

# Data Ecosystem in the SKA Regional Centre Network: the Italian contribution

Vincenzo Galluzzi  
On behalf of itSRC

*USC-C General Assembly - INAF Large Projects (13<sup>th</sup> March 2026)*

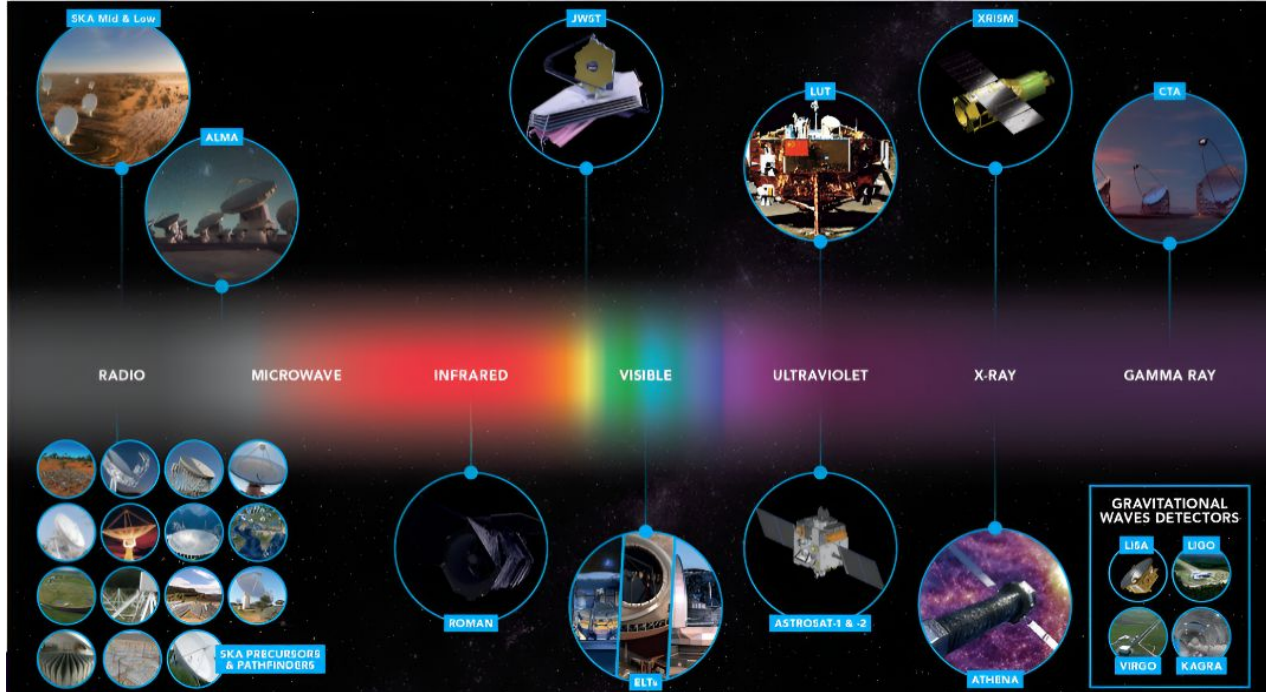
it | SRC

SKAO Regional Centre Italy



# New telescopes and big data

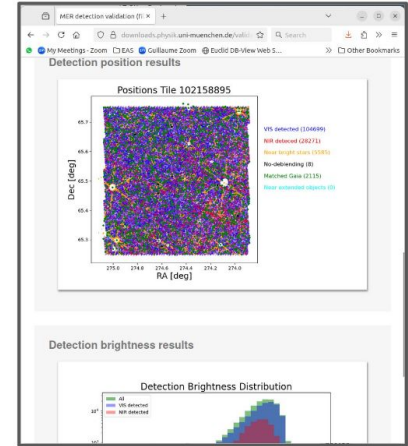
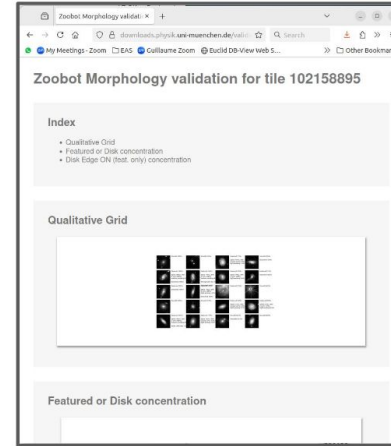
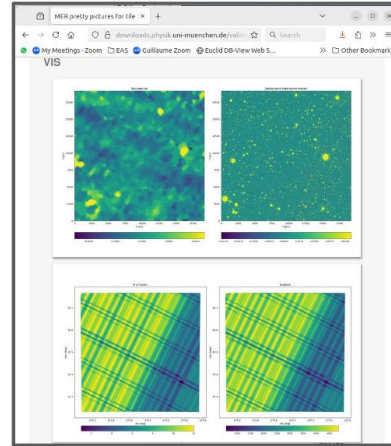
New facilities are coming online, providing unprecedented performances (sensitivity and fov, hence survey speed, resolutions, flexibility and commensality), resulting in data rates for science data products at level of several PBs/year.



- Variety of data types
- Big data volume per single entity
- High potential for reusability
- High potential for interoperability
- Higher scientific potential calls for an higher number of potential users



# Bringing code to data: science platforms, a few examples



# Bringing code to data: science platforms, a few examples



## Portal

Discover data in the browser



[Learn more about the portal.](#)

## Notebooks

Process and analyze LSST data with Jupyter notebooks in the cloud



[Learn more about notebooks.](#)

## APIs

Learn how to programatically access data with Virtual Observatory interfaces



# Bringing code to data: science platforms, a few examples



**CTAO**

Welcome to **ctao**

Sign in with your ctao credentials

oates

.....

Sign in

Forgot your password?

Or sign in with

eduGAIN

Not a member?

Apply for an account

Register an account with eduGAIN

[Info and Privacy Policy](#)

Search Res **Basket** Jobs Query Store Profile

Crab analysis

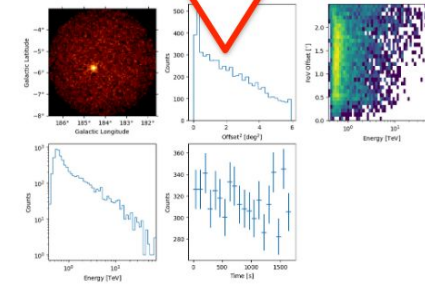
Obs. id: 23523	Target Name: Crab Nebula	T_min: 04/12/2004 22:08:10 TT	<input type="button" value="View"/>	<input type="button" value="Delete"/>
Obs. id: 23526	Target Name: Crab Nebula	T_min: 04/12/2004 22:54:04 TT	<input type="button" value="View"/>	<input type="button" value="Delete"/>
Obs. id: 23559	Target Name: Crab Nebula	T_min: 06/12/2004 23:05:15 TT	<input type="button" value="View"/>	<input type="button" value="Delete"/>
Obs. id: 23592	Target Name: Crab Nebula	T_min: 08/12/2004 21:55:00 TT	<input type="button" value="View"/>	<input type="button" value="Delete"/>

Other Baskets

ADASS demo

Create New Basket

New basket name

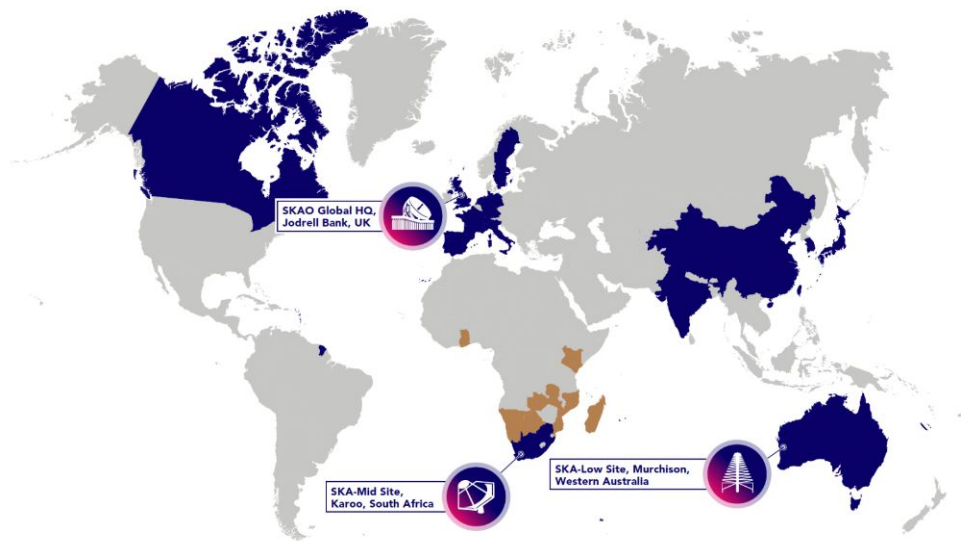


→ Assess the relevance of the basket



# The SKAO Milestones

- December 2011: Establishment of SKA Organization (Italy founding member)
- 15 January 2021: SKA Observatory enters into force (Italy confirmed among 6 founding members)
- 1 July 2021: SKAO Construction begins (Italy gets first contracts)



Phase	SKA-Mid (Dishes)	SKA-Low (Stations)	Key Milestone / Status
AA0.5	4	4	First operational fringes & verification
AA1	8	18	End-to-end system testing
AA2	68	64	Comparable to VLA/LOFAR sensitivity
AA*	144	307	"Staged Delivery" (Current Funded Baseline)
<b>AA4</b>	<b>197</b>	<b>462*</b>	<b>Full Design Baseline</b>

■ SKAO Partnership - includes SKAO Member States\* and SKAO Observers (as of June 2022)



■ African Partner Countries

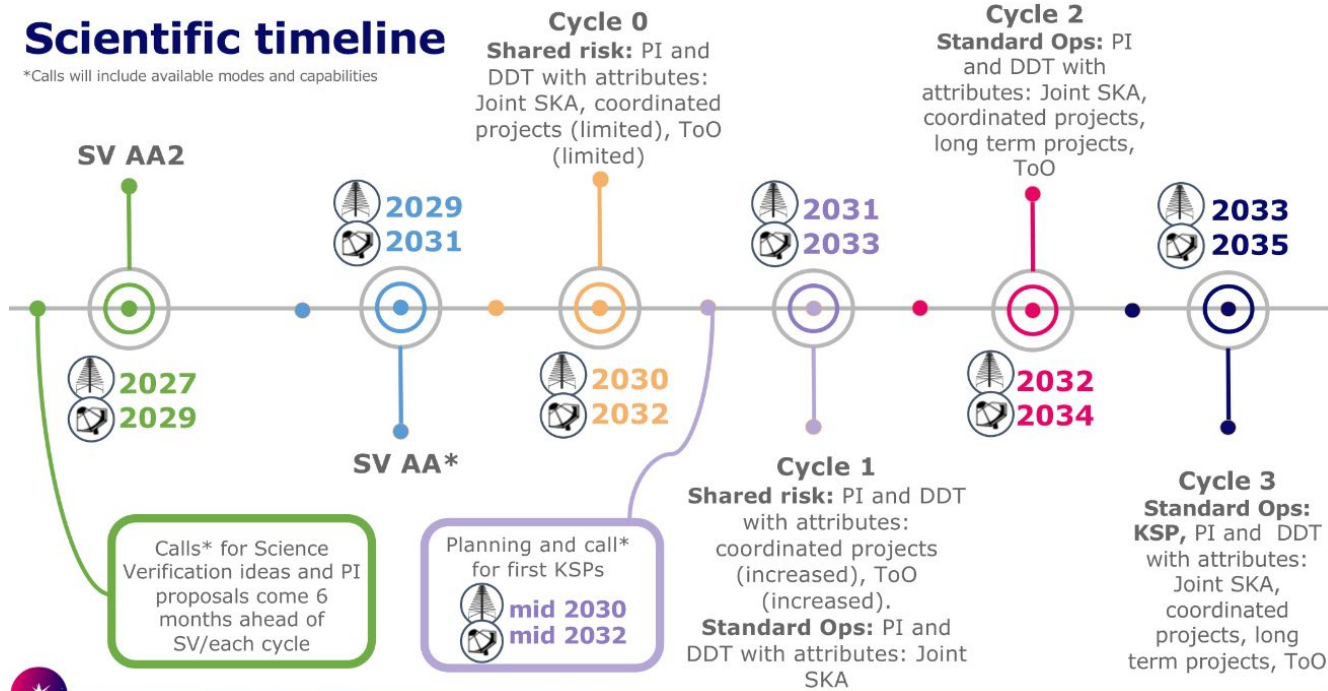


# The SKAO Milestones

- December 2011: Establishment of SKA Organization (Italy founding member)
- 15 January 2021: SKA Observatory enters into force (Italy confirmed among 6 founding members)
- 1 July 2021: SKAO Construction begins (Italy gets first contracts)

## Scientific timeline

\*Calls will include available modes and capabilities



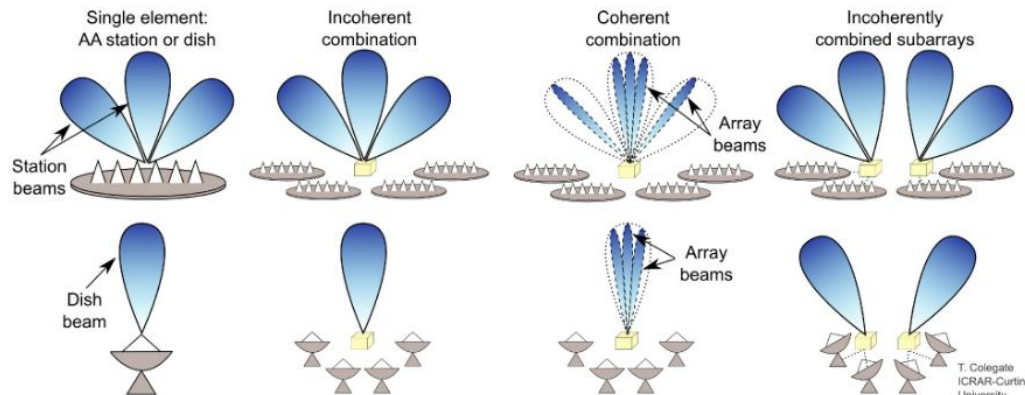
# The extreme case of SKA

Examples on how flexible the observing modes can be for both SKA-LOW and MID.

Telescope	Search		Timing		Bandwidth (Max)
	Beams	Subarrays	Beams	Precision (1 sigma)	
SKA1-Mid	1500	up to 16	16 (8 on B5)	5 ns	300 MHz
SKA1-Low	500	up to 16	16	10 ns	300 MHz

Possible simultaneously:

- imaging
  - VLBI
  - pulsar search
  - pulsar timing
- via commensal / sub-arrays



# The extreme case of SKA

SKA capabilities both in terms of fov and spatial/temporal/energy-resolutions will easily produce multi-TB data for a single SDP product associated to an Execution Block (EB)

Nominal frequency	110 MHz	300 MHz	770 MHz	1.4 GHz	6.7 GHz	12.5 GHz
Range [GHz]	0.05-0.35	0.05-0.35	0.35-1.05	0.95-1.76	4.6-8.5	8.3-15.4
Telescope	Low	Low	Mid	Mid	Mid	Mid
FoV [arcmin]	327	120	109	60	12.5	6.7
Max. Resolution [arcsec]	10	3.5	0.55	0.3	0.06	0.03
Max. Bandwidth [MHz]	300	300	700	810	3900	2 x 2500
Cont. rms, 1hr [ $\mu$ Jy/beam] <sup>a</sup>	26	14	4.4	2	1.3	1.2
Line rms, 1hr [ $\mu$ Jy/beam] <sup>b</sup>	1850	800	300	140	90	85
Resolution range for Cont. & Line rms. [arcsec] <sup>c</sup>	12-600	6-300	1-145	0.6-78	0.13-17	0.07-9
Channel width [kHz]	5.4	5.4	13.4	13.4	80.6	80.6
Spectral zoom windows x narrowest bandwidth [MHz]	4 x 3.9	4 x 3.9	4 x 3.1	4 x 3.1	4 x 3.1	4 x 3.1
Finest zoom channel width [Hz]	226	226	210	210	210	210

Table Notes:

(a) Line sensitivity assumes fractional bandwidth per channel of  $Dn/n = 10^{-4}$  ( $>10^{-6}$  will be possible)

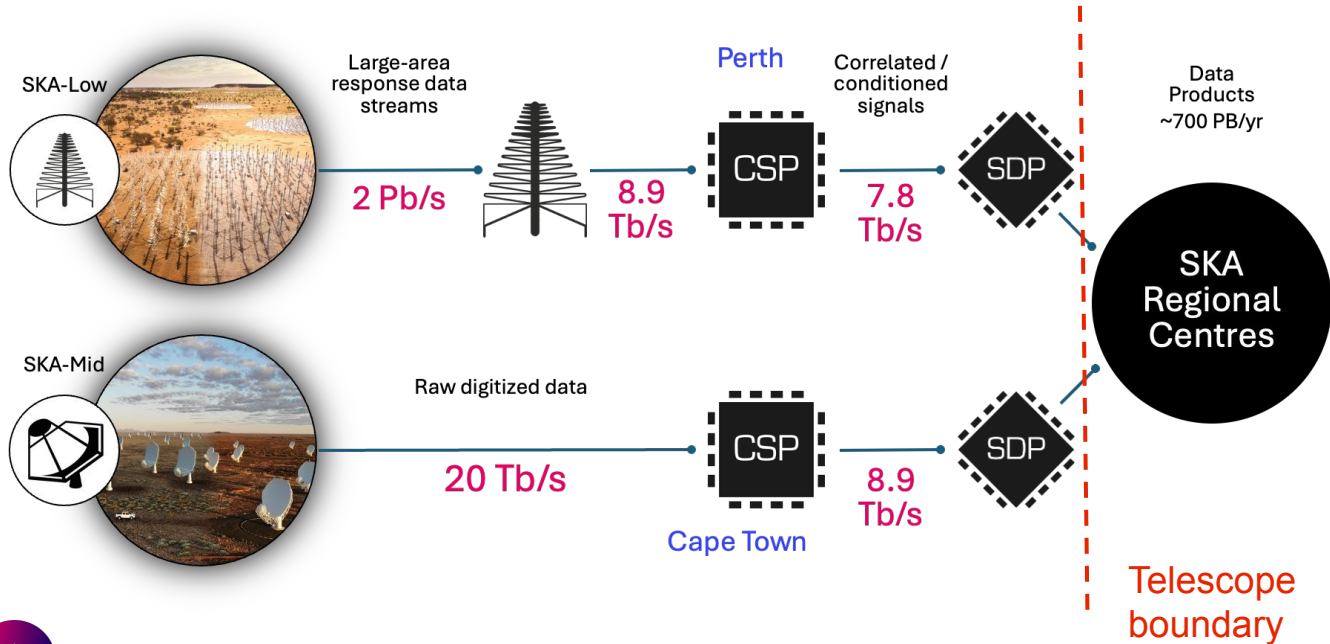
(b) Continuum sensitivity assumes fractional bandwidth per channel of  $Dn/n = 0.3$

(c) The sensitivity numbers apply to the range of beam sizes given by Min. and Max. beam sizes



# The extreme case of SKA

- Raw data can only survive few days in the system, and SDP will deliver up to 700 PB/year (foreseen for SKA-phase 1) to SRCNet.
- SKA Telescope construction and SRCNet development are necessarily coupled.



# What are the SKA Regional Centres?

SRC | Net

SKAO Regional Centre Network

**Science Gateway**, giving access to  
**Science enabling tools and applications**  
running on federated compute and storage

enabling users to discover data in the global SKA archive, develop workflows, perform analyses and collaborate, while supporting and training the community

addresses the "orders of magnitude" data problem and other challenges posed by interfaces (towards SKAO, and national HPC centres, community engagement)

**maximise the scientific productivity and impact of the SKA**



# SRCNet core services



**Science Enabling Applications**  
Analysis Tools, Notebooks,  
Workflows execution  
Machine Learning, etc



**Data Discovery**  
Discovery of SKA data from the  
SRCNet, local or remote,  
transparently to the user

**Support to Science Community**  
Support community on SKA data  
use, SRC services use, Training,  
Project Impact Dissemination



**Data Management**  
Dissemination of Data to SRCs  
and Distributed Data Storage



**Distributed Data Processing**  
Computing capabilities provided  
by the SRCNet to allow data  
processing

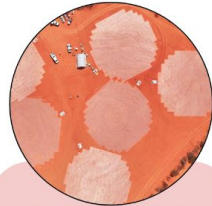
**Visualization**  
Advanced visualizers for SKA  
data and data from other  
observatories

**Interoperability**  
Heterogeneous SKA data from  
different SRCs and other  
observatories



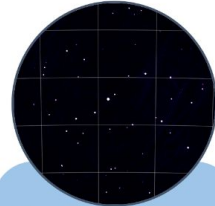
# SKA Telescope and SRCNet Science Delivery Roadmaps

## SKAO milestones



### Construction

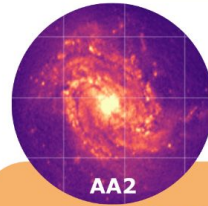
- Building antennas, dishes, roads etc!
- Followed by Assembly, Integration and Verification



### Commissioning

- SKAO activity
- Collaborative across system verification and science commissioning

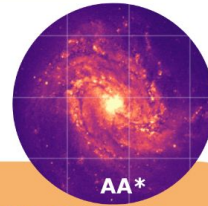
Community involvement starts



AA2

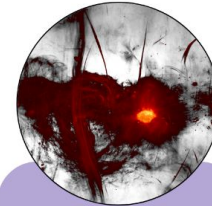
### Science Verification

- A full dress rehearsal of the end-to-end system for every mode of operation
- Modes and capabilities rolled out over time
- Data will be publicly available for scrutiny
- Number 1 goal is verification, science secondary
- Feedback crucial, for SKAO and SRCNet



AA\*

Credit: I. Heywood, SARAO



### Cycle 0

- Shared-risk PI projects
- SRCNet resources ready for user
- proprietary periods



Now



Now



SRCNet real users doing real things!!! →

Now

First half 2027

First half 2029

2030

Ramping up

First half 2029

First half 2031

2032

SRCNet Development project is to be extended by the end of 2030



# Science Delivery missions

## 1. User Representation

- **Mission:** Act as the **voice of the user community** for SRCNet functionality.
- **Key Activities:**
  - Gathering, consolidating, and translating **user requirements and feedback** into actionable work items.
  - Prioritising development based on **scientific impact**.
  - Conducting **market research** on similar tools and emerging technologies.

## 2. User Support

- **Mission:** Provide a **single point of entry for user assistance** via the unified SKA Helpdesk.
- **Key Activities:**
  - Answering questions on **SKA data, processing, access, and proposal processes**.
  - Supporting the use of **SRCNet services and scientific software**.
  - Generating and maintaining **user documentation, FAQs, and training materials** (tutorials/videos).

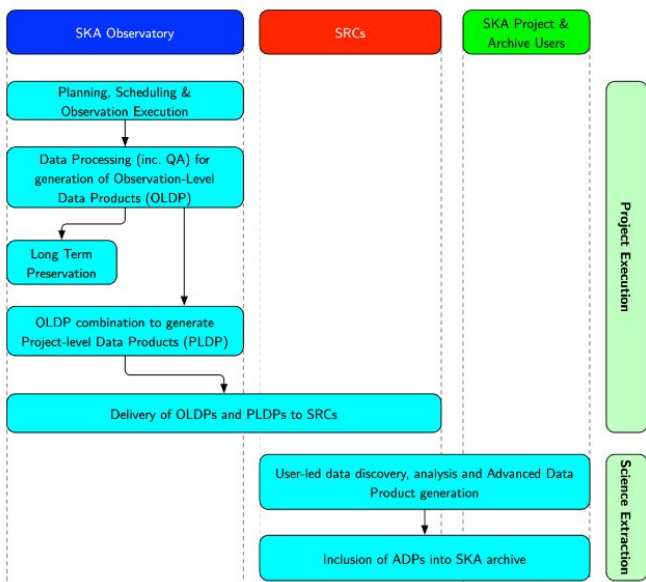
## 3. Workflows

- **Mission:** Enable and support **scientific data processing** on SRCNet.
- **Key Activities:**
  - Creating, maintaining, and compiling **template workflows** for common use cases.
  - Distributing **reusable, interoperable software libraries**.
  - Supporting the integration and execution of **SKAO pipelines** (PLDPs) and **user-generated workflows** (ADPs).



# Data product types identification/classification

- Relevant documents for Observatory Data Products (ODPs) are:
  - SKA-TEL-SKO0001818, SKAO Science Data Products: a summary
  - Science Data Processor anticipated data products: A quick guide for SWG members

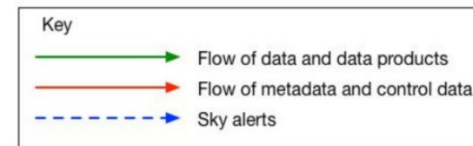
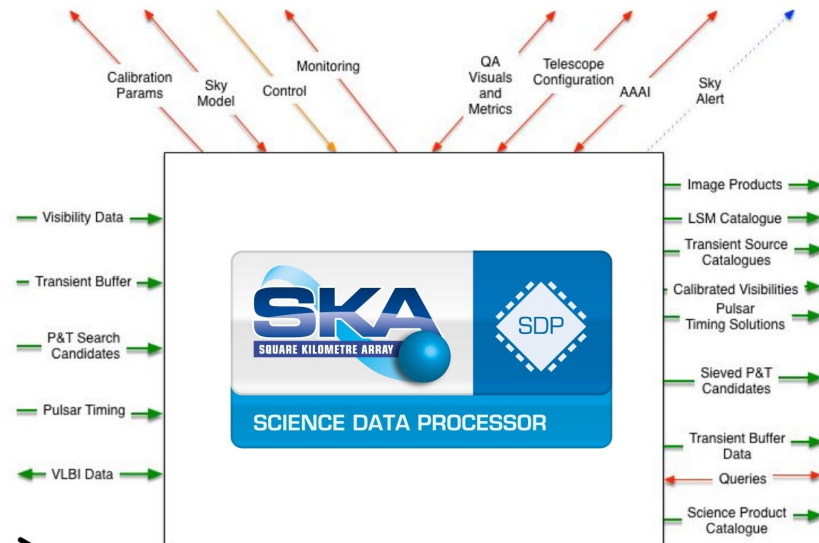
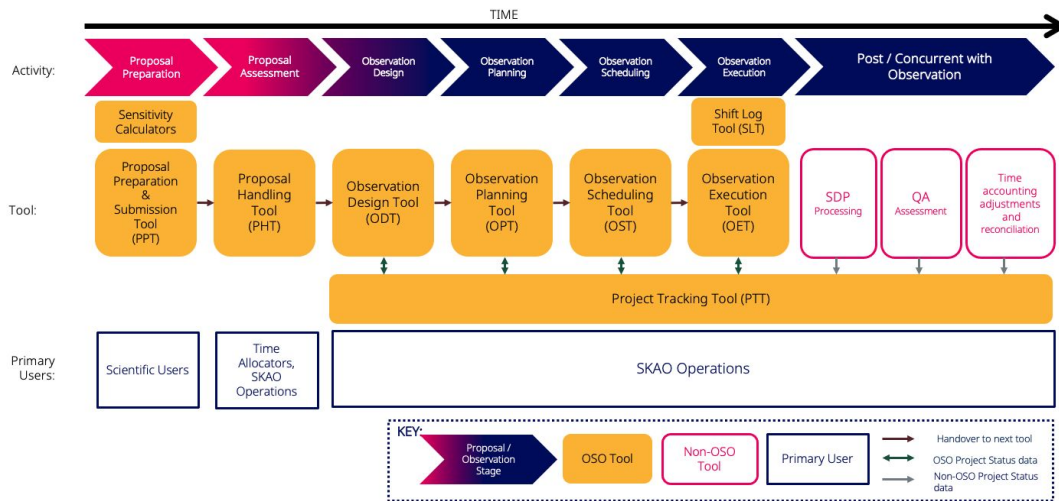


- ❑ Image Products 1: Image Cubes
- ❑ Image Products 2: uv gridded visibilities
- ❑ Calibrated visibilities
- ❑ LSM Catalogue
- ❑ Imaging Transient Source Catalogue
- ❑ Pulsar Timing Solutions
- ❑ Transient Buffer Data
- ❑ Sieved Pulsar and Transient Candidates
- ❑ Science Alerts Catalogues
- ❑ Science Product Catalogue



# Towards SKA: data product types identification/classification

- Relevant documents for Observatory Data Products (ODPs) are:
  - [SKA-TEL-SKO0001818](#), SKAO Science Data Products: a summary
  - [Science Data Processor anticipated data products: A quick guide for SWG members](#)



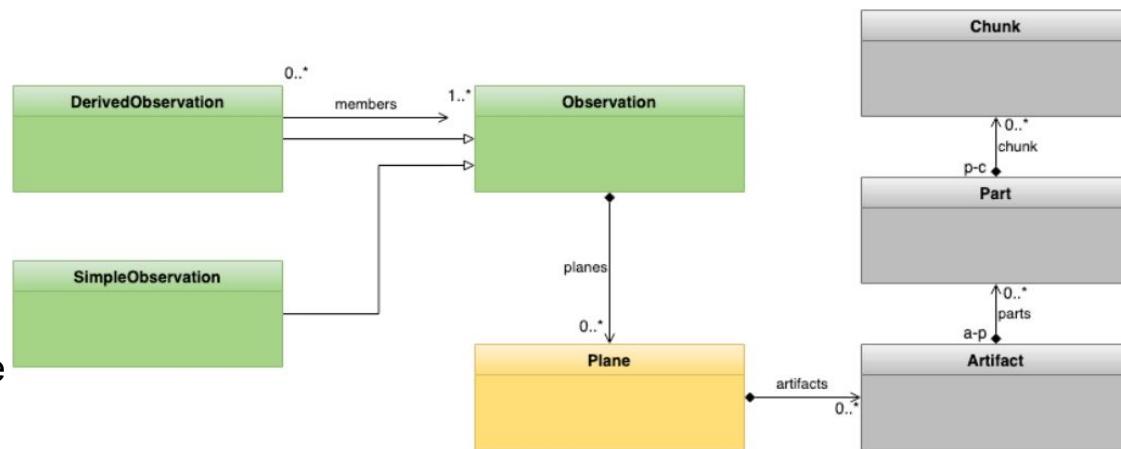
Data and metadata fluxes through SDP  
(cf. Fig. 3 of [SKA1 SDP High Level Overview](#))

# Towards the SKA Science Data Archive: CAOM

- The Common Archive Observation Model (CAOM) is an open source metadata framework, developed as a collaborative effort among STScI, ESAC and CADC.
- Designed to describe and organize observational metadata in archives.
- Enables efficient search and retrieval of astronomical data.
- Tracks data provenance, relationships, and essential metadata.

## Hierarchical Structure:

- **Observation:** A data acquisition event (telescope + instrument).
- **Plane:** Describes data products (image, cube, spectrum). Includes calibration level (raw, processed).
- **Artifact:** Physical storage (files, database entries).
- **Part:** a logical subcomponent of an Artifact (e.g. a FITS extension).
- **Chunk:** a single data array using WCS



## Prototype 4: Visualization of SKA data with high volume of users and high amount of data

Instantiation and deployment of visualisation services such as VisIVO, Aladin and CARTA or other visualisers for a high user demand and data volume environment, using flexible local computational resources. This includes storage sharing, data access, federated authentication, and integration with workload manager.

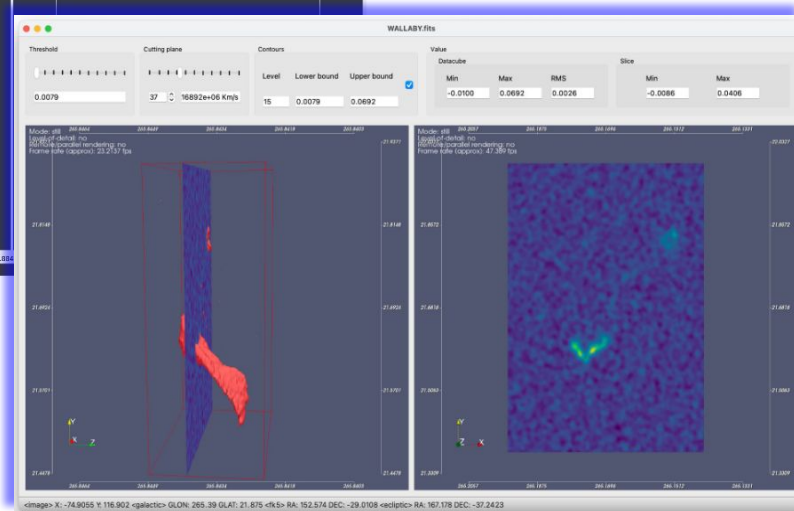
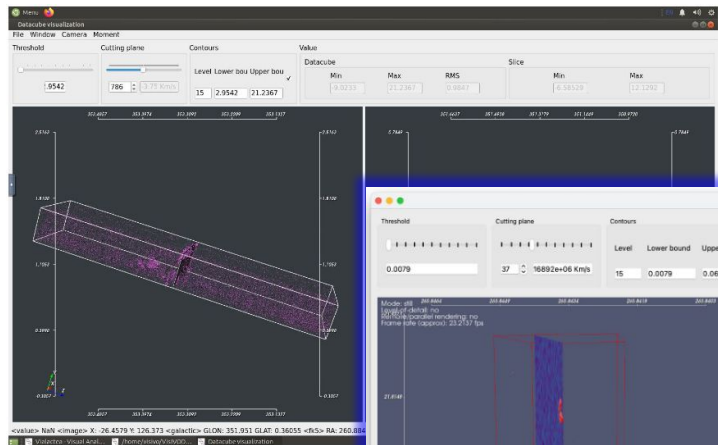
Also, includes the analysis for local and remote data locations (coupling with prototype 1, i.e. *Data products replication, distribution, and synchronization across multiple locations*)

Team member	Role	SRC
Fabio Vitello	<a href="#">Product Owner</a>	ITSRC
Giuseppe Tudisco	<a href="#">Scrum Master</a>	ITSRC
Eva Sciacca	Member	ITSRC
Vincenzo Galluzzi	Member	ITSRC
Alessandra Zanichelli	Member	ITSRC
Robert Butora	Member	ITSRC
Matthieu Baumann	Member	FRSRC
Marco Molinaro	Observer	ITSRC
François Bonnarel	Member	FRSRC
Thomas Boch	Observer	FRSRC
Nicolas Moreau	Observer	FRSRC
Kechil Kirkham	Observer	SASRC



# Visualization prototype

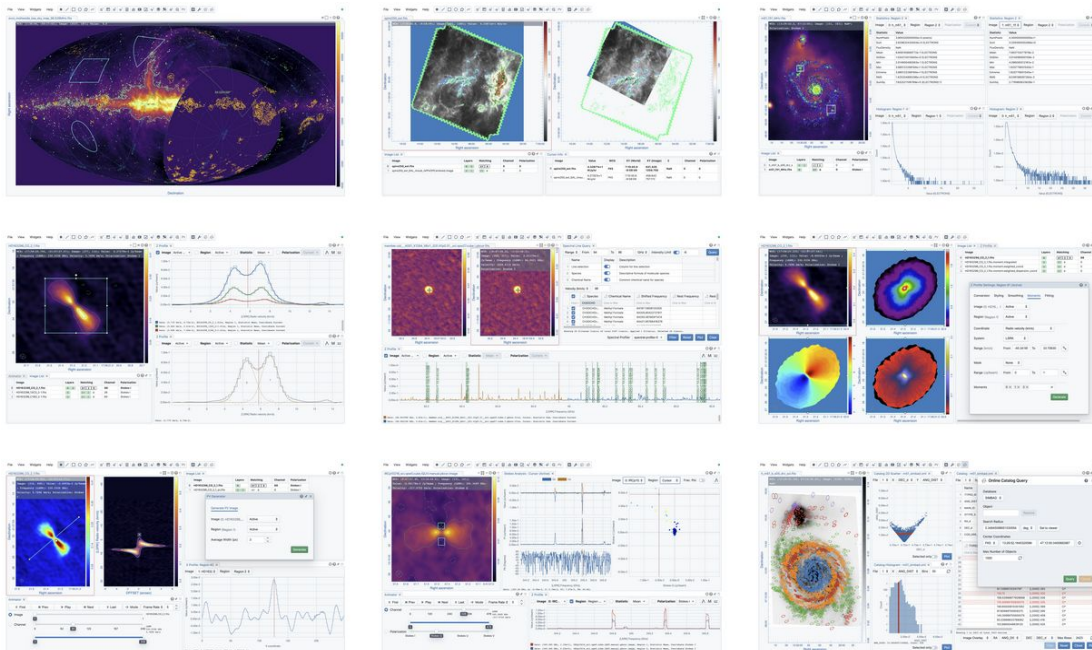
Improve existing visualizations (VisIVO, Aladin and CARTA) tools based on use cases collected by WG6 (e.g. adopting a client-server architecture, optimization for HPC environment, pythonic interface, enhancing 3D visualization features, enabling VO services exploitation for discovery and data access, e.g. SODA)





# Visualization prototype

Improve existing visualizations (VisIVO, Aladin and CARTA) tools based on use cases collected by WG6 (e.g. adopting a client-server architecture, optimization for HPC environment, pythonic interface, enhancing 3D visualization features, enabling VO services exploitation for discovery and data access, e.g. SODA)



## CARTA



Academia Sinica Institute of  
Astronomy and Astrophysics  
(ASIAA)



Inter-University Institute for  
Data Intensive Astronomy (IDIA)



National Radio Astronomy  
Observatory (NRAO)



Department of Physics,  
University of Alberta



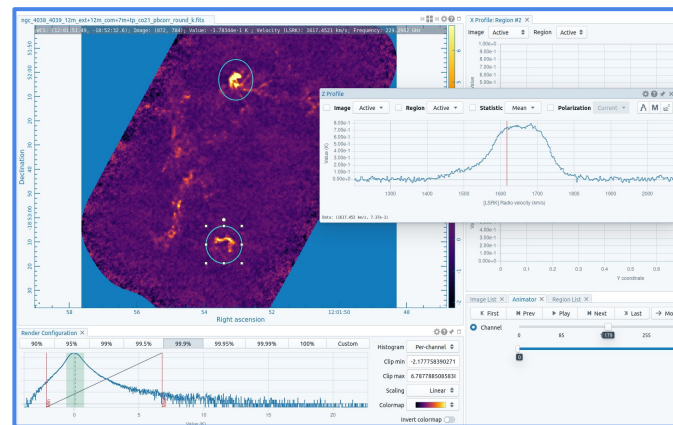
# Visualization prototype

- Improve existing visualizations (VisIVO, Aladin and CARTA) tools based on use cases collected by WG6
- Containerization or pythonic interface for using them inside Science Platforms (e.g. CANFAR and Azimuth)

The screenshot shows the Science Portal interface. At the top, there is a navigation bar with the CANFAR logo and links for Documentation, Services, About, Open Source, Support, and a user profile for vgaluzzi. The main content area is divided into several sections:

- Active Sessions:** A list of two active sessions. The first is 'carta1' (skaha/cartalatest) started on 2024-12-06 06:33 UTC, with 7M memory and 0.001/4 CPU cores. The second is 'notebook1' (skaha/astroml/latest) started on 2024-12-06 08:57 UTC, with <none> memory and <none>/2 CPU cores.
- New Session:** A form to create a new session. The 'Standard' tab is selected. Fields include: type (notebook), project (skaha), container image (astroml/latest), session name (notebook1), memory (8), and # cores (2). There are 'Launch' and 'Reset' buttons.
- Platform Load:** A section showing resource usage. Available CPU is 797.0 / 1798.0. CPU usage is shown as a bar chart. Available RAM is 509.67GB / 6632.67GB. Memory usage is shown as a bar chart. Running instances are 441, broken down by session, desktopApp, and headless. The last update is 2024-12-06 08:53 UTC.

## CARTA



# Visualization prototype

- Improve existing visualizations (VisIVO, Aladin and CARTA) tools based on use cases collected by WG6
- Containerization or pythonic interface for using them inside Science Platforms (e.g. CANFAR and Azimuth)

The screenshot displays the 'iris' dashboard interface. The top navigation bar includes the 'iris' logo, 'Other Clouds', 'Cloud Metrics', 'Documentation', and 'Sign out (vgalluzzi)'. A left sidebar contains navigation options: 'Platforms', 'Identity Provider', 'Quotas', 'Project metrics', 'Advanced', 'Switch tenancy', and 'SSH public key'. The main content area is titled 'ska-src-cambridge' and features a grid of platform cards. Each card includes a star icon, a 'READY' status, an icon (Tux penguin or Slurm logo), a name, a description, and a 'Details' button. The cards are:

- bb23-1gs**: Linux Workstation (with SSH access), Web console, Monitoring, Created 2 mesi fa, Expires tra 9 mesi.
- jw2**: Kubernetes, Kubernetes Dashboard, Monitoring, Created 27 giorni fa, Expires tra 2 mesi.
- la25-small**: Linux Workstation (with SSH access), Web console, Monitoring, Created 2 mesi fa, Expires tra 9 mesi.
- la25-small-v2**: Linux Workstation (with SSH access), Web console, Monitoring, Created 1 mese fa, Expires tra 5 mesi.
- mvdw-lofarint**: Slurm (CephFS home), Open OnDemand, Monitoring, Created 2 mesi fa, Expires tra 9 mesi.
- rucio-data-handling-1gs**: Linux Workstation (with SSH access), Web console, Monitoring, Created 15 giorni fa, Expires tra 11 mesi.
- test-di-service**: Linux Workstation (with SSH access), Web console, Monitoring, Created 2 mesi fa, Expires tra 2 anni.
- test-jr**: Linux Workstation (with SSH access), Web console, Monitoring, Created 11 giorni fa, Expires tra 2 anni.

At the top right of the grid, there are buttons for 'New platform' and 'Refresh'.



# Test workflows

- Plot of cylindrically-averaged power spectrum (Figure 9)
- Plot of spherically-averaged power spectrum (Figure 10)

## SWF-001 - Visualize Imaging Data Products

## SWF-002 - Catalogue Search and Crossmatch

## SWF-003 - Smoothing and Resampling

## SWF-005 - Retrieve and import archival data including postage stamps

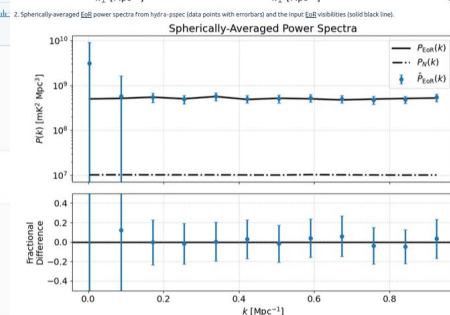
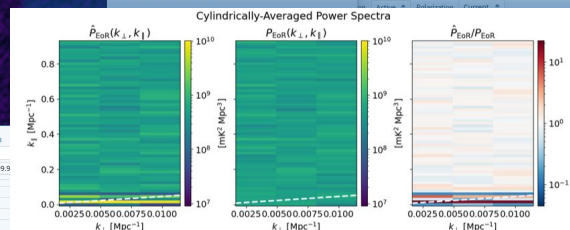
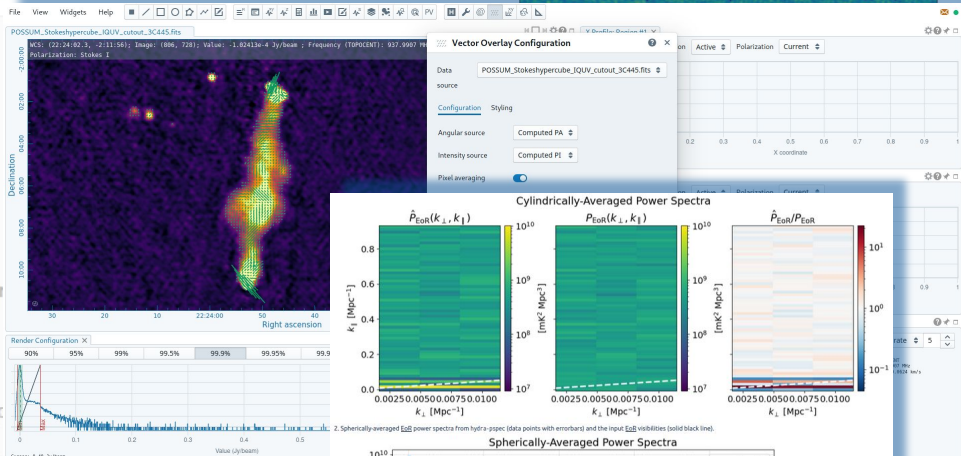
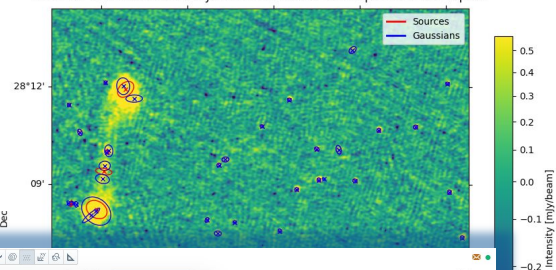
## SWF-006 - Change sky projection

## SWF-008 - Systematics investigation and EoR power spectrum estimation

## SWF-010 - Continuum source finding

## SWF-012 - Image cube inspection

Cutout of LoTSS mosaic with PyBDSF source and Gaussian positions and ellipses



```
In [1]: import numpy as np
from astropy import units
from astropy.cosmology import LambdaCDM
from astropy.cosmology import Planck15
from astropy.units import Quantity
from astropy.time import Time
from pyuvdata import UVData
from pyuvdata.analysis import utils
from pyuvdata.analytic_beam import AiryBeam
from scipy.signal.windows import blackmanharris

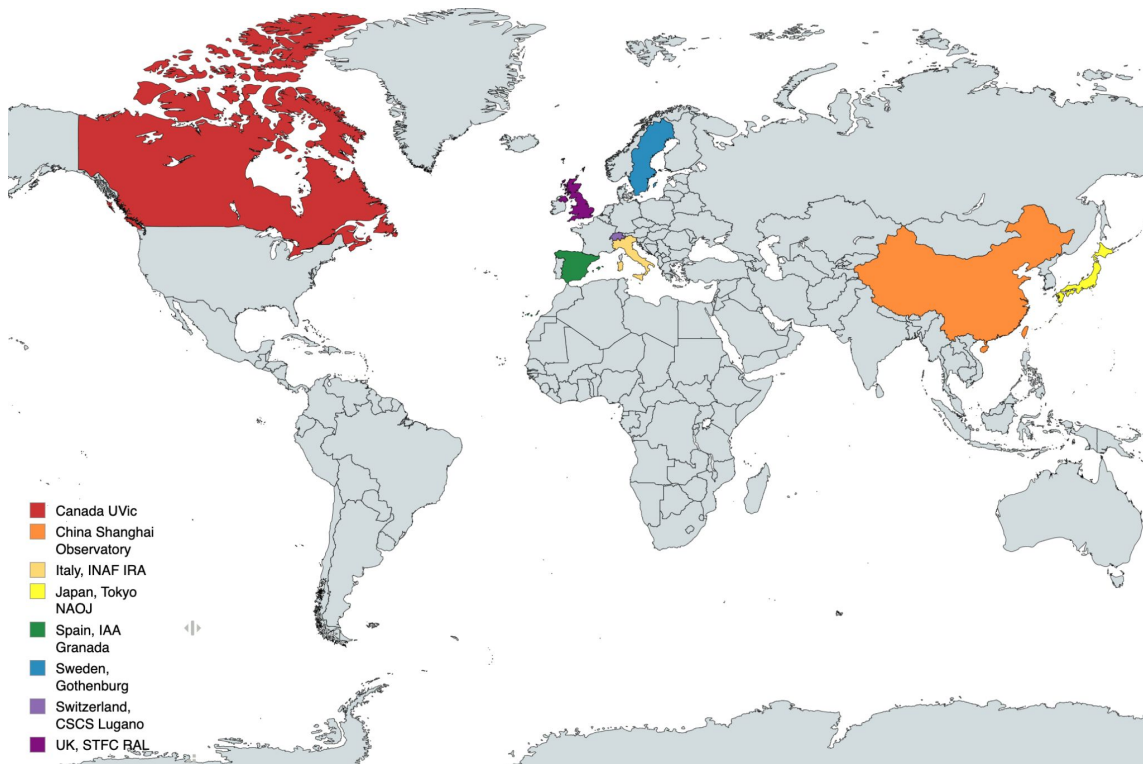
import matplotlib.pyplot as plt
from matplotlib.colors import LinearSegmentedColormap
from mpl_toolkits.axes_grid1 import ImageGrid
```

```
In [2]: plt.rcParams["font.size"] = 14
plt.rcParams["image.interpolation"] = "None"
plt.rcParams["figure.facecolor"] = "w"
```

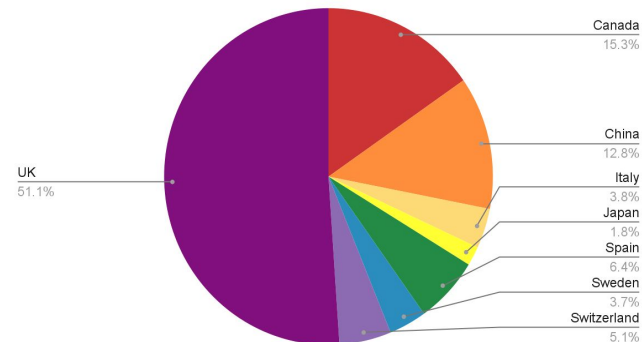


# SRCNet v. 0.1 sites

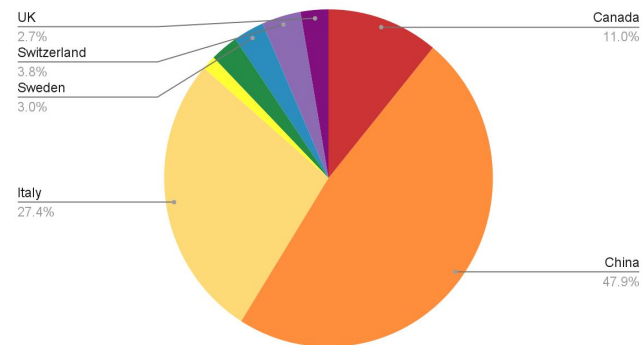
- 8 sites contributed with computing and storage resources to SRCNet0.1, for 2025 testing period
- 7.8 PBytes total storage; 0.5 PFLOPS computing resources



Storage fraction for SRCNet0.1



Compute fraction for SRCNet0.1



# Detailed Science Cases and Science Workflows

- Project Scientist worked a number of DSC (starting from those for AA2 Low)
- Starting from PI28 (September 2025) the Support Scientist (Teal) Team worked out:

Science Working Group	Key	No DSCs
Cosmology	CO	1
Cradle of Life	CoL	4
Epoch of Reionisation	EoR	3
Extragalactic Continuum	EC	7
Extragalactic Spectral Line	ESL	0
Gravitational Waves	GW	1
High Energy Cosmic Particles	HECP	0
HI Galaxy Science	HIGS	1
Magnetism	MAG	5
Pulsars	PUL	4
Our Galaxy	OG	4
Solar, Heliospheric, Ionospheric Physics	SHI	2
Transients	TR	2
VLBI	VLBI	0
Total number of cases		34

**DSC-035** Timing a know strong pulsar with Low

**DSC-037** Cable reflection systematics for EoR science (*Look for the presence of cable reflections which could prohibit EoR power spectrum detection. Identify any new/unknown systematics in the process*)

**DSC-043** Associated HI absorption in a high redshift galaxy

**DSC-057** Low-frequency continuum observation of an equatorial deep field

The work included creating the gitlab repository and necessary documentation (jupyter notebook and tutorials, e.g. for interactive visualization)

During last SRCNet Workshop Project Scientists walked through a selection of different SWFs by exploiting resources from sweSRC, espSRC and ukSRC.



# DSC-057: Low-frequency continuum observation of an equatorial deep field

Files available	Description
cube_11ch.fits	11 Channel <u>FITS</u> cube of simulated data. <u>Frequency</u> range $1.035 - 2.396 \times 10^8 \text{ Hz}^{-1}$ .
lofar-catalogue-filtered-sp5984.fits	PyBDSF output catalogue of the given mosaic.
lofar-mosaic-cropped-sp5984.fits	Cropped <u>LOFAR</u> mosaic which represents the <u>SV</u> data.

Step 1 Login and data access

Step 2 Initial data inspection

Step 3 Generate extra image data products

Step 4 Source findings and catalogue generation

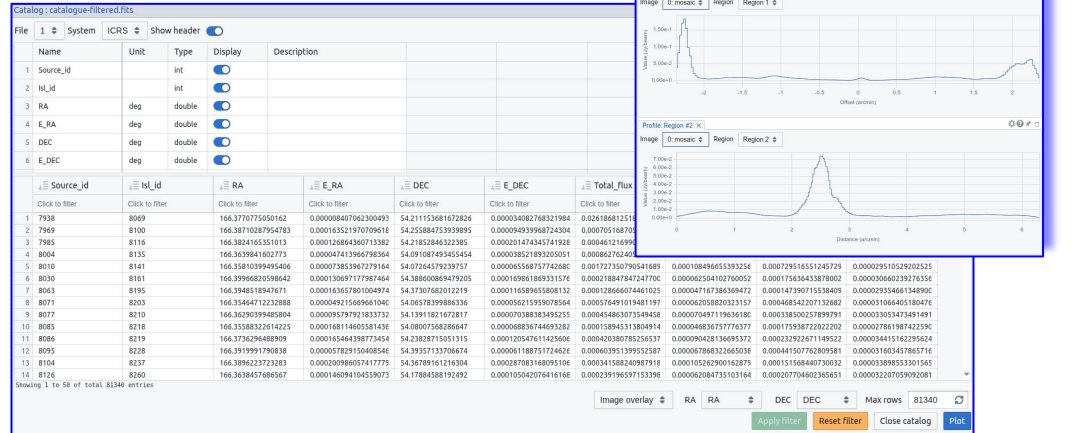
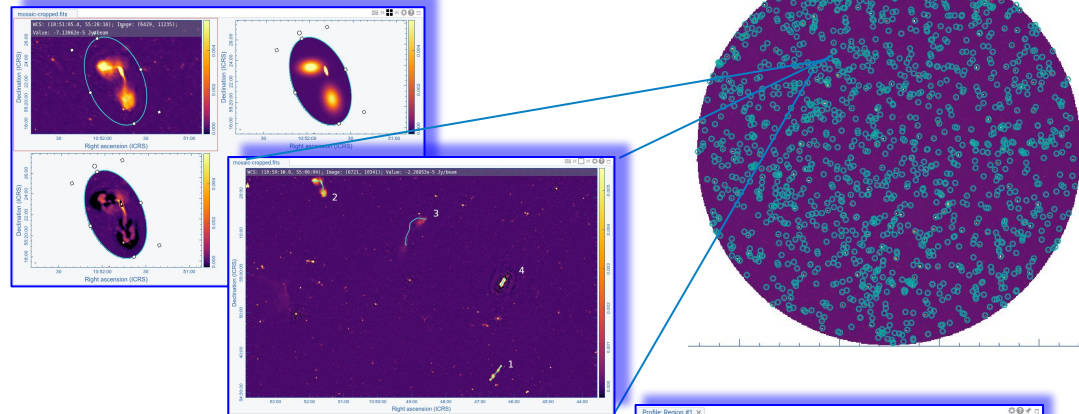
Step 5 Source catalogue cross-matching

Step 6 Radio spectrum analysis

Step 7 Differential source counts

Step 8 Final verification assessment

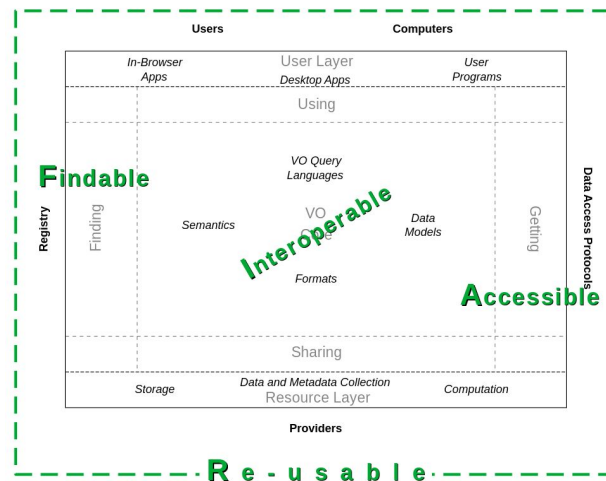
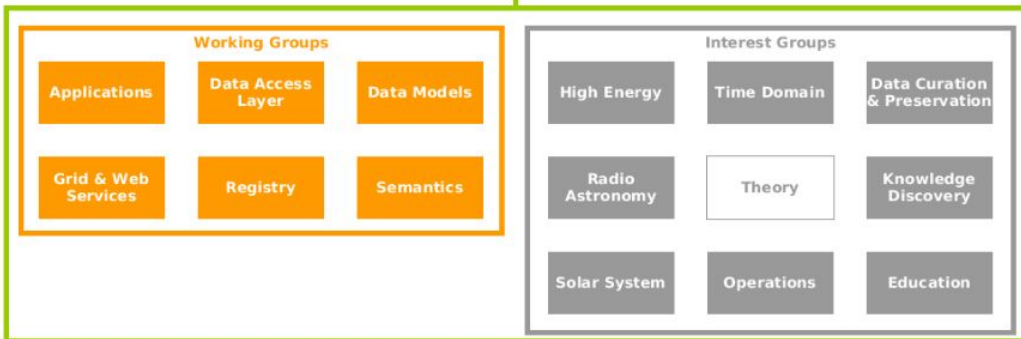
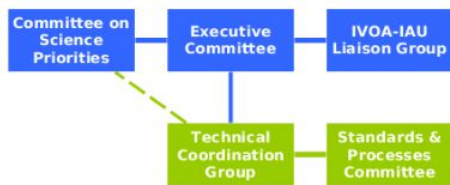
Step End ADP ingested in the SKA archive and SRCNet are freed



# The International Virtual Observatory Alliance

The Virtual Observatory (VO) is the vision that astronomical datasets and other resources should work as a seamless whole.

The International Virtual Observatory Alliance (IVOA) is an organisation that debates and agrees the technical standards that are needed to make the VO possible.



# IVOA Radio Interest Group

*(Chair: Mark Kettenis - JIVE)*



Considered data variety and complexity in the radio domain, as well as little integration and utilization within the VO framework, the IVOA Radio Interest Group (RIG) was established in May 2020.

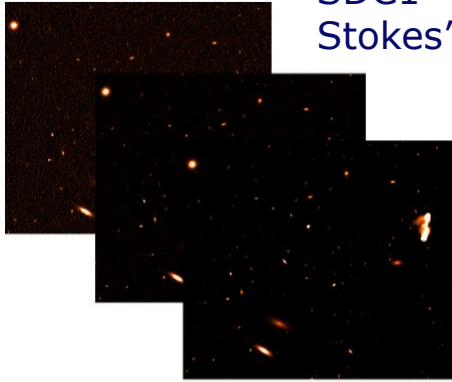
**The Radio Interest Group aims to bridge the gap between radioastronomy specific data needs and the Virtual Observatory framework by:**

- **Defining and promoting radio-specific data handling within the VO.** This includes developing use cases for data exploration, access, and visualization, and identifying necessary metadata standards.
- **Serving as a central point of contact.** The group fosters collaboration between radio astronomy projects and the IVOA, encouraging the adoption of VO standards.
- **Facilitating knowledge sharing.** Through dedicated sessions at IVOA meetings, the group promotes discussion and development related to radio astronomy data.



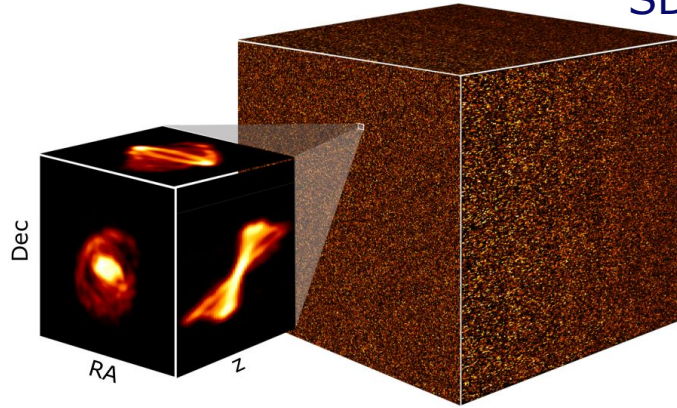
# SKA Data Challenges (SDCs)

SDC1 – Continuum Stokes' I

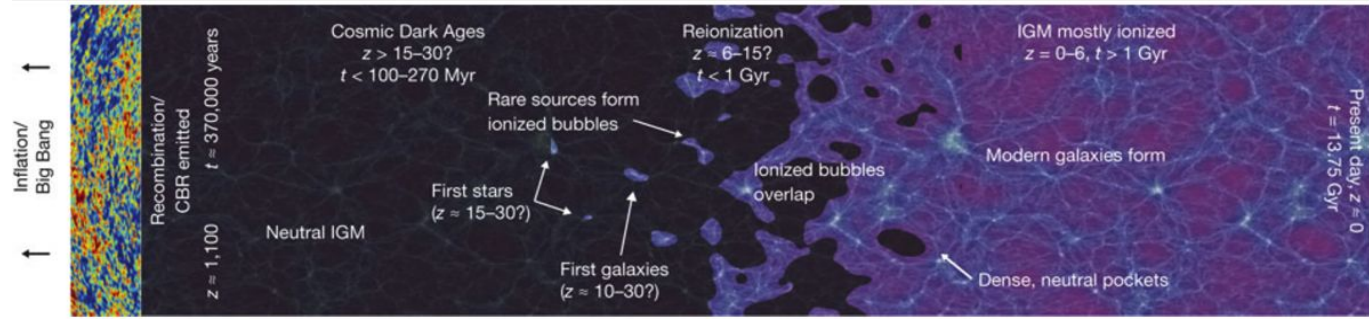


Zoom-in of the 1.4 GHz maps, showing the same region of the sky with different telescope integration: 8, 100, 1000 h left to right.

SDC2 – HI survey



SDC3 a,b – EoR

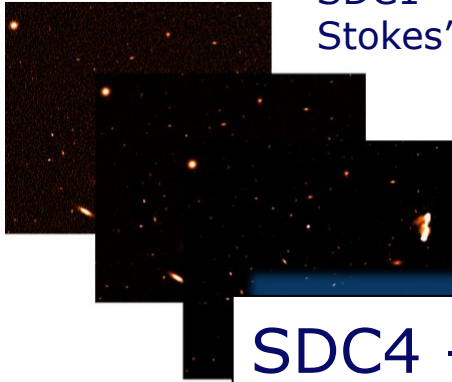


1. Removal of foreground emission from Galactic and Extragalactic sources ('[Foregrounds](#)', or SDC3a)
2. Inference of important parameters of the Epoch of Reionisation ('[Inference](#)', or SDC3b)



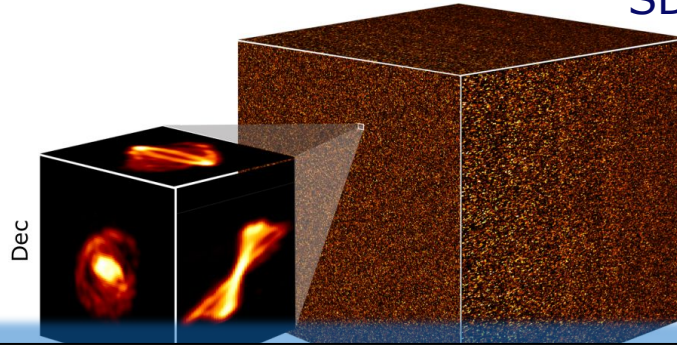
# SKA Data Challenges (SDCs)

SDC1 – Continuum  
Stokes' I



Zoom-in of the 1.4 GHz maps, showing different telescope integration: 8, 100,

SDC2 – HI survey



SDC4 – Cosmic Magnetism is still collecting expression of interests

SDC3 a,b – EoR

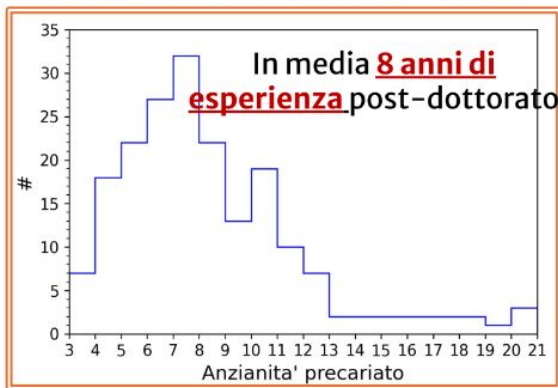


1. Removal of foreground emission from Galactic and Extragalactic sources ('[Foregrounds](#)', or SDC3a)
2. Inference of important parameters of the Epoch of Reionisation ('[Inference](#)', or SDC3b)

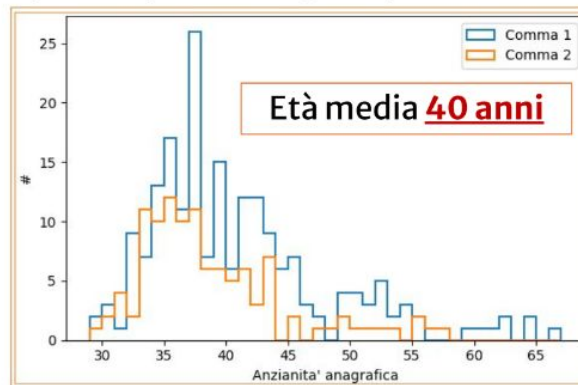


# La situazione del personale precario in INAF è **INSOSTENIBILE!**

**1.200 Tempo Indeterminato** Vs **650** precari: più di 1 precario ogni 2 persone di ruolo



Plot di un campione rappresentativo dei precari INAF al 31/12/2024



Dei **650**, **287** possono essere stabilizzati:  
**173** tramite chiamata diretta (comma 1)  
**114** tramite concorsi riservati (comma 2)

Entro l'anno, l'attuale situazione determinerà l'esodo di > 100 lavoratori altamente qualificati e il MUR se ne lava le mani

È **URGENTE** che INAF **PROCEDA ORA** con le **STABILIZZAZIONI TRAMITE MADIA**: unica soluzione per questa emergenza



Molti colleghi (972) hanno già firmato, per sostenerci e aggiungere il nome alla lista del QR,  
contattaci a [retestabilizzandi1.inaf@gmail.com](mailto:retestabilizzandi1.inaf@gmail.com)



Thank you!

*We recognise and acknowledge the  
Indigenous peoples and cultures that have  
traditionally lived on the lands on which  
our facilities are located.*

**SKAO**

[www.skao.int](http://www.skao.int)