SPECTRUM

Current and future needs in Radio Astronomy Computing

M. Iacobelli (ASTRON) | J. Wagg (Observatoire de la Côte d’Azur)

Oct 2nd, 2024

Funded by
the European Union

SPECTRUM is funded by the European Union - Grant Agreement Number 101131550

Data intensive radio astronomy

Over the past decades, radio astronomy has evolved significantly

Quest for deeper sensitivities, higher resolutions, wider fields of view & exploration of new portions of the spectrum

→ push to build larger and more complex facilities

→ fundamental challenges associated with

- data handling
- data editing

LOFAR - LOFAR ERIC

MEERKAT - SARAO

ASKAP - CSIRO

Data intensive radio astronomy

Over the past decades, radio astronomy has evolved significantly

Quest for deeper sensitivities, higher resolutions, wider fields of view & exploration of new portions of the spectrum

→ push to build larger and more complex facilities

→ fundamental challenges associated with

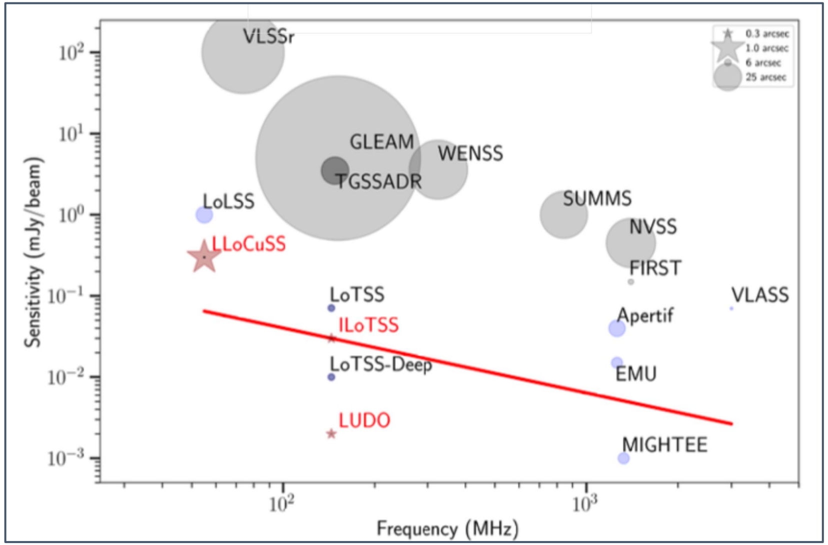
- data handling
- data editing

Avoiding loss of information (especially) at low ν requires:

- transport & storage of large amounts of data
- complex algorithms & workflow chains for data editing

Need for customised science archive services to enable

- data discovery & science exploitation



Astronomers ~~will be~~ are the rare resource (ADASS 2014)

Daniel Durand (ADASS 2015): "The next frontier for science archives is to make the content of the data searchable"

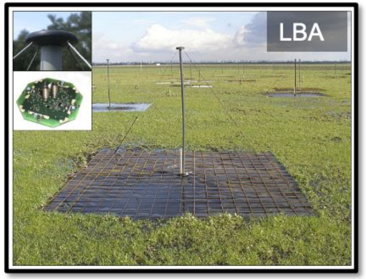
The LOw Frequency ARray (LOFAR) key facts

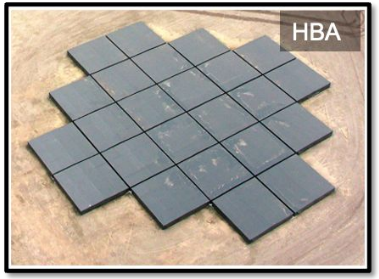
Array of 52 dipole antenna stations operating in the frequency range 10-250 MHz

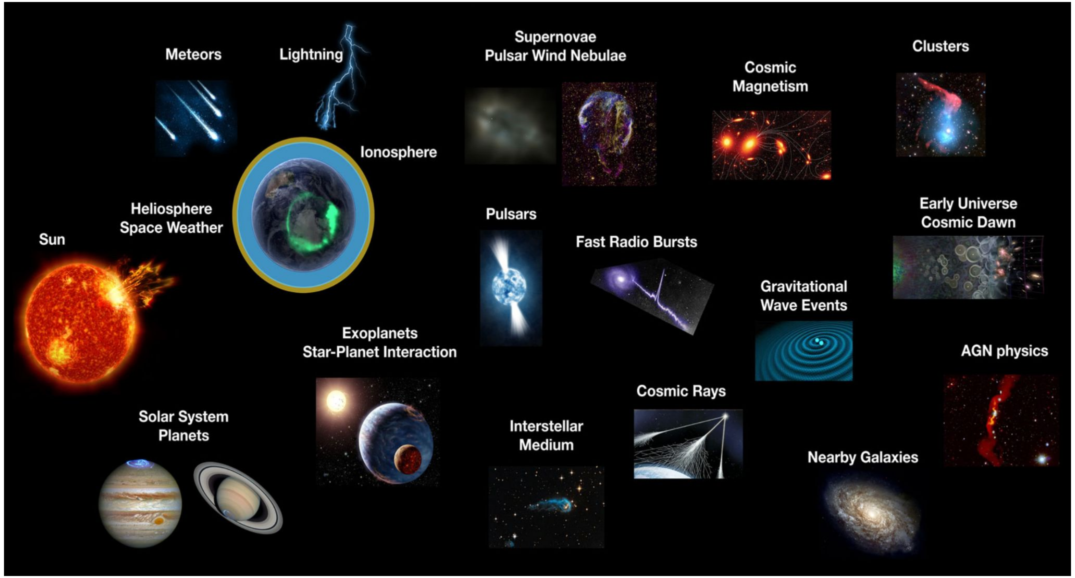
Low Band Antenna (LBA; 4800 dipole pairs, 96 LBA per station, Area ~75200 m^2; 10-90 MHz)

High Band Antenna (HBA; 47616 dipole pairs, 48/96 tiles per station in NL/EU, Area ~57000 m^2; 110-250 MHz)

Several observing modes (imaging, BF, BF+IM, TBB) → 96 MHz bandwidth (multi-beam option)







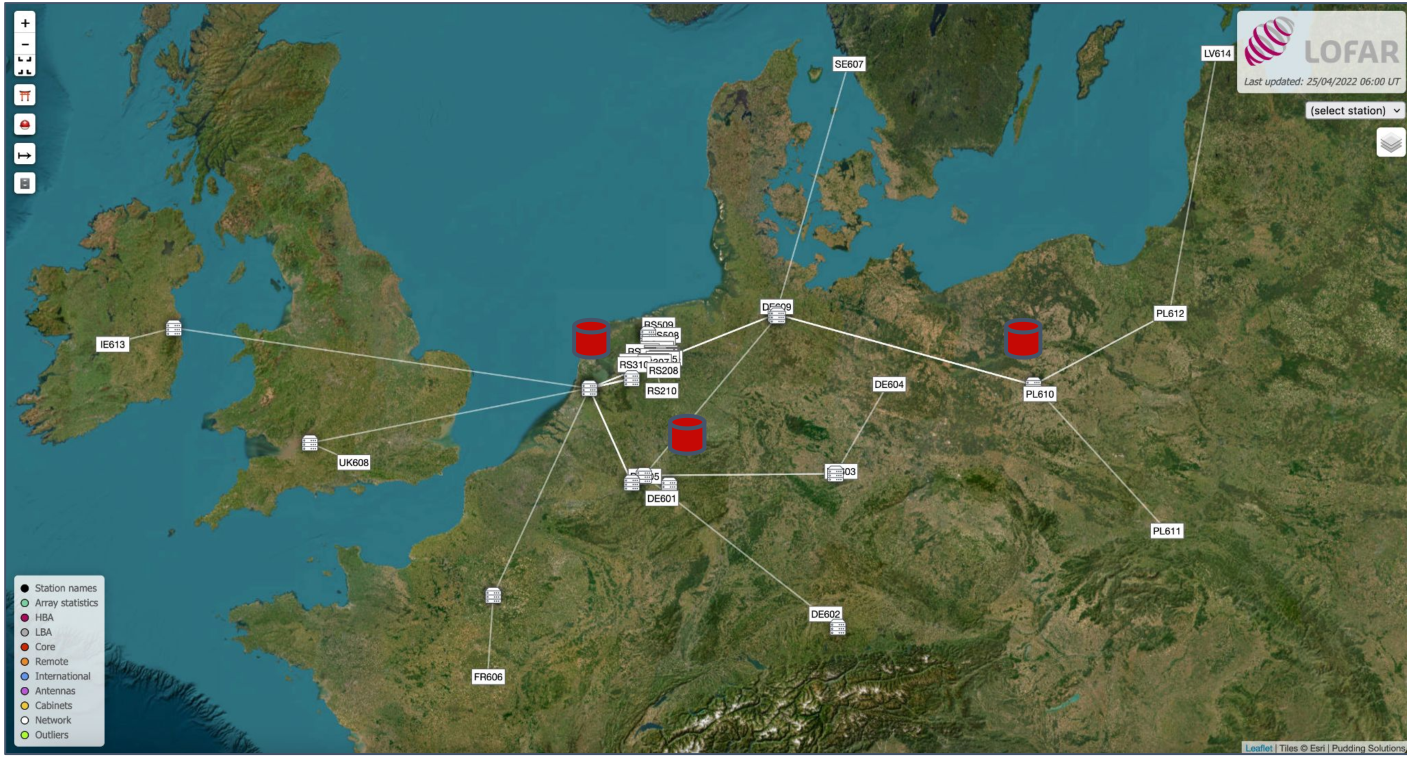
Distributed data flow & computing infrastructure

52 stations across Europe

Science archive with 3 sites

- storage
- computing

Access to resources with the Grid paradigm



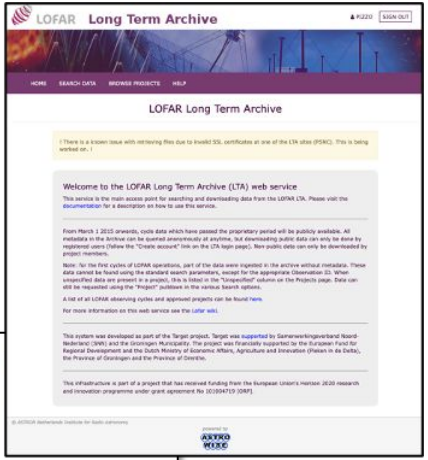
Distributed data flow & computing infrastructure

stations: data collection

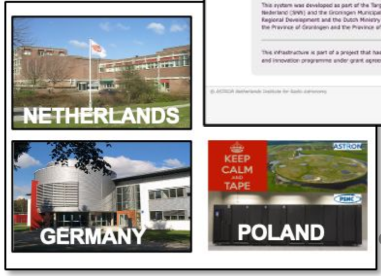
COBALT correlator:
online processing

CEP4 compute cluster

Intermediate data generated through offline processing



Ingest into science archive sites.
dCache system.



Transport, processing and storage of large amounts of data :

- Data flow from all antennas combined: 1.7 Tbyte/s
- To COBALT from station after beamforming: 28 Gbyte/s
- Correlator output to disk: between 2-10 Gbyte/s
- Data storage challenges: ~80 TB/h
- Data transfer to the archive: ~10 TB/h

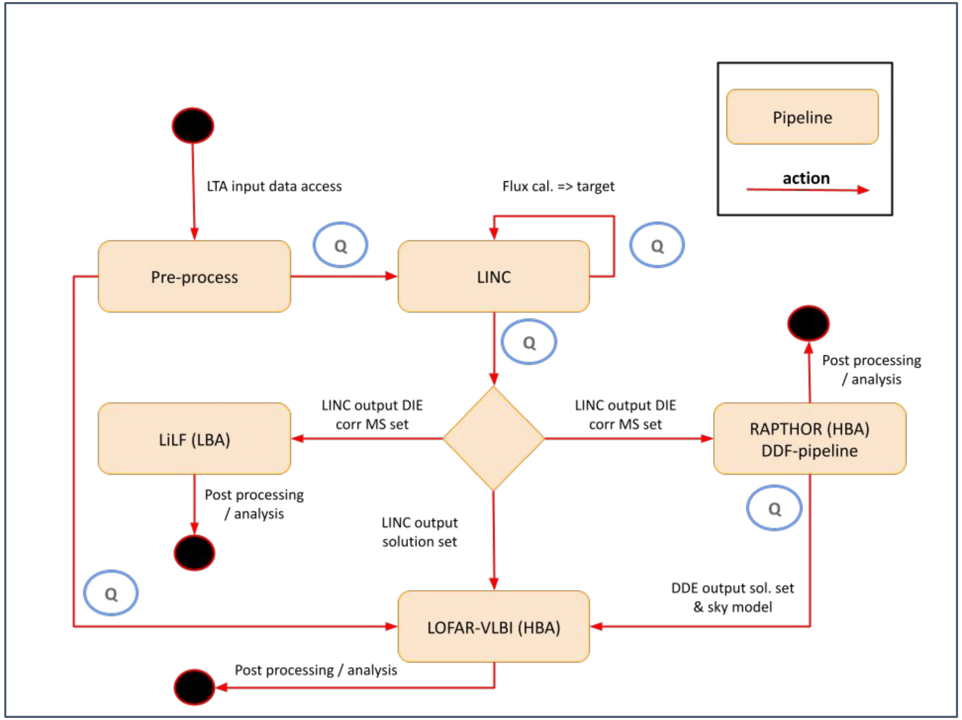
Science archive now:

- ~60 PB offline stored data collection
- mixed state of reduction & science readiness

Interferometric data analysis workflow

LOFAR pipelines designed to perform an incremental data editing

- complexity (self-) calibration strategies → demanding hardware requirements
- data complexity and sizes O(TB) → demanding storage requirements O(10TB)

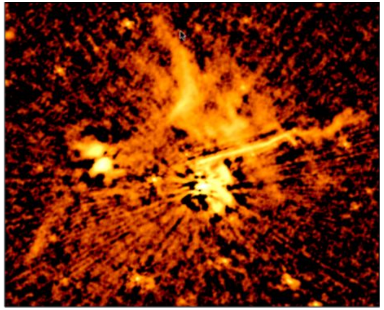


Interferometric data analysis workflow

Direction Independent Errors (DIE) calibration & imaging

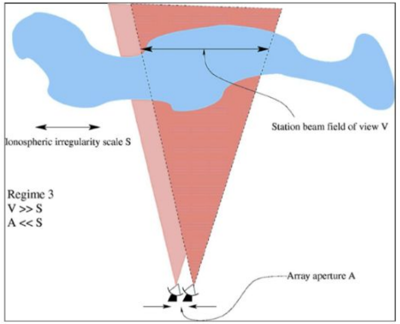
- Correct for instrumental errors

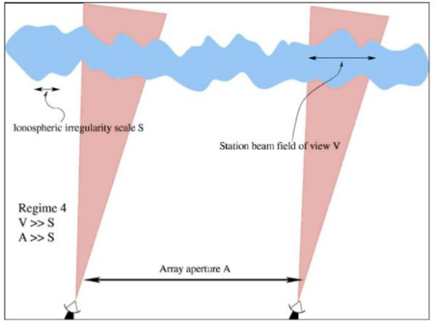
Scientifically limited



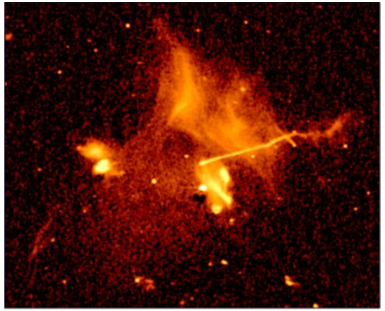
Direction Dependent Errors (DDE) calibration & imaging

- data dominated by dispersive delays caused by the ionosphere
  - severity scales as $\varphi \propto \nu$^-1
- The ionosphere as seen by LOFAR strongly varies across the Field of View





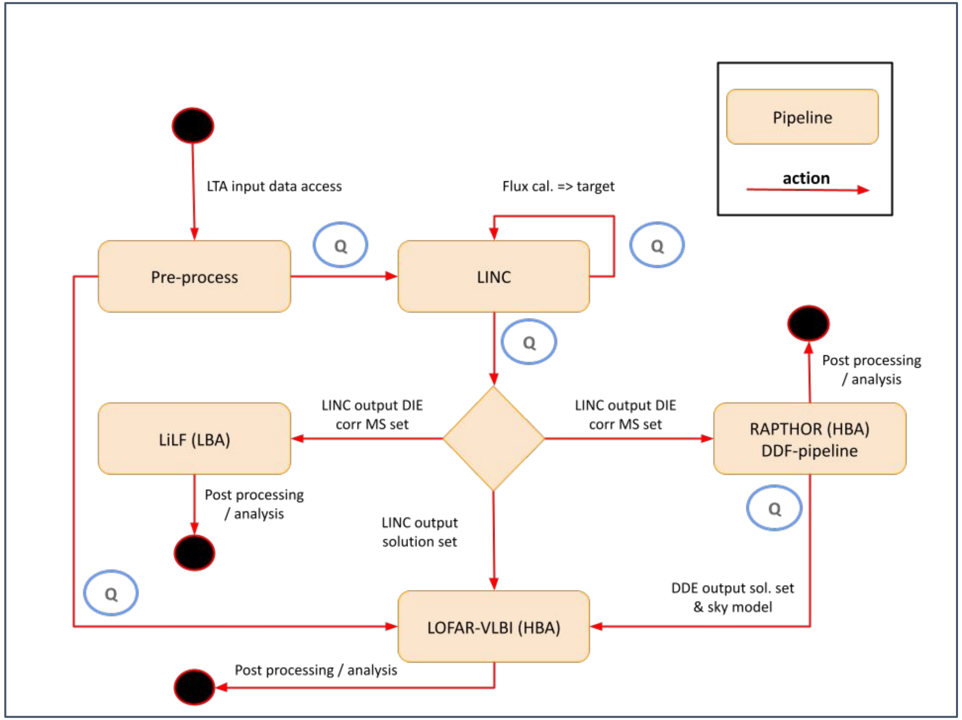
Rich in science



Interferometric data analysis workflow

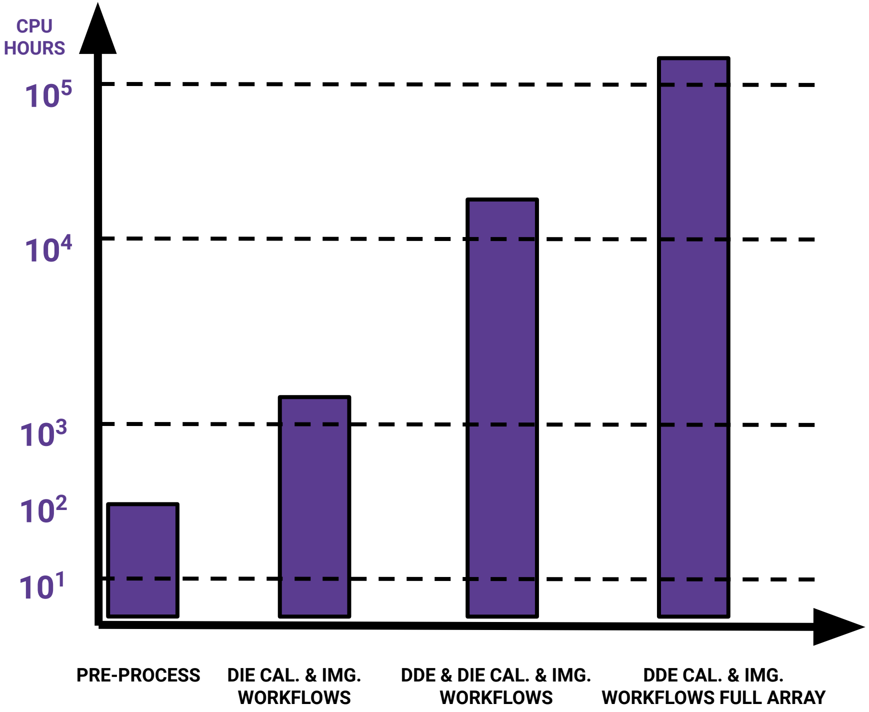
(Some of the) challenges in LOFAR calibration & imaging

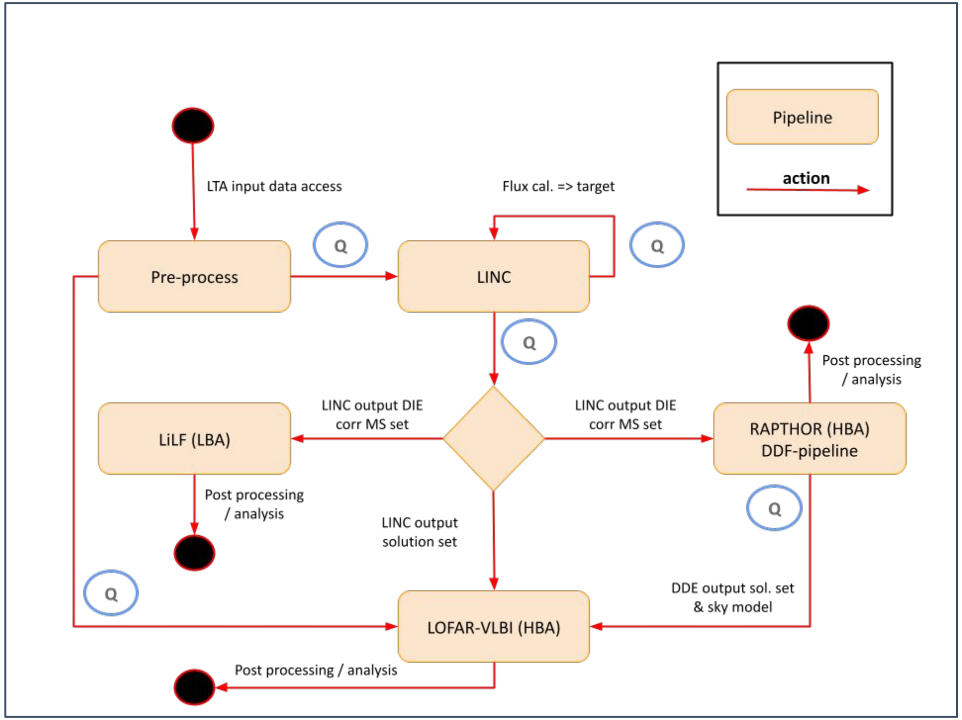
- Large data volumes [CAL | IMG]
- Low S/N regime → calibration errors [CAL]
- Large fractional bandwidth
- Requires multi-frequency approaches [CAL | IMG]
- Large FOV
- Direction-dependent calibration approaches needed [CAL]
- Large w-values [IMG]
- Deconvolution complex [IMG]



Interferometric data analysis workflow

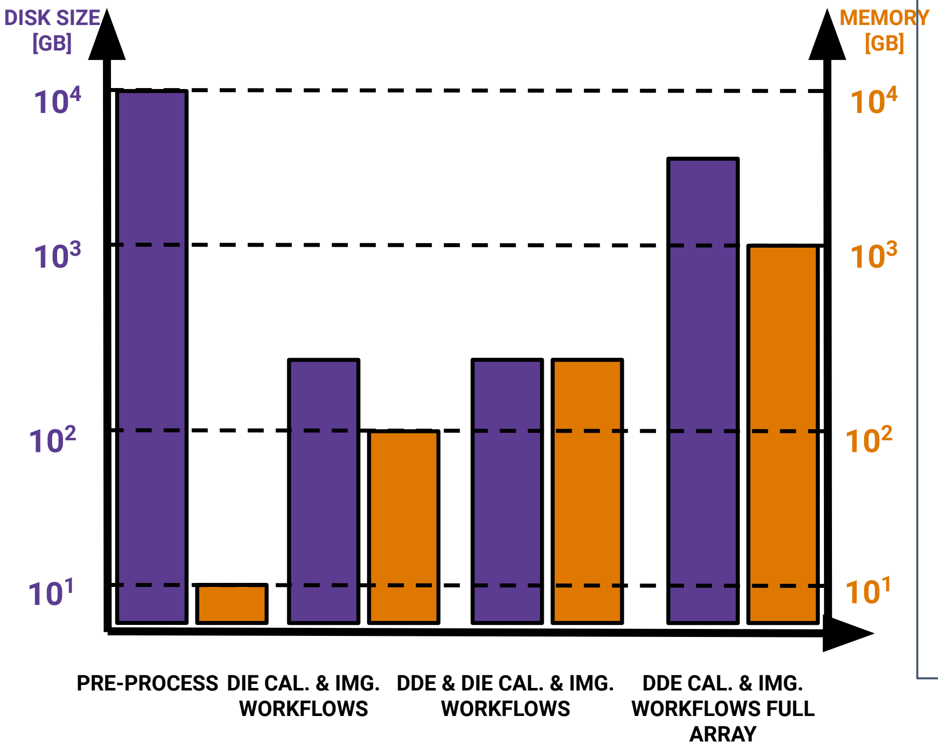
Computing budget of a data(set) processing chain

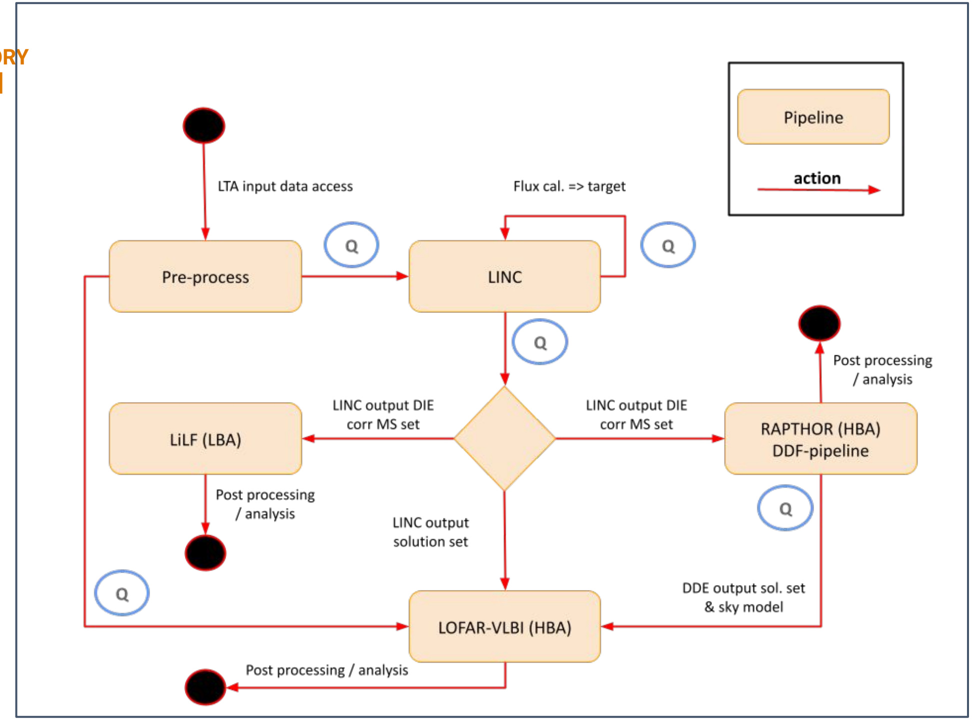




Interferometric data analysis workflow

Computing budget of a data(set) processing chain





How to manage the logistics of data intensive facility ?

Sustainability (environmental & financial) of science operations (aggregated estimates)

- storage O(10-100) PB
- computing O(1000) Mcore hours
- efficient mechanisms to access E-infra resources

Use of tools to support distributed computing vs HPC centres specs.

- multi node / cores jobs processing
- CWL standard

Mitigation measures

- Hardware: use of GPU or FPGA for selected workflows
- Algorithms: res/devs on deterministic and/or Machine Learning for calibration and/or imaging

The Square Kilometre Array (SKA)

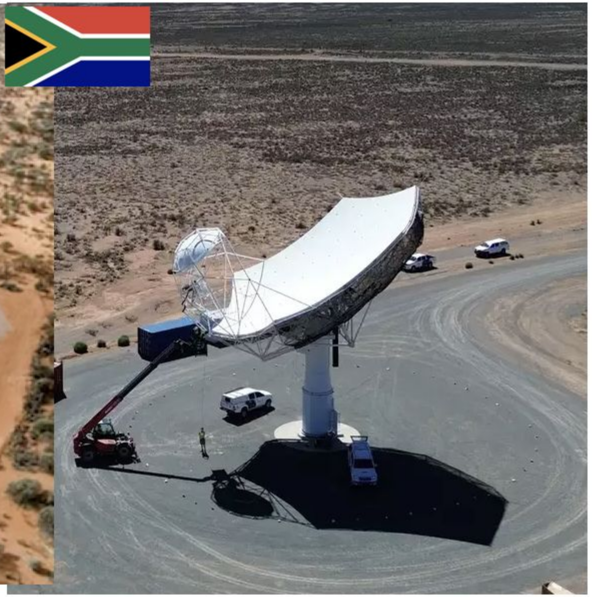
SKA LOW

SKA MID

The Square Kilometre Array (SKA)

SKA LOW

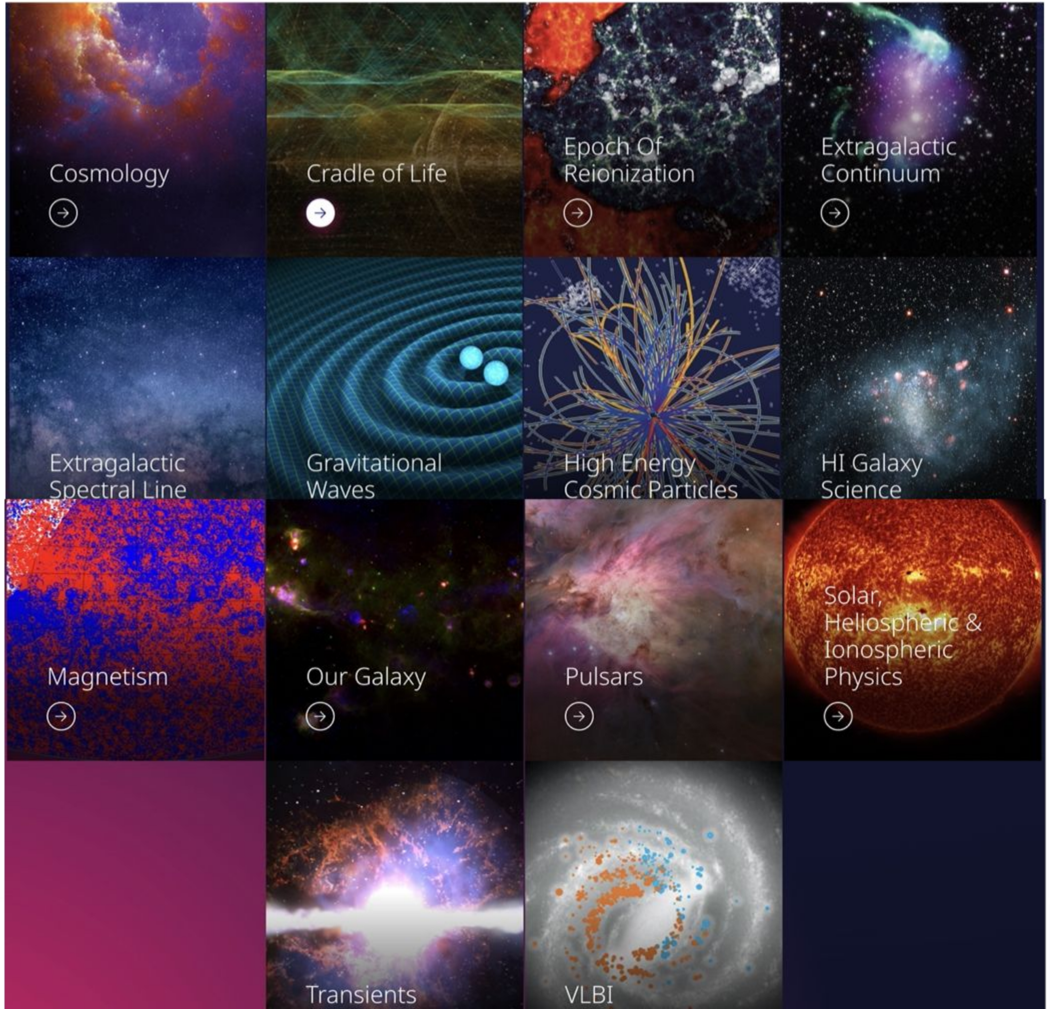
SKA MID



Global SKA Community

The SKA Observatory is being established as an Intergovernmental Organisation, taking over from the SKA Organisation.
It will undertake the construction and operation of the SKA telescopes.

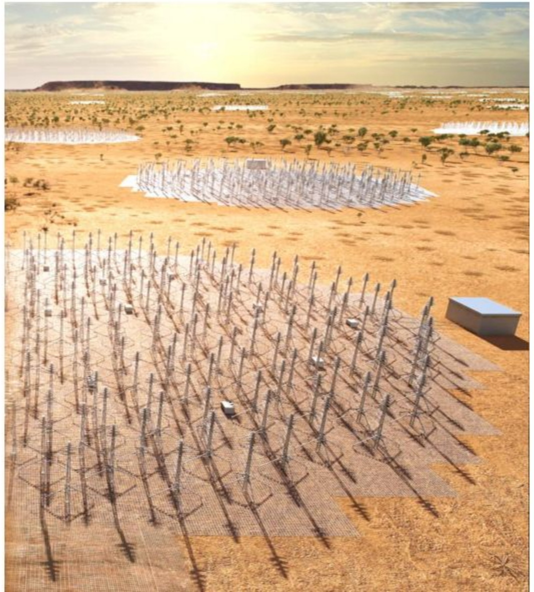
SKA Science


Cosmology
Cradle of Life
Epoch Of Reionization
Extragalactic Continuum
Extragalactic Spectral Line
Gravitational Waves
High Energy Cosmic Particles
HI Galaxy Science
Magnetism
Our Galaxy
Pulsars
Solar, Heliospheric & Ionospheric Physics
Transients
VLBI

SKA Design Overview

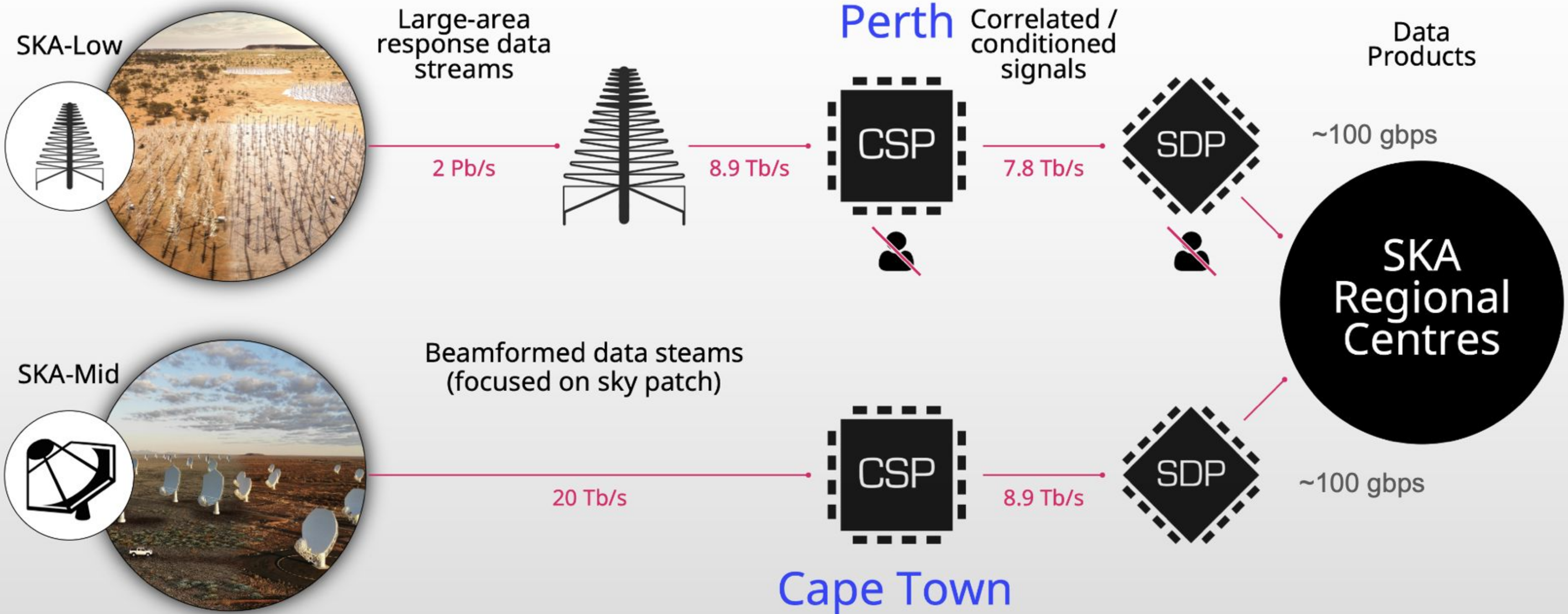
3 sites (AUS, RSA, UK-HQ) 2 telescopes (LOW, MID) **one Observatory (SKAO)**
Construction activities: began July 1, 2021

SKA-Low: ~131,000 low-freq dipoles,
50 – 350 MHz
65 km max. baseline (>11″ @ 110 MHz)
Murchison, Western Australia

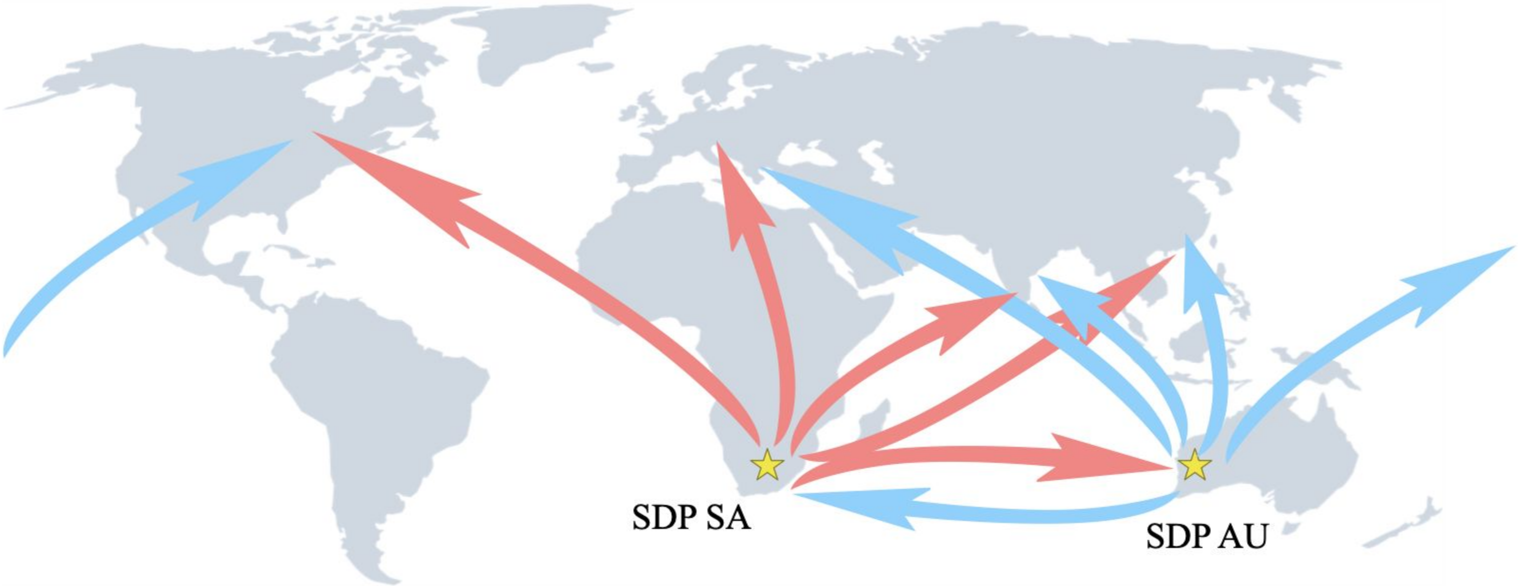


SKA-Mid: 133x15m dishes+ MeerKAT l
0.35 – 15 GHz
150 km max. baseline (>0.2″ @ 1.7 GHz)
Karoo, South Africa

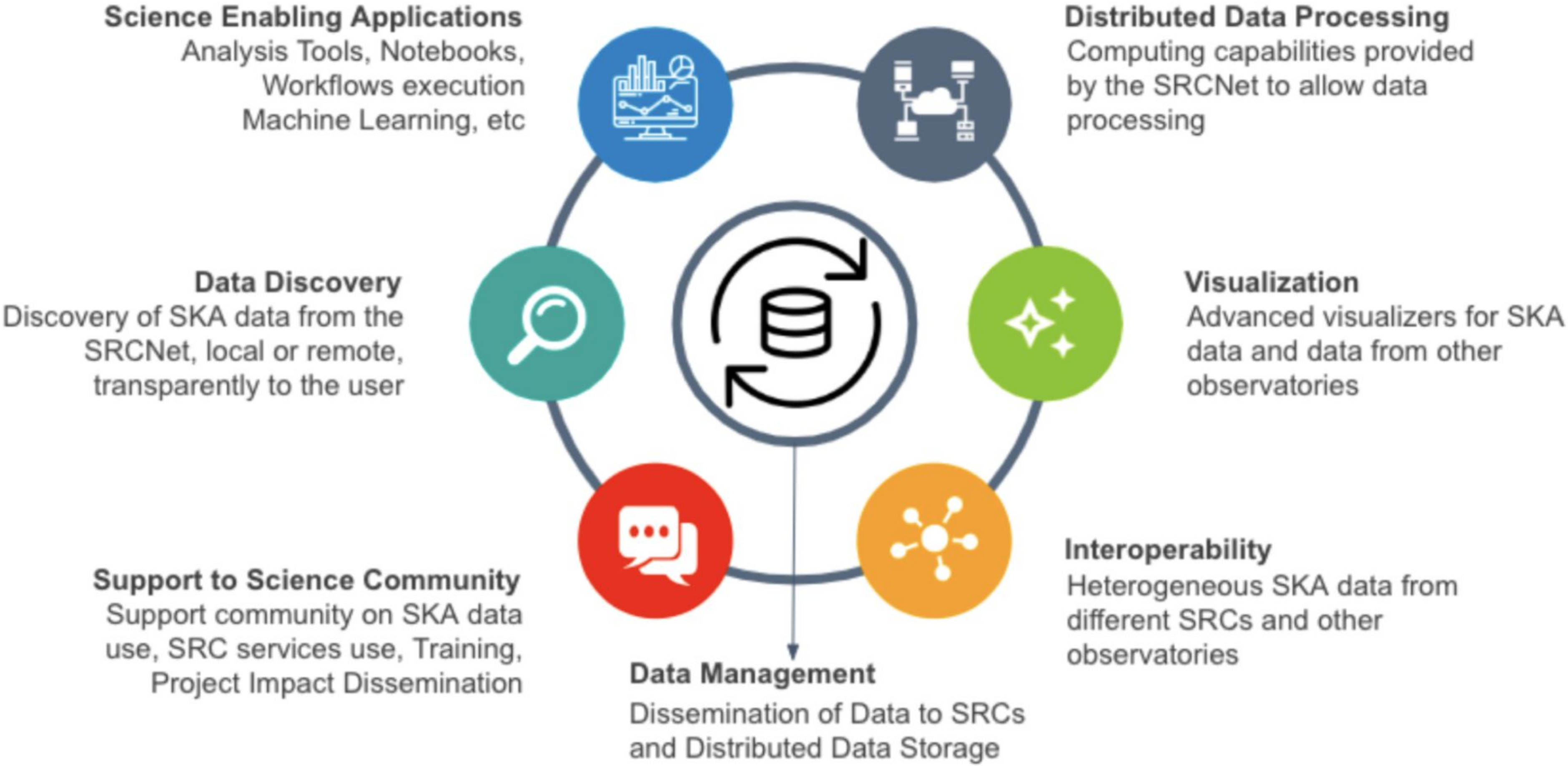
SKA Regional Centres: SKAO data processing stages



SKA Data Flow



Observatory Data Products flow from the Science Data Processors in Perth and Cape Town to Science Regional Centres around the globe


Science Enabling Applications
Analysis Tools, Notebooks, Workflows execution Machine Learning, etc
Distributed Data Processing
Computing capabilities provided by the SRCNet to allow data processing
Data Discovery
Discovery of SKA data from the SRCNet, local or remote, transparently to the user
Visualization
Advanced visualizers for SKA data and data from other observatories
Support to Science Community
Support community on SKA data use, SRC services use, Training, Project Impact Dissemination
Interoperability
Heterogeneous SKA data from different SRCs and other observatories
Data Management
Dissemination of Data to SRCs and Distributed Data Storage

SPECTRUM

Thank you!
Questions?

spectrumproject.eu

linkedin.com/company/spectrum-project-eu

Funded by
the European Union

SPECTRUM is funded by the European Union - Grant Agreement Number 101131550