape



Combining SKA and Optical Data Infrastructure



SKA = Square Kilometer Array

A screenshot of the Table Access Protocol (TAP) Query interface. The window title is "Table Access Protocol (TAP) Query". It has a toolbar with icons for search, refresh, help, and close. Below the toolbar are tabs for "Select Service", "Use Service", "Resume Job", and "Running Jobs". The main area is divided into "Metadata" and "ADQL Text". The "Metadata" section has a "Find:" field and radio buttons for "Name", "Descrip", and "Or". A tree view on the left shows a hierarchy of tables, with "gleamx_dr1.idr1_subset" selected. A table view on the right shows columns with their names, types, units, indexed status, and descriptions. The "ADQL Text" section has a "Mode:" dropdown set to "Synchronous" and a large text area for writing queries. At the bottom, there are "Examples" and "Info" buttons, and a "Run Query" button.

Name	Type	Unit	Indexed	Description
Name	VARCHAR		<input checked="" type="checkbox"/>	IAU Name
background_wide	REAL	Jy/beam	<input type="checkbox"/>	Background l
local_rms_wide	REAL	Jy/beam	<input type="checkbox"/>	Local noise l
ra_str	VARCHAR		<input checked="" type="checkbox"/>	Right Ascens
dec_str	VARCHAR	d:m:s	<input checked="" type="checkbox"/>	Declination J;
RAJ2000	REAL	deg	<input checked="" type="checkbox"/>	Right Ascens
err_RAJ2000	REAL	deg	<input type="checkbox"/>	Uncertainty in
DEJ2000	REAL	deg	<input checked="" type="checkbox"/>	Declination J;
err_DEJ2000	REAL	deg	<input type="checkbox"/>	Uncertainty in
peak_flux_wide	REAL	Jy/beam	<input type="checkbox"/>	Peak flux in
err_peak_flux_wide	REAL	Jy/beam	<input type="checkbox"/>	Uncertainty in
int_flux_wide	REAL	Jy	<input type="checkbox"/>	Integrated fl
err_int_flux_wide	REAL	Jy	<input type="checkbox"/>	Uncertainty in

- › Through ASTRO 3D Centre of Excellence, Data Central and Australian SKA Regional Centre (AusSRC) teams meet monthly to discuss DIA
- › Gave rise to pilot project on table cross-matching based on sky locations
- › Uses Table Access Protocol – “TAP”
- › Upload your own tables. Cross match with Data Central tables. Retrieve results. All from your Python script!
- › Successfully used by SKA pathfinder missions
- › Potential next step is to move to imaging, building on Data Central’s “Aggregation Service”
- › Long-term goal is to add 3D capabilities

Combining SKA and Optical Data Infrastructure



SKA = Square Kilometer Array

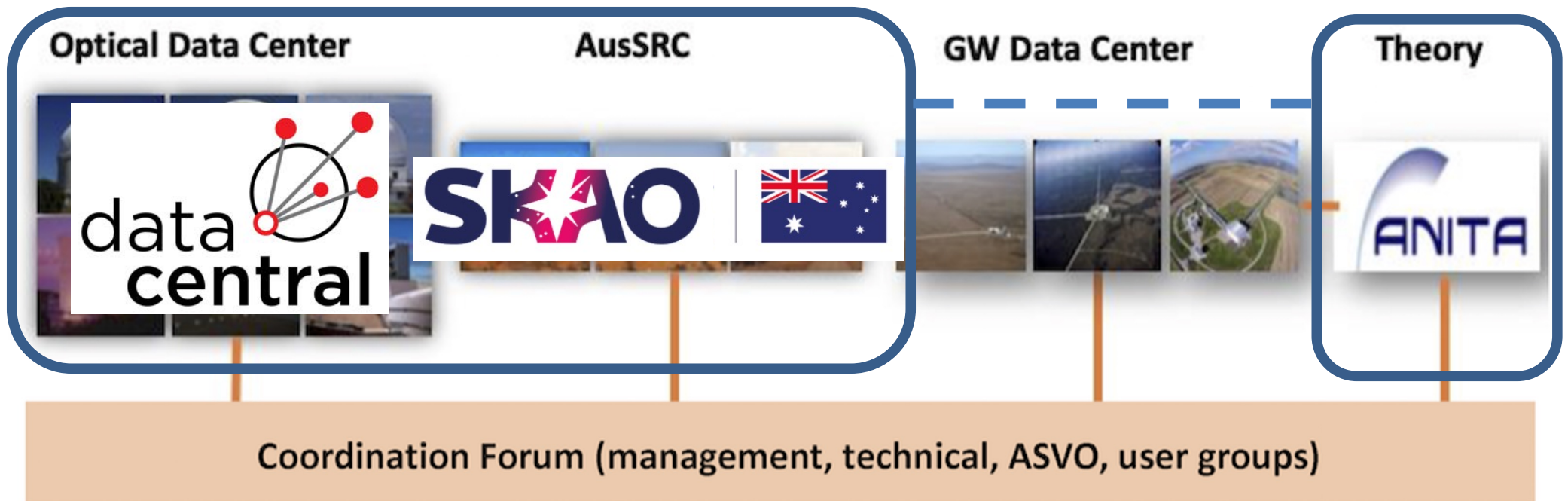
The screenshot displays the Data Central interface. At the top, a 'Table Access Protocol (TAP) Query' window is visible. Below it, the 'Data Aggregation Service' logo and user information (SAMP, Richard McDermid, Overlay, Transparency: 100%) are shown. The main area features a sky map visualization with various colored markers (red, green, blue, orange) overlaid on a dark background. To the right, a table of query results is displayed with columns: dist_arcsec, object_id, raj2000, dej2000, e_raj2000, and e_dej2000. Below the table, there is a search bar and a section for 'Images' with columns: assoc_id, access_uri, access_format, access_estsize, and calib_level.

dist_arcsec	object_id	raj2000	dej2000	e_raj2000	e_dej2000
0.2392946491	303122008	307.888299	-7.088353	36	36
7.5834696687	303122009	307.888127	-7.086249	36	36
16.3376512347	303122011	307.888229	-7.083806	187	101
16.771304069	303122007	307.886796	-7.092733	1313	212
30.8869813678	303122026	307.895966	-7.092431	36	83

assoc_id	access_uri	access_format	access_estsize	calib_level
le_dr8		image/fits	11981	3

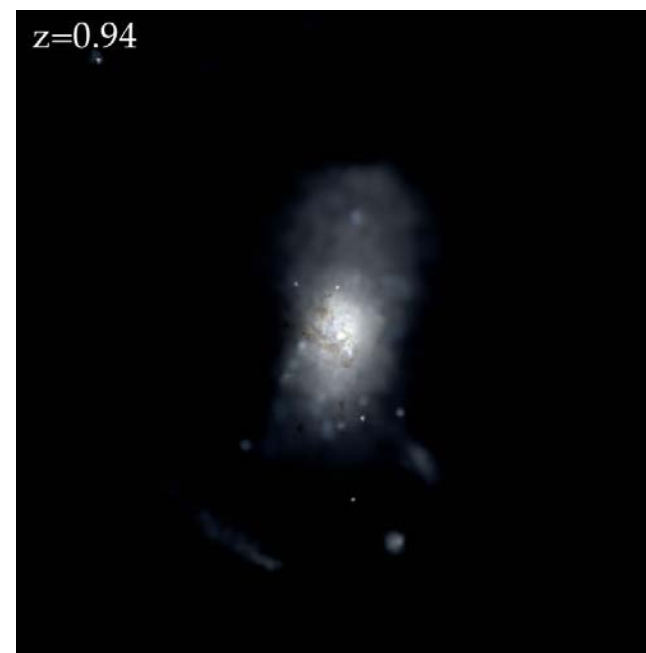
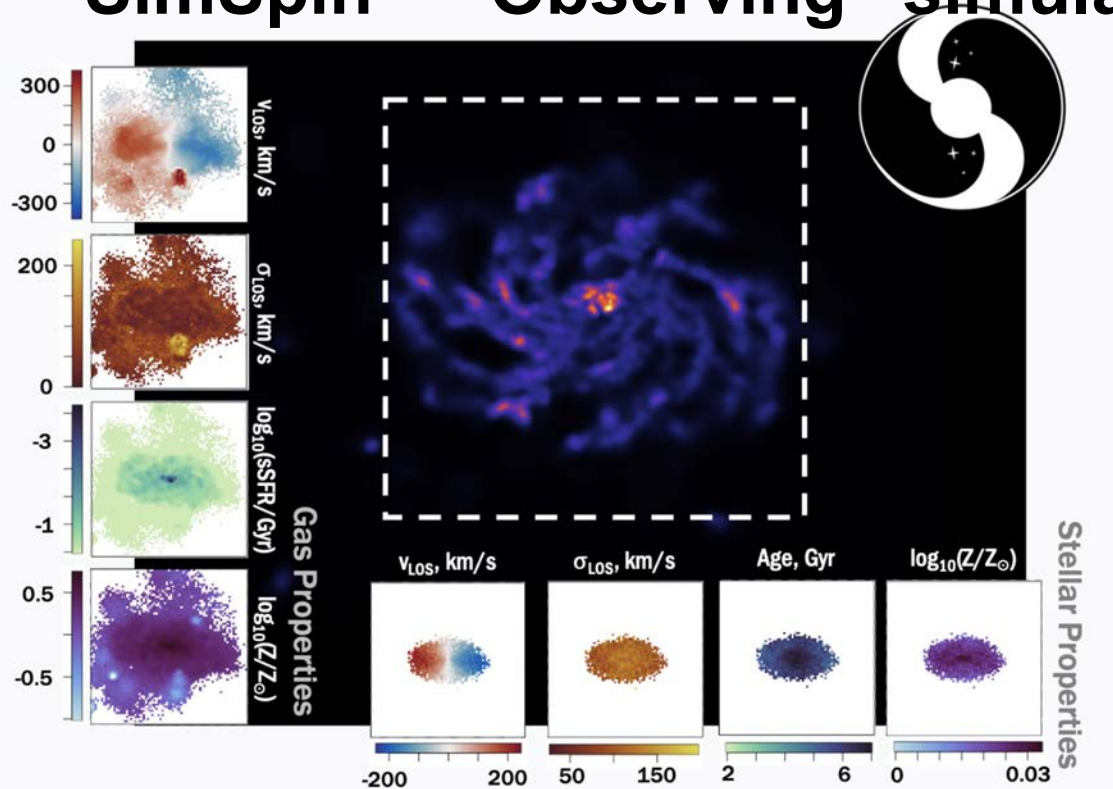
- > Through ASTRO 3D Centre of Excellence, Data Central and Australian SKA Regional Centre (AusSRC) teams meet monthly to discuss DIA
- > Gave rise to pilot project on table cross-matching based on sky locations
- > Uses Table Access Protocol – “TAP”
- > Upload your own tables. Cross match with Data Central tables. Retrieve results. All from your Python script!
- > Successfully used by SKA pathfinder missions
- > Potential next step is to move to imaging, building on Data Central’s “Aggregation Service”
- > Long-term goal is to add 3D capabilities

Australian Data Landscape



Forward Modelling

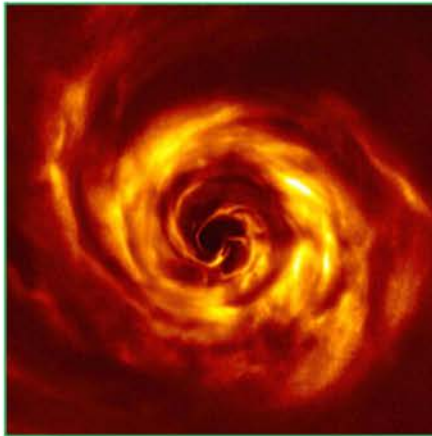
SimSpin – “Observing” simulations



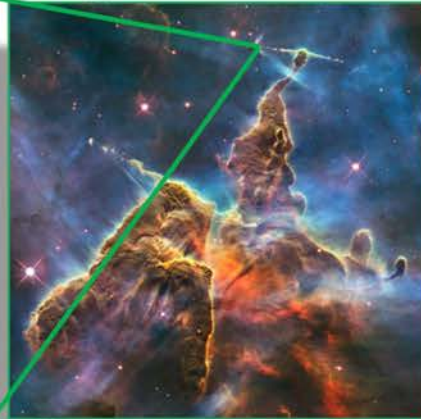
<https://simspin.datacentral.org.au/app/>

Complexity is observed on all scales in the cosmos

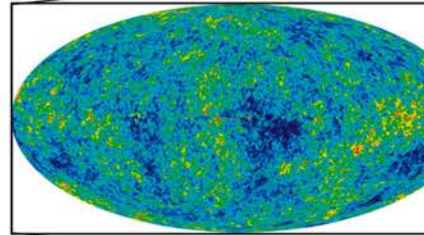
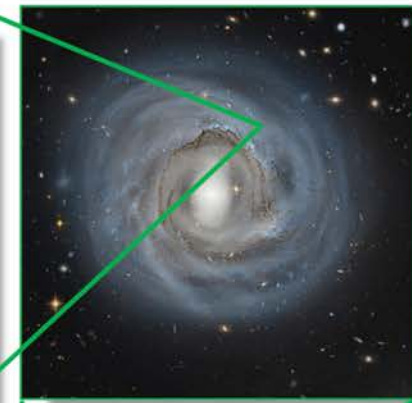
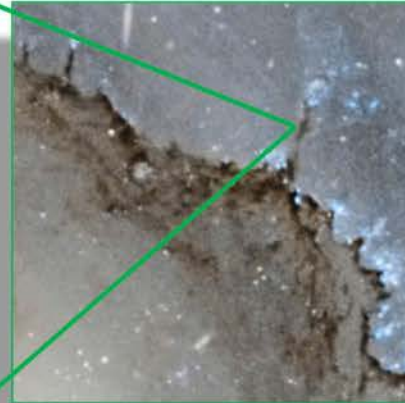
AB Aurigae - Sphere



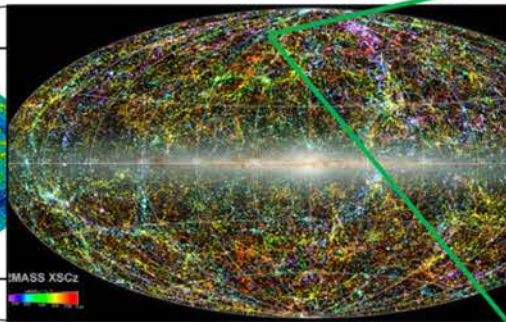
Eagle Nebula - HST



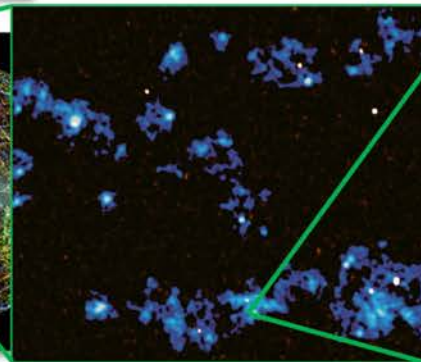
NGC4921 - HST



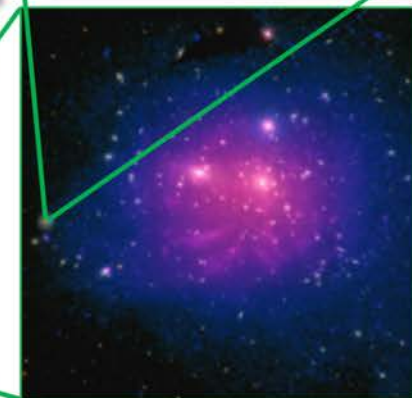
WMAP



2MASS-RS



SSA22 - MUSE

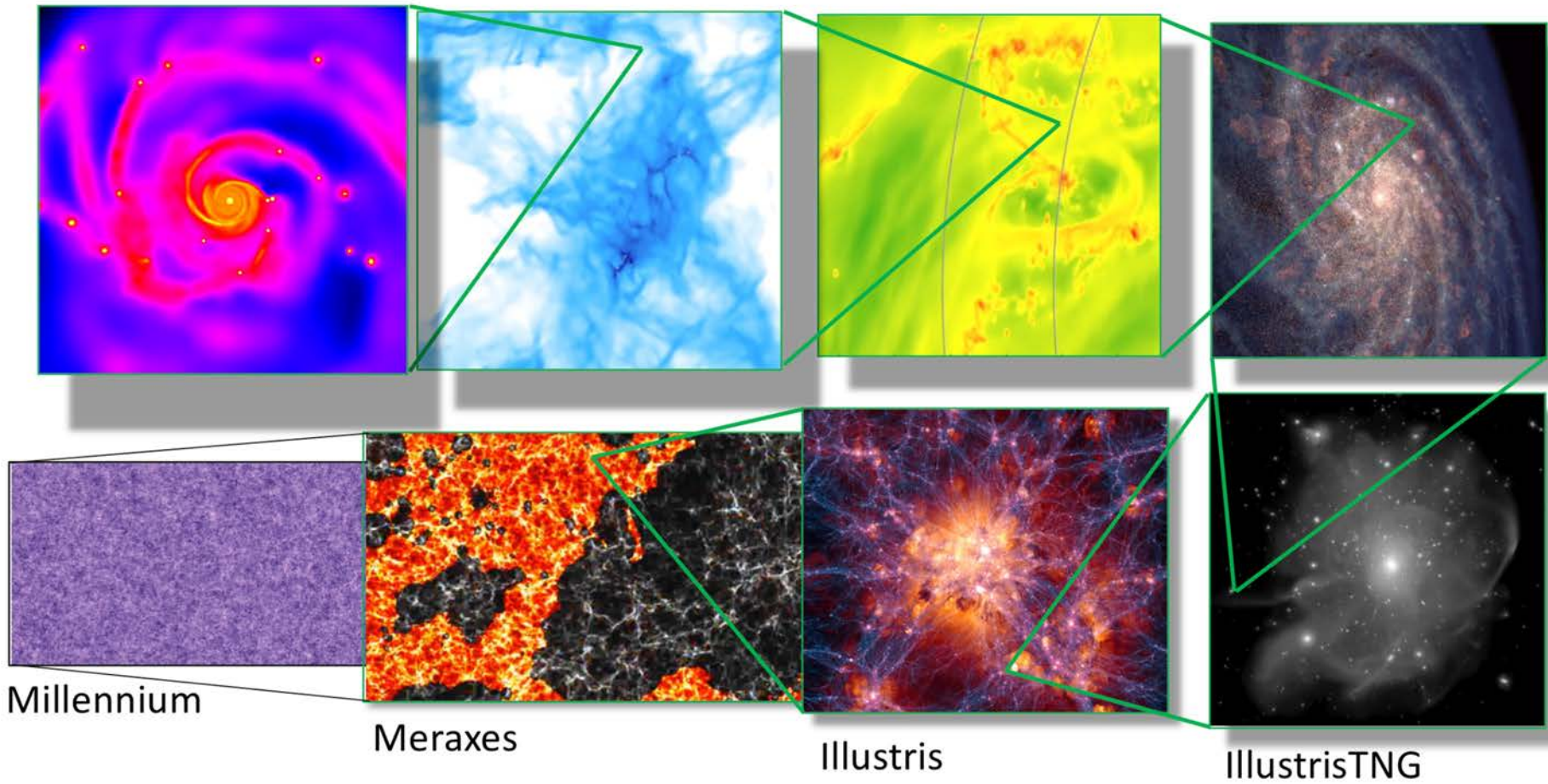


Coma - Chandra/ESO

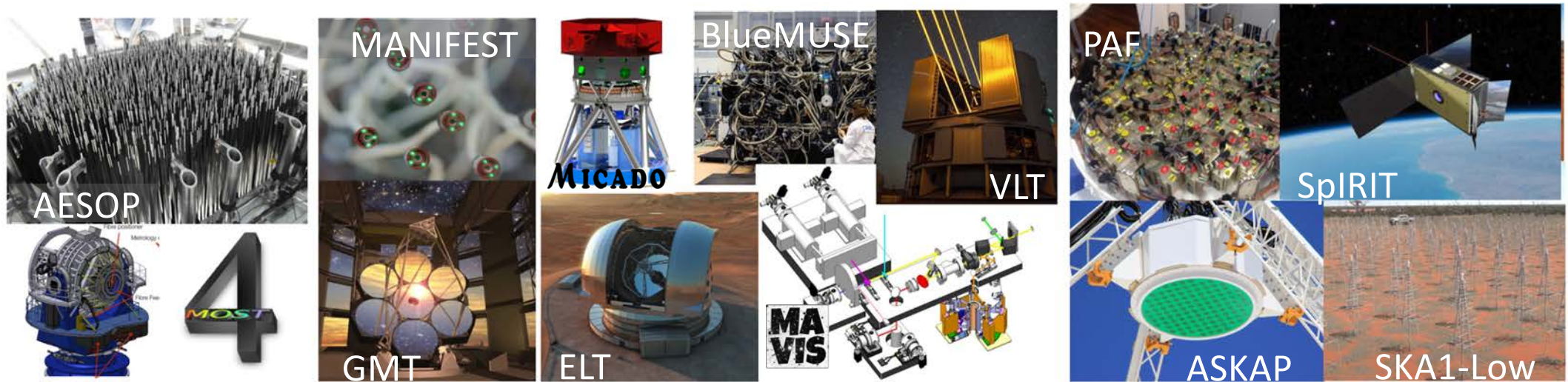
Complexity is simulated on all scales in the cosmos

Lufkin+2004

Mark Krumholz Group



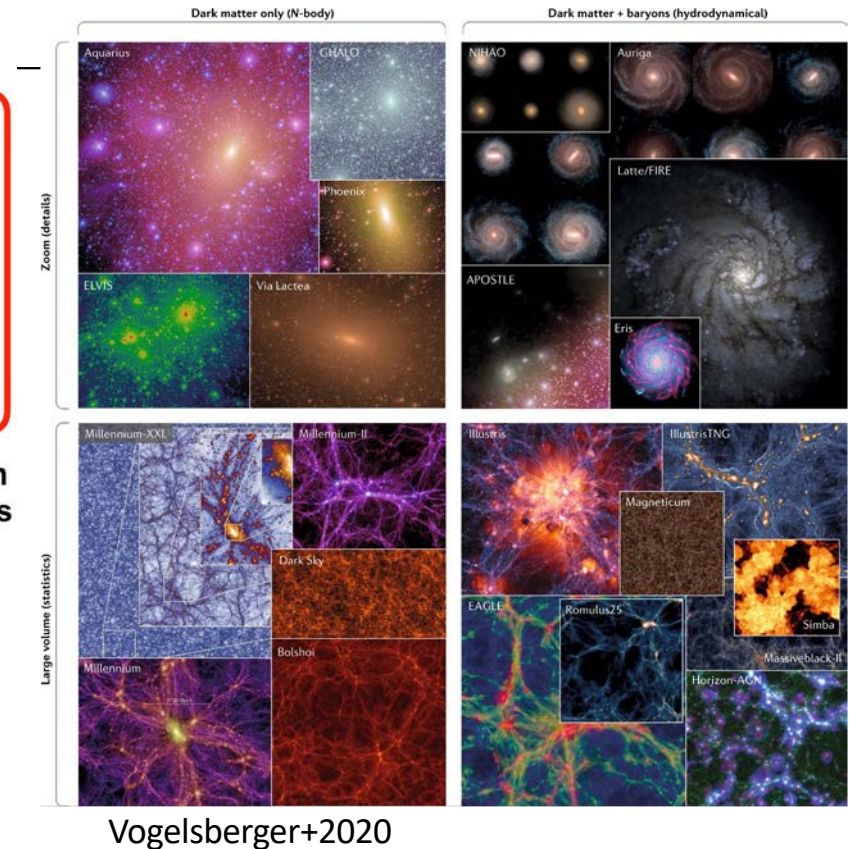
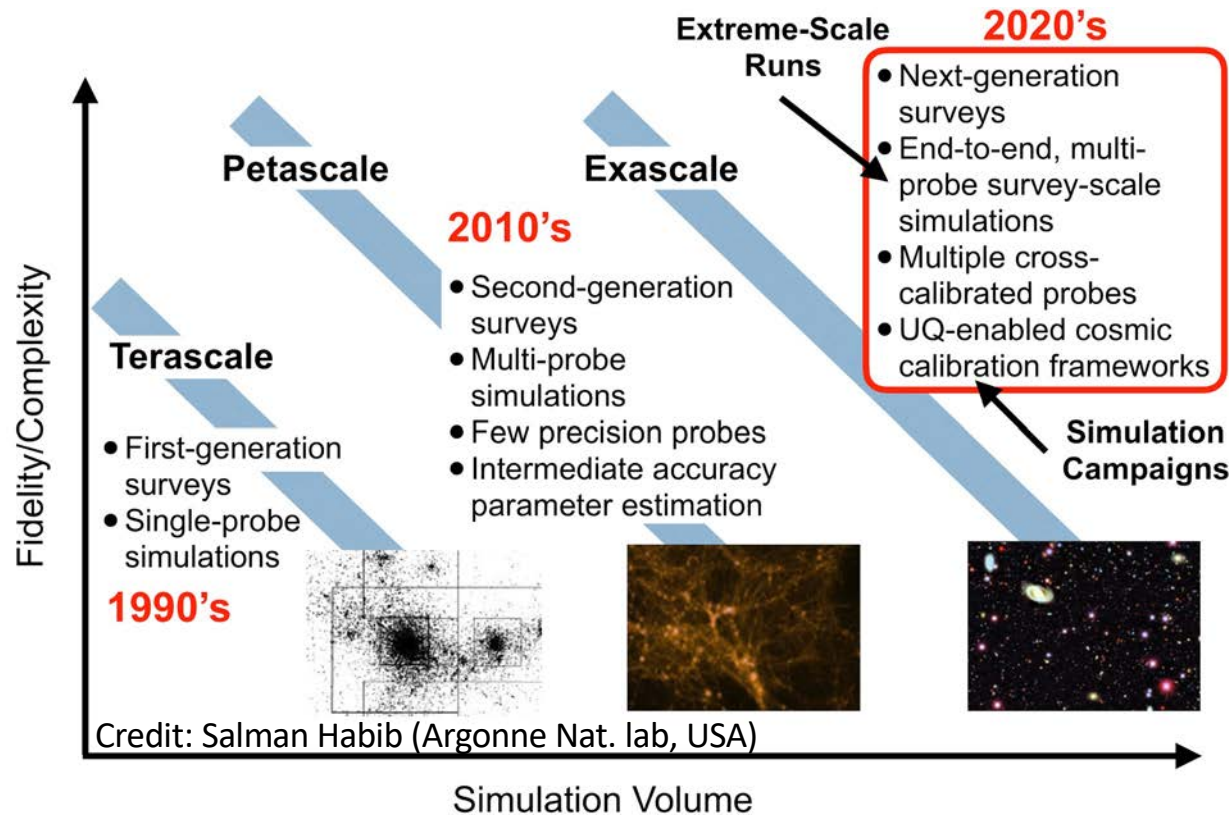
Existing and Future Observational Technologies



Technology involving Australian leadership

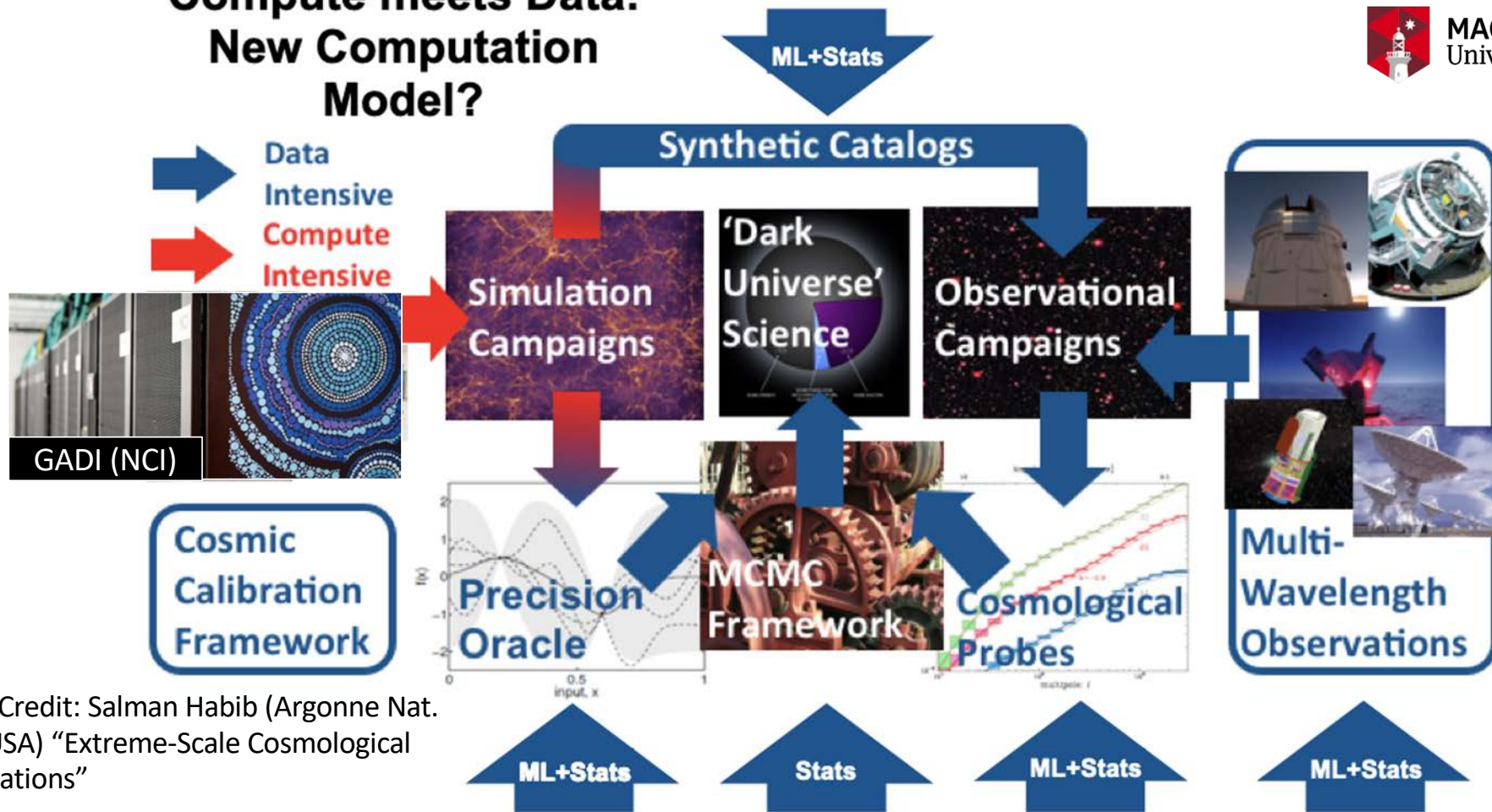
- Dedicated facilities provide multi-wavelength surveys: **Big and Complex Data**
- Will outstrip institution-scale resources - need 'compute-near-data' facilities/methods
- Data accessibility is key for collaboration, and broadening participation / inclusivity

Existing and Future Simulation Methodologies



- Next-Gen simulations will break sub-grid boundaries, for cosmological volumes
- More focus on forward-modelling of higher-order, multi-probe observations

Compute meets Data: New Computation Model?



Slide Credit: Salman Habib (Argonne Nat. lab, USA) "Extreme-Scale Cosmological Simulations"

- Confronting models and observations will need new ways of working
- Dramatically stronger focus on **Statistics** and **Data Science** techniques

Training Complex Thinkers



Future Role of Astronomy and Data

- Use astronomy as a training vehicle in complex problem solving
- Develop innovative, data-literate, and creative scientists
- Build authentic relationships with high-tech and data-centric industries
- Provide students and ECRs with experience in a non-academic setting, broadening career horizons

Summary

-
- Astronomy is a data-intensive science, and faces issues of data volumes, complexity, physical scales, and skills training
 - Major investments in observing facilities need to be matched to data, compute, and human capital development
 - International leadership will require new skill sets in data science, and meaningful partnership with industry (data and technology sectors)
 - Multi-disciplinary Centre of Excellence would allow a broad range of science goals to be tackled using commonality between these new methodologies
 - What is the role of data science, statistics, and computing in such an astronomy-driven Centre of Excellence?
 - Discussions are currently under way – white papers being written