ASTRON

Netherlands Institute for Radio Astronomy

NWO

Science Data Centres for Radio Astronomy: from LOFAR to SKA

Michiel van Haarlem

Head of NL SKA Office

@SKA_NL

Monday 13 January 2020

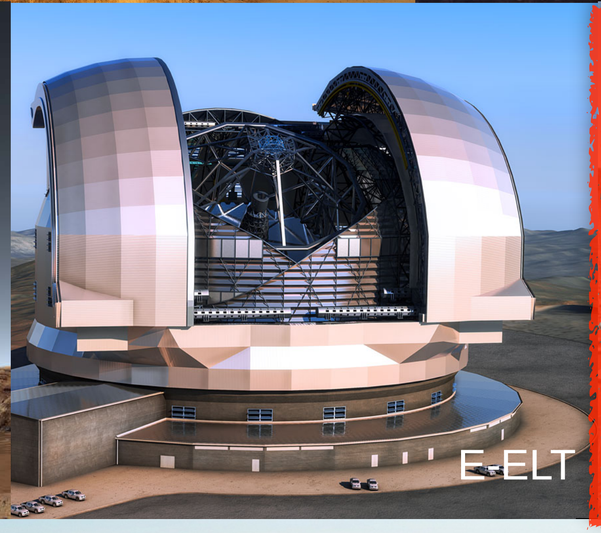
The Science Cloud – Towards a Research Data Ecosystem for the next Generation of Data–intensive Experiments and Observatories


JVLA

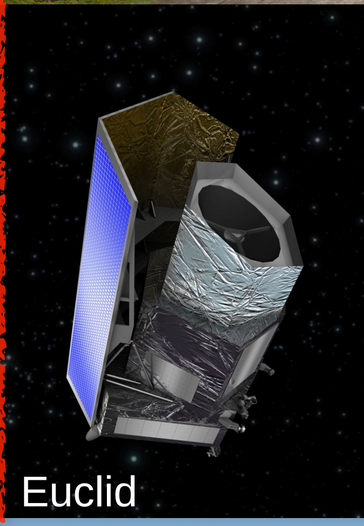

ALMA


LOFAR

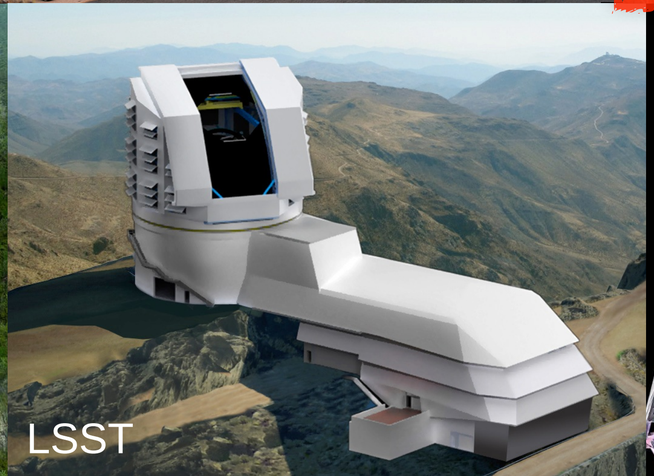

TMT

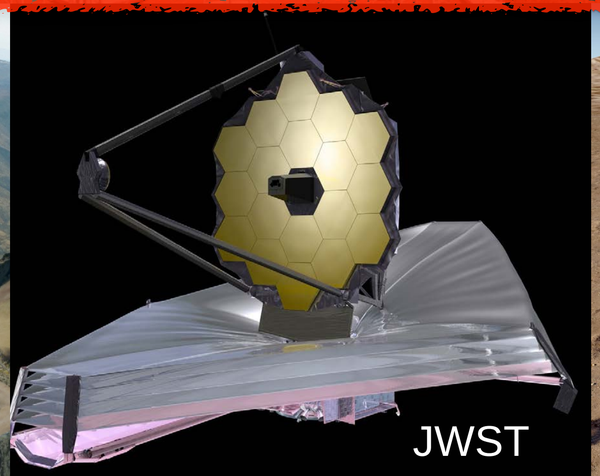

E-ELT

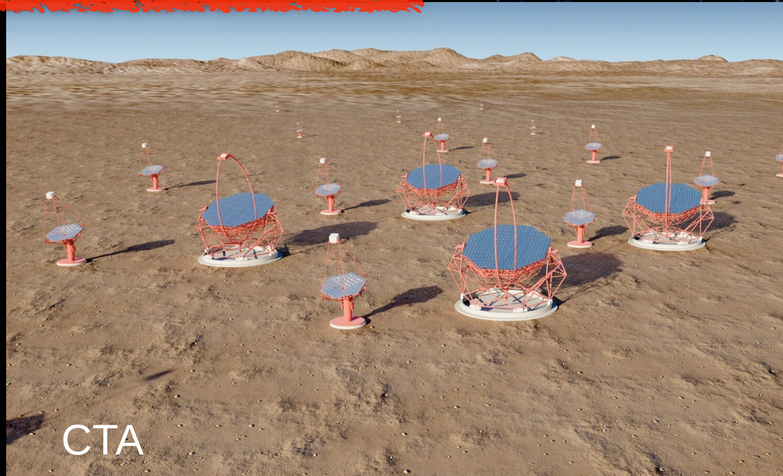

SKA


Euclid

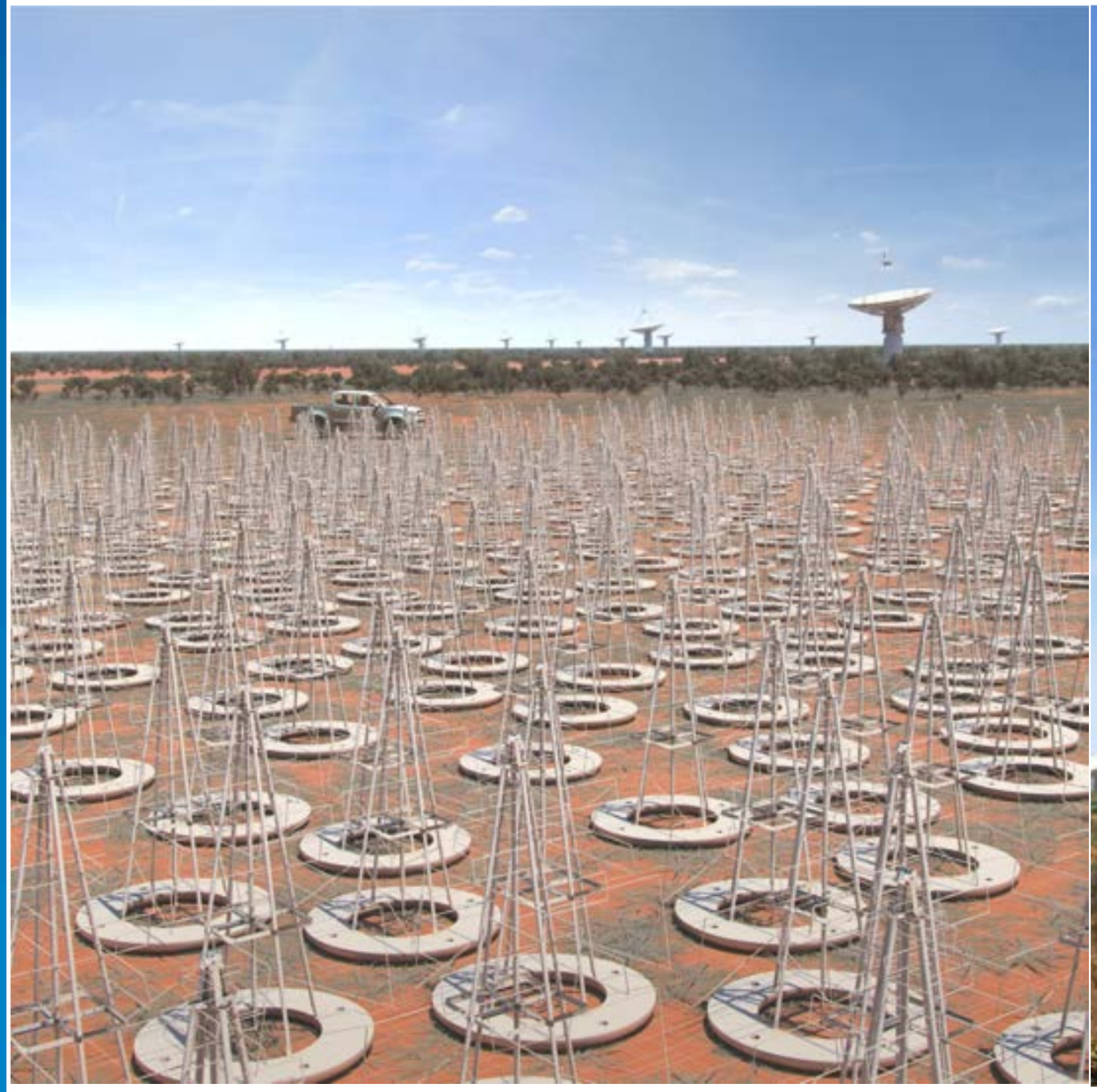

FAST


LSST


JWST

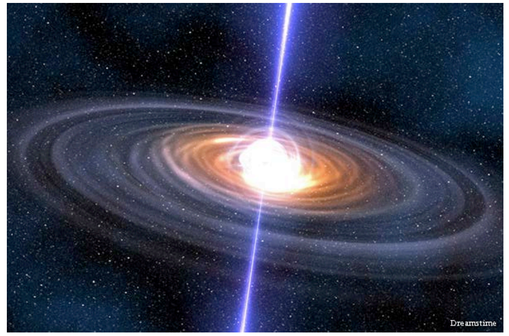

CTA

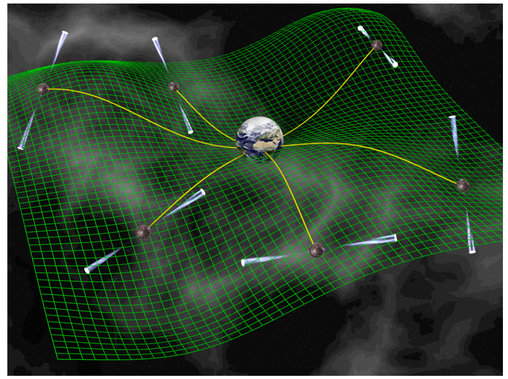
Square Kilometre Array in Australia and South Africa

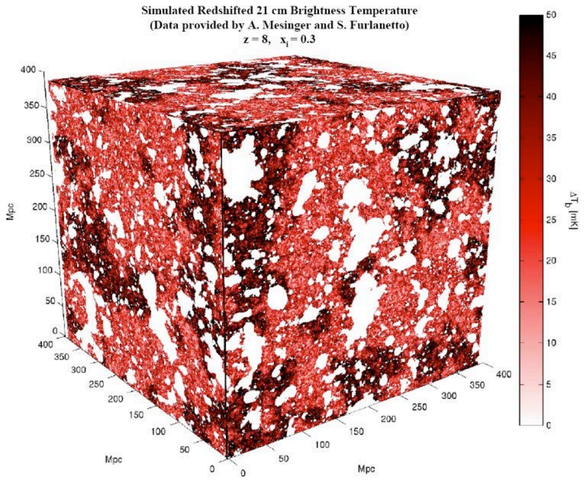


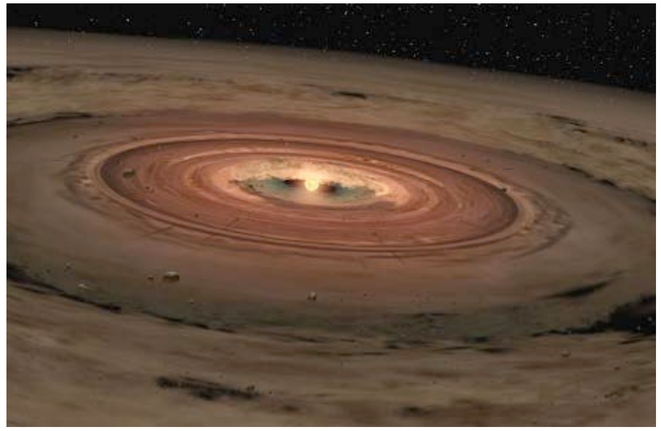
SKA Headline Science

- Pulsar Surveys and Timing - Study of Gravitational waves
- Cradle of Life and Astrobiology - How do solar systems form and where could life emerge?
- Galaxy Evolution and Cosmology - How do galaxies get their gas and form stars?
- Cosmic Magnetic Fields - When did ordered magnetic fields in galaxies form
- Cosmic Dawn and Epoch of Reionisation - When did the first galaxies form and begin to reionise the Universe
- Radio transients and *Exploration of the Unknown*
- Full Science Case available at: http://bit.ly/SKA_Science_2014









SKA Observatory Convention signed on 12 March 2019

Members: AU, UK, ZA, CA, CN, NZ, IN, NL, IT, SE, DE, ES, FR

Members of the SKA Organisation
Host Countries: Australia, South Africa, United Kingdom

African Partner Countries

SKA Observatory will be established as an Intergovernmental Organisation in 2020, taking over from the SKA Organisation. It will undertake the construction and operation of the telescope.

As of March 2019, confirmed SKA Observatory members are





SKA1-mid - the SKA's mid-frequency instrument

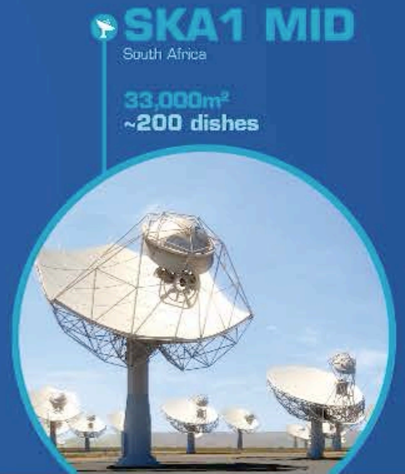
Frequency range:
350 MHz to
14 GHz

~200 dishes
(including 64 MeerKAT dishes)

Total collecting area:
33,000m²

Maximum distance between stations:
150km

How SKA1-mid compares with the Janksy Very Large Array (JVLA), the current best similar instrument in the world.

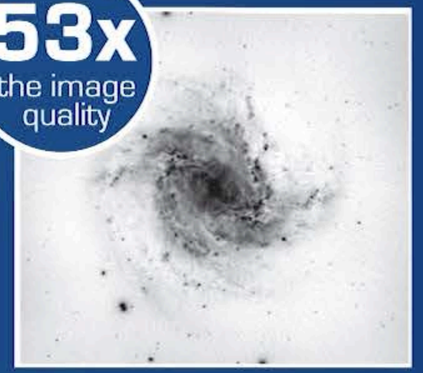


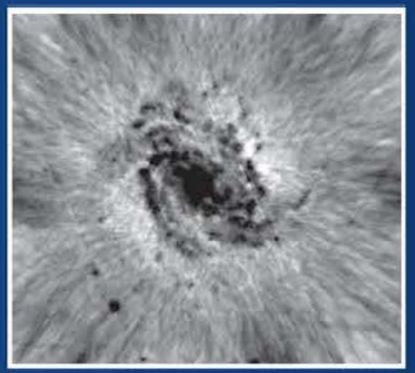
4x the resolution

5x more sensitive

60x the survey speed

53x the image quality









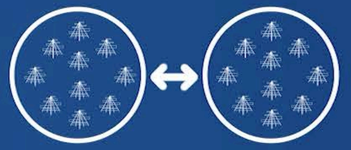

SKA1-low - the SKA's low-frequency instrument

Location:
Australia

Frequency range:
50 MHz to
350 MHz

~130,000
antennas spread between
500 stations

Total collecting area:
0.4km²

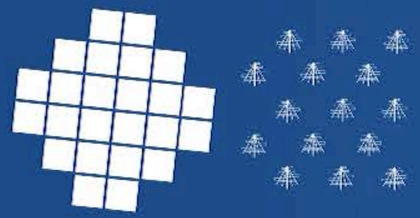


Maximum distance between stations:
>65km

How SKA1-low compares with the LOw Frequency ARray (LOFAR), the current best similar instrument in the world

SKA1 LOW
Australia

419,000m²
~130,000 antennas

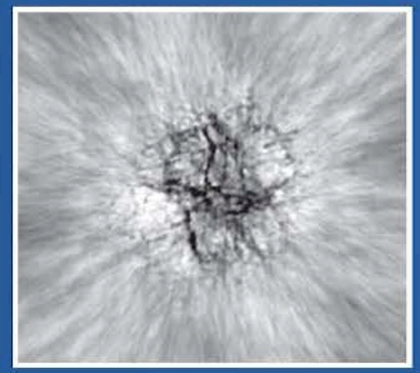


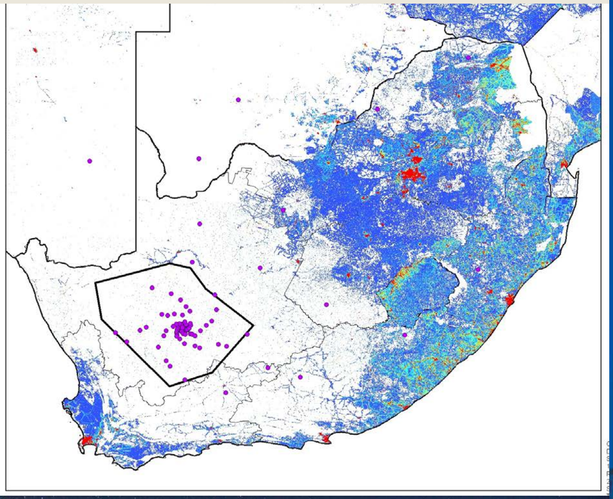
25% better resolution

8x more sensitive

135x the survey speed

68x the image quality





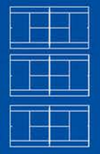
SKA1 MID - the SKA's mid-frequency instrument

The Square Kilometre Array (SKA) will be the world's largest radio telescope, revolutionising our understanding of the Universe. The SKA will be built in two phases - SKA1 and SKA2 - starting in 2018, with SKA1 representing a fraction of the full SKA. SKA1 will include two instruments - SKA1 MID and SKA1 LOW - observing the Universe at different frequencies.

Frequency range:
350 MHz to
14 GHz

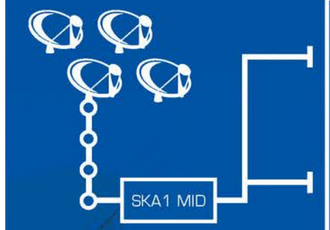
~200 dishes
(including 64 MeerKAT dishes)

Total collecting area:
33,000m²



or
126 tennis courts

Maximum distance between dishes:
150km



Total raw data output:
2 terabytes
per second
62 exabytes
per year

Enough to fill
340,000
average laptops with content **every day**

Compared to the JVLA, the current best similar instrument in the world:

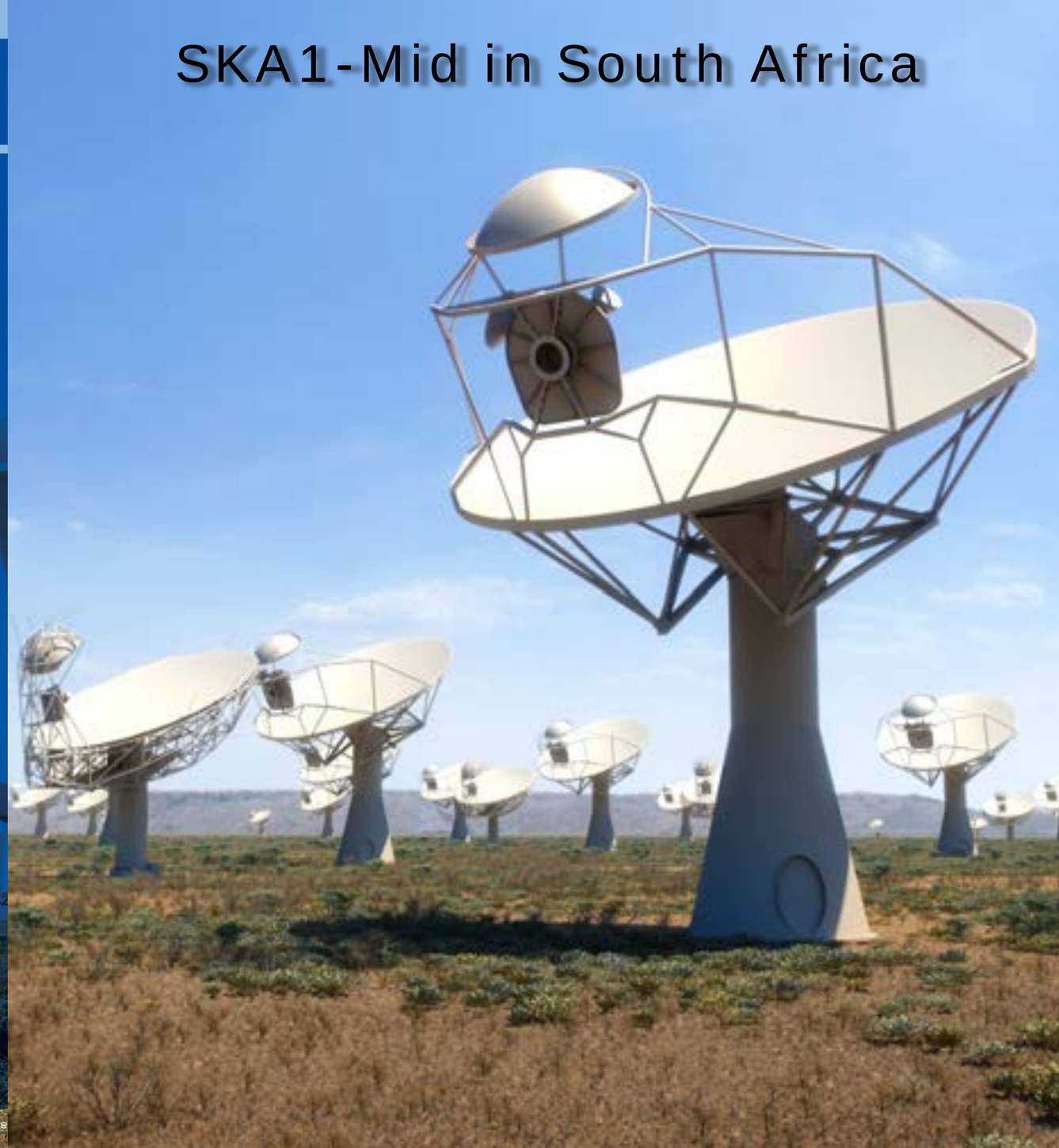
4x the resolution

5x more sensitive

60x the survey speed

www.skatelescope.org · Square Kilometre Array · @SKA_telescope · YouTube The

SKA1-Mid in South Africa



SKA1-Low in Australia

SKA1 LOW - the SKA's low-frequency instrument

The Square Kilometre Array (SKA) will be the world's largest radio telescope, revolutionising our understanding of the Universe. The SKA will be built in two phases - SKA1 and SKA2 - starting in 2018, with SKA1 representing a fraction of the full SKA. SKA1 will include two instruments - SKA1 MID and SKA1 LOW - observing the Universe at different frequencies.

Location: Australia

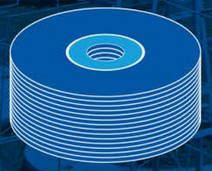
Frequency range:
50 MHz to **350 MHz**

~130,000
antennas spread between **500 stations**

Total collecting area:
0.4km²

Maximum distance between stations:
65km

Total raw data output:
157 terabytes per second
4.9 zettabytes per year



Enough to fill up
35,000 DVDs
every second

5x
the estimated global internet traffic in 2015
(source: Cisco)

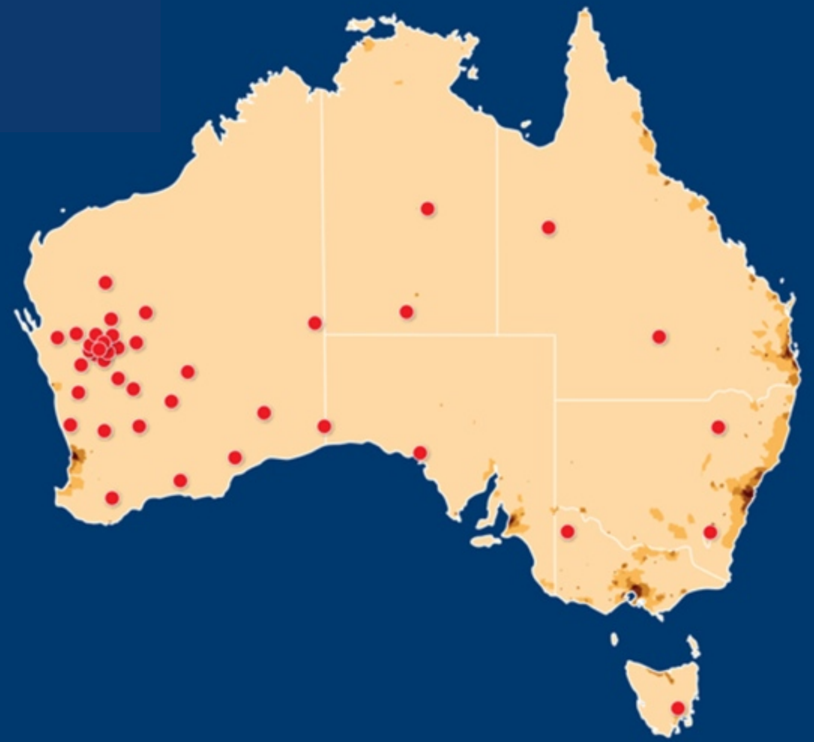
Compared to LOFAR Netherlands, the current best similar instrument in the world

25% better resolution
8x more sensitive
135x the survey speed

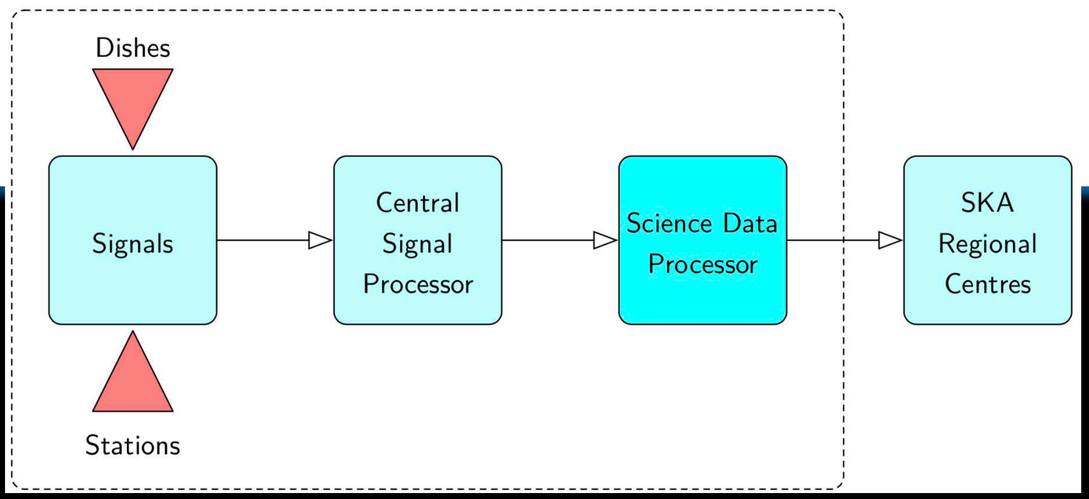
www.skatelescope.org Square Kilometre Array @SKA_telescope YouTube The

Shire of Murchison:

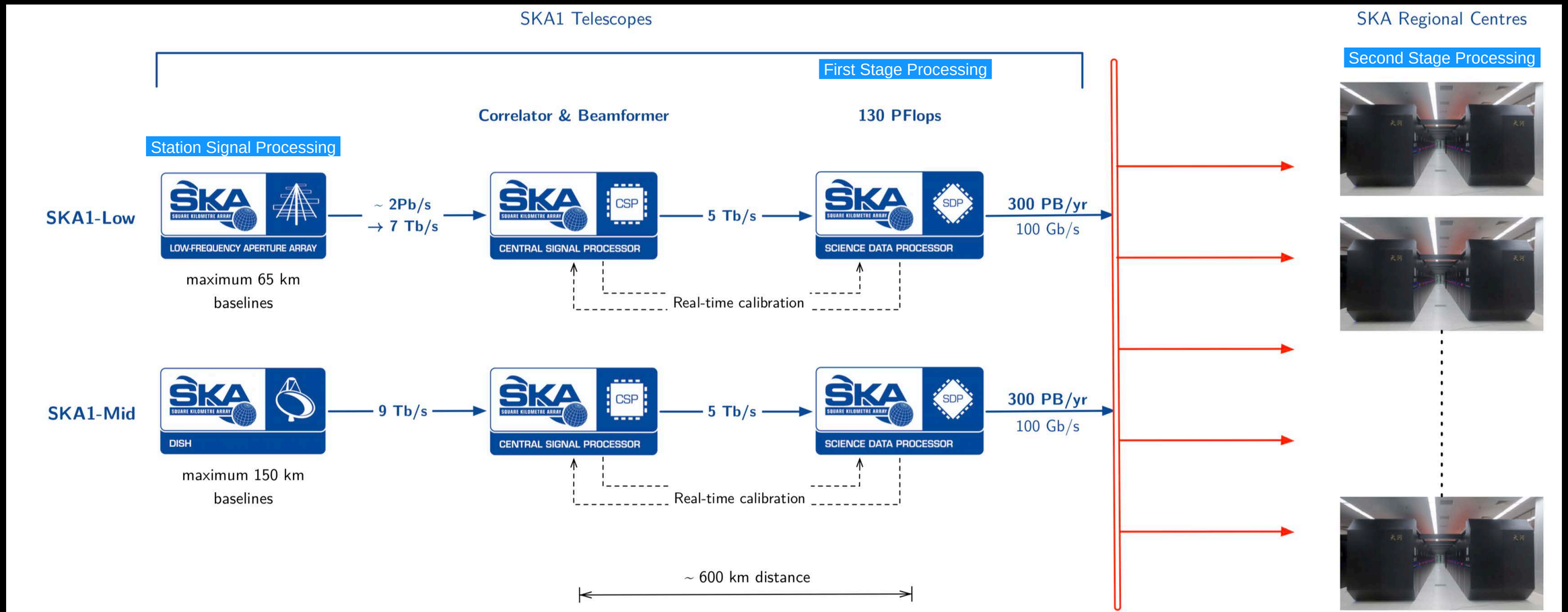
- **50,000 km²** Size of the Netherlands
- **0 gazetted towns**
- **29 sheep/cattle stations**
- **110 population**



SKA Data Flow



Credit: Chrysostomou (SKAO)



SKA First Stage Processing

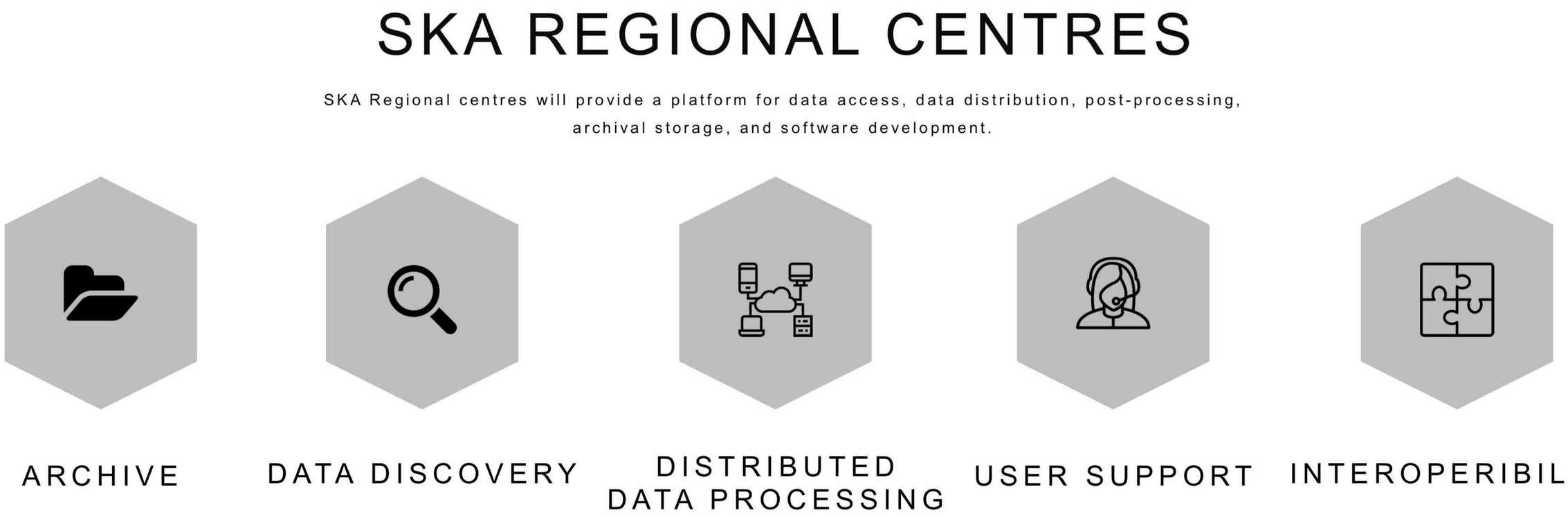
Parametric Model Pipelines

- Real-time imaging pipelines
  - Ingest: receive and pre-process visibilities from CSP
  - RCAL: real-time calibration
  - FastImg: fast imaging for slow transient detection
- Batch imaging pipeline
  - ICAL: iterative self calibration (including direction-dependent calibration)
  - DPrepA: preparation of continuum image data products
  - DPrepB: preparation of coarse spectral image data products
  - DPrepC: preparation of fine spectral image data products
  - DPrepD: preparation of calibrated averaged visibilities data products (EoR projects)
- Non-imaging pipelines
  - Pulsar search and timing, single-pulse transient detection

Observatory Data Products

- Image data products
  - Image cubes
  - Gridded visibilities
- Non-image data products
  - **Science Data Model** - more on this in a moment
  - Calibrated averaged visibilities (designed for EoR projects)
  - Transient source catalogue
  - Pulsar timing solutions
  - Sieved pulsar and transient candidates
  - Transient buffer data

Functions of SKA Regional Centre

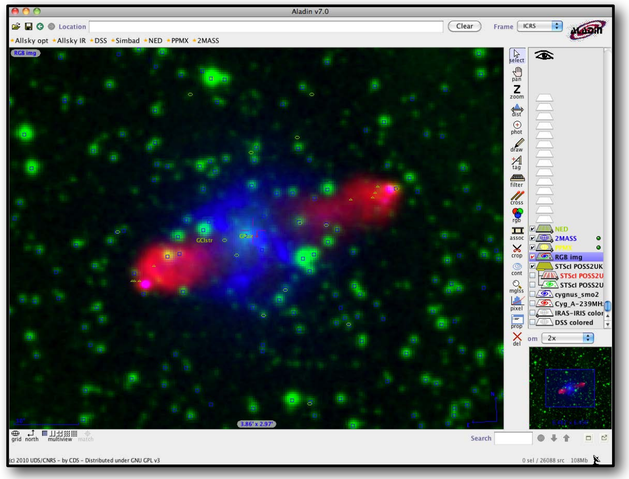


Credit: Joshi & Scaife

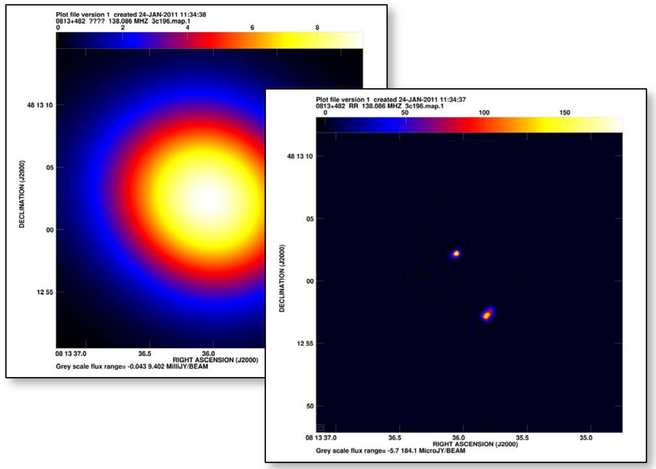
Regional Centre Functionality

Data Discovery

- Observation database
- Associated metadata
- Quick-look data products
- Flexible catalog queries
- Integration with VO tools
- Publish data to VO



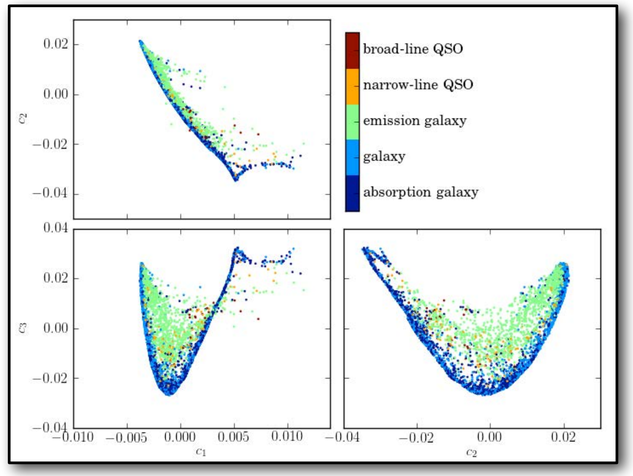
Data Processing



- Reprocessing and calibration
- High resolution imaging
- Mosaicing
- Source extraction
- Catalog re-creation
- DM searches

Data Analysis

- Multi-wavelength studies
- Catalog cross-matching
- Light-curve analysis
- Transient classification
- Feature detection
- Visualization



SKA Regional Centre Steering Committee

Members: Quinn, Peter (Chair, Australia), An, Tao (China), Barbosa, Domingos (Portugal), Bolton, Rosie (SKA), Chrysostomou, Antonio (SKA), Conway, John (Sweden), Gaudet, Séverin (Canada), van Haarlem, Michiel (Deputy Chair, Netherlands), Klockner, Hans-Rainer (Germany), Andrea Possenti (Italy), Simon Ratcliffe (South Africa), Scaife, Anna (UK), Lourdes Verdes-Montenegro (Spain), Vilotte, Jean-Pierre (France), Wadadekar, Yogesh (India)

- SRCSC Mission: to define and create a long-term operational partnership between the SKA Observatory and an ensemble of independently-resourced SKA Regional Centres.
- SRC High Level Functions: Data Flow, Data Processing, Data Curation, User Support, Commonality and Resource Management
- Some Key Principles: data placement driven by optimising science, integrated resource allocation process, access through IVOA services

AENEAS

Advanced European Network of E-infrastructures for Astronomy with the SKA

Design and specification of a distributed, European SKA Regional Centre to support the astronomical community in achieving the scientific goals of the SKA

EC Horizon 2020 (€3 million)

13 countries, 28 partners, SKAO, host countries, e-infrastructures (EGI, GÉANT, RDA), NREN's

Three year project (2017-2019)

- Computing and Processing Requirements
- Data Transport and Optimal European Storage Topologies
- Data Access and Knowledge Creation
- User Services

Final deliverable: preliminary ESDC Design and Implementation Plan

WP2: ESDC Design & Governance

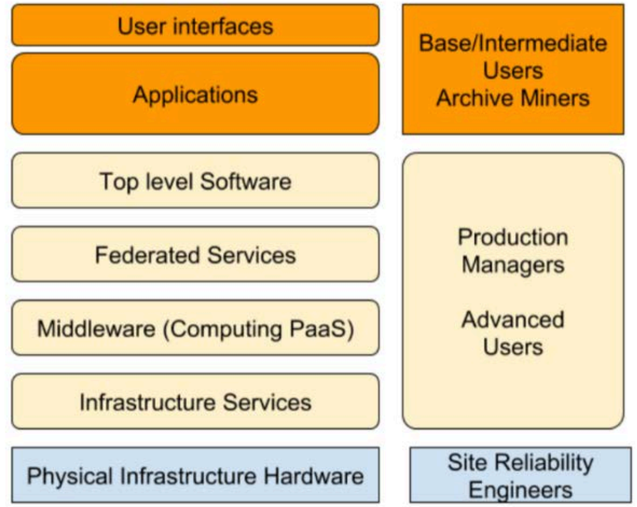
Survey of Potential Providers

- *Over 50 expressions of interest*
- *Mixture of scientific institutes, infrastructure providers, and industrial partners*

- *ESDC Requirements based on those developed by SRCCG*
- *Final deliverable: preliminary ESDC Design and Implementation Plan*
- *User input needed!*

WP3: Computing and Processing Requirements

- Analysis of compute load, data transfer and data storage anticipated as required for SKA Key science
- Suggested solutions to address each of the key software areas associated with running a distributed ESDC
- Initial System Sizing



Reprocessing and post-processing

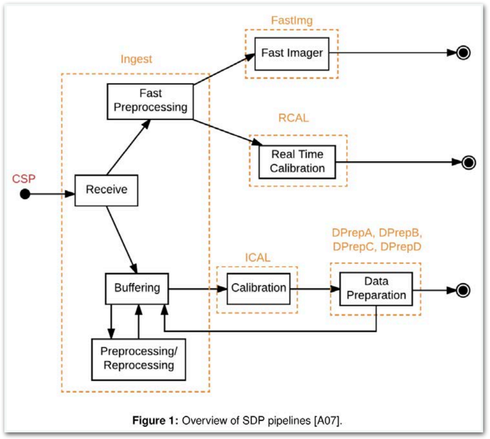


Figure 1: Overview of SDP pipelines [A07].

Processing is being examined in terms of (1) compute load; (2) memory requirements; (3) potential for distribution; (4) suitability of platform.

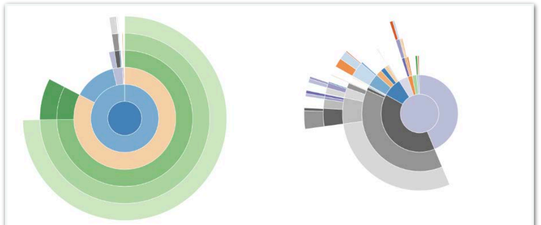
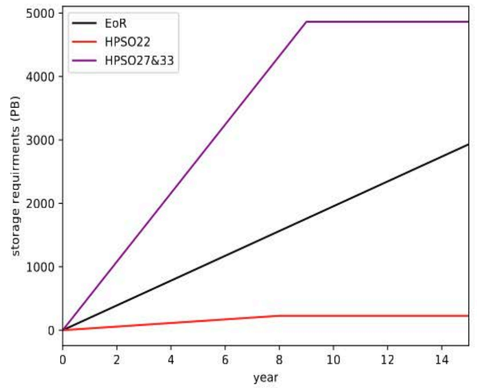
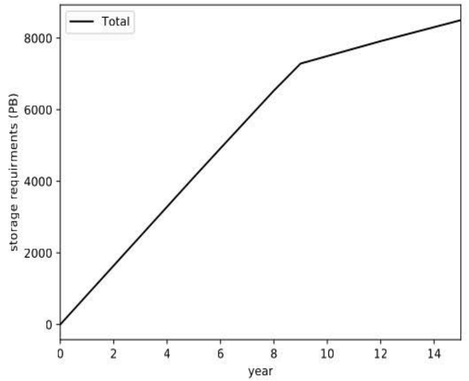


Figure 9: Visualisation of processing time breakdown for PyRMSynth. (a) The overall breakdown, dominated by `write_output_files` (green); (b) the breakdown beneath `compute_dirty_image`, dominated by the FFT (grey). See text for details.

Minimum for HPSOs ~13 PFlops

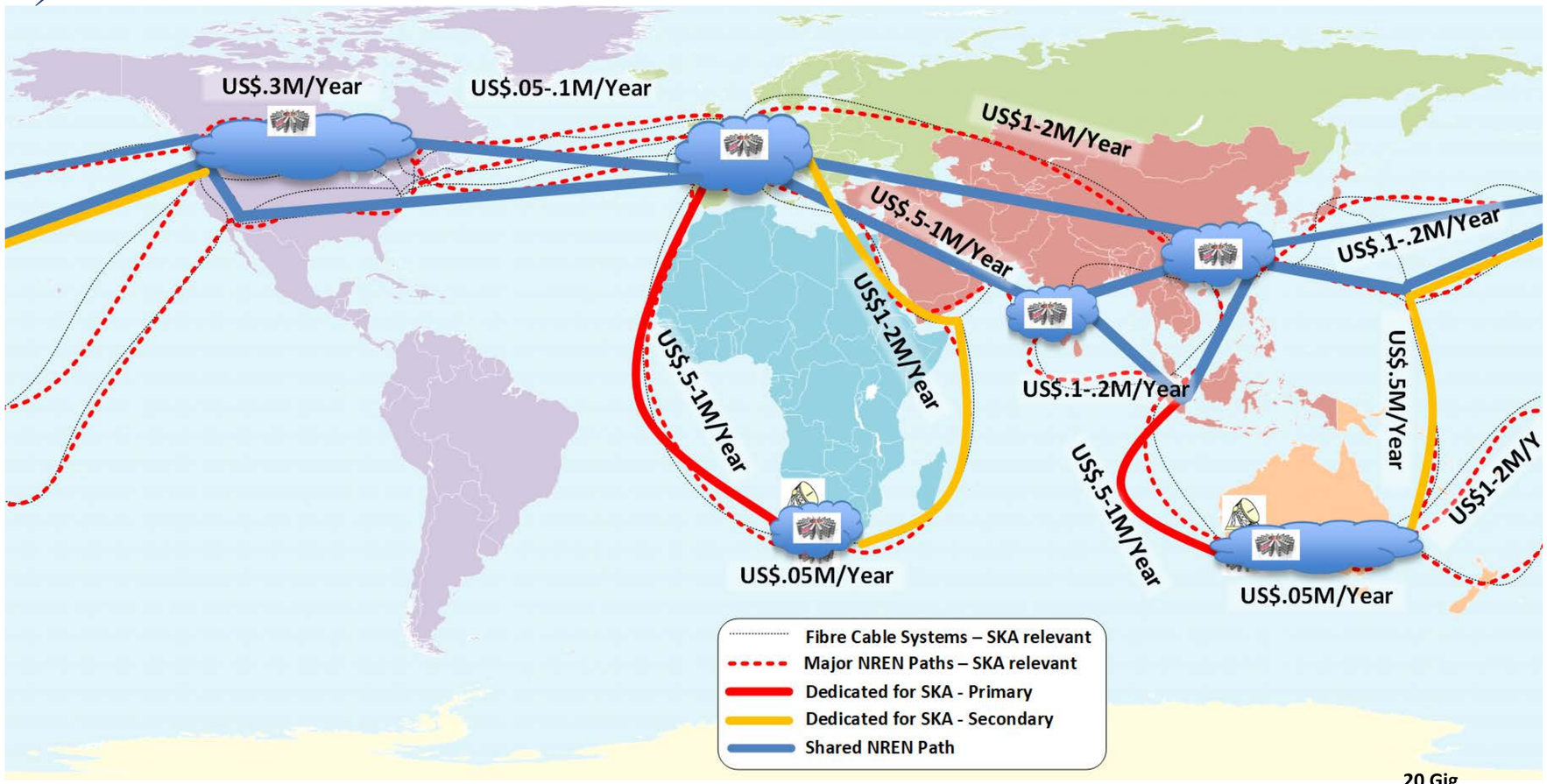
Storage estimates for HPSOs

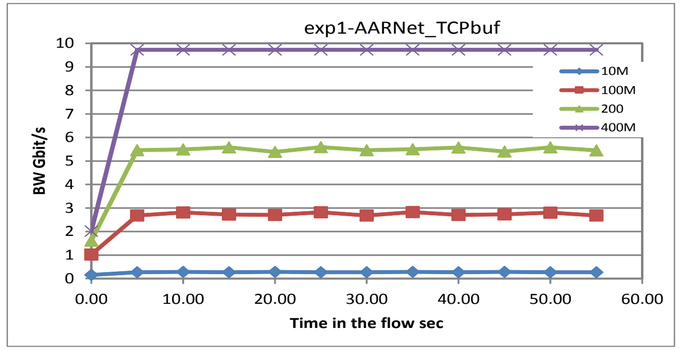




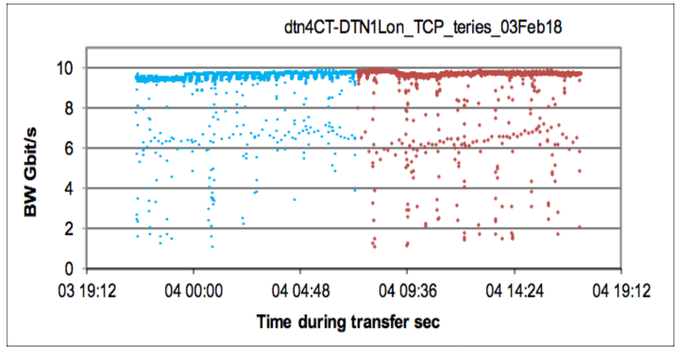
10 ExaBytes over first 15 years of SKA operations

WP4: Data Transport and Optimal Storage Topologies



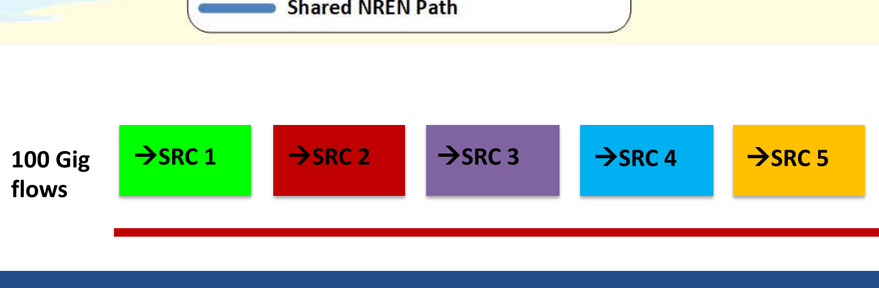


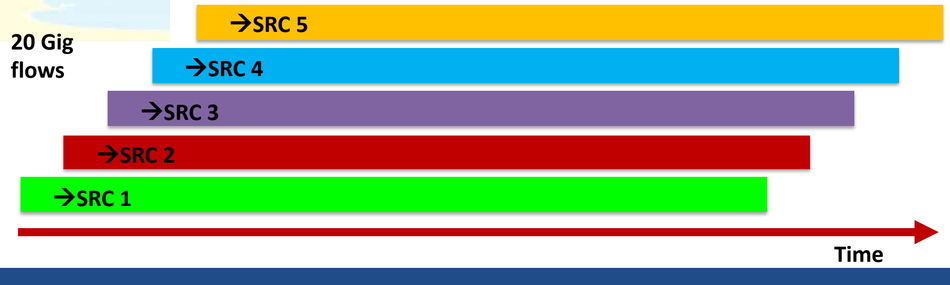
GÉANT London to AARNet Canberra



SANReN Cape Town to GÉANT London

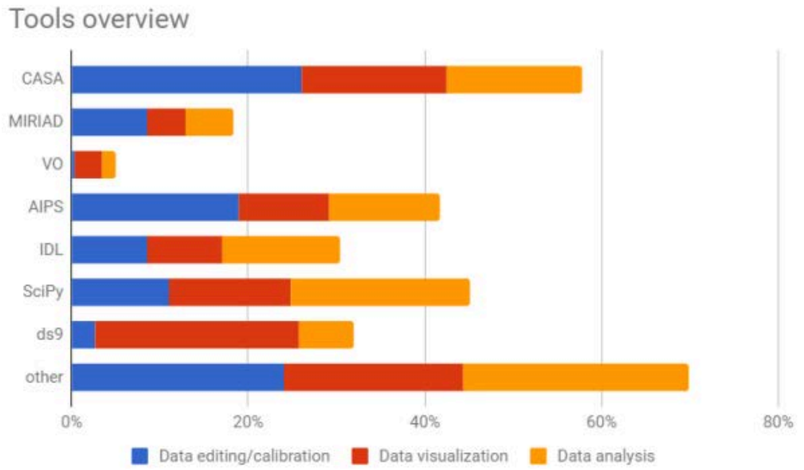
What fraction of the SKA archive will we host in Europe?



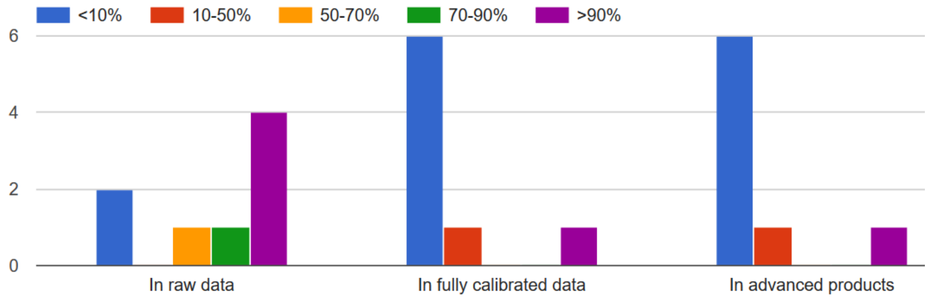


WP5: Data Access and Knowledge Creation

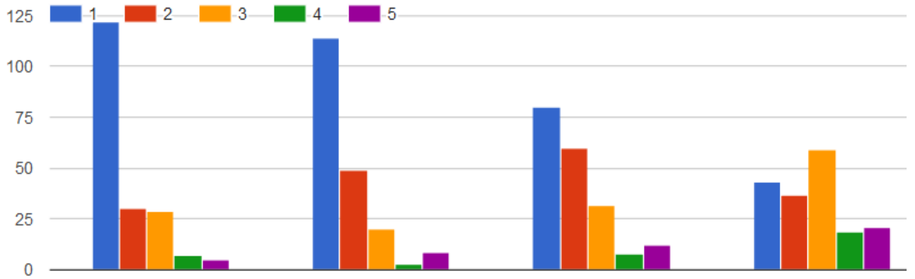
- Surveys of Astronomical Facilities and of their User Communities
- Gap analysis
- Recommendations on the design of user interfaces
  - for data discovery, access, and retrieval
  - data processing, re-processing, analysis and visualization



What is the fraction of data distributed through archives in raw, calibrated and advanced format?



What data products would you like to find in a facility archive (1=necessary, 5=useless)?



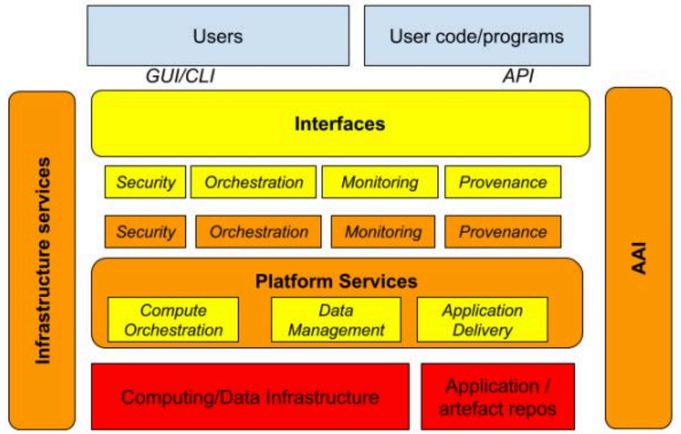
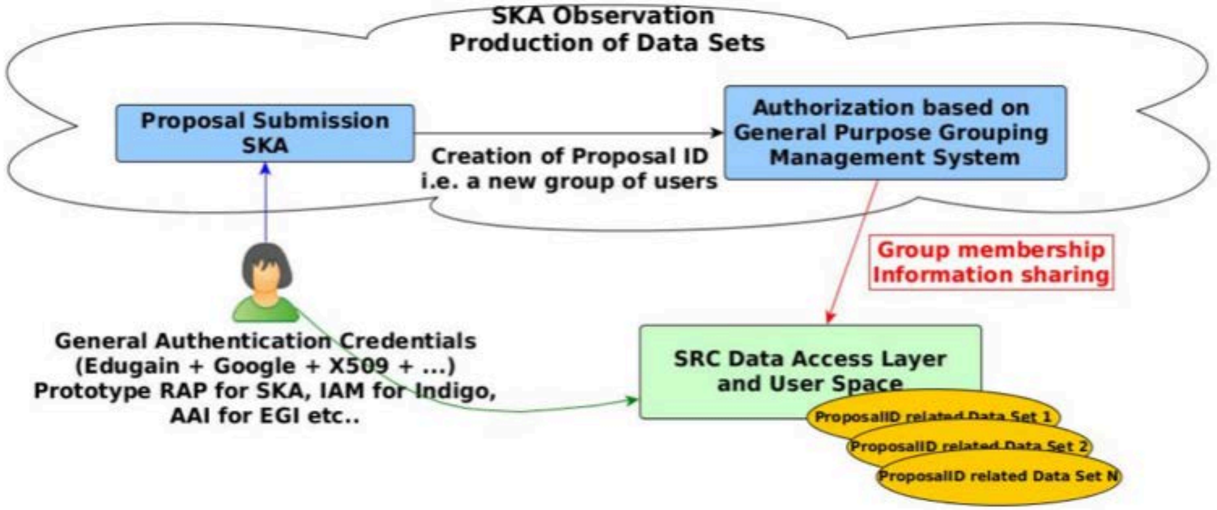


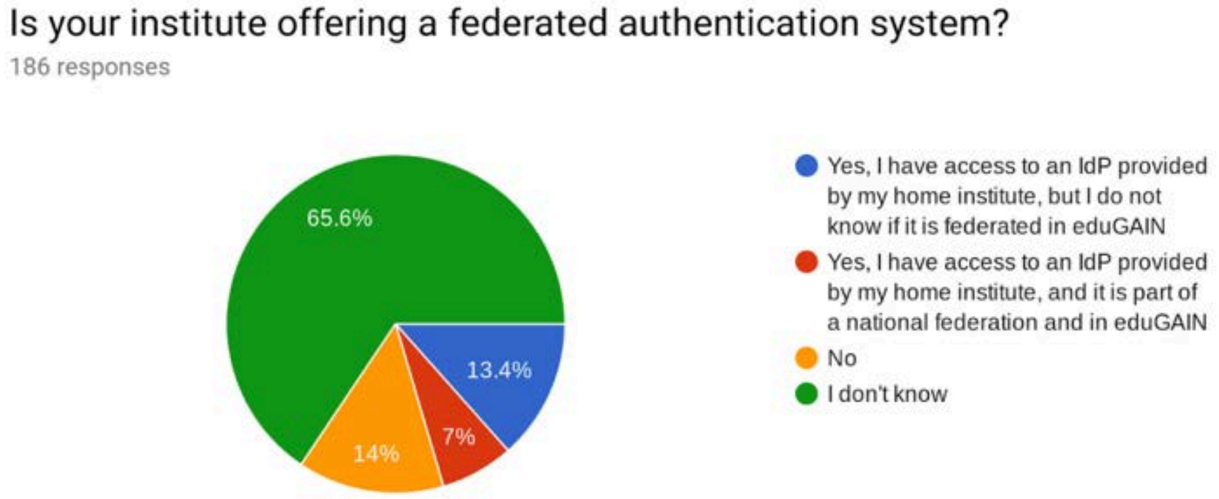
Figure 5.2.1: SG modular components to manage the exploitation of computing/data resources by the user community and user applications.

WP6: User Services

- Authentication and Authorisation Infrastructure
  - Federated Access for Research
  - Exploration of Technologies
  - Proposed AAI Architecture

- Framework for designing and implementing a Service Portfolio for the ESDC and SKA
  - validate users' requests for data access;
  - keep accounts of computing and storage resources for each user or user group;
  - minimize data movement between sites.





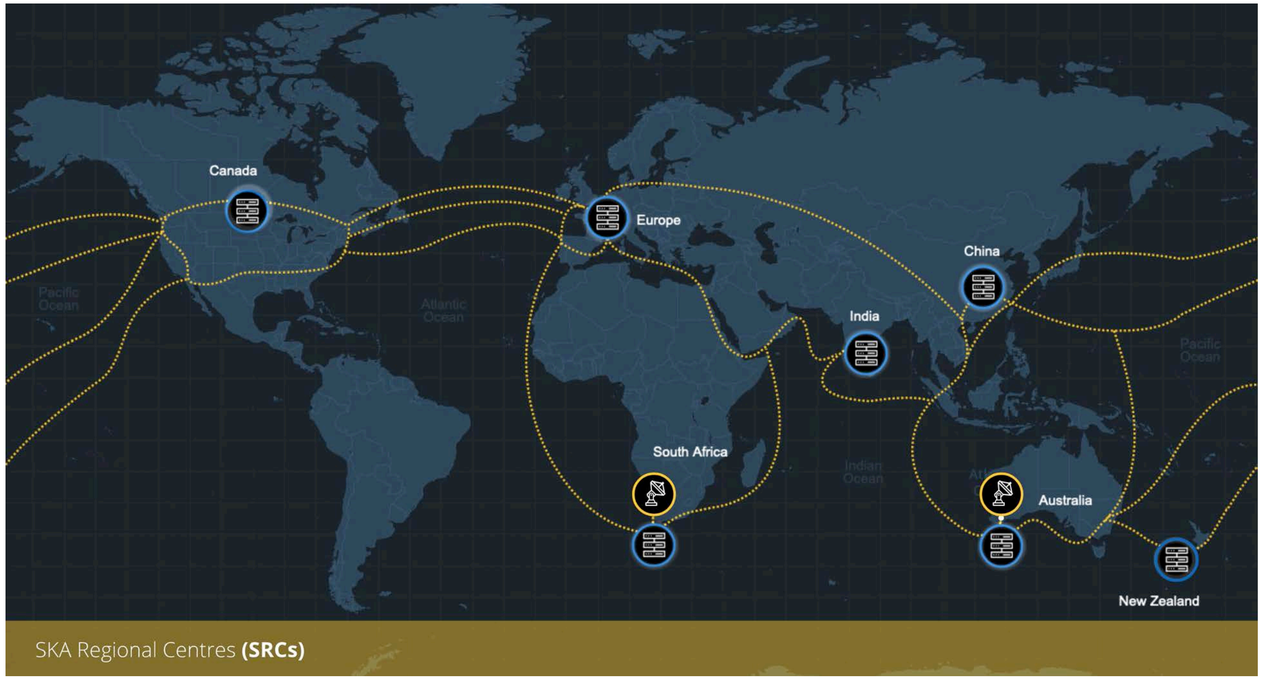
AENEAS results

Available through web site: www.aeneas2020.eu

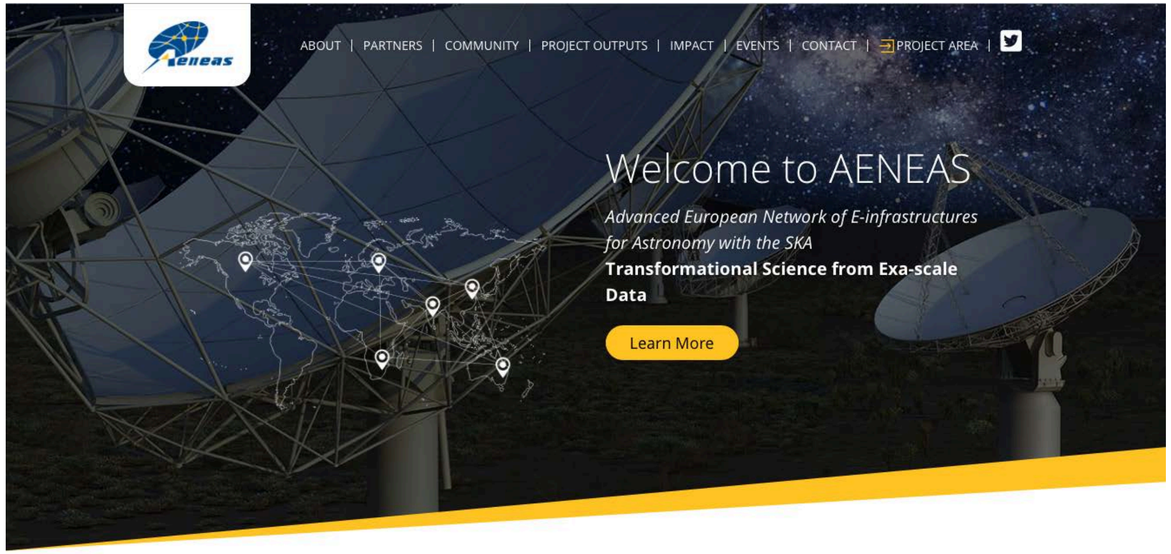
Key Project Findings

Deliverables & Milestones

Presentations



SKA Regional Centres **(SRCs)**



News & Announcements

A new treaty paves the way forward for the Square Kilometre Array

Mar 13, 2019 | News

Representatives from the founding member states of the Square Kilometre Array gathered in Rome yesterday to sign a treaty establishing the SKA Observatory as an intergovernmental organization that will oversee the delivery and operation of the world's largest radio...
read more

A step closer to a comprehensive design for the European SKA Regional Center

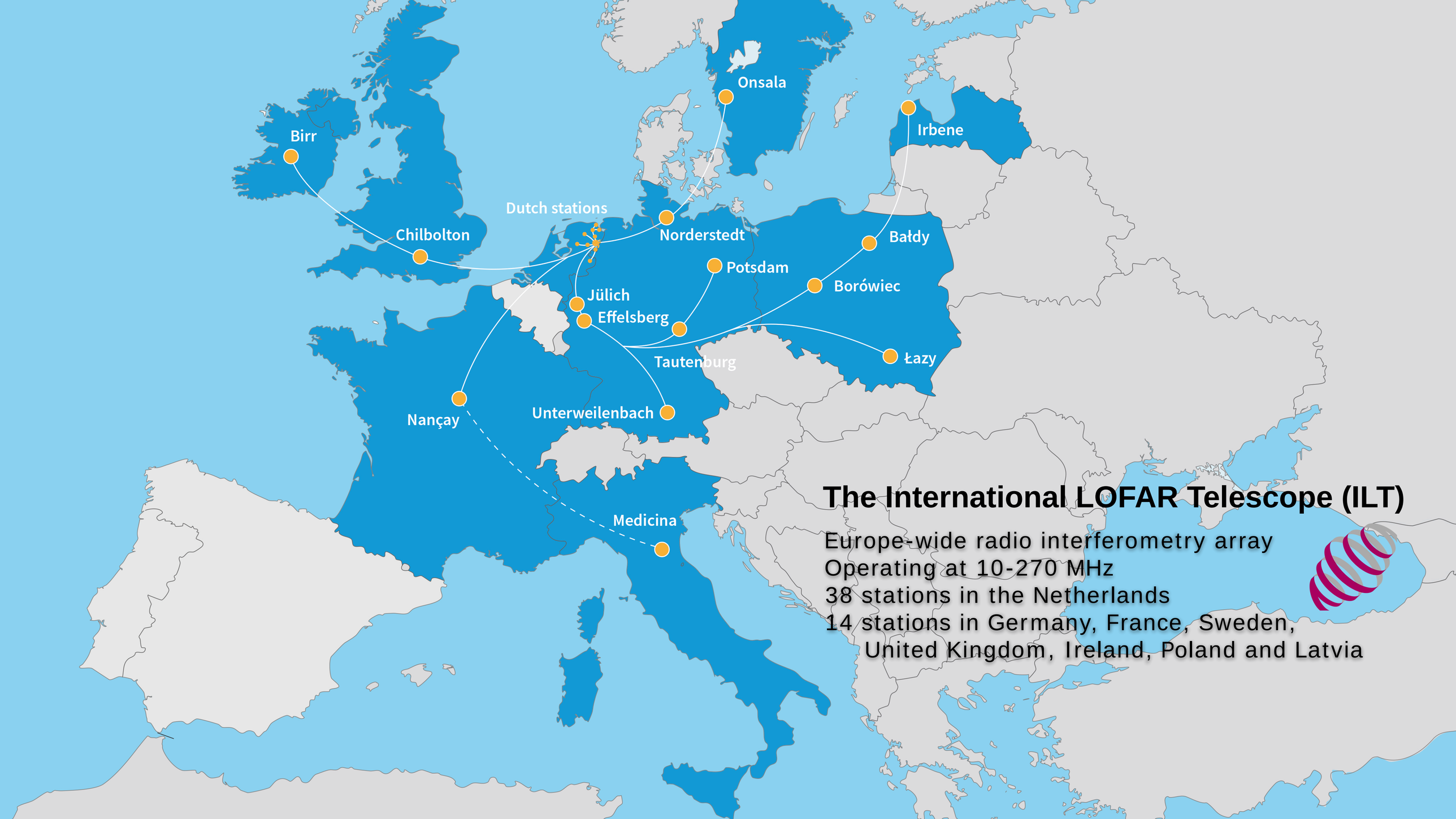
Mar 12, 2019 | News

The AENEAS team gathered at the University of Manchester last week for its 4th all-hands meeting. In addition to the usual project updates and presentations on regional center activities beyond Europe, much of the meeting was dedicated to focused discussions that...
read more

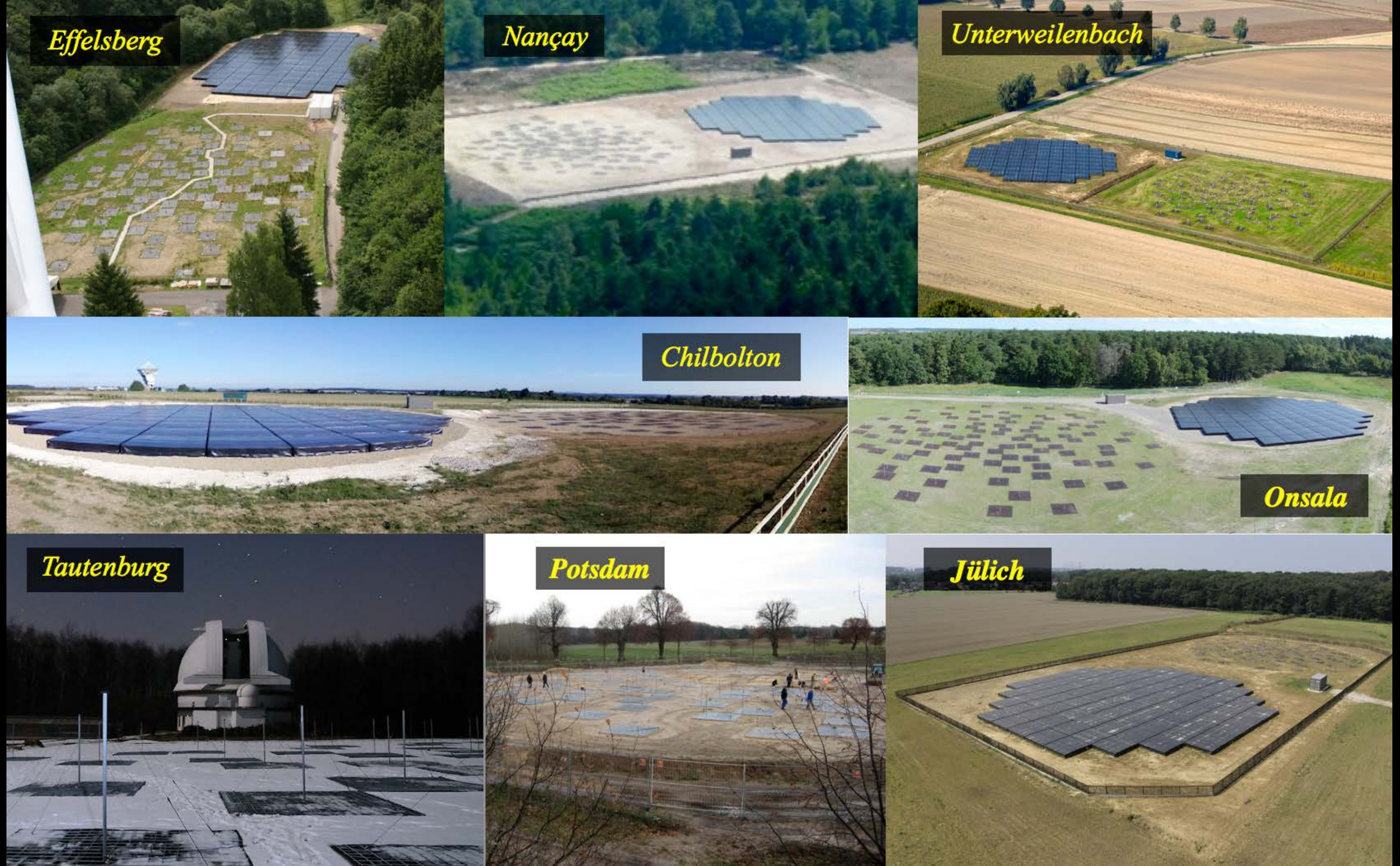
AENEAS and SKA collaborators convene in Manchester, UK

Feb 15, 2019 | News

The AENEAS team is convening to Manchesker, UK for the 4th all-hands meeting on March 5-7. Along with the usual updates from team members, contributed talks by colleagues at the SKA office and partners from other SKA regional center, the meeting will...
read more

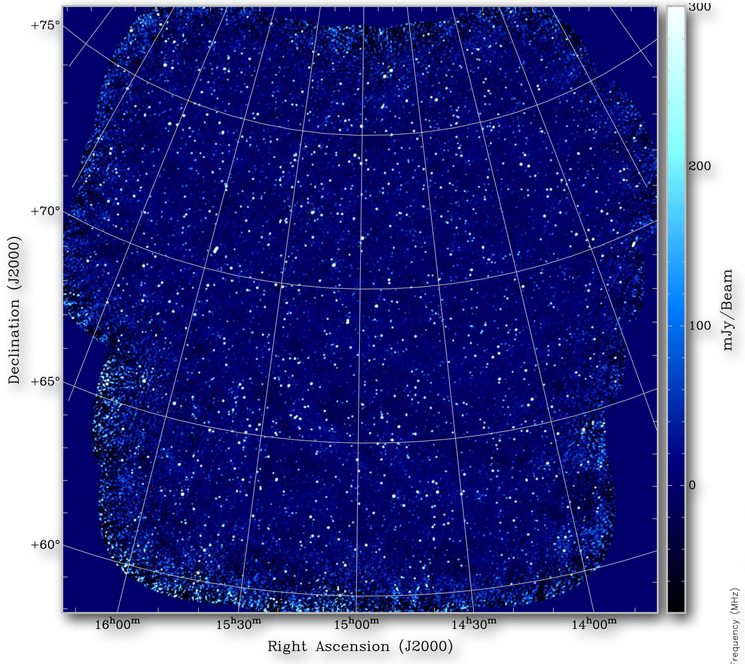

Onsala
Irbene
Birr
Dutch stations
Chilbolton
Norderstedt
Bałdy
Potsdam
Borówiec
Jülich
Effelsberg
Tautenburg
Łazy
Nançay
Unterweilenbach
Medicina
The International LOFAR Telescope (ILT)
Europe-wide radio interferometry array
Operating at 10-270 MHz
38 stations in the Netherlands
14 stations in Germany, France, Sweden,
United Kingdom, Ireland, Poland and Latvia

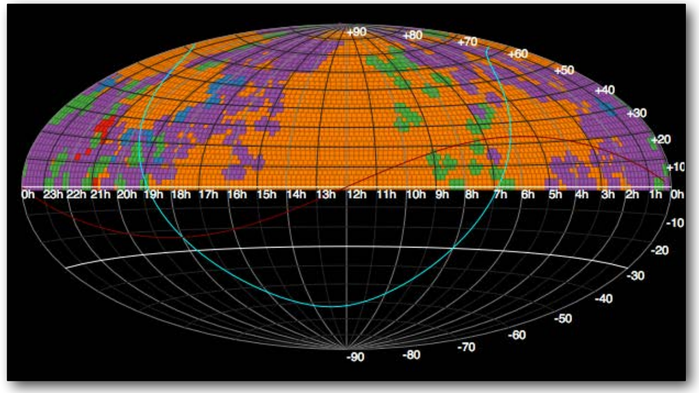

2011 © AEROPHOTO EELDE

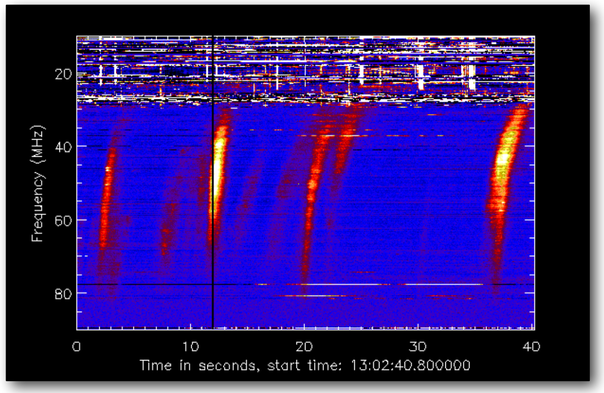

Effelsberg
Nançay
Unterweilenbach
Chilbolton
Onsala
Tautenburg
Potsdam
Jülich

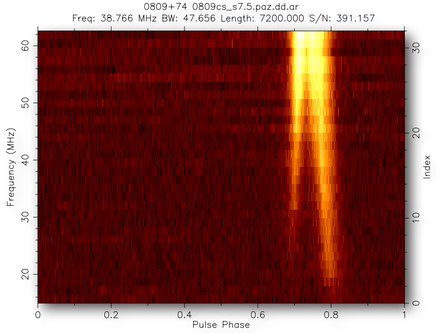
LOFAR Science Products

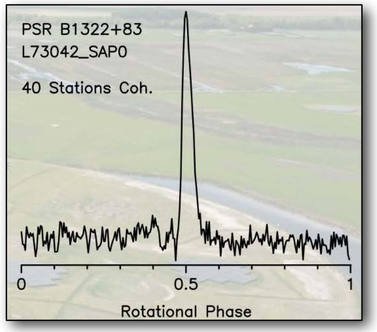
- Velocity *(Raw data rates of ~13 Tbits/s, correlated ~10 TB/hr)*
- Volume (*100 TB visibilities, 1 TB cubes, 1 PB catalogues*)
- Variety *(raw telemetry, uv data, beam-formed data, 2D-3D-4D-5D cubes, RM cubes, light-curves, catalogues, etc.)*

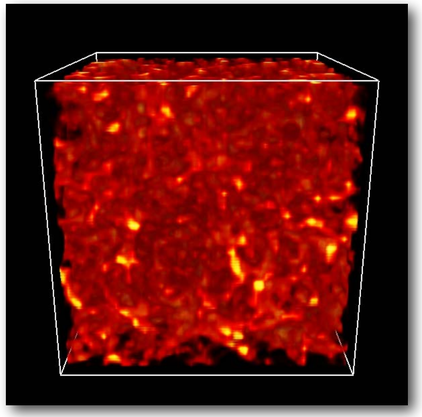






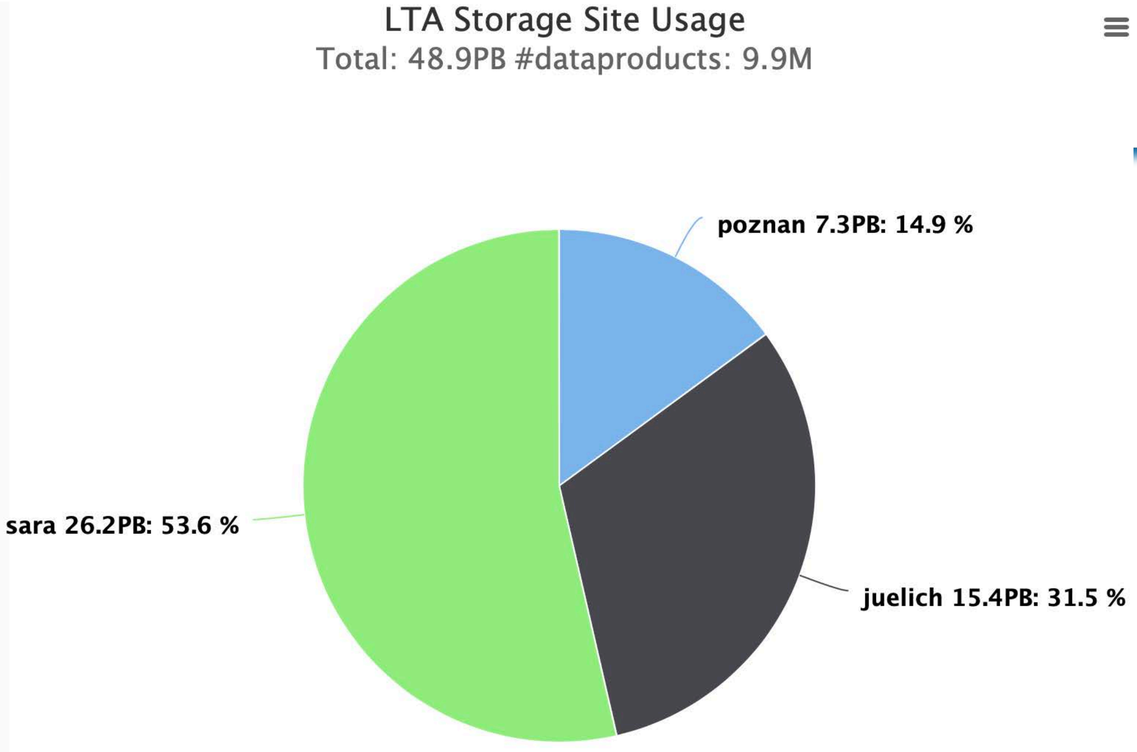


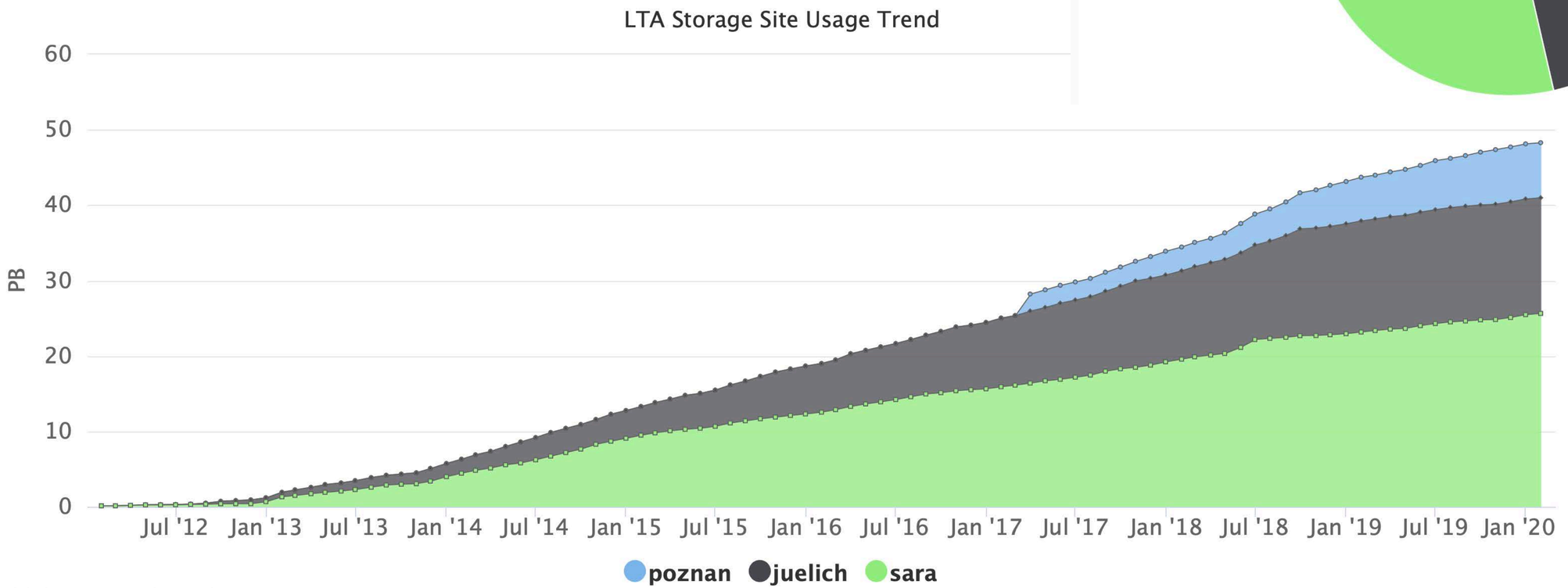




LOFAR Long Term Archive

- Current size archive ~49 PB
- Annual growth ~7 PB/year
- Data Stored: SURFsara (Amsterdam), Jülich and Poznan

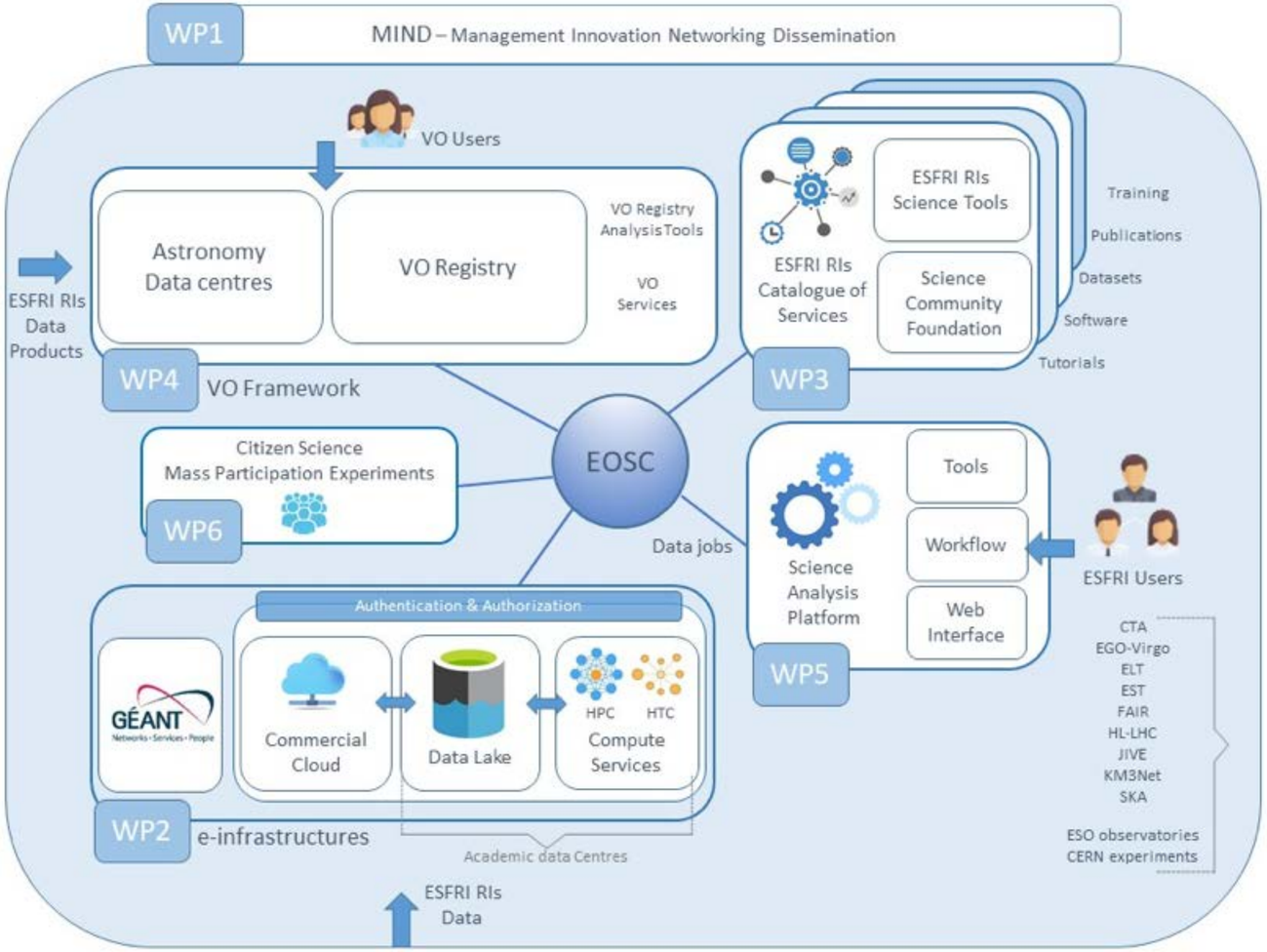




ESCAPE

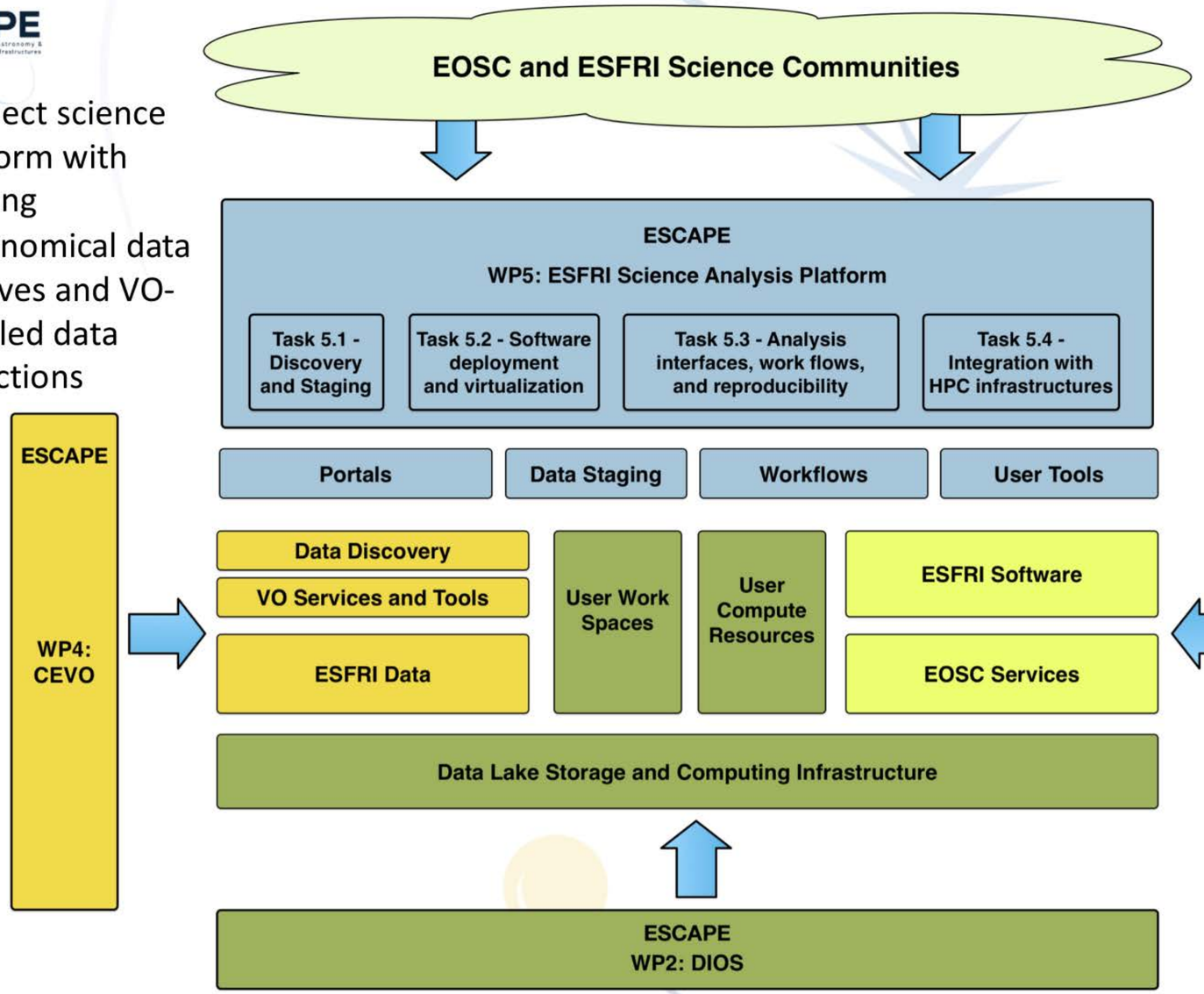
European Science Cluster of Astronomy & Particle physics
ESFRI research infrastructures

ESCAPE aims to address the Open Science challenges shared by ESFRI facilities as well as other pan-European research infrastructures in astronomy and particle physics



- EC H2020 (16 M€, 2019-2023)
- Partners include SKA, CTA, KM3Net, EST, ELT, HL-LHC, FAIR, CERN, ESO, JIVE
- Led by CNRS, 32 different EU institutions
- ASTRON leading Science Platform WP
- Work kicked off in February 2019

Connect science platform with existing astronomical data archives and VO-enabled data collections



Integration with Data Lake - distributed computing and storage

Access to software & services in ESCAPE-EOSC catalogue

Supporting

Summary

SKA Regional Centres

- Receiving ~600-700 PB of SKA data per year from ~2027
- Major effort required to prepare - using pathfinder data
- Funding these plans may be a challenge
- Global initiatives and collaboration essential
- Integrates well with European Open Science Cloud
- Expect great new scientific opportunities (e.g. multi-messenger)


The Square Kilometre Array