GÉANT

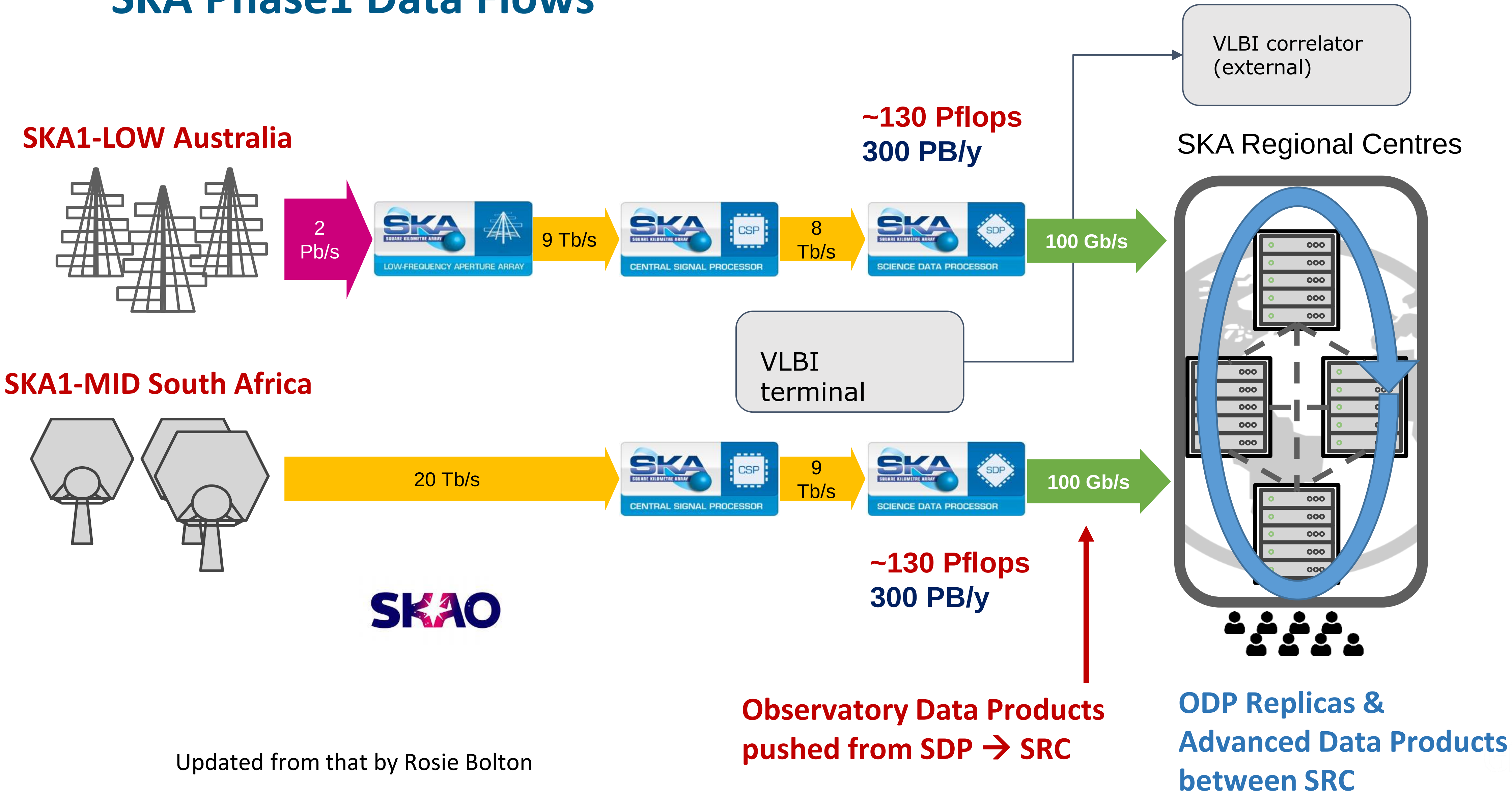
Report from the Technical Working Group on Global Network Architecture for SKA Connectivity

Richard Hughes-Jones

SKA-NREN Forum Meeting 6

26 February 2024

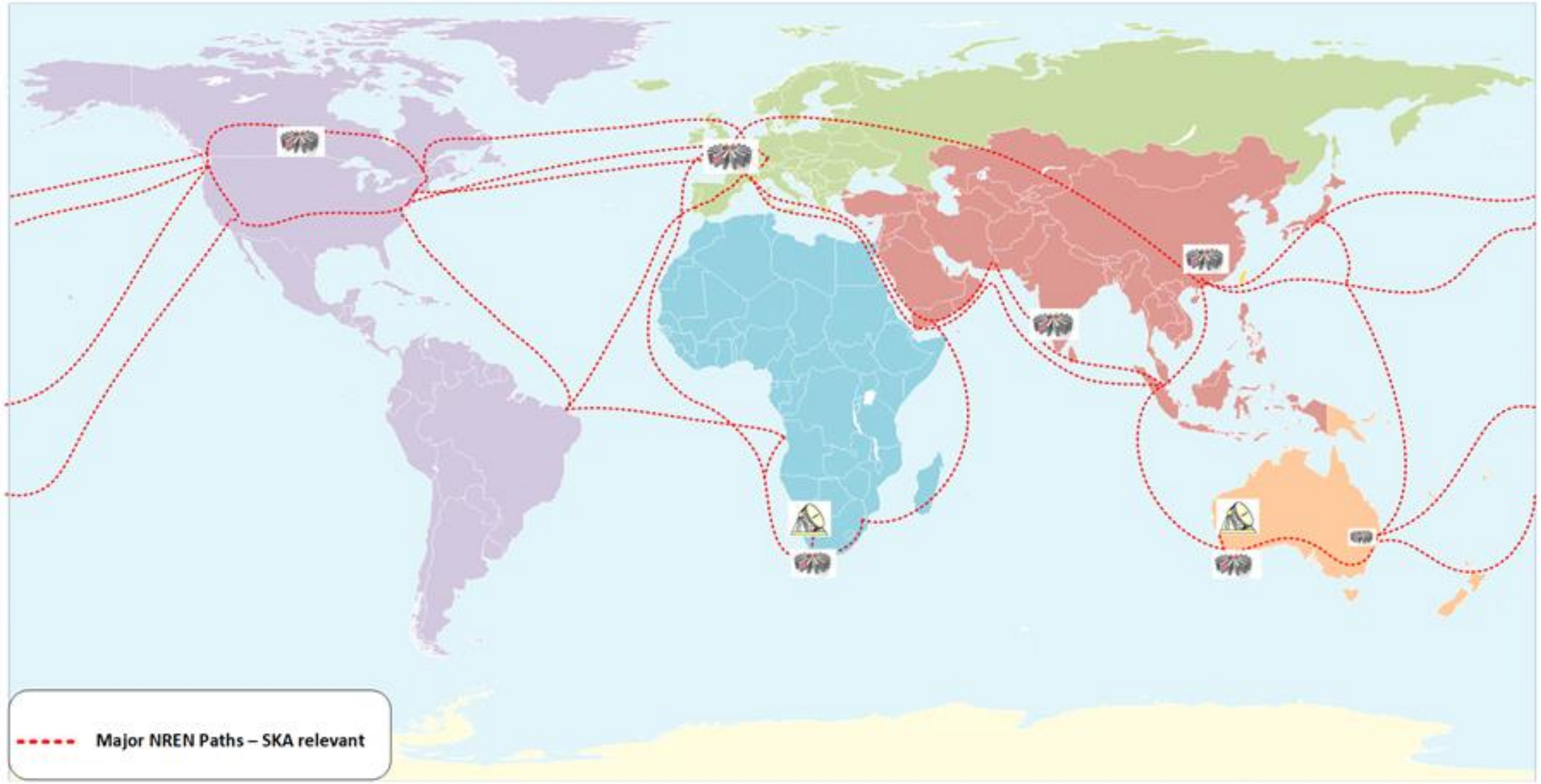
SKA Phase1 Data Flows

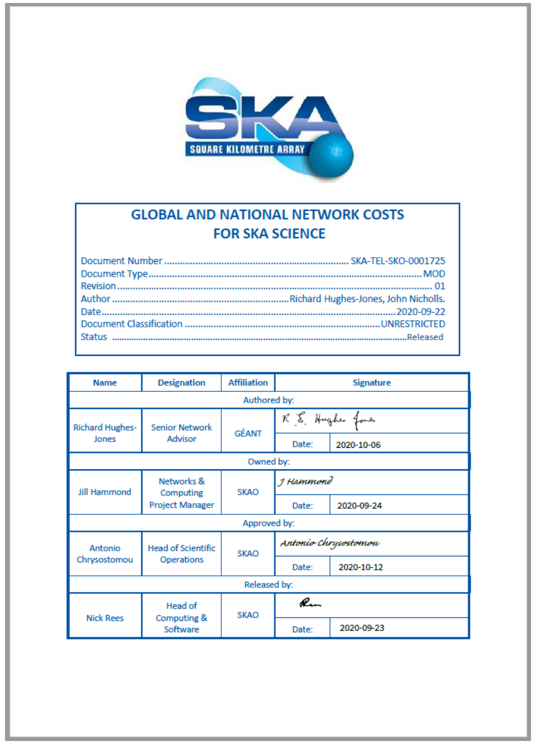


Updated from that by Rosie Bolton

Fibre and Cable Systems and major NREN paths

- The 2020 intercontinental fibre cable systems used by the international research and education community.
- Document produced for the SKA Regional Centres Coordination Group
  John Nicholls (AARNet) & Richard Hughes-Jones (GÉANT)





SQUARE KILOMETRE ARRAY

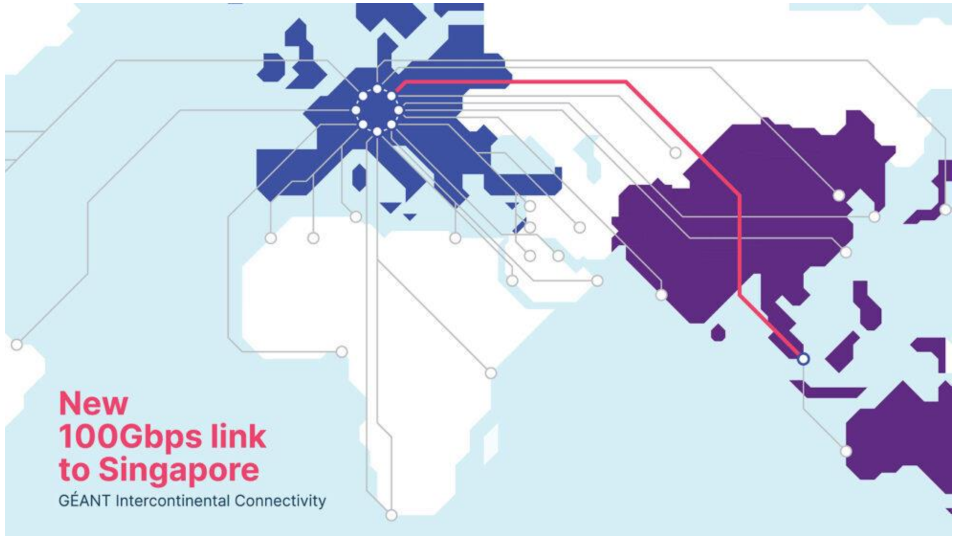
GLOBAL AND NATIONAL NETWORK COSTS
FOR SKA SCIENCE

Document Number SKA-TEL-SKO-0001725
Document Type .... MOD
Revision 01
Author .... Richard Hughes-Jones, John Nicholls.
Date 2020-09-22
Document Classification .... UNRESTRICTED
Status .... Released

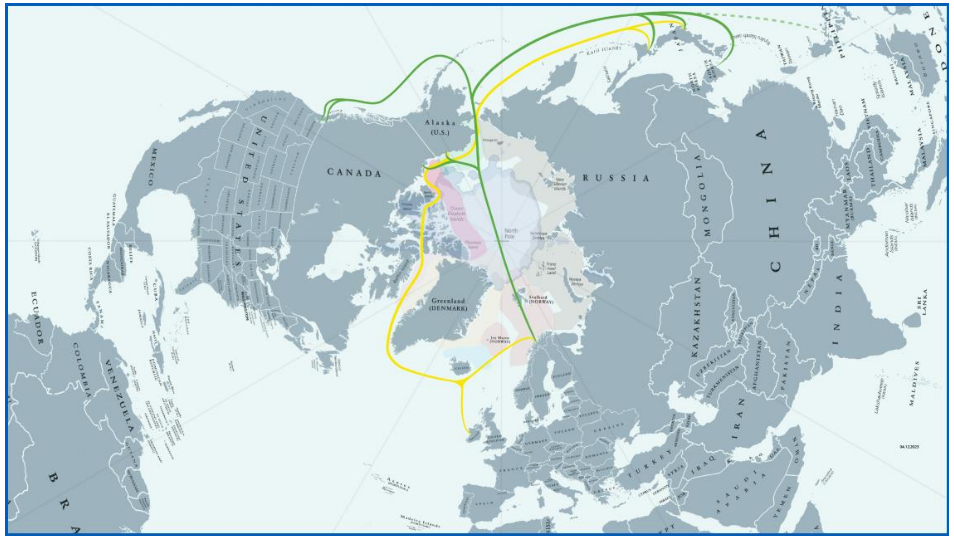
Name	Designation	Affiliation	Signature	
Authored by:				
Richard Hughes-Jones	Senior Network Advisor	GÉANT	R. E. Hughes-Jones	
			Date:	2020-10-06
Owned by:				
Jill Hammond	Networks & Computing Project Manager	SKAO	J Hammond	
			Date:	2020-09-24
Approved by:				
Antonio Chrysostomou	Head of Scientific Operations	SKAO	Antonio Chrysostomou	
			Date:	2020-10-12
Released by:				
Nick Rees	Head of Computing & Software	SKAO	[signature]	
			Date:	2020-09-23

Network Evolution

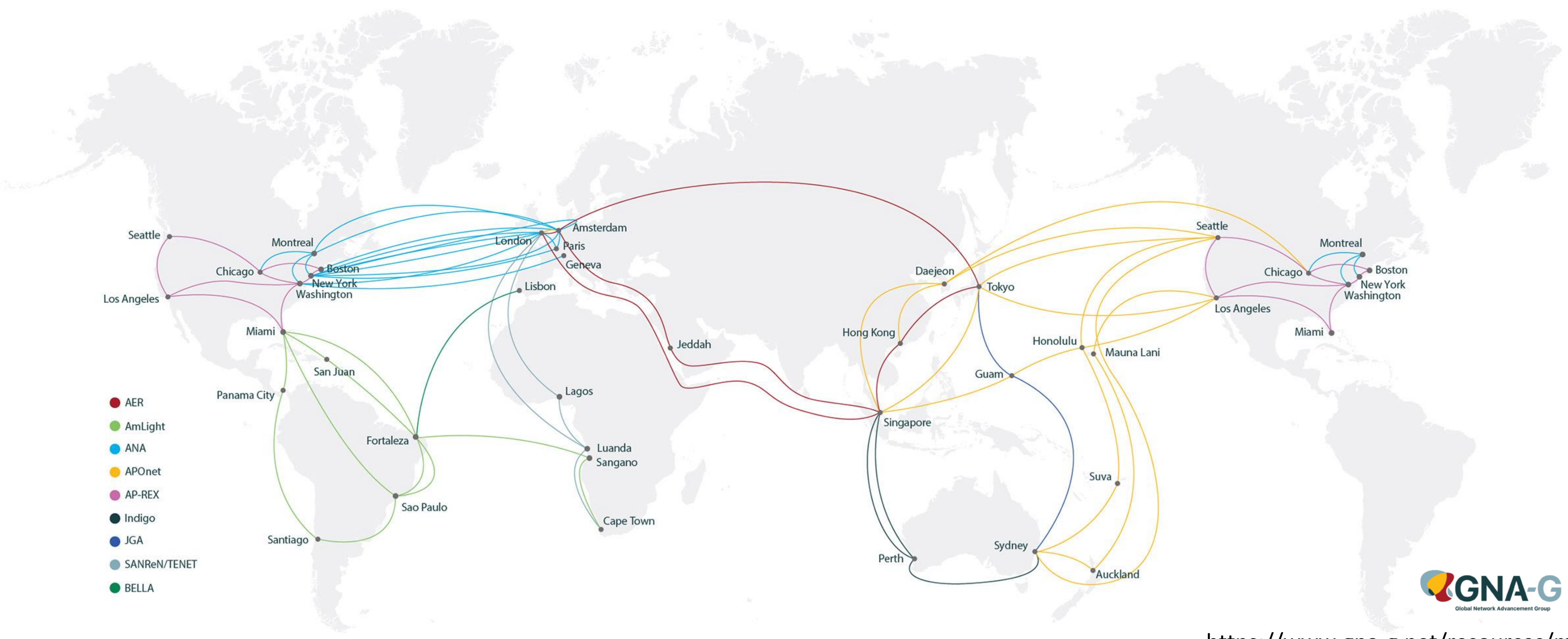
- EC funded GN5-IC1 project
  - New 100G link Marseille to Singapore along the Asia Africa Europe-1 (AAE-1) cable.
  - Under a 7-year Indefeasible Rights of Use (IRU).
  - Tender for trans-Atlantic spectrum in progress with US partners.



- Nordic NRENs & NORDUnet
  - A resilient submarine cable system between Europe and Asia traversing the Arctic Ocean by 2030
  - Far North Fiber is intended to link Europe and Japan
  - Polar Connect passing under the ice cap of the North Pole towards North America and East Asia
  - EC funded Vision 2030 report



Global Research and Education Network GREN

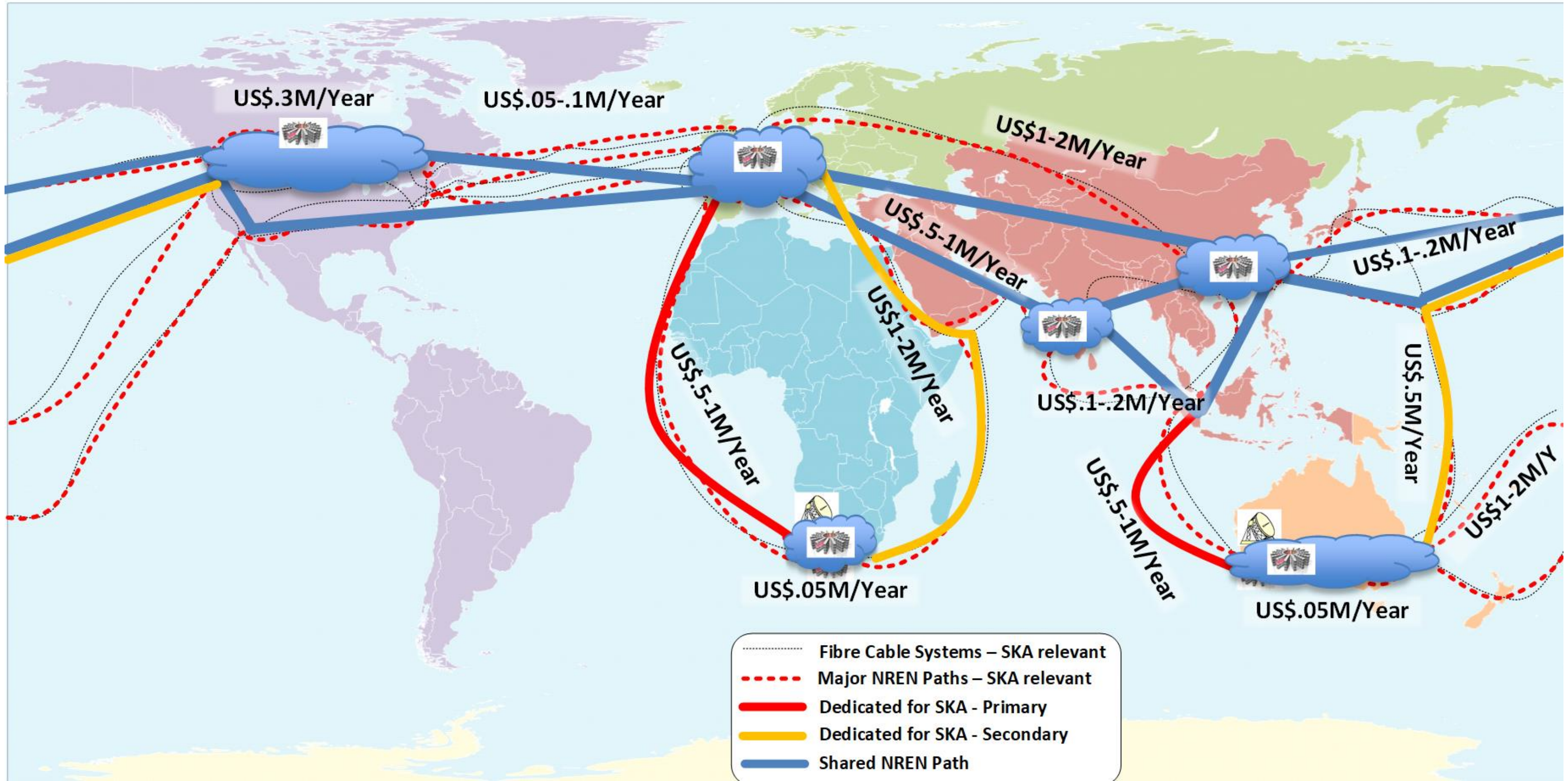


GNA-G
Global Network Advancement Group

https://www.gna-g.net/resources/maps/

Global Network & Paths of Interest to SKA

- Dedicated 100 Gigabit Primary paths (red lines) & Backup paths (yellow lines) from both telescopes
- Use of the academic network infrastructure shared between user communities (blue lines).
- 1 PetaByte/day pushed by SDP from each Telescope → 100 Gigabit/s for the Full Design
- Costs based on 10 to 15 year IRU per 100 Gbit circuit projected to 2025 prices



- Primary 100G bandwidth USD 1.7 – 2.3 M per year.
- Backup 100G bandwidth USD 2.3 – 3.3 M per year When required.
- SKAO agreed to funding the operational costs of these paths.
- Funding for the shared network infrastructure follows that for other science communities.

Main Data Flows

- From the Telescopes to the SRCs for the 1st replica of Observatory Data Products (ODP).
- Between SRCs to create the 2nd replica of the ODPs
- Between SRCs to create a 2nd replica of the Advanced Data Products (ADP).

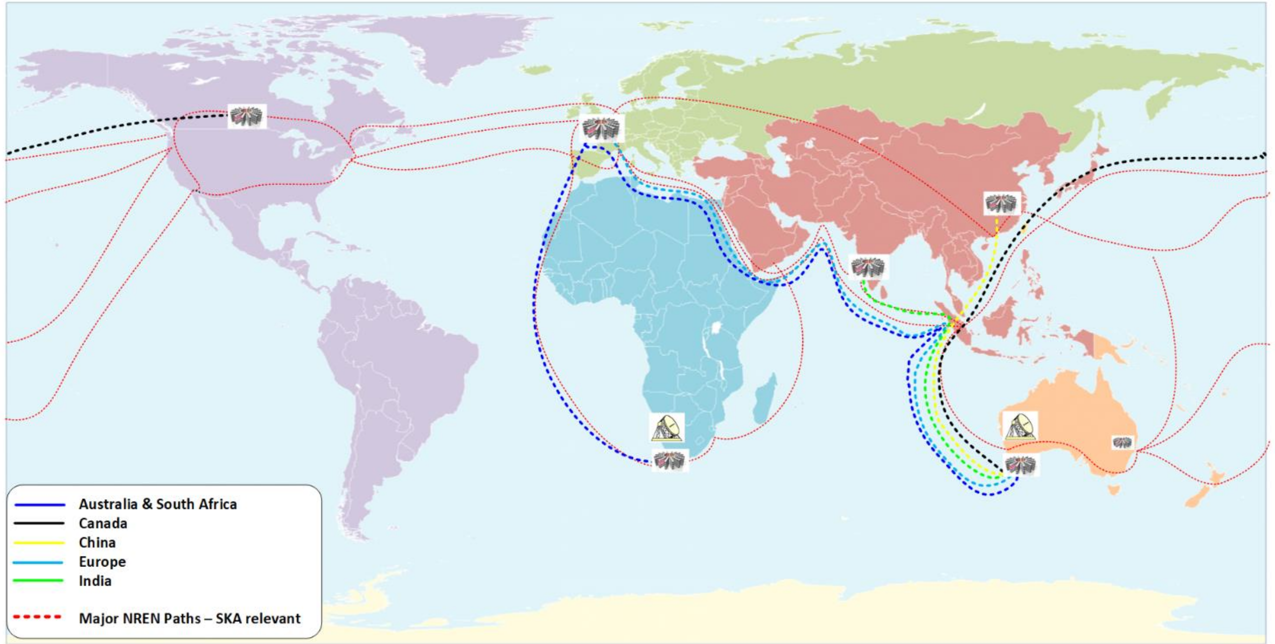
- In terms of storage there will also be an archive copy of the ODPs and ADPs stored at the SRCs.

Global Paths of the Data Flows Pushed to the SRC for the 1st Replica

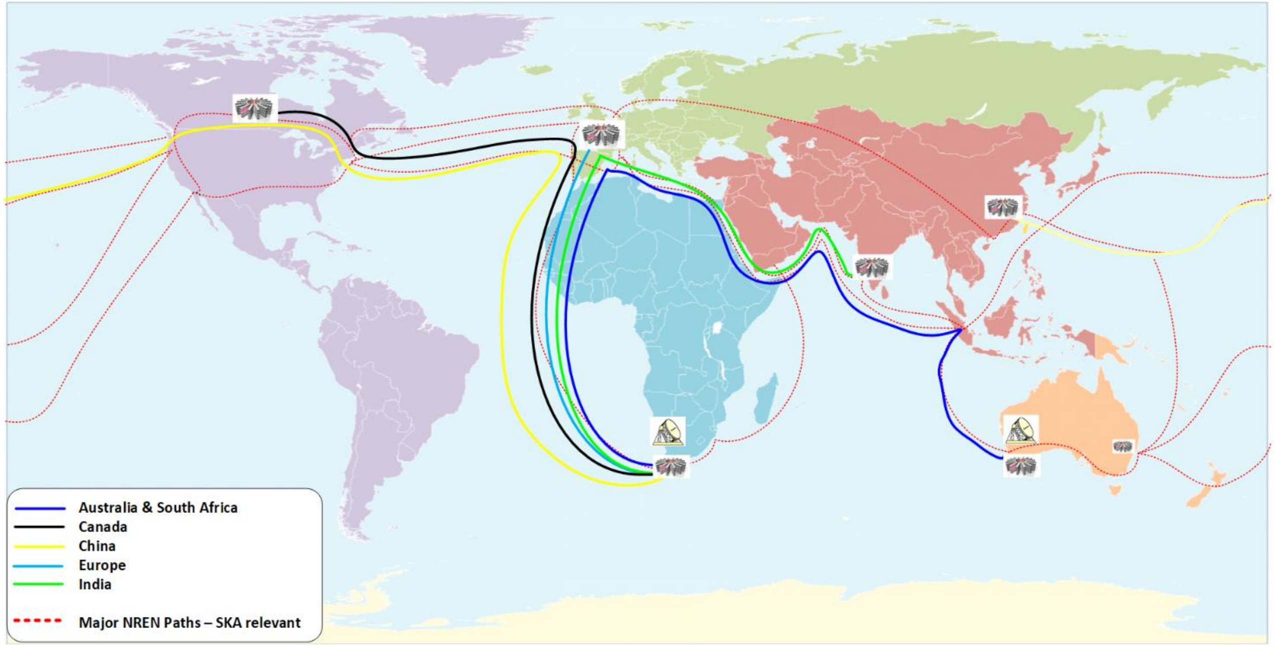
- Five flows on the submarine cable from Perth to Singapore .
- Then join the general purpose routed IP academic network.
- Single flows on the routes to Canada, China and India, Australia is local, and
  two 20 Gbit/s flows would be carried to London to reach SRCs in Europe and South Africa.

- Five flows on the submarine cable from Cape Town to London.
- Then join the general purpose routed IP academic network.
- Different submarine cables used to reach India and Australia, Europe is local, and
  two 20 Gbit/s flows cross the Atlantic to SRC in Canada and China.

SKA1-LOW Australia

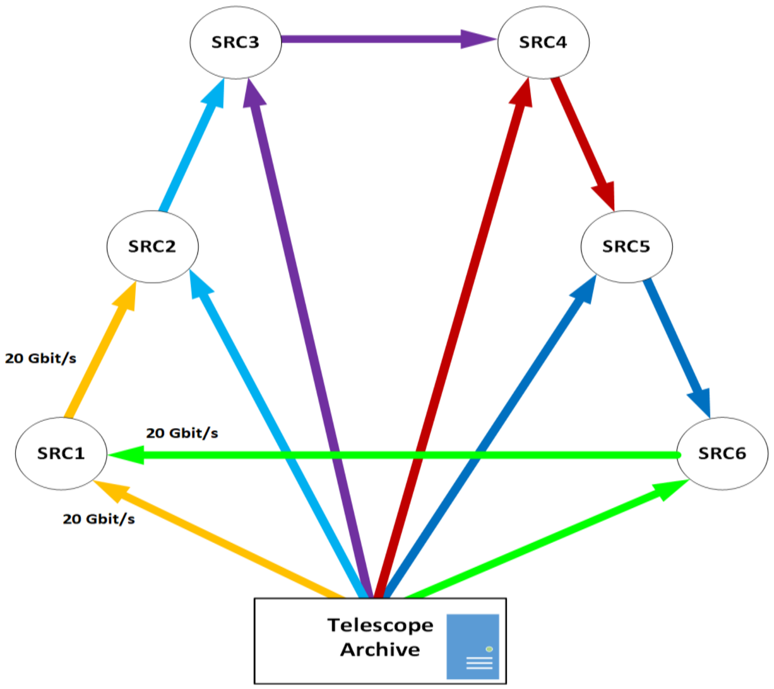


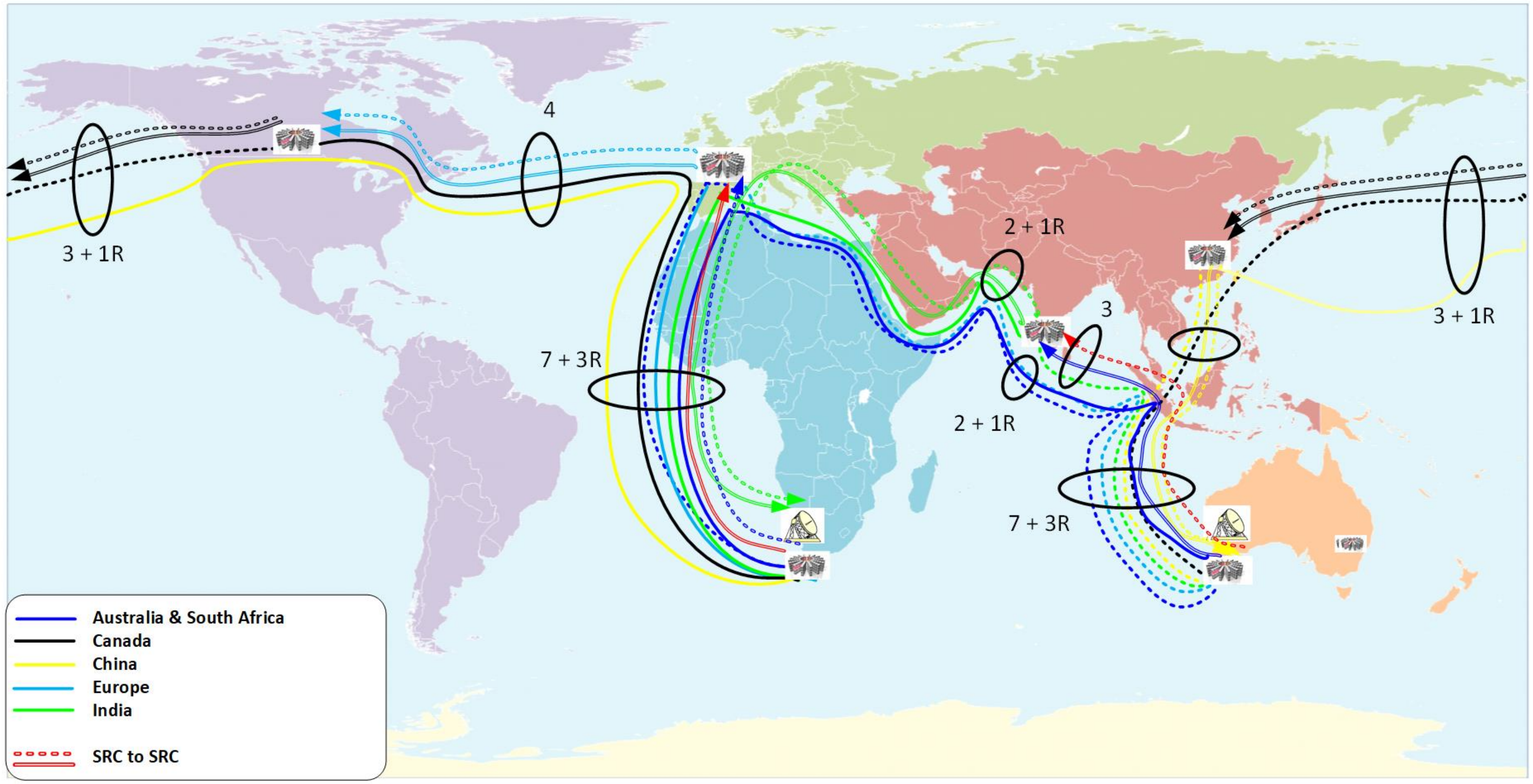
SKA1-MID South Africa



Global Data Flows if the SRC Re-distribute data 2nd Replica

- Each SRC accepts its fraction of the Observatory Data Products and re-distributes to another SRC.
- SRC has 20 Gbit/s flow from the telescope & a second continuous 20 Gbit/s flow from another SRC.
- Each SRC sends out a 20 Gbit/s flow.
- Makes substantial use of the shared academic network.





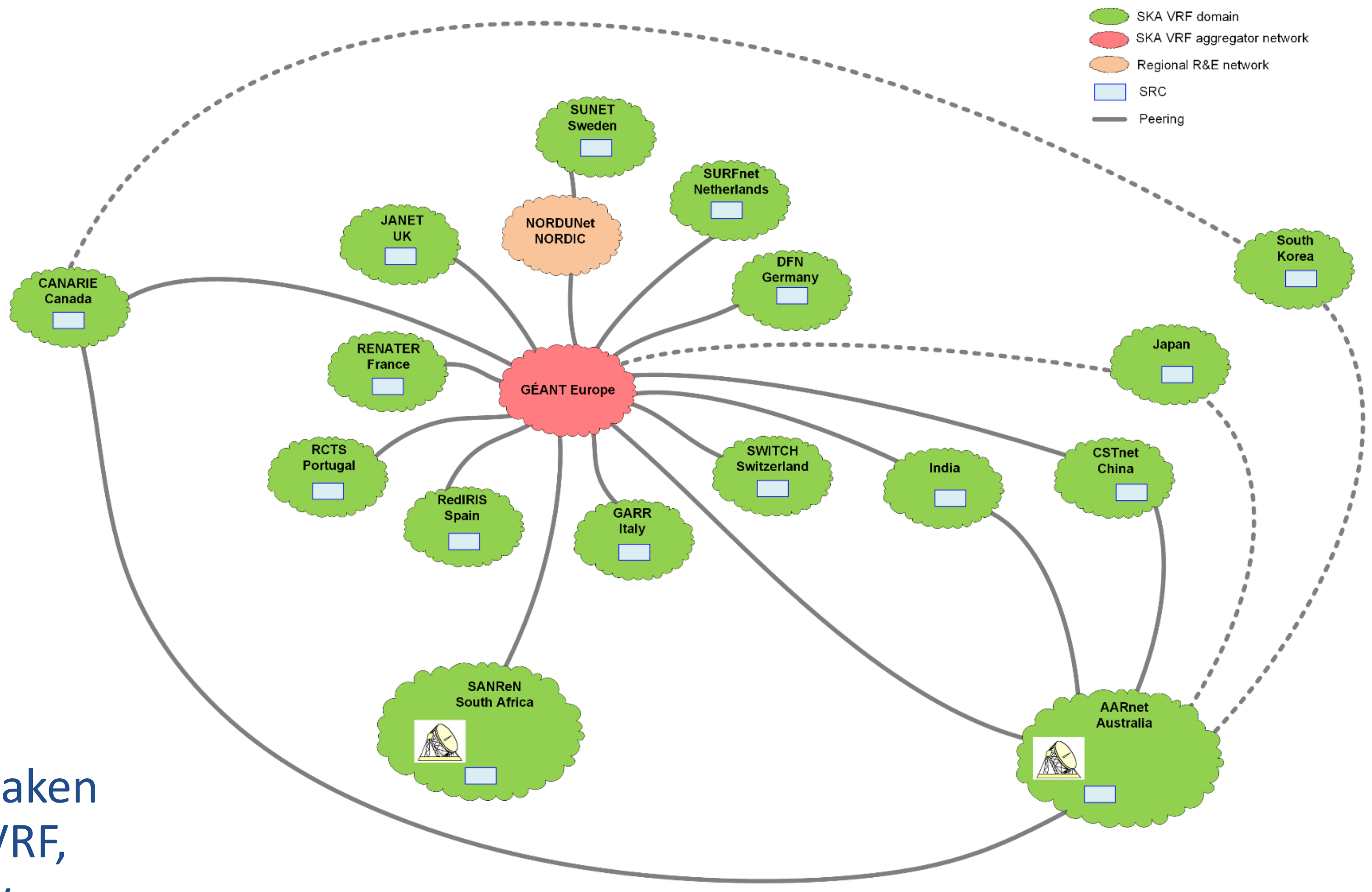
SKA-NREN Forum Technical Working Group
Global Network Connectivity & Architecture for SKA

- Build on the experience with LHCONE and the AENEAS project.
- All Telescope to SRC and SRC to SRC data transfer traffic to operate on VRF peered over the shared academic network.
- For high performance data transfers the data transfer node servers (DTN's) should i be located in a site "Science DMZ" with ACLs for site policy.
- The DTN nodes should be tuned for high RTT latency data transfers
  - kernel parameters defining maximum TCP buffer size
  - Queueing discipline
  - NIC ring buffer size
- All SRCnet data transfer traffic, at least, to use IPv6 only.
- Use larger MTU sizes. Specifically: 9000 byte Jumbo frames.
- For network monitoring SRCnet deploy a mesh of "perfSONAR" nodes with at least one perfSonar system per SRC site. (https://www.perfsonar.net)

Global Network Connectivity for SKA

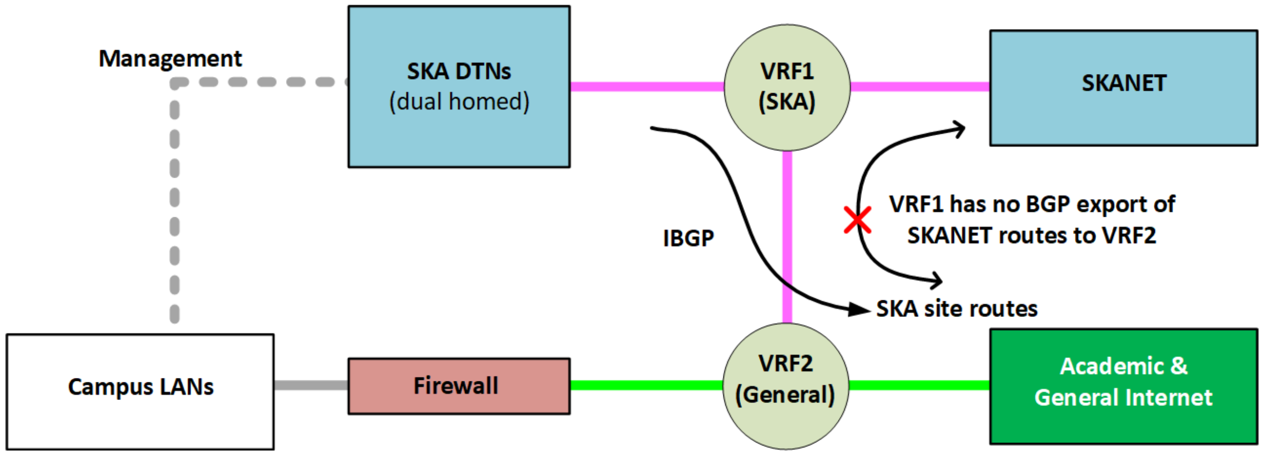
- Global VRF based overlay with peering linked over the shared academic network.

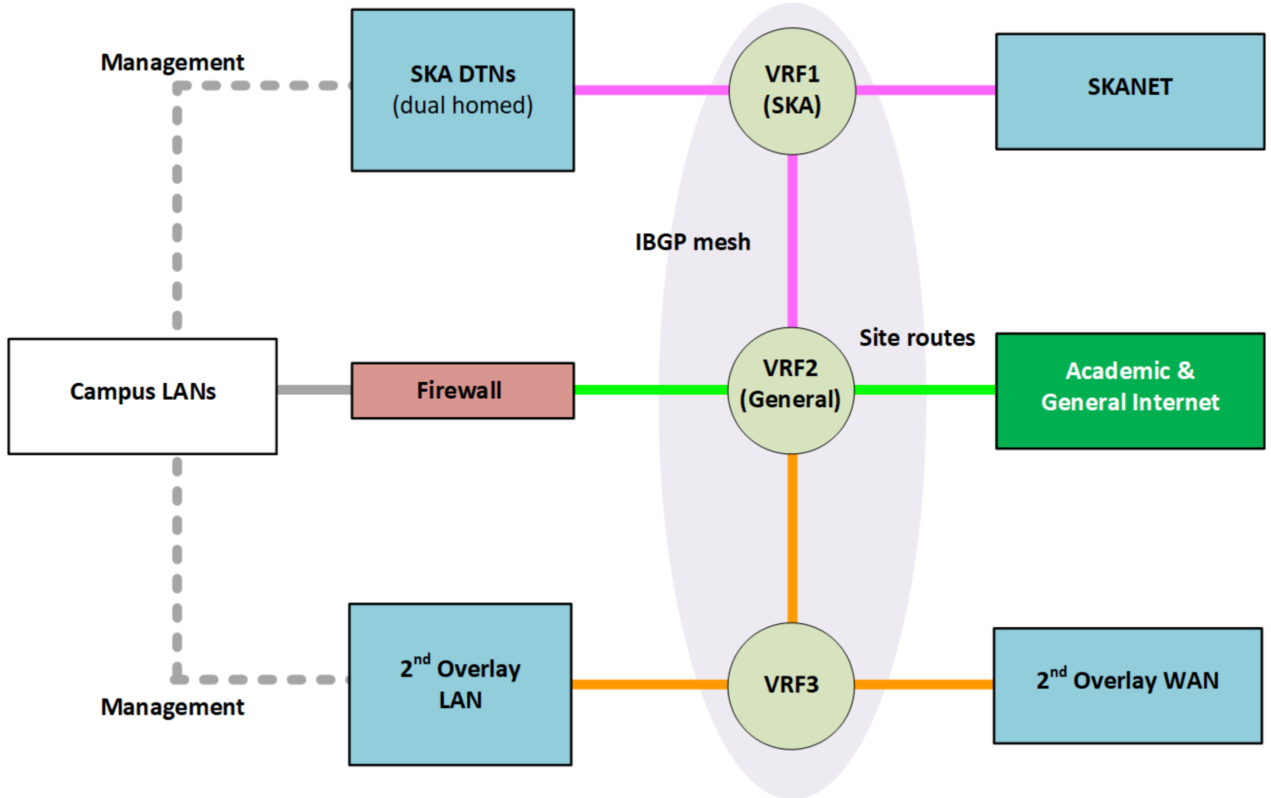
- Isolation of SKA traffic from other users
- Easier for NRENs to implement the routing, policies and monitoring
- SKA traffic can be engineered
  - Use specific paths & routes
- Layer 3 routing provides isolation
  - any network configuration issues
  - strictly limits broadcast storms
- Layer 3 will re-route traffic as long as there is an alternative network path
- Configuration actions have to be undertaken by the NREN and a Site to join the SKA VRF, which provides an extra layer of security.



Network Considerations for a Site DMZ and Tuned DTNs

- Need for high performance Data Transport Node hardware
  - Tuned for RTT ~300 ms
  - Network – disk transfer rate ~20 Gbit/s (Typical disk-to-disk transfer rate ~6 Gbit/s)
- Flexible but secure ACLs give a high performance DMZ connected to the VRF
- Separates the SKA data & Campus LAN traffic and implements site policies
- Uncongested site access link





Use of Jumbo Frames

- Makes a big improvement to data transfers:
  - Throughput ≥ * 2 (~3.5 to Aus)
  - Reduces recovery time
  - Transfers more stable and less re-transmitted packets

- Concern about mixing Jumbo & 1500 Byte PCs
- Tests show 9000 to 1500 Byte MTU transfers do work for TCP.

- Care needed when configuring sites & PCs:
  - Path MTU discovery (PMTUD ICMP) needs to work
  - net.ipv4.tcp_mtu_probing=1 for IPv4

Network test data

•Iperf (Raul from Jisc)

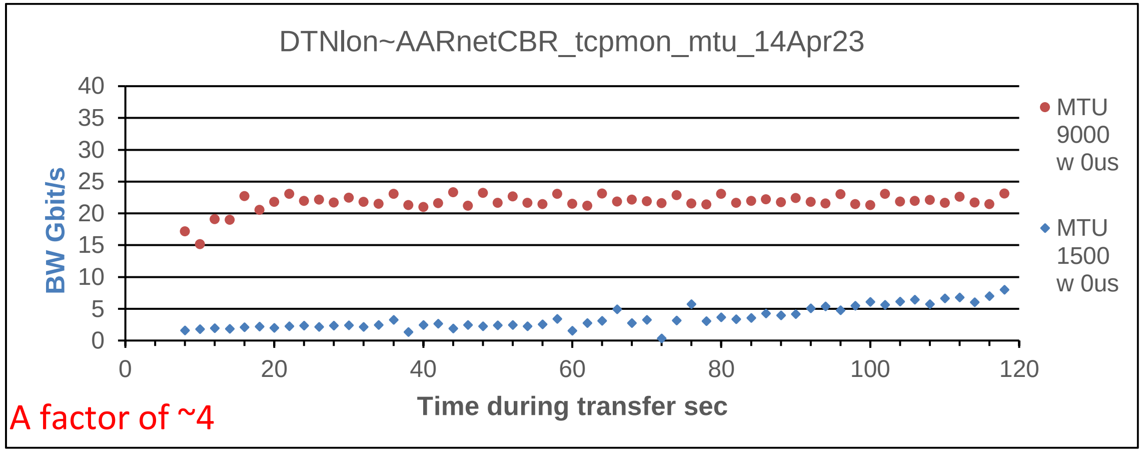
Source	Destination	RTT	9000	1500
SURF (NL)	RNP (Brazil)	100ms	31 Gbit/s	20 Gbit/s
Jisc (London)	BNL (USA)	100ms	14 Gbit/s	6 Gbit/s
SURF (NL)	Jisc (London)	7.2 ms	23 Gbit/s	6 Gbit/s

•Tcpmon (Richard Hughes-Jones – Geant)

Source	Destination	RTT	9000	1500
London	Cambridge	3ms	37 Gbit/s	15.8 Gbit/s
London	AARnet	262ms	21 Gbit/s	3.4 Gbit/s

4 Chris Walker LHCOPN-LHCONE April 2023 Jisc

Throughput Europe to Australia RTT 262 ms



Data Distribution Across Countries Based on a Fair-share Model

- Allocate data to the SKA countries in proportion to their funding for SKA construction.

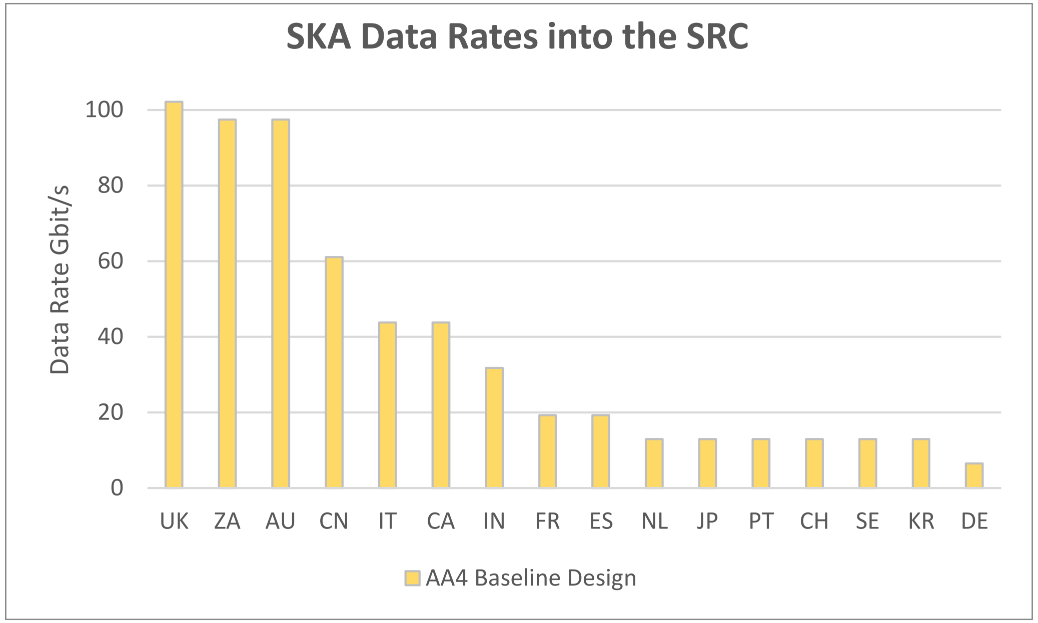
- Data flows included:
  - Movement of ODPs from telescope to SRC for the 1st replica
  - Movement of ODPs between SRCs for the 2nd replica
  - Movement of ADPs between SRCs to create the 2nd replica

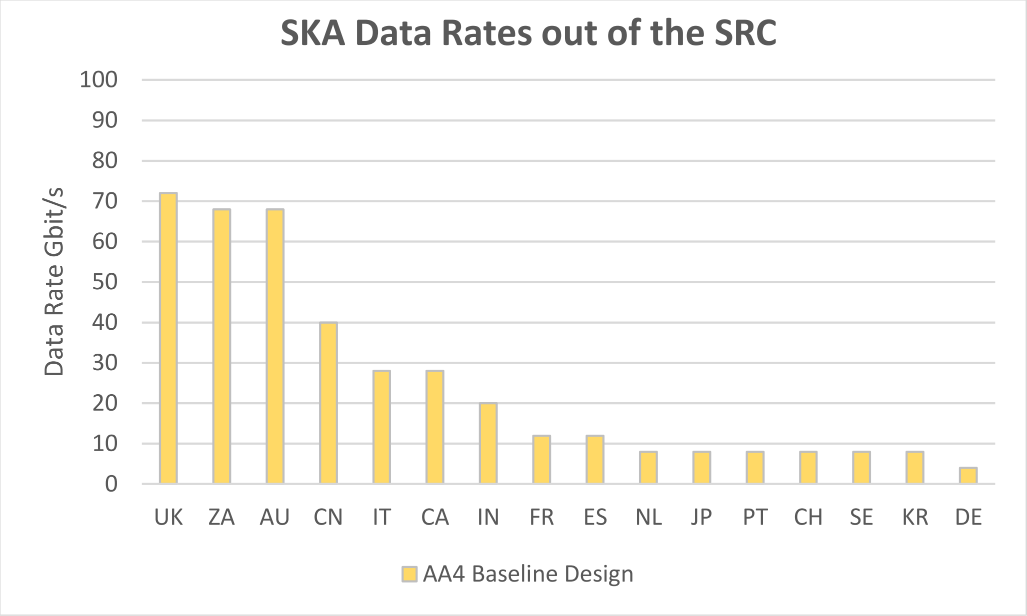
- Does not consider :
  - If compute architecture at a site matches the data product requirement.
  - A data cube will need to be at one location.
  - Amount of storage available at a site.

SRC Data Rates For AA4 Baseline Design

- AA4 (Low 100 Gbit/s Mid 100 Gbit/s)
  - Almost 100 Gbit/s into UK ZA AU
  - Rates into many countries 10 – 20 Gbit/s.

- Model gives an indication.
  - Likely that SRCs will need at least 20 Gbit/s for SKA data.
  - Tuned Long-haul disk-to-disk transfers ~5-6 Gbit/s
  - Plan for multiple concurrent file transfers.

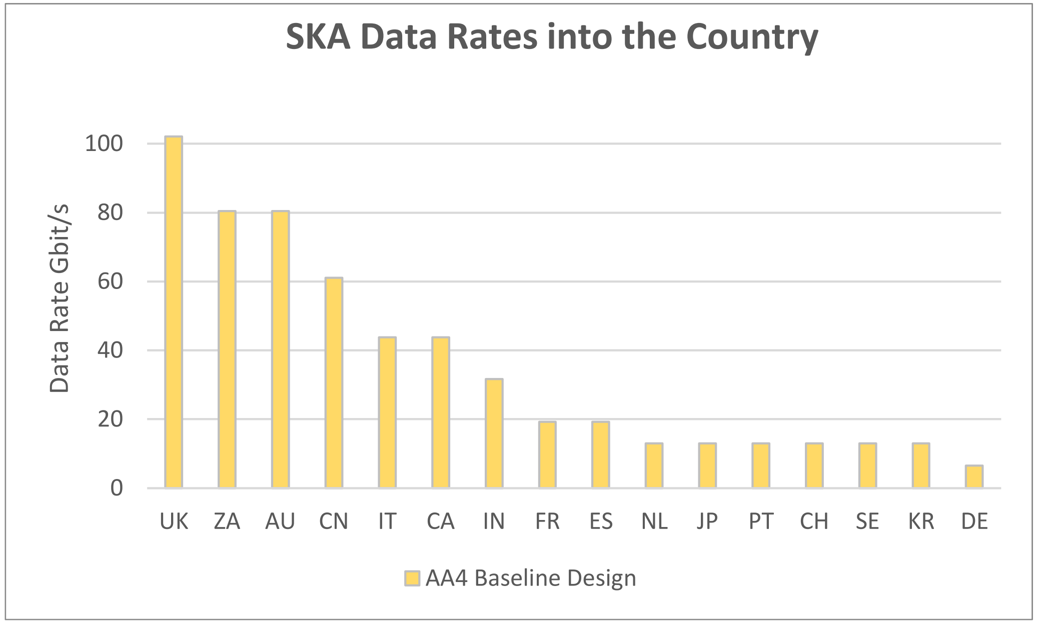


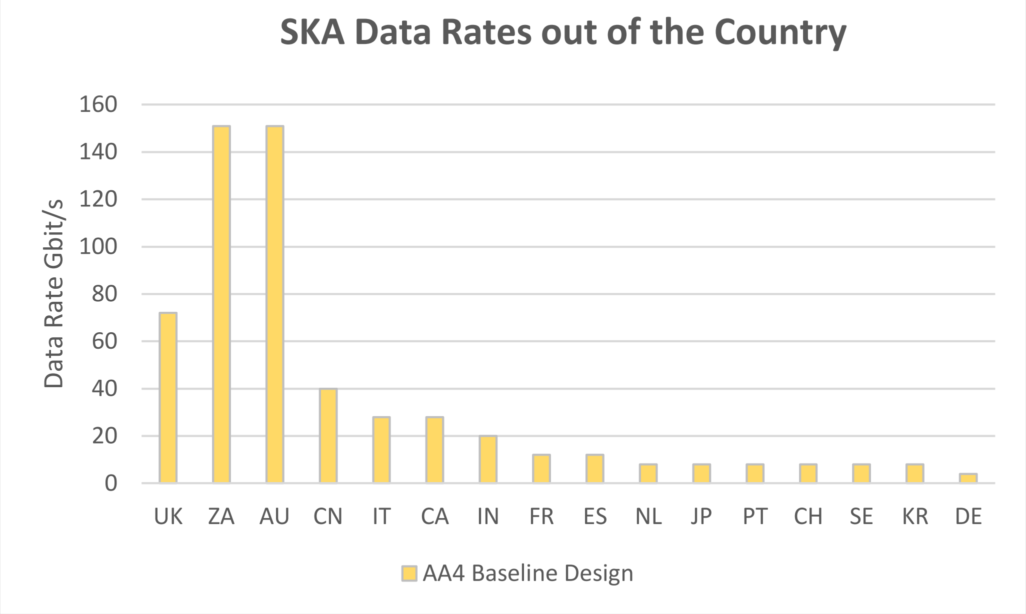


Country Data Rates For AA4 Baseline Design

- AA4 (Low 100 Gbit/s Mid 100 Gbit/s)
  - Almost 100 Gbit/s into UK
  - Telescopes to SRC in a host countries on local NREN.
  - The ~150 Gbit/s out of the host countries includes ODPs from the telescopes.
    Rates into many countries 10 – 20 Gbit/s.

- Model gives an indication.
  - A data cube will need to be at one location.
  - ODPs to Europe ~40% (19% to the UK)
  - Expect significant traffic between AU & ZA





Data Rates for Array Configurations for AA2, AA*, AA4 (Design baseline)

- Based on data from
  “SKAO staged delivery, array assemblies and layouts” SKAO-TEL-0002299 dated Nov 2023

- Considered a plausible scenario using the following types of observations:

Low

- Image cube. Size dependent on the max. baseline length and the sensitivity on number of stations
  - AA* baselines similar, sensitivity denominates
- Calibrated smoothed visibility data (EoR). Size is dependent on the number of baselines

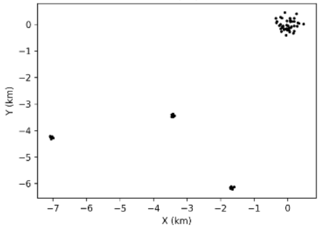
Mid

- Image cube. Size dependent on the max. baseline length and the sensitivity on number dishes

A Plausible Scenario for Data Rates from the Telescopes

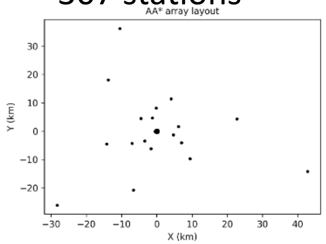
Low

AA2
64 stations



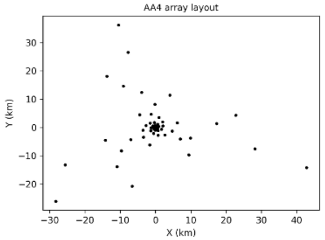
8 km baseline
Baseline ratio 0.11
Station ratio 0.125

AA*
307 stations



70 km baseline
Baseline ratio 1
Station ratio 0.6

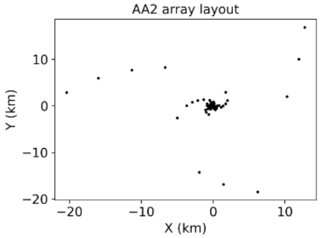
AA4 Design baseline
512 stations



70 km baseline
Baseline ratio 1
Station ratio 1

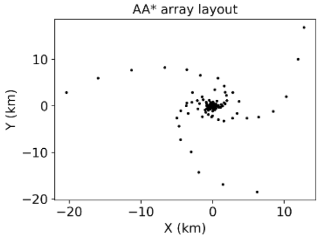
Mid

AA2
64 dishes



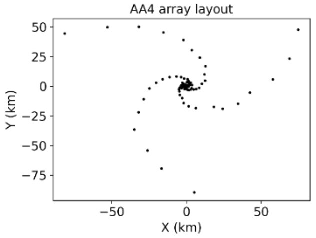
35 km baseline
Baseline ratio 0.23
Dish ratio 0.32

AA*
144 dishes



35 km baseline
Baseline ratio 0.23
Dish ratio 0.73

AA4 Design baseline
197 dishes



150 km baseline
Baseline ratio 1
Dish ratio 1

	AA2	AA*	AA4 (Design baseline)
Timescale Low	Oct 2026	Jan 2028	After 2030?
Low data rate Gbit/s	2-5	50	100
Timescale Mid	Mar 2027	Dec 2027	After 2030?
Mid data rate Gbit/	10	10	100

These numbers are indicative but not endorsed by SKAO.
A document with more robust estimation of rates
Expected Summer 2024

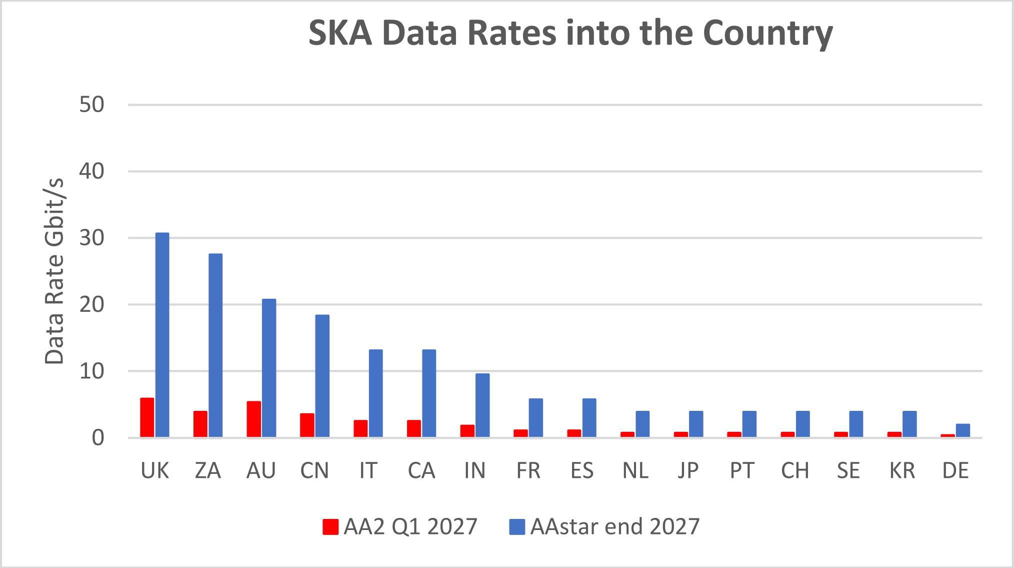
Country Data Rates

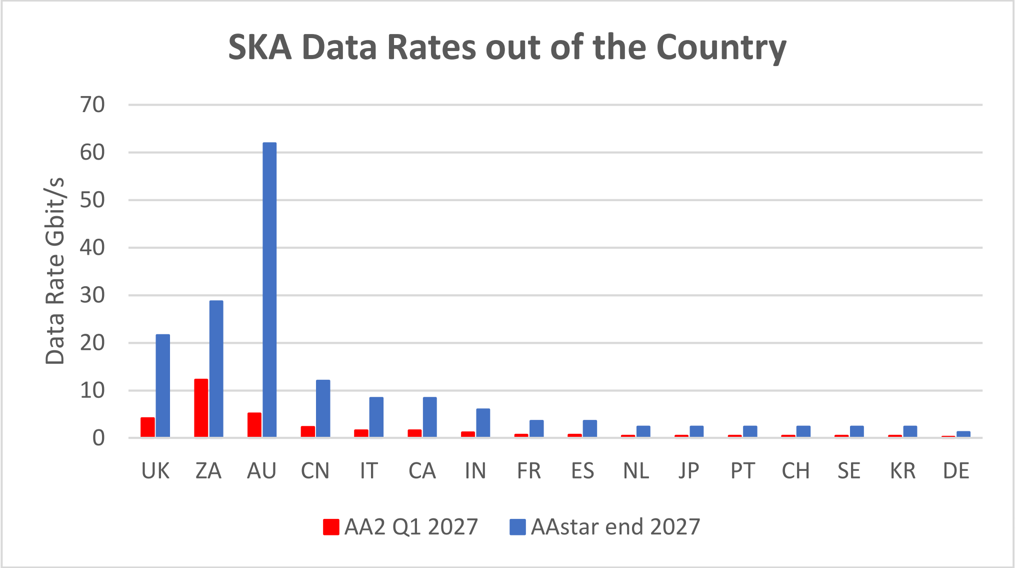
- Used the estimated bandwidth requirements for the different AAs.

- AA2 (Low 1.5 Gbit/s Mid 10 Gbit/s)
  - All countries have low rates.
  - Best to plan for disk-to-disk site transfers with each flow 5 – 10 Gbit/s.

- AA* (Low 50 Gbit/s Mid 10 Gbit/s)
  - 30 Gbit/s into the SRCs of UK ZA AU.
  - Rates into many countries 5-20 Gbit/s.
  - Data rate out of AU ~60 Gbit/s (inter-continental).

- Model gives an indication.
  - Tuned Long-haul disk-to-disk transfers ~5-6 Gbit/s
  - Plan for multiple concurrent file transfers.
  - Likely that SRCs will need ~ 10-20 Gbit/s for SKA data.
  - ODPs to Europe ~40% (19% to the UK)
  - Expect significant traffic between AU & ZA





Summary from SKA-NREN Forum Technical Working Group
Global Network Connectivity & Architecture for SKA

- All Telescope to SRC and SRC to SRC data transfer traffic to operate on peered VRFs.
- NRENs, and the sites, design the network implementations.
- For high performance data transfers the data transfer node servers (DTN's) should be located in a site "Science DMZ".
- The DTN nodes should be tuned for high RTT latency data transfers.
- All SRCnet data transfer traffic, at least, to use IPv6 only.
- Use larger MTU sizes. Specifically: 9000 byte Jumbo frames.
- Monitor the network with a mesh of "perfSONAR" nodes for SRCnet.

Implement for SRCnet 0.1. Avoid disruption. Provide the infrastructure for smooth growth.

Questions ?

Advanced European Network of
for Astronomy with the SKA