The SKA Observatory and the SKA Regional Centres: two challenges for the scientific research

Andrea Possenti

INAF – Osservatorio Astronomico di Cagliari

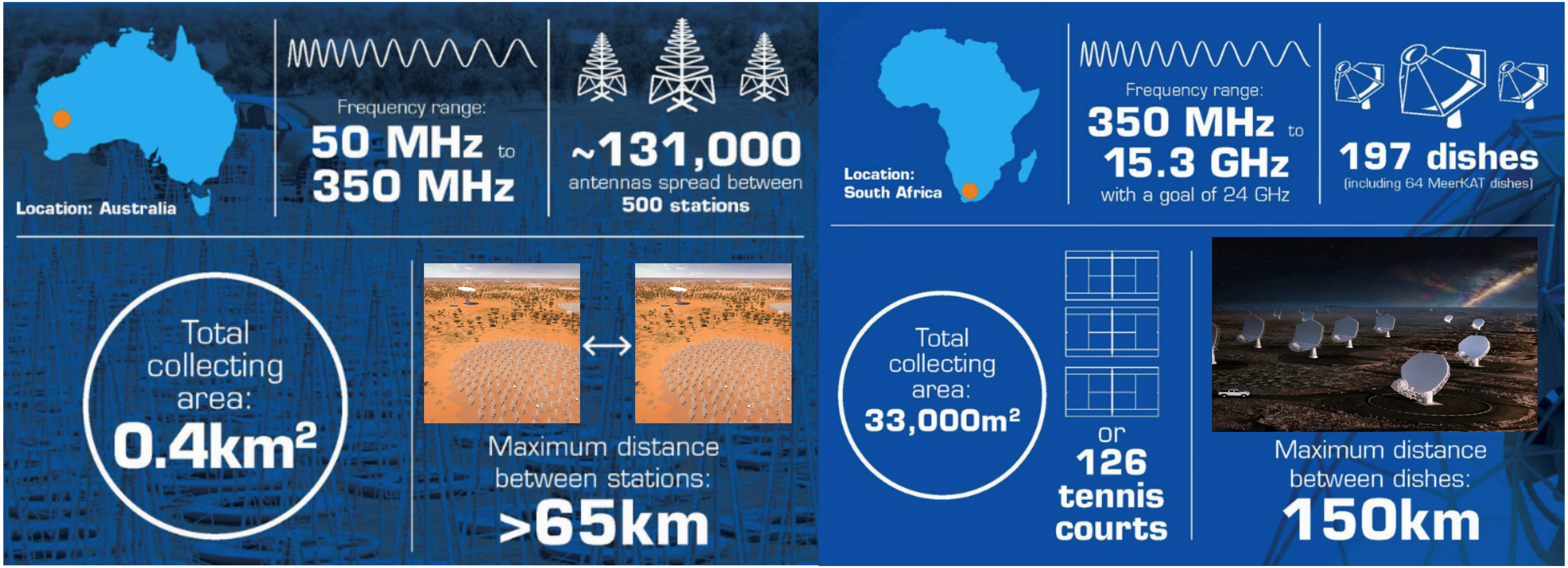
WORKSHOP GARR 2021

NET MAKERS

INAF
ISTITUTO NAZIONALE
DI ASTROFISICA

The SKA Observatory (SKAO)

SKAO in one slide



SKA-LOW in Western Australia

131000 antennas / 50-350 MHz

500 beams on sky

SKA-MID in South Africa

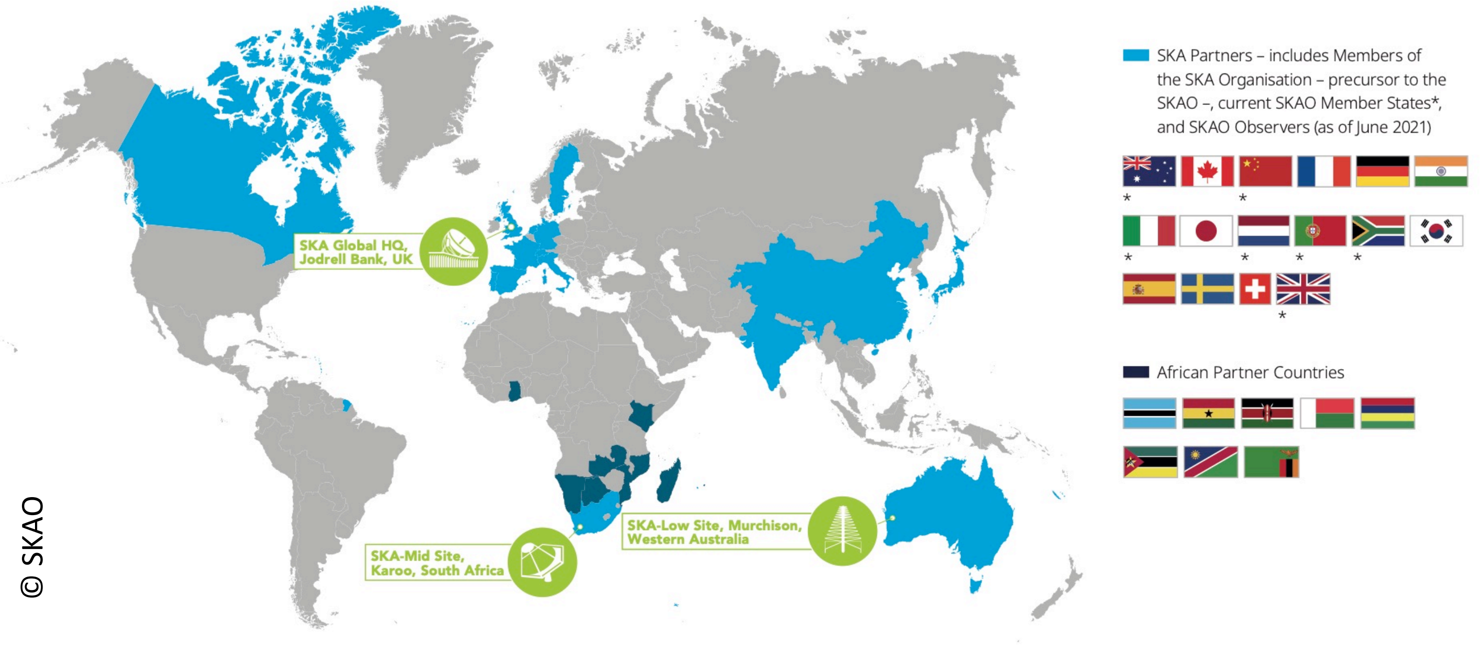
197 dishes / 0.35-1.05 GHz / 0.95-1.96 GHz

1500 beams on sky

SKAO members and funds

Established a cost ceiling for SKA1 capital expenditures of [2020 value] :

€ 1280 Million + € 600 Million running cost until 2031 (including €150 Million design effort)



© SKAO

Italian investment in construction € 100 Million

SKAO Governance

Signature for the IGO (inter-governmental organisation) occurred in Rome on 12 March 2019

© SKAO

IGO Council operational since 4 February 2021 and signed the start of the construction on 29 June 2021

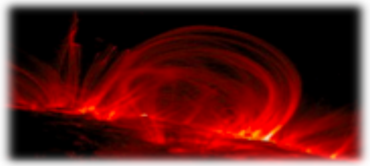
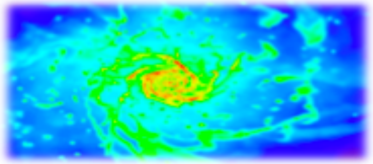
© SKAO

SKA expected sensitivity

Transformational SKA science I

	SKA1	SKA2
The Cradle of Life & Astrobiology	Proto-planetary disks; imaging inside the snow/ice line (@ < 100pc), Searches for amino acids.	Proto-planetary disks; sub-AU imaging (@ < 150 pc), Studies of amino acids.
	Targeted SETI: airport radar 10^4 nearby stars.	Ultra-sensitive SETI: airport radar 10^5 nearby star, TV ~10 stars.
Strong-field Tests of Gravity with Pulsars and Black Holes	1st detection of nHz-stochastic gravitational wave background.	Gravitational wave astronomy of discrete sources: constraining galaxy evolution, cosmological GWs and cosmic strings.
	Discover and use NS-NS and PSR-BH binaries to provide the best tests of gravity theories and General Relativity.	Find all ~40,000 visible pulsars in the Galaxy, use the most relativistic systems to test cosmic censorship and the no-hair theorem.
The Origin and Evolution of Cosmic Magnetism	The role of magnetism from sub-galactic to Cosmic Web scales, the RM-grid @ 300/deg2.	The origin and amplification of cosmic magnetic fields, the RM-grid @ 5000/deg2.
	Faraday tomography of extended sources, 100pc resolution at 14Mpc, 1 kpc @ z ≈ 0.04.	Faraday tomography of extended sources, 100pc resolution at 50Mpc, 1 kpc @ z ≈ 0.13.
Galaxy Evolution probed by Neutral Hydrogen	Gas properties of 10^7 galaxies, <z> ≈ 0.3, evolution to z ≈ 1, BAO complement to Euclid.	Gas properties of 10^9 galaxies, <z> ≈ 1, evolution to z ≈ 5, world-class precision cosmology.
	Detailed interstellar medium of nearby galaxies (3 Mpc) at 50pc resolution, diffuse IGM down to N_H < 10^17 at 1 kpc.	Detailed interstellar medium of nearby galaxies (10 Mpc) at 50pc resolution, diffuse IGM down to N_H < 10^17 at 1 kpc.

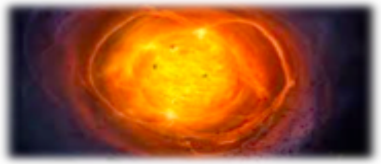


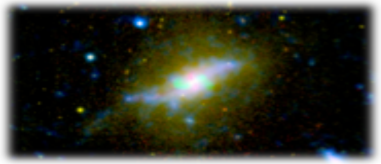



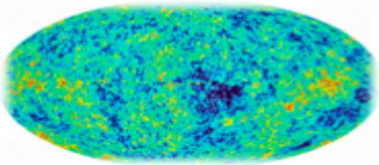
Courtesy: Robert Braun (SKAO)

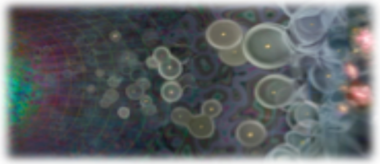
Transformational SKA science II

	SKA1	SKA2
The Transient Radio Sky	Use fast radio bursts to uncover the missing "normal" matter in the universe.	Fast radio bursts as unique probes of fundamental cosmological parameters and intergalactic magnetic fields.
	Study feedback from the most energetic cosmic explosions and the disruption of stars by super-massive black holes.	Exploring the unknown: new exotic astrophysical phenomena in discovery phase space.
Galaxy Evolution probed in the Radio Continuum	Star formation rates (10 M_Sun/yr to z ~ 4).	Star formation rates (10 M_Sun/yr to z ~ 10).
	Resolved star formation astrophysics (sub-kpc active regions at z ~ 1).	Resolved star formation astrophysics (sub-kpc active regions at z ~ 6).
Cosmology & Dark Energy	Constraints on DE, modified gravity, the distribution & evolution of matter on super-horizon scales: competitive to Euclid.	Constraints on DE, modified gravity, the distribution & evolution of matter on super-horizon scales: redefines state-of-art.
	Primordial non-Gaussianity and the matter dipole: 2x Euclid.	Primordial non-Gaussianity and the matter dipole: 10x Euclid.
Cosmic Dawn and the Epoch of Reionization	Direct imaging of EoR structures (z = 6 - 12).	Direct imaging of Cosmic Dawn structures (z = 12 - 30).
	Power spectra of Cosmic Dawn down to arcmin scales, possible imaging at 10 arcmin.	First glimpse of the Dark Ages (z > 30).





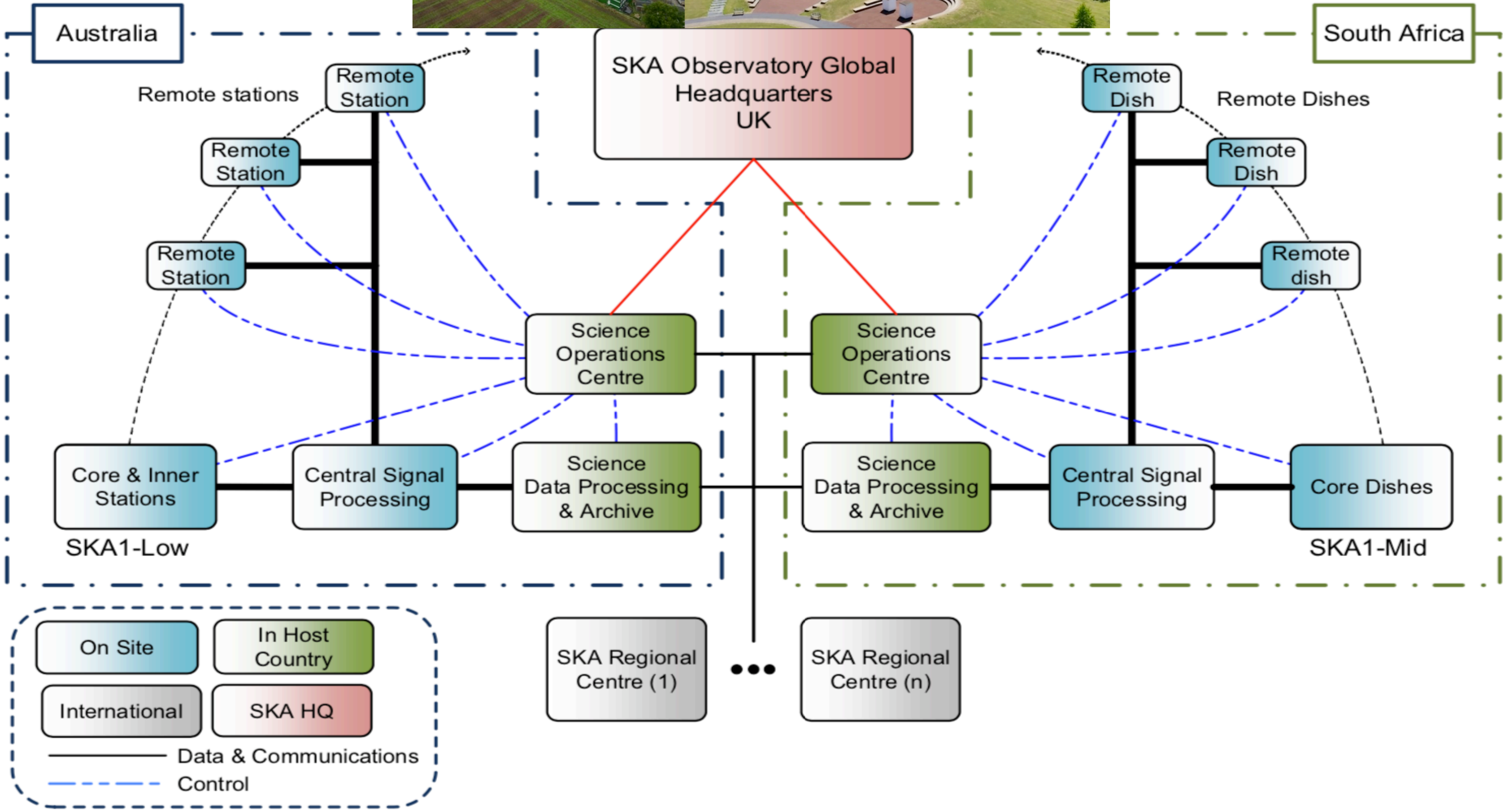




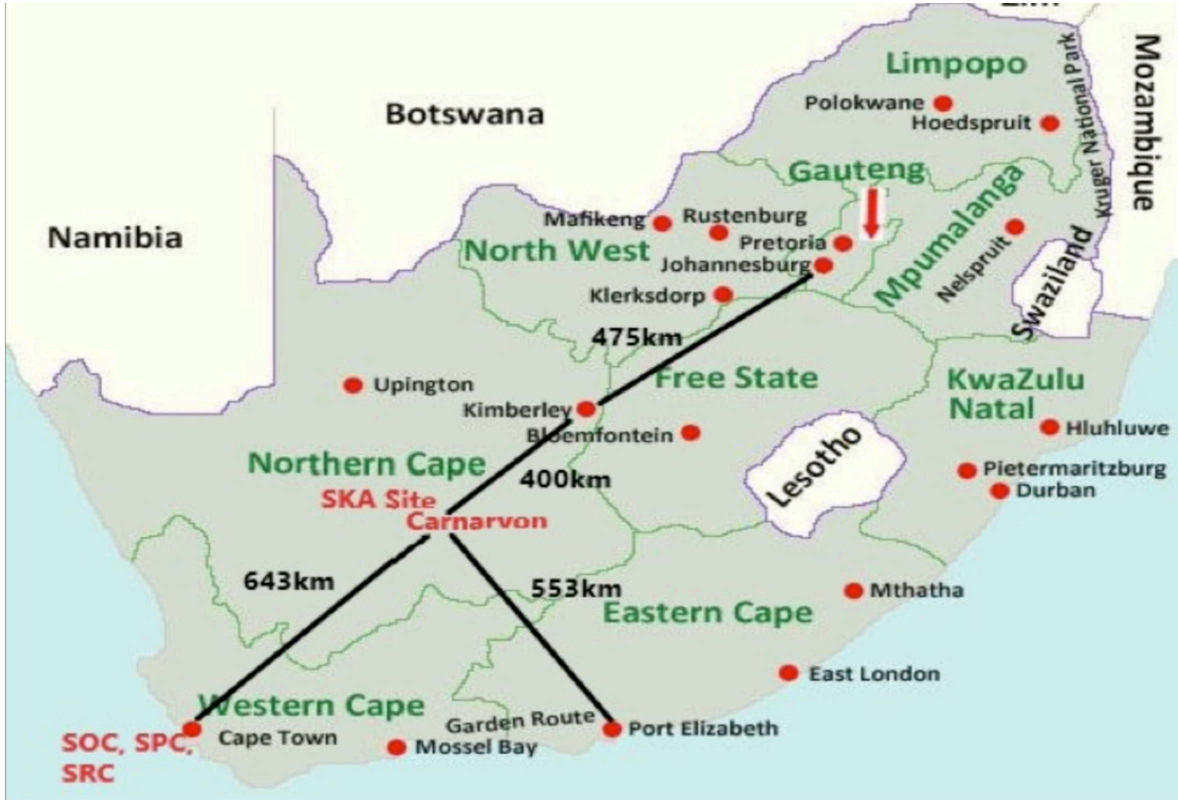
Courtesy: Robert Braun (SKAO)

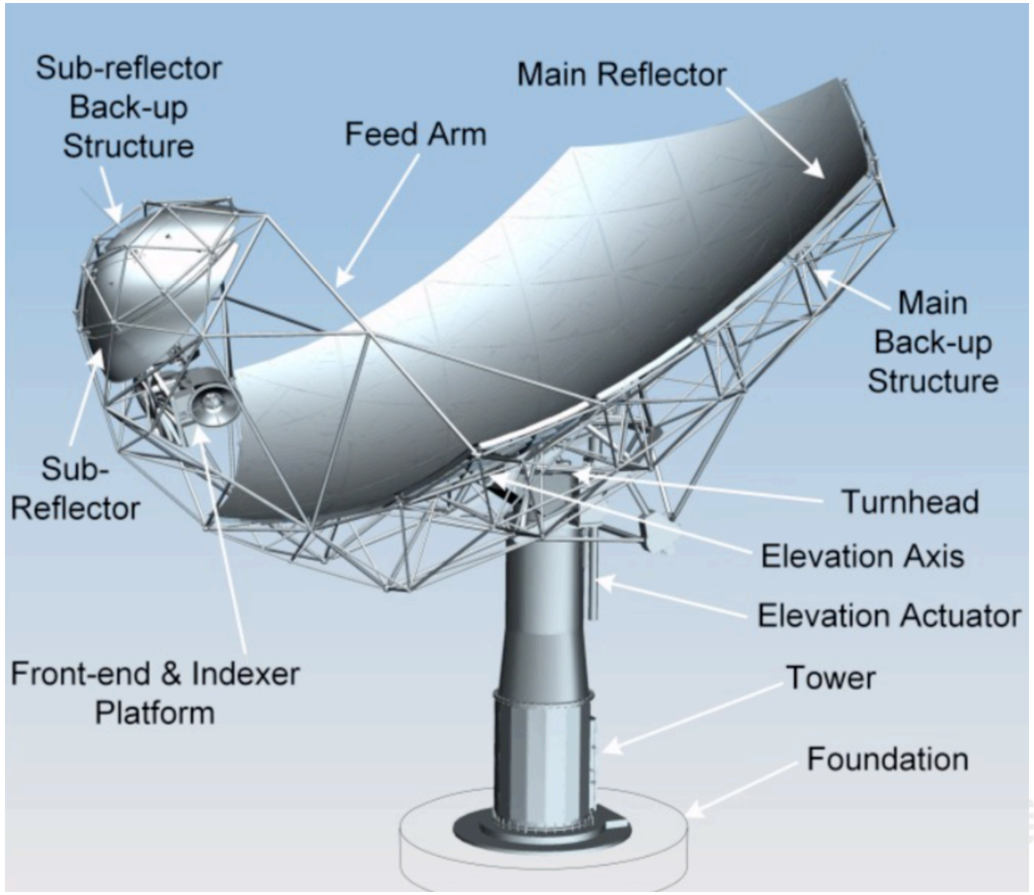
SKAO global view

© SKAO

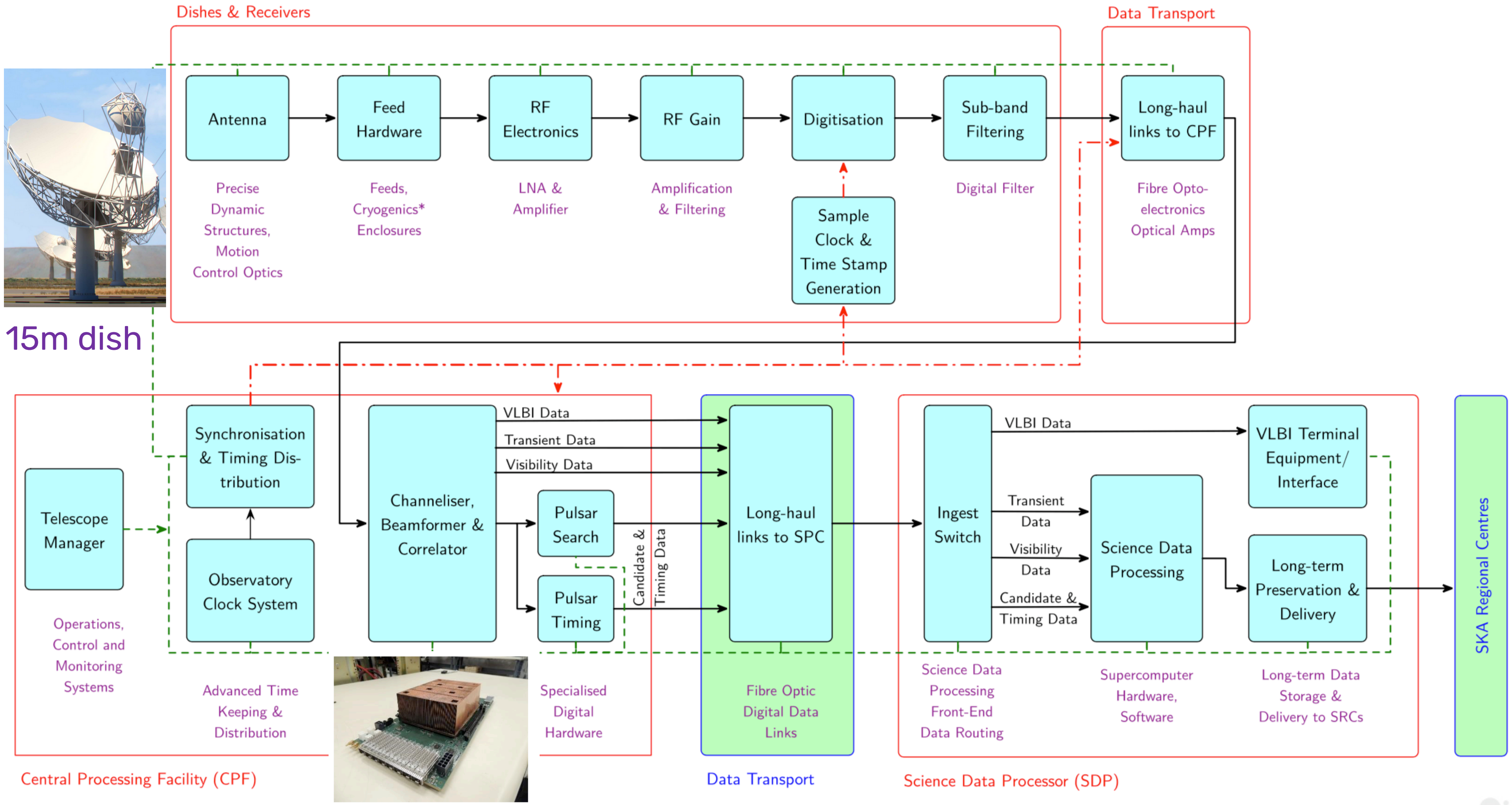


SKA – MID location and antenna





SKA-MID Block diagram



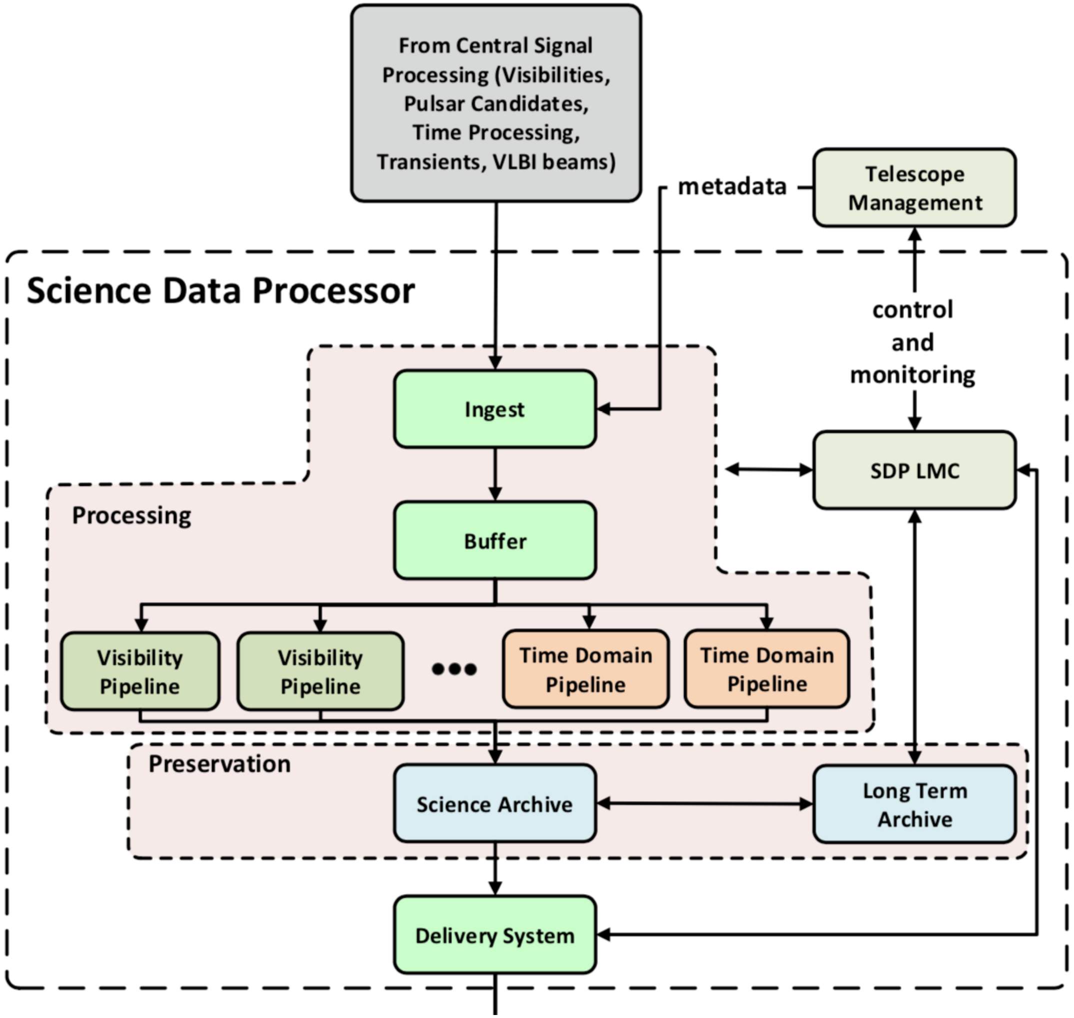
FPGA correlator

Adapted from : Construction Proposal (SKAO)

The Science Data Processor

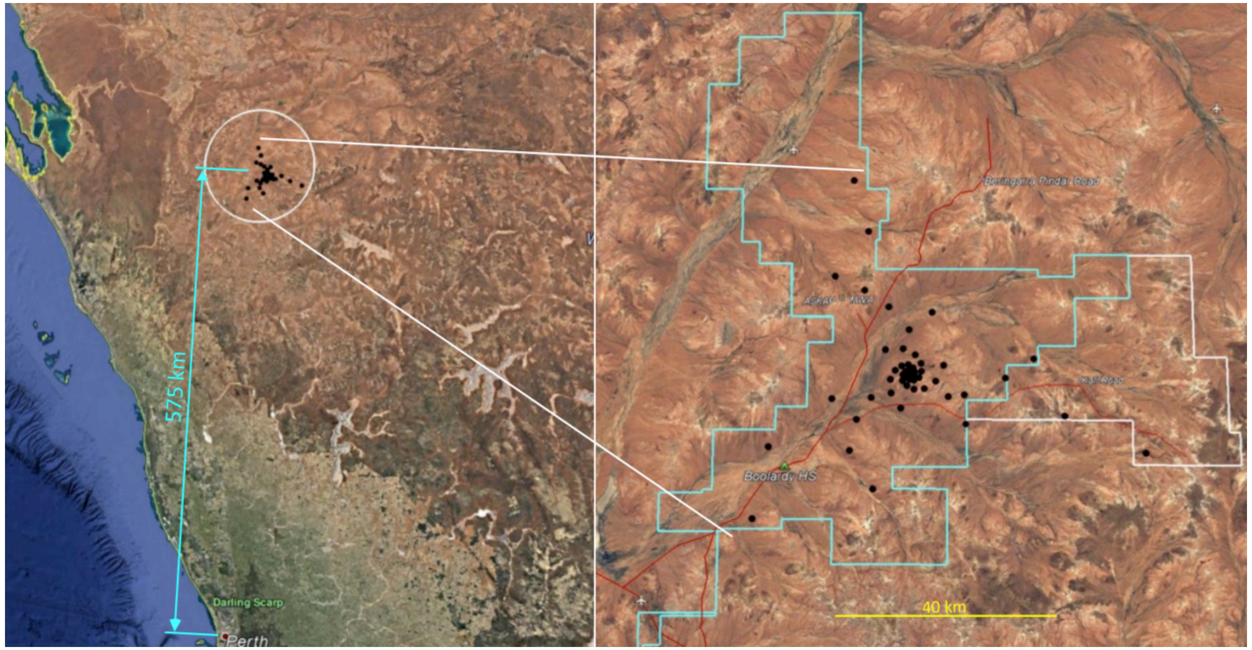
Two identical machines at the two antenna hosting Countries (at CapeTown for SKA – MID and at Perth for SKA – LOW)

> 130 Pflop/s each

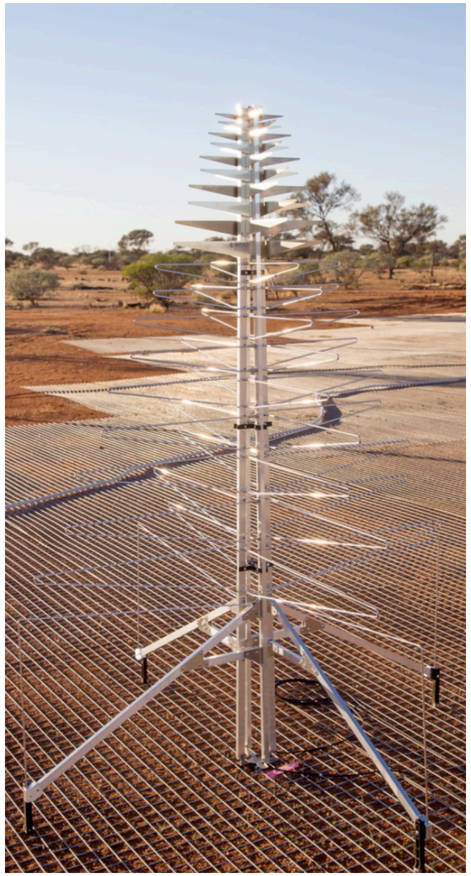


Adapted from : Construction Proposal (SKAO)

SKA – LOW location and antenna



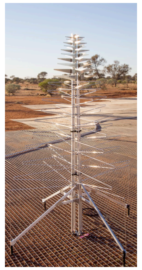
© Construction Proposal (SKAO)



© Construction Proposal (SKAO)

SKA – LOW block diagram

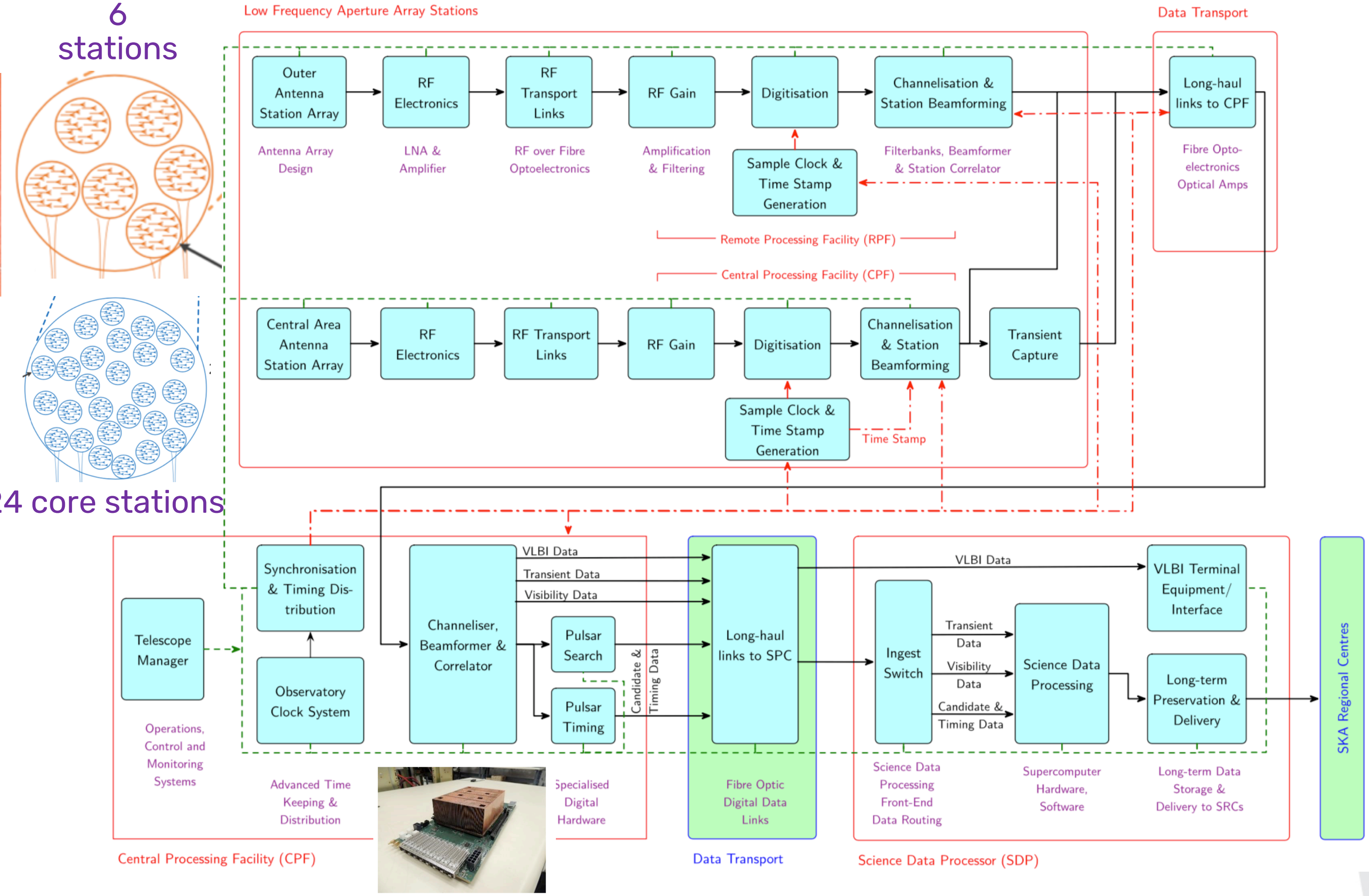
log-periodic antenna



station of 256 antennas

6 stations

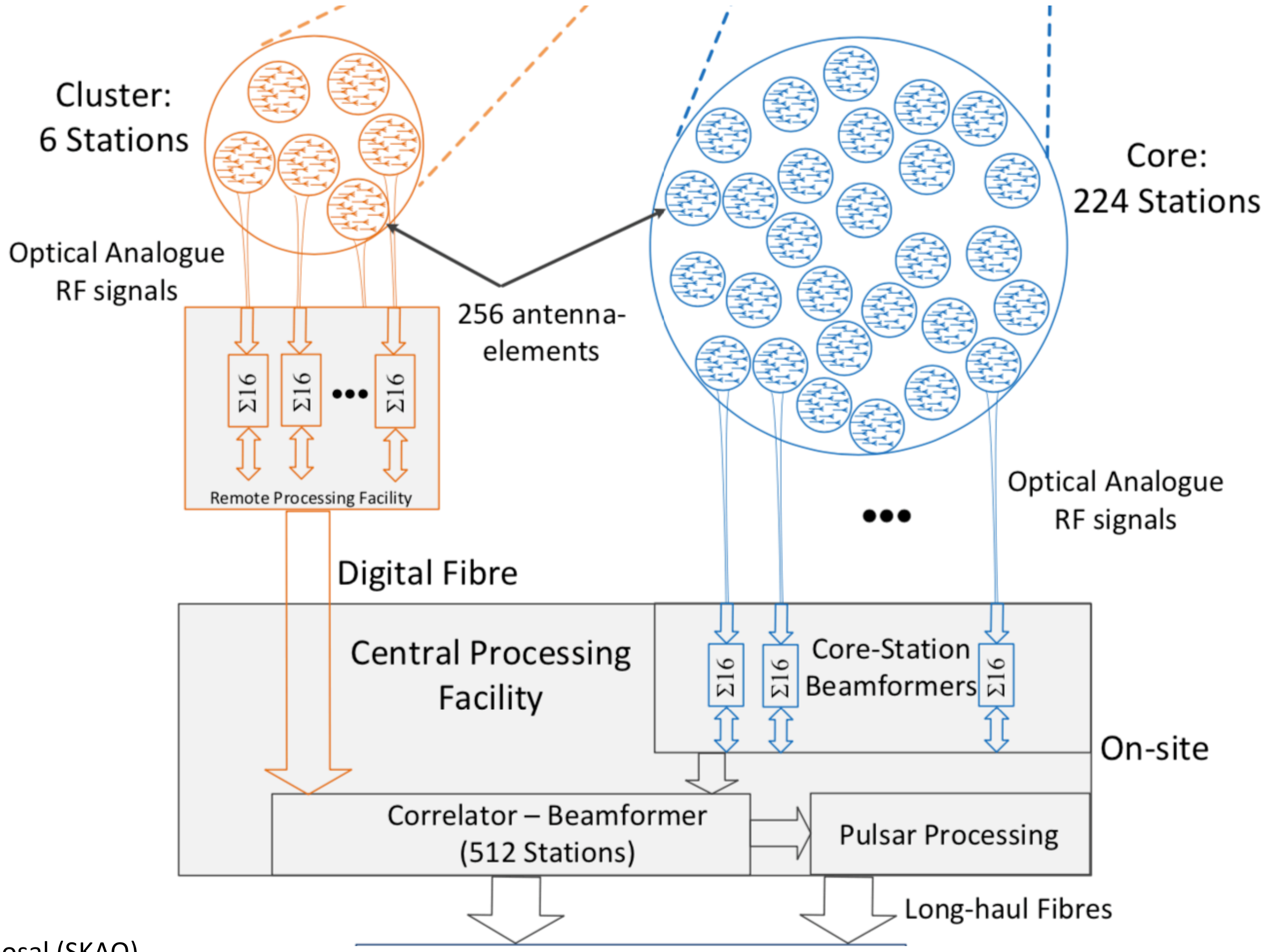
224 core stations



FPGA correlator

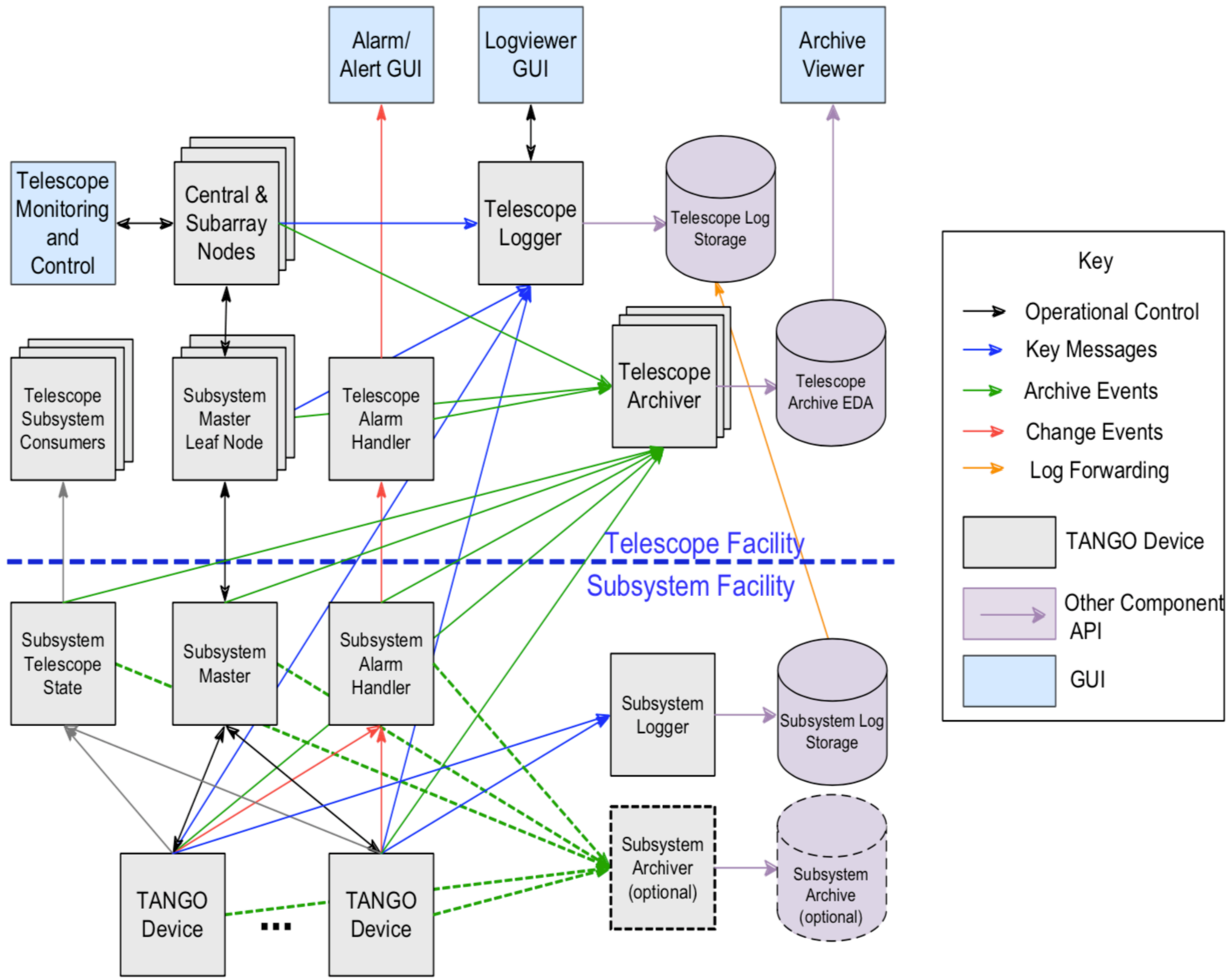
Adapted from : Construction Proposal (SKAO)

512 clusters of 256 antennas: all data to CPF

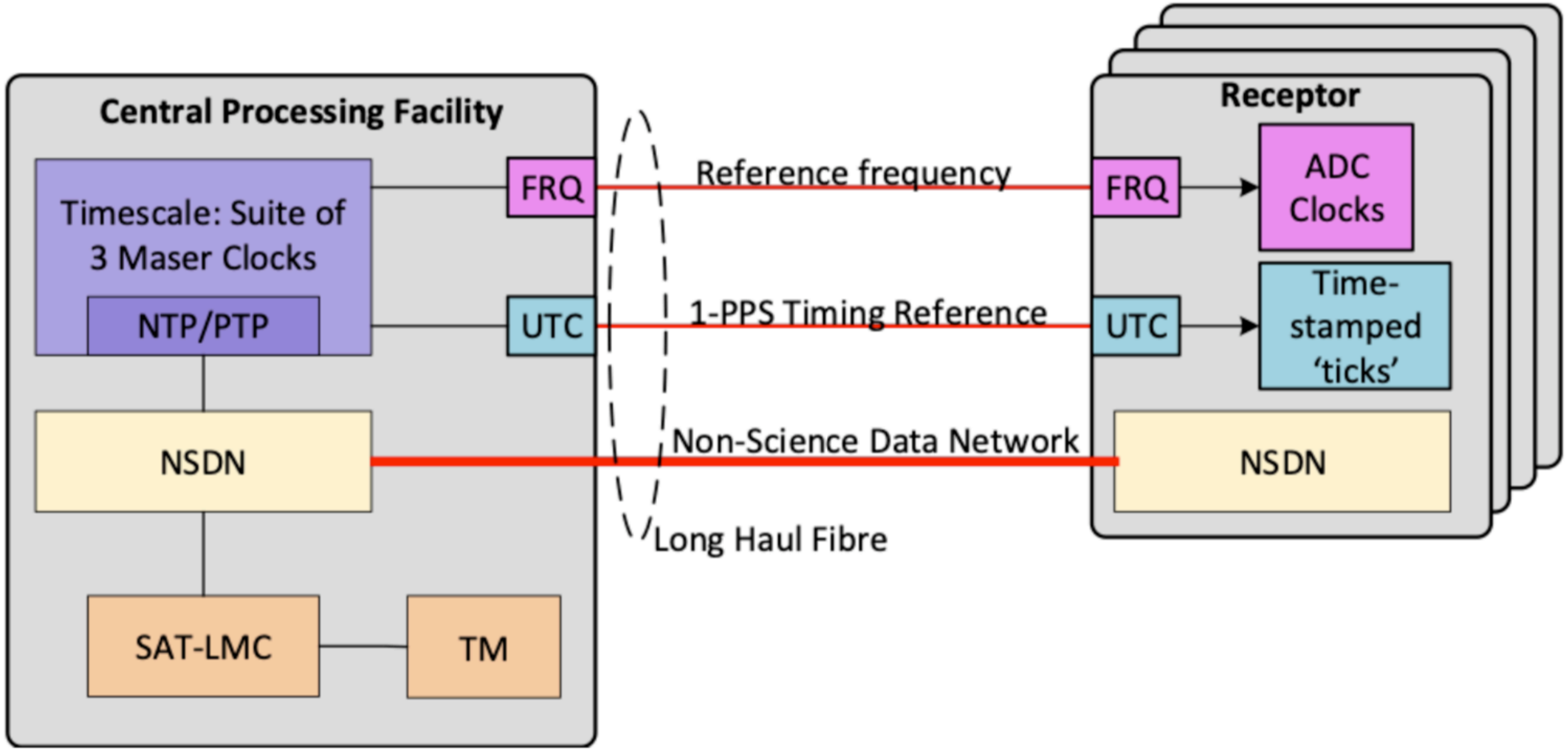


Adapted from : Construction Proposal (SKAO)

The TANGO framework to control the telescope



Time and frequency keeping and distributing



synchronisation accuracy with UTC over a time-duration of at least 10 years:

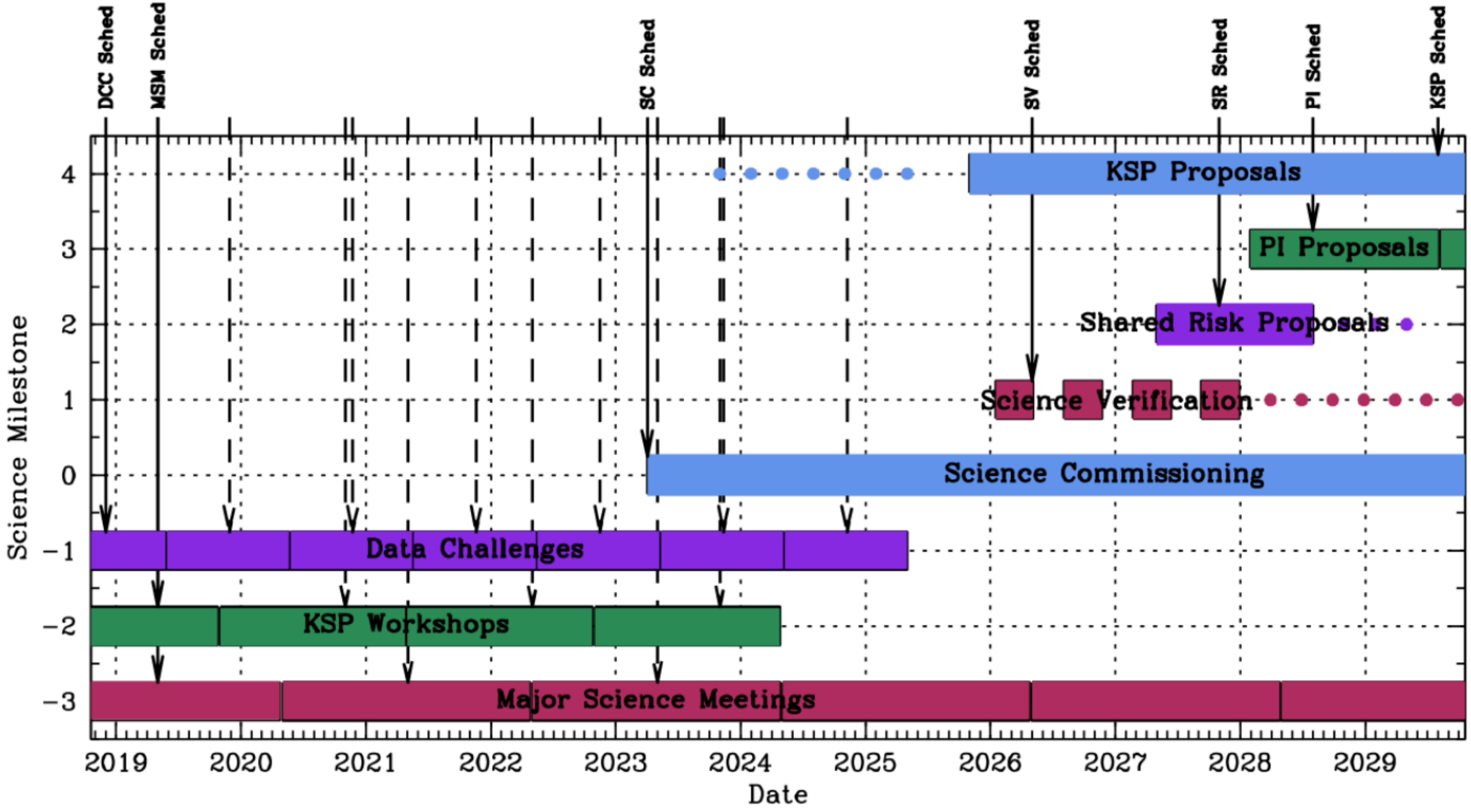
5 ns for SKA-MID

9 ns for SKA-LOW

the required accuracy for the UTC ***distribution*** is

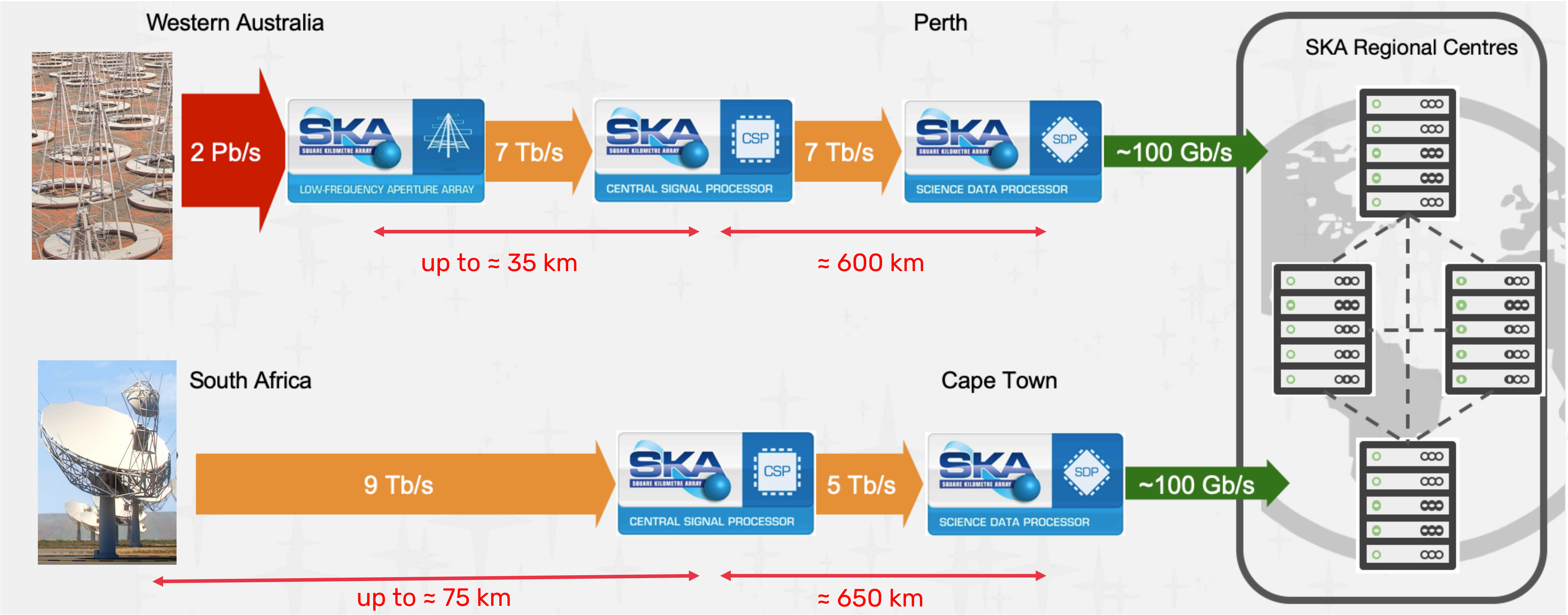
2 ns (rms) for both SKA-LOW and SKA-MID

Timeline for SKAO



Courtesy: Robert Braun (SKAO)

The flow of the SKAO “Data Rate”

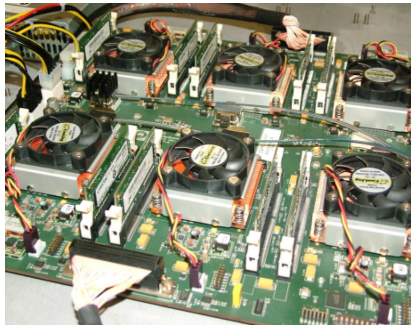


SKA output Data rates are extraordinary for a science experiment
Approximately 50-70 times higher output than LSST, or CTAO. In line with "High Luminosity" LHC at CERN

Adapted from Rosie Bolton (SKAO)

The SKAO Data Flow

CSP: Central Signal Processor



e.g. FPGAs in the ASKAP correlator

5 Tb/s + 7Tb/s
data buffer of 2 weeks

SDP: Science Data Processor

e.g. SDP prototype, Cambridge

9 Tb/s + 7 Tb/s data buffer of 2 minutes

≈ 300 PB/yr of delivered data for each of the two sites

SRC: SKA Regional Centre network

SRC
SRC
SRC
SRC
SRC

Distributed facilities

SKA portal

USERS

Courtesy: Philippa Hartley (SKAO)

SKAO REGIONAL CENTRE NETWORK

WORKSHOP GARR 2021 NET MAKERS

The SKA Regional Centres (SRC)

A short story of the Ska Regional Centres (SRCs)

July 2016: the SKA Board created the concept:

"The SKA Observatory will coordinate a network of SKA Regional Centres that will provide the data access, data analysis, data archive and user support interfaces with the user community"

November 2018: the SKA Board launched the start of the activities, via the set up of a SC:

"The mission of the SRC Steering Committee (SRC-SC) is to define and create a long-term operational partnership between the SKA Observatory and an ensemble of independently-resourced SKA Regional Centres".

November 2021: ***SRC-SC: 14 member Countries + 2 observer Countries + SKAO***

Australia – Canada – China – France – Germany – India – Italy – The Netherlands – Portugal – South Africa – Spain – Sweden – Switzerland – United Kingdom – SKA Observatory – Japan (observer) – Korea (observer)

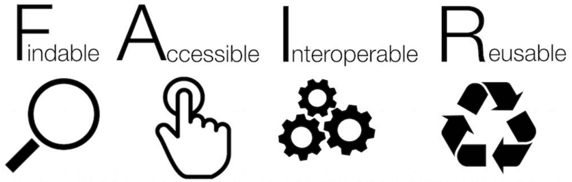
SKAO REGIONAL CENTRE NETWORK

WORK SHOP GARR 2021 NET MAKERS

The responsibilities of the SKA Observatory and of the Ska Regional Centres (SRCs)

- The SKA Observatory and the SRCs will be jointly responsible for:

- a) maximizing the quality of SKA data delivered to users;
- b) the production of Advanced Data Products;
- c) storing, archiving and curation of the primary SKA output data and of the Advanced Data Products;
- d) ensuring that the approved science program can be accommodated within available resources;
- e) ensuring the availability of a platform of distributed services across computational and data infrastructures to support the user community to deliver SKA science, under the FAIR principles.

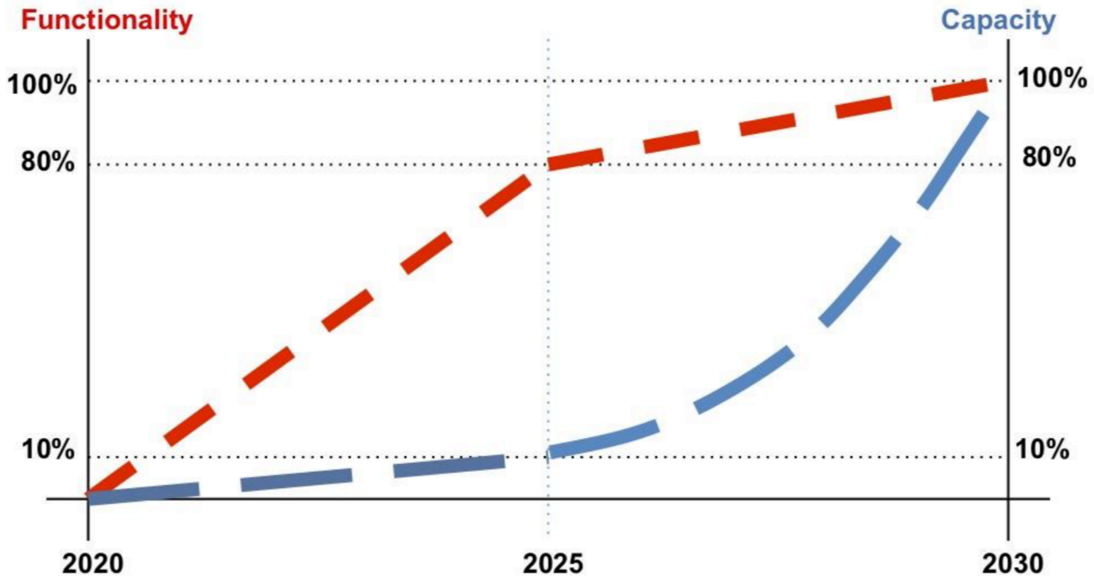


SKAO REGIONAL CENTRE NETWORK

WORKSHOP GARR 2021 NET MAKERS

Still under discussion …

- Governance and Operations
- National Participation
- Baseline Functionalities



Ramp-up of the SRC network

First guess on estimated budget

Some reference specifics for the whole SRC network at regime ≈ 2029

Data Flow PB/yr	Processing PFlop/s	Network mean speed Gb/s
710	22	100

Some reference costs for the whole SRC network at regime ≈ 2029

Data (M€/yr)	Processing (M€/yr)	Network (M€/yr)	Personnel (M€/yr)
18	2.4	5	10

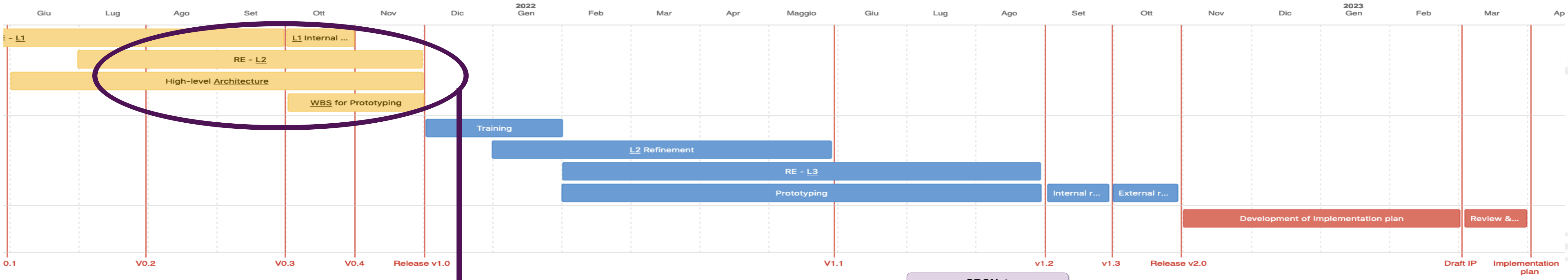
Allowing for the current uncertainties in the design
the likely cost for the whole international SRC network at regime
will be in the 20-40 M€/year range …

… including ≈ 100 FTE of personnel

The SKA Regional Centres Working Groups

- WG0: SRC Network Architecture
- WG1: Data Logistics Working Group
- WG2: Operations Working Group
- WG3: SW, Federated Computing and Data Software Services
- WG4: SW, Science Archive-VO-FAIR
- WG5: Compute Working Group
- WG6: Science User Engagement

The timeline for the SRC initial prototyping

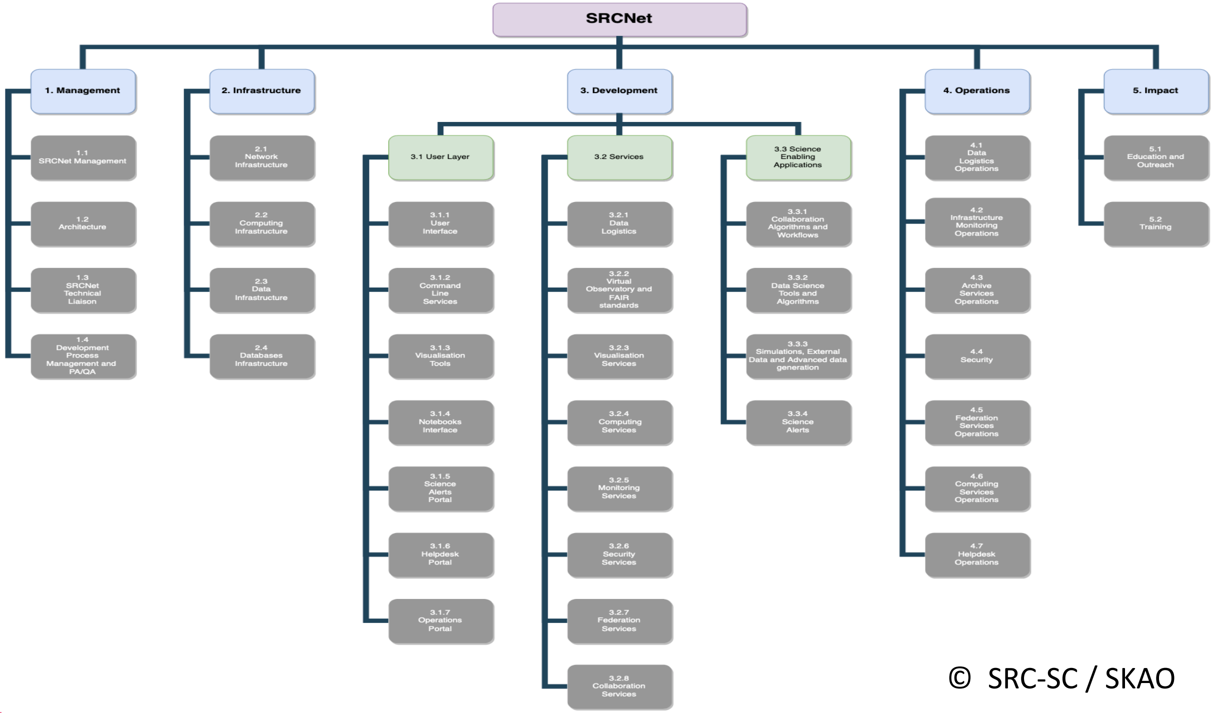


Now mostly working on:

* Collection of requirements

* Work Breakdown Structure (WBS)

Expertise in the Working Groups will be distributed in the WBS



SKAO REGIONAL CENTRE NETWORK

WORKSHOP GARR 2021 NET MAKERS

Italian involvement in current SRC activities

≈ 100 Italian scientists are members of the SKA Science Working Groups!

Develop requests and imagine solutions to the USE CASES for the SRC network

+

Staying at the frontline in ADAPTING to the new way for doing data reduction and computation in the SKA era

≈ 15 INAF members are involved in the SRC core activities (+ 10 consultant)

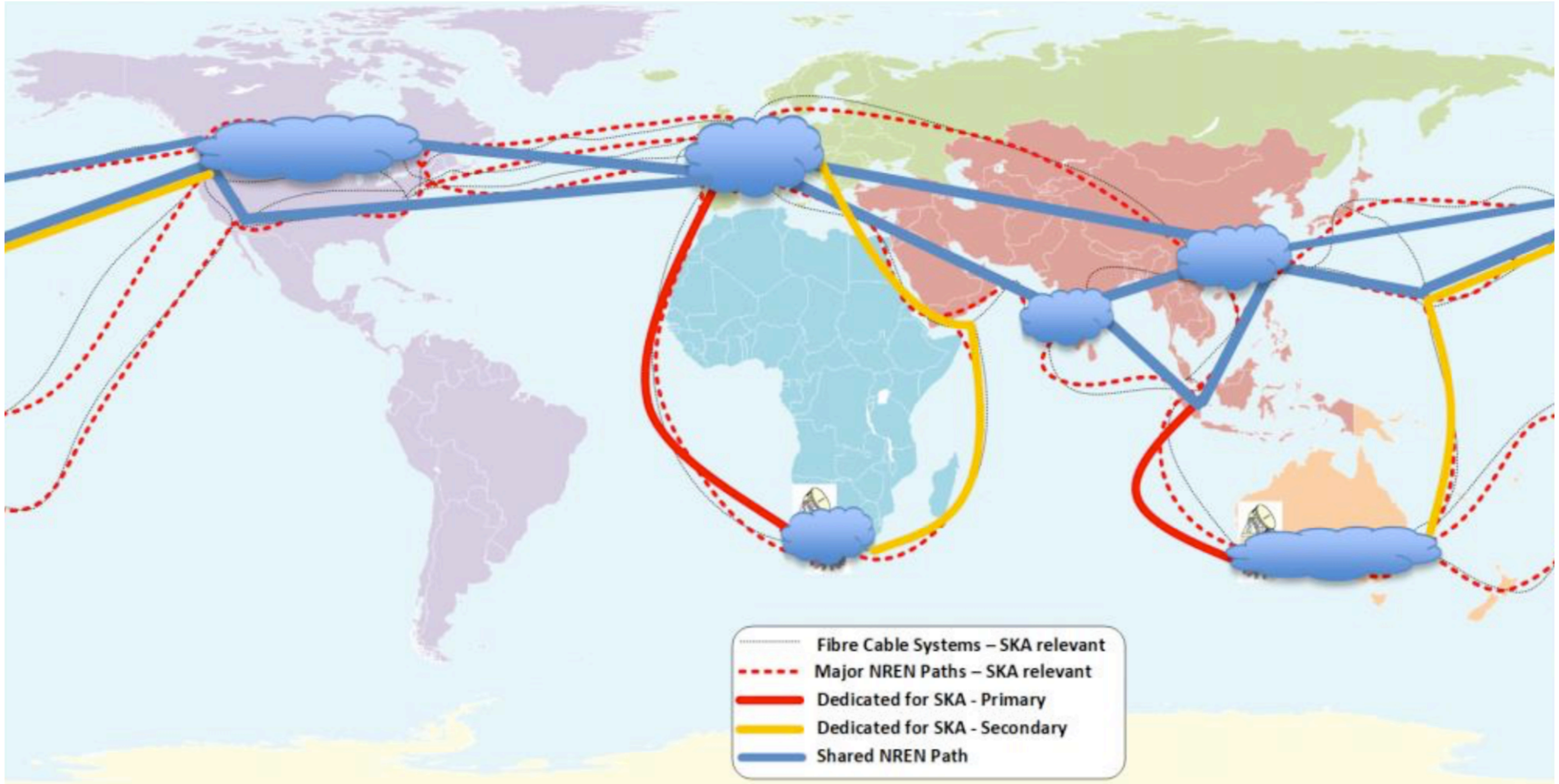
Working Group	Theme	Italian participants
WG0	SRC Network Architecture	1 core member
WG1	Data Logistics Working Group	1 core member
WG2	Operations Working Group	1 core member + 3 consultant
WG3	SW, Federated Computing and Data Software Services	3 core members + 1 consultant
WG4	SW, Science Archive-VO-FAIR	1 core member + 4 consultant
WG5	Compute Working Group	1 task leader + 2 core members
WG6	Science User Engagement	1 chair + 2 task leader + 2 core members + 6 consultant

Some of the IT-related challenges of SKAO data

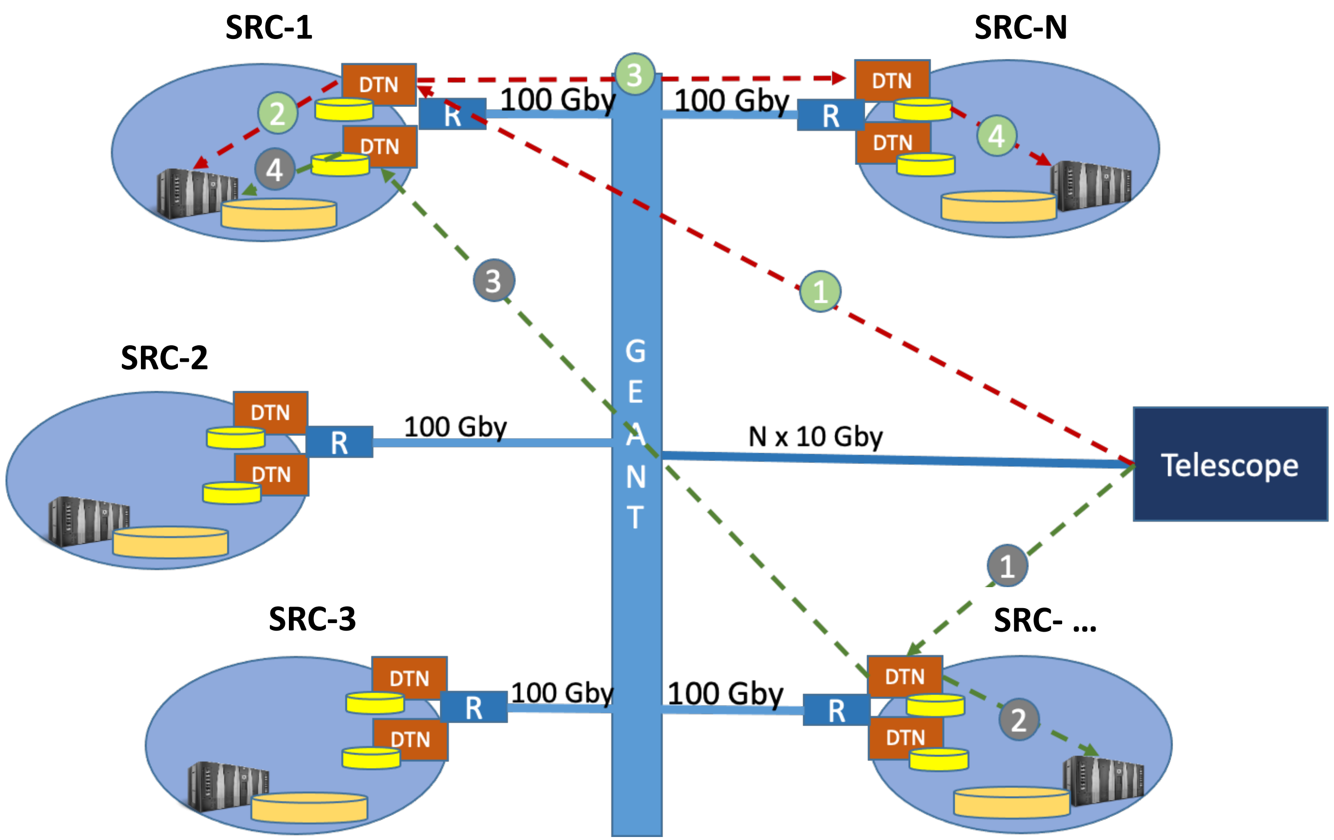
- Calibration
- Polarization Calibration
- RFI excision in presence of very large number of frequency channels
- Huge data volumes to transport
- Unprecedented number of sources per pointing to extract and characterize
- Data visualization
- Data archiving
- .....

Early discussion about data network for SKA+SRC science

network uses the ***general academic network*** infrastructure as well as ***dedicated links***



Data products workflow under discussion



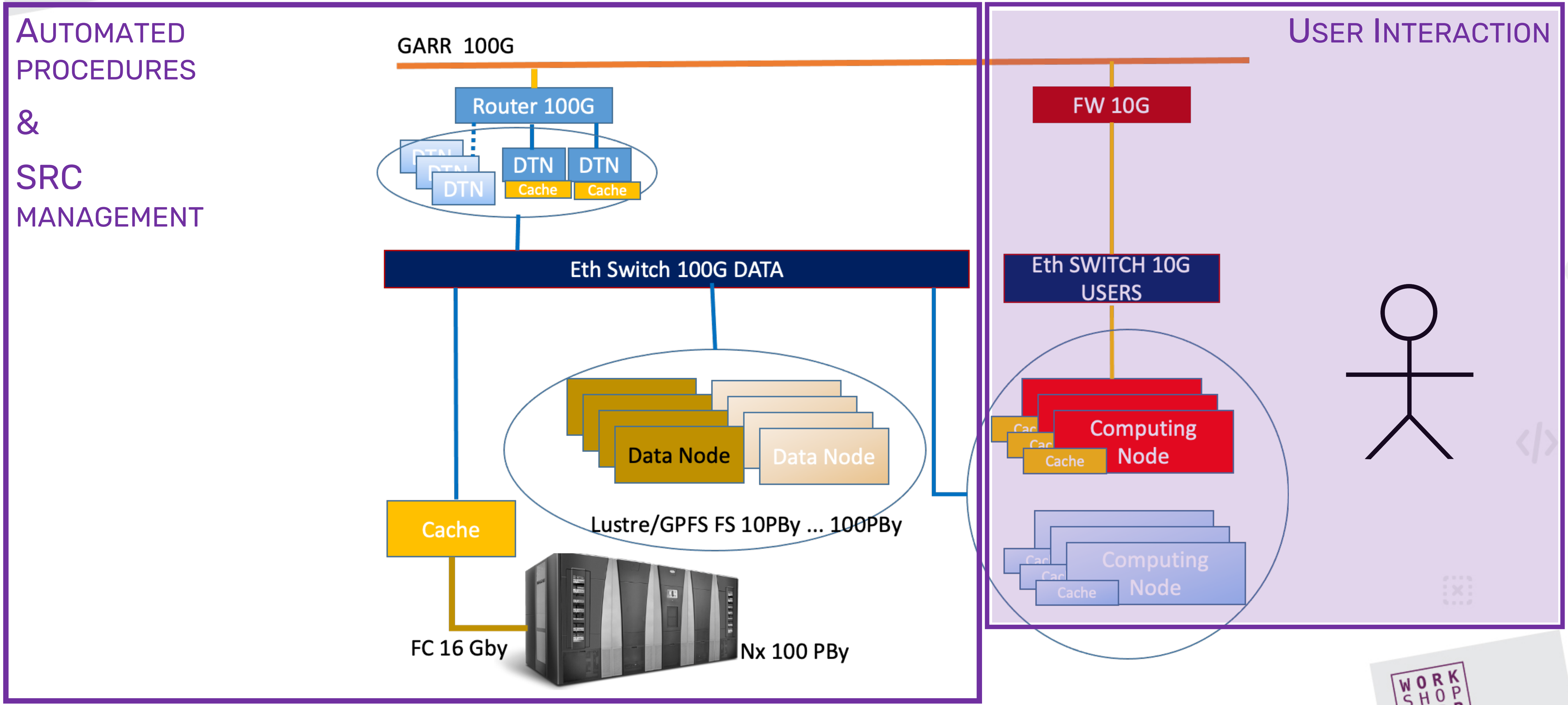
Data products will be transferred from Telescopes to DTNs

Currently investigating different software and filesystem options to manage data

ceph RUCIO

Courtesy: Matteo Stagni (INAF)

Local SRC architecture under discussion



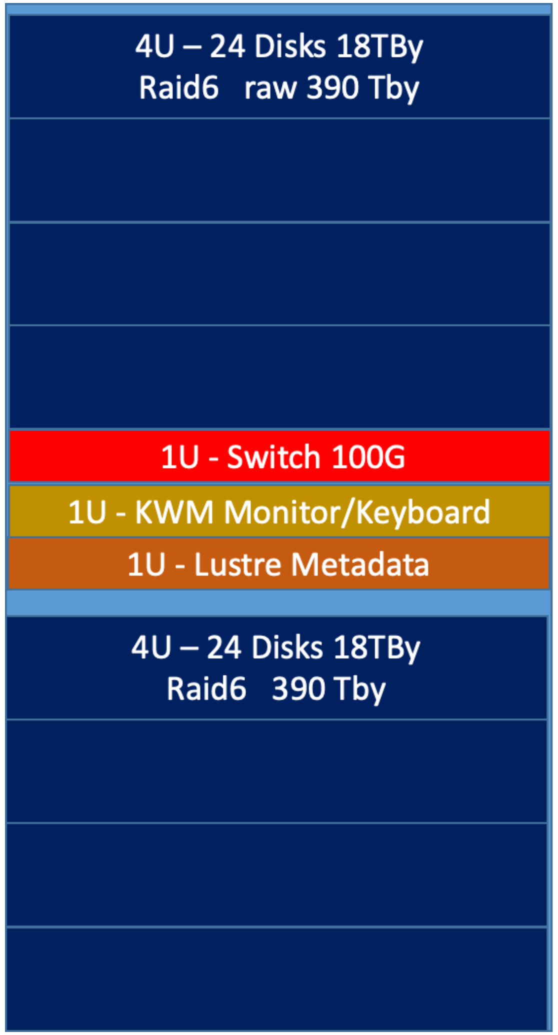
Courtesy: Matteo Stagni (INAF)

A possible rack for a data node of a protoSRC

Server rack sample - 42 U

- Available space 3.17 PB
- Estimated cost 175~200 K EUR
- Estimated energy consumption

  3~4 KW/H



Courtesy: Matteo Stagni (INAF)

Current thoughts for a Common Platform functionalities

The functionalities would include:

- ***Integration of (heterogeneous) storage resources*** provided by different facilities in order to offer a common space for storing data (similar to the Data-lake concept of ESCAPE project), which would form the underlying infrastructure of a distributed SRC.

- ***Integration of (heterogeneous) computing resources*** provided by different facilities in order make efficient use of the resources

- ***Uniform access to the SRC services*** (data, computational resources and tools), irrespective of their geographical location

- ***AAI services for the users***, so that they do not require different credentials to access different resources/SRCs

Current thoughts for a Common Platform basics services

Basics services necessary for working in a distributed way would be implemented.

- An environment allowing the implementation of ***workflows / notebooks*** in a collaborative way
- An execution ***engine able to capture provenance***, and store it following standards, each time a workflow is executed to generate ADPs (advanced data products) from the ODPs, or other variations (e.g. comparing numerical simulations with ADPs for a specific science case)
- A ***data archive which includes provenance*** information gathered from the instrument and the subsequent workflow (item above).
- A service to ***access the data following VO standards*** allowing interoperability and multi-wavelength/multi-messenger science
- A ***catalogue of workflows/pipelines/code*** that can be customised to facilitate reuse

Italian expected aims in the SRC context

✓ 1. SCIENCE WISE: The identification of a kernel of "modi operandi" in the interactions among the various actors to secure an efficient and always developable science-needs driven system

✓2. HUMAN CAPITAL WISE: The possibility for the regional communities to obtain access to the system and keep a role of management/development of that <u>at the minimum</u> in proportion to the local investments

✓3. INFRASTRUCTURE WISE: The establishment of a SRC network with a significant pole located in Italy

Perspective plan

Start with a Tier-3/Tier 2 protoSRC by 2023

… and progressively attain, ***by 2029***, a ***Tier-1*** size infrastructure with capability of ***≈ 3+ Pflops*** and ***≈ 70+ PBy/yr*** of storage connected at ***100G*** with the other poles

A most likely location for the Tier-1 will be the Technopolo in Bologna, where there will be also the Leonardo 270 Pflops system, the ECMWF, the INFN and the CINECA

Technopolo - Bologna

WORKSHOP GARR 2021 NET MAKERS

Thanks for looking into ...

...the radio astronomy future