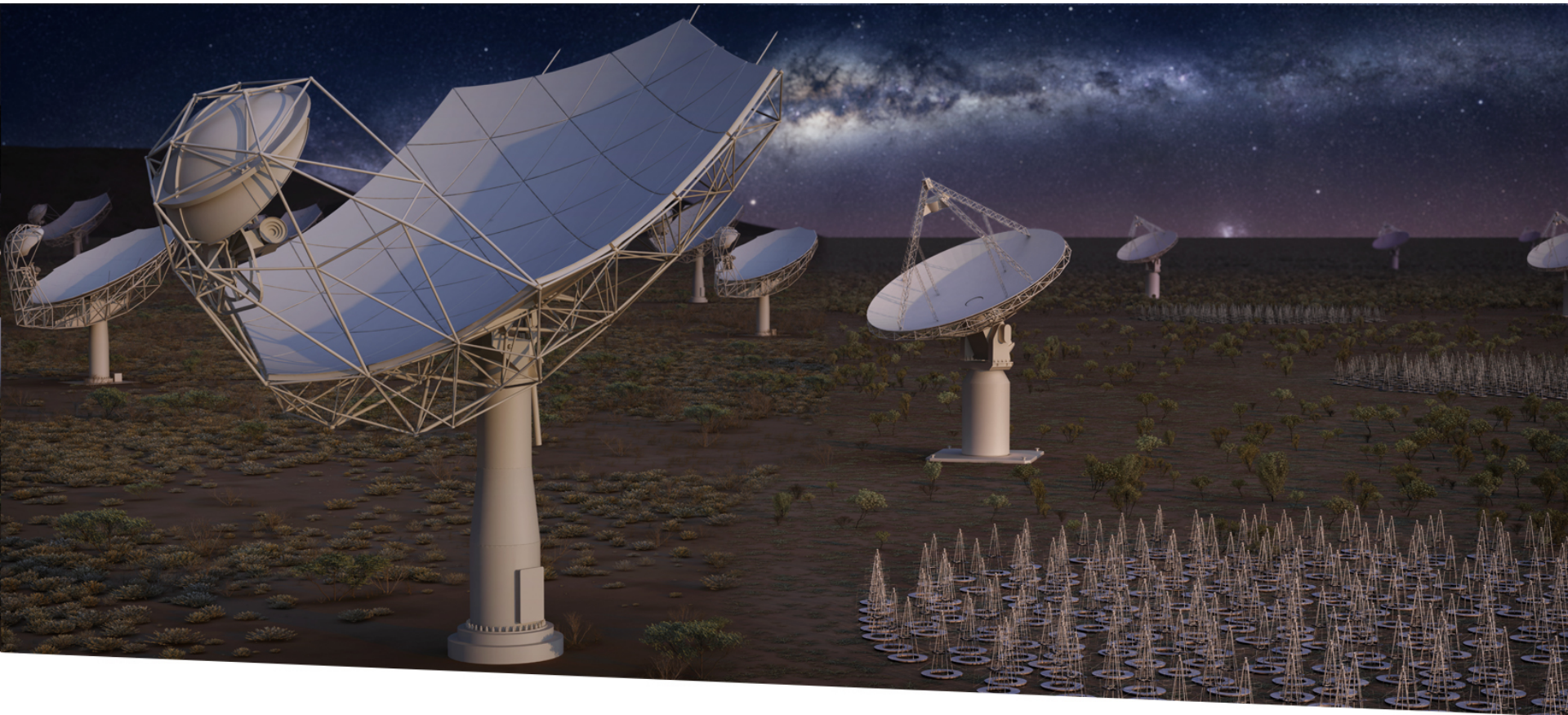# The CERN-SKA Collaboration



**SQUARE KILOMETRE ARRAY**

Exploring the Universe with the world's largest radio telescope

**Miles Deegan, SKA Organisation**

# SKA Project Overview

# Square Kilometre Array
## 3 sites; 2 telescopes + HQ
## 1 Observatory

Design Phase: late 2013 – 2018, €200M; 600 scientists+engineers

Phase 1
Construction: 2019 – 2024
Construction cost cap: €674.1M (inflation-adjusted)
Operations cost: under development (power is a big part)

MeerKat integrated
Observatory Development Programme (€20M/year planned)
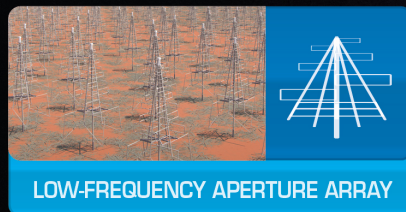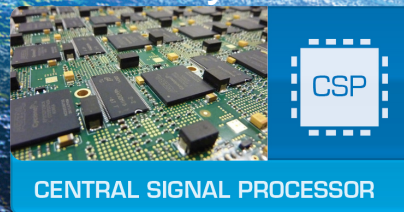SKA Regional Centres out of scope of centrally-funded SKAO.

Phase 2: start mid-2020s
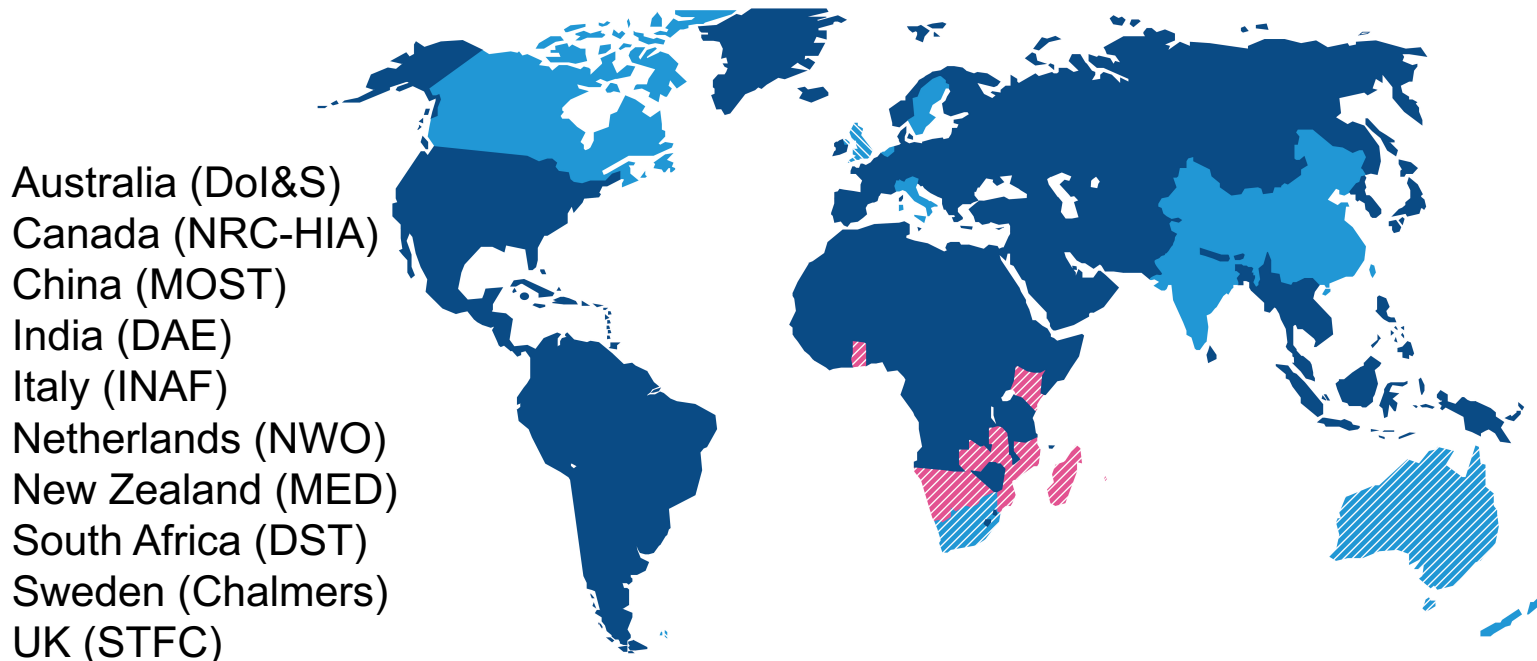~2000 dishes across 3500km of Southern Africa
Major expansion of SKA1-Low across Western Australia + further telescopes after that. Computing requirements are eye-watering

# International Design Teams

- Project Management, System Engineering and Architecture based at Jodrell Bank, UK - SKA Office is the Design Authority, will lead procurement and then the operational phase
- >600 scientists & engineers in institutes and industry in Member countries of the SKA



WIDE BAND SINGLE PIXEL FEEDS

TELESCOPE MANAGER

CENTRAL SIGNAL PROCESSOR

SIGNAL AND DATA TRANSPORT

SCIENCE DATA PROCESSOR

DISH

MID-FREQUENCY APERTURE ARRAY

LOW-FREQUENCY APERTURE ARRAY

ASSEMBLY, INTEGRATION & VERIFICATION

INFRASTRUCTURE AUSTRALIA

INFRASTRUCTURE SOUTH AFRICA

# SKA Organisation: 10 countries, more to join

Australia (DoI&S)
Canada (NRC-HIA)
China (MOST)
India (DAE)
Italy (INAF)
Netherlands (NWO)
New Zealand (MED)
South Africa (DST)
Sweden (Chalmers)
UK (STFC)

Interested Countries:
- France
- Germany
- Japan
- Korea
- Malta
- Portugal
- Spain
- Switzerland
- USA

Contacts:
- Mexico
- Brazil
- Ireland
- Russia



Full members

SKA Headquarters host country

SKA Phase 1 and Phase 2 host countries

African partner countries
(non-member SKA Phase 2 host countries)

This map is intended for reference only and is not meant to represent legal borders

SKA
SQUARE KILOMETRE ARRAY

# Future SKA governance structure

- SKA Org is currently a not-for-profit limited by guarantee UK company
- We aim to transition to being an Inter-Governmental Organisation
- IGO = 'Convention' agreed between governments
  - Government commitment: Long-term political stability, funding stability
  - A level of independence in structure
  - Availability of 'supporting processes' through Privileges and Immunities from members: functional support for project
  - 'Freedom to operate', specifically through procurement process, employment rules and so on



CERN/Science Source/Photo Researchers, Inc.

# Operations Scope
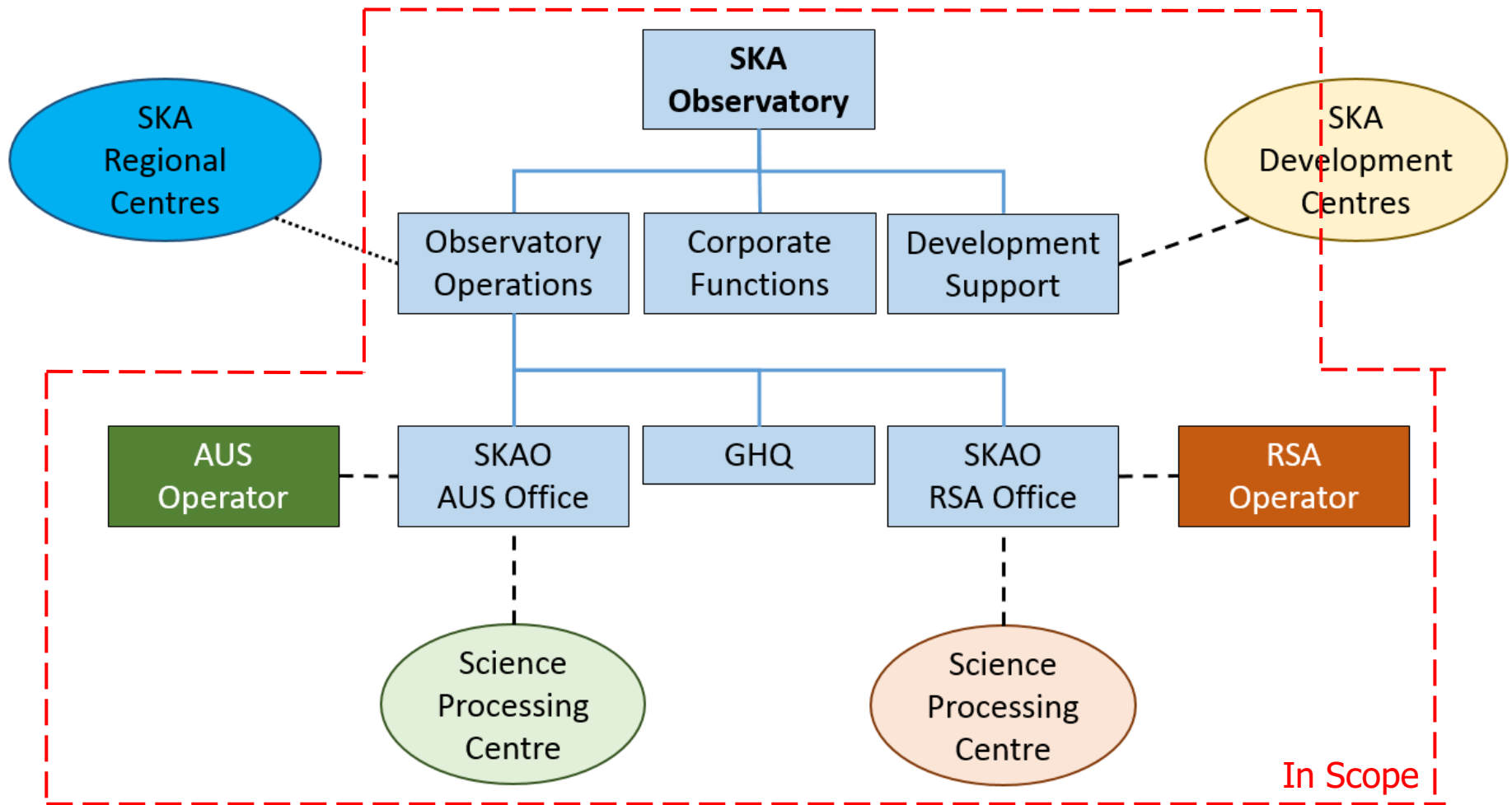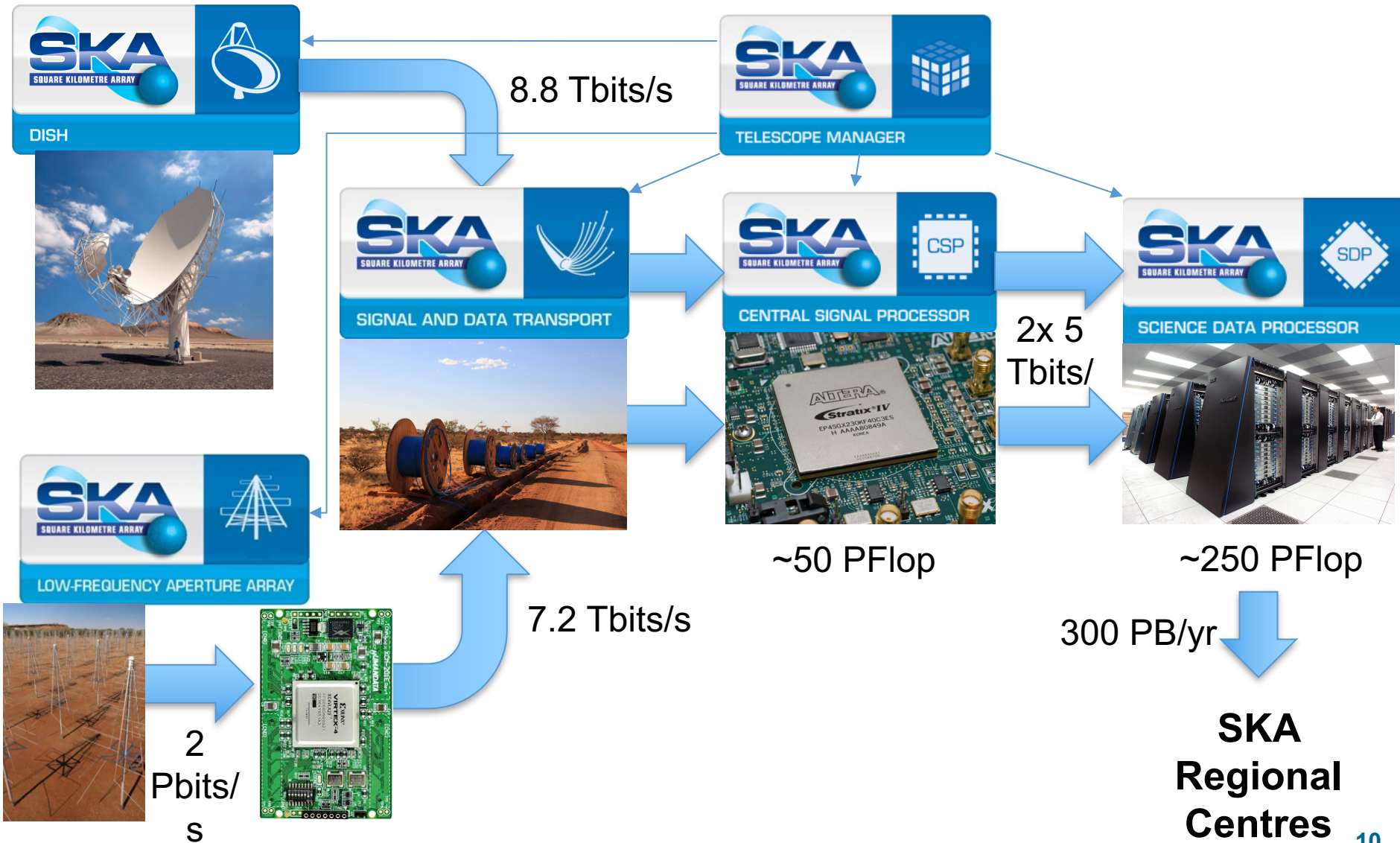
# Project Status

- Expect signature process for the IGO to start before the end of this year
- Element-level and System-level CDRs due to taken place over the next 12 months or so
  - Carried out against current *design baseline not deployment baseline*
- Construction proposal will be put to (proto) SKA Council we hope in early 2019 and construction starts soon after

# SKA Computing Overview

# System Data Flows



8.8 Tbits/s

2x 5 Tbits/

7.2 Tbits/s

2 Pbits/s

~50 PFlop

~250 PFlop

300 PB/yr

**SKA Regional Centres**

- SKA is a software telescope
  - Very flexible and potentially easy to reconfigure
  - Major software development, maintenance and ongoing improvement challenge for 50 years

- HPC piece – Science Data Processor
  - SDP needs ~25 PFLOPs sustained over a scheduling period
  - Current estimate is that SDP needs 250 PFLOPs peak
    - Memory bandwidth is ~200 PB/sec
    - Low arithmetic intensity
    - Roofline models, detailed parametric modelling and prototyping to confirm
  - Real-time as well as batch; complex buffering and data management; all within a live running experiment
  - 50 GF/Watt – Exascale design point

- + Pulsar Search (CSP) is an additional 50 PFLOPs signal processing w/ FPGAs & GPUs
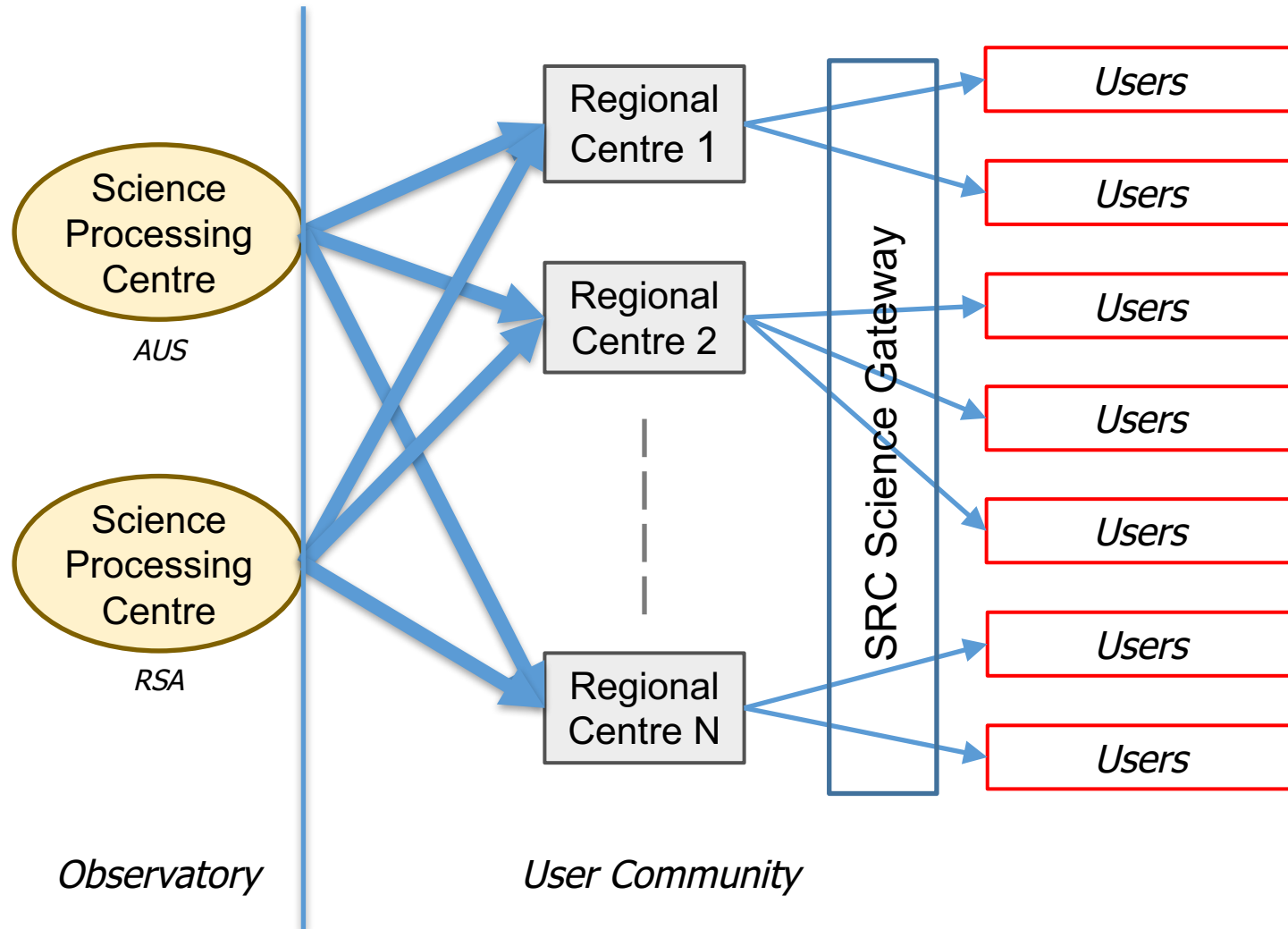
# Software

- Budget of ~€90M on manpower for software development across the whole telescope.
    - Need professional practices for development, testing, integration and deployment.
    - Need to unify the processes across the world-wide team of developers.
    - Need world-leading expertise in a number of areas.
    - Engaged SEI (CMU); Construction will be done using SAFe® methodology
    - Challenge: deal with geographic spread of developers and eliminate fractional FTEs

- Delivered system will not be static
    - SDP hardware and software will be updated regularly
    - Key input for development will be the scientific and software community through the regional centres (more later)

# SKA Regional Centres Overview

- SKA Board agreed the principle of SRCs
- Not part of SKA, or funded by SKA, but:
    - Essential to get most out of the science products produced by SDP
    - Coordinated with assistance from SKAO and accredited with SKAO
- Requirements analysis, technical work underway – SKA Regional Centre Coordination Group (SRCCG)
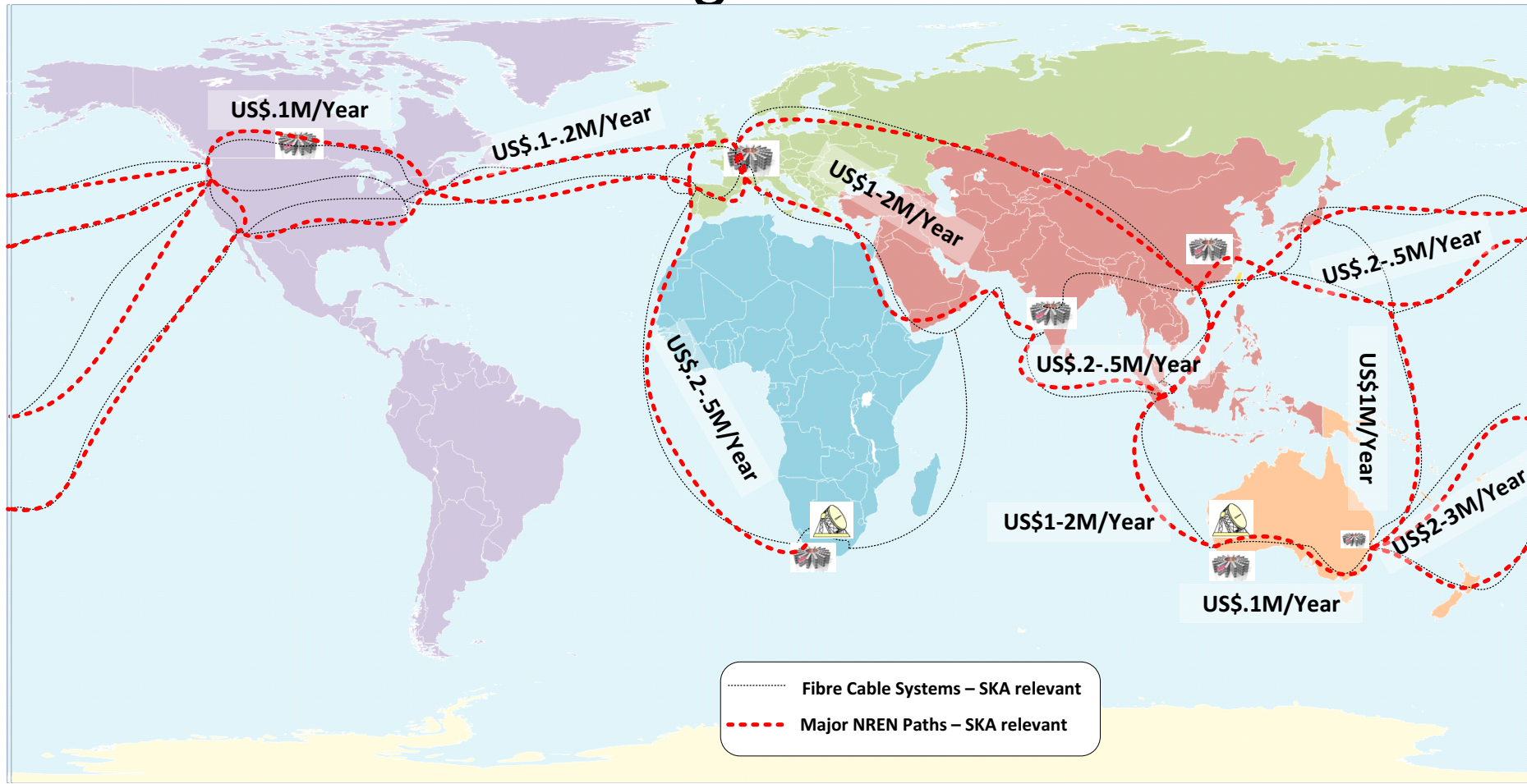    - Personnel from member countries plus CERN and LSST;F2F in Granada yesterday

# Regional Centre Concept – some similarity with WLCG, ARCs
~300 PB/annum

# Regional Centre Network

- 10 year IRU per 100Gbit circuit early 2020s onwards

- Guesstimate of Regional Centre locations

# What will SRCs do?

- Provide a nexus for resources in a member country or region
  - Scientific and software expertise
  - Access to computing resources -> advanced data products
    - but direct ownership is not a requirement
  - How much compute and storage? TBD

- Provide access to data
  - Ensure security and adherence to SKA data policies
  - Play a defined role in hosting and distributing data products
  - Perhaps process precursor and pathfinder data

- Provide local (time zone) user support, proposal access, information, training and outreach activities

- Liaison with SKA Observatory and NRENs
  - Need to ensure sufficient and affordable network capacity is procured and provisioned in a timely fashion

# SRCs cont.

- So, clearly plenty in common with WLCG

- Developments so far:

  - EU H2020 – AENEAS (meeting in Granada)

  - Sino-Australia collaboration – ERIDANUS;

  - NRC, Canada;

  - India

- In the EU we clearly need close EOSC/EDI (PRACE etc) engagement as a potentially significant consumer *and* contributor

# SKA Data Flows

❖ SKA faces computing challenges similar to HEP - CSP, SDP – not just SRCs

  ❖ Data processing, data reduction and distribution



8.8 Tbits/s

Similar to the HEP Software triggers (Looking at GPU and FPGA as well)

Performance and data production are similar to the central reconstruction resources

Very similar in speed and data reduction factor to the hardware triggers in HEP (Currently implemented as ASICs)

~50 PFlop

~250 PFlop

2 Pbits/s

300 PB/yr

**SKA Regional Centres**

CERN openlab

Maria Girone, CERN openlab CTO

# CERN-SKA Collaboration

- Recognition on both sides of potential synergies and requirements
    - Various ad-hoc interactions between communities
    - Reviews and panels etc.
    - Planning a CERN-SKA "Big data" workshop in the UK Alan Turing Inst. in 9-10/4/2018
- On 13 July CERN and SKAO DG's signed a collaboration agreement on computing, data management, etc.
    - Recognizing that both HL-LHC and SKA will be Exabyte-scale scientific experiments on a 10-year timescale



**COLLABORATION AGREEMENT**
**KN3644**

Between

**THE EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH** (hereinafter referred to as "CERN") an Intergovernmental Organization with its seat at Geneva, Switzerland, represented by its Director-General, Dr Fabiola Gianotti,

And

**THE SKA ORGANISATION** (hereinafter referred to as "SKAO") with its headquarters at Jodrell Bank, Manchester, United Kingdom, represented by its Director-General, Professor Philip Diamond,

hereinafter individually and collectively referred to as the "Party" and "Parties" respectively,

**CONSIDERING THAT:**

1) Both Parties are constructing scientific instruments which will be capable of collecting scientific data at the Exabyte scale in the next decade;

2) The acquisition, storage, management, distribution and analysis of scientific data at such a scale represent technological and management challenges that are unique and unprecedented in science;

3) These data will be analysed by globally distributed scientific collaborations;

4) The computational and storage resources needed by the Parties and their respective scientific collaborations will, in many countries, be common;

5) The challenges faced by the Parties represent several areas that can potentially be addressed collaboratively;

**HAVE AGREED AS FOLLOWS:**



**CERN COURIER**

Aug 11, 2017

**SKA and CERN co-operate on extreme computing**

Big-data co-operation agreement

# Initial topics of collaboration

- Bi-annual collaboration meetings
    - First was held last week at SKAO HQ
    - (+ some joint meetings at SC17)
- A position paper/roadmap to be produced quickly to focus on Exabyte-scale data infrastructure needs
    - To influence funding agencies, vendors, long term public infrastructures, resource providers
- Explore collaboration opportunities on common aspects of networking, storage, computing, etc.
    - Computing/Openstack already very active
    - Discussing storage and data management
- Investigate work with industry via CERN Openlab
- Joint projects to demonstrate/prototype concepts for regional centres and computing models
- Other items as needed

# Roadmap ideas

- Idea is to have a high level document (~10 pages) rather quickly
- Position: we (HL-LHC and SKA) are the RI's facing the Exascale challenge now
  - We are at the forefront – but others are coming soon
  - Want a common and coordinated approach to access HPC and open science resources
- A 10-year roadmap to cover the period of HL-LHC and SKA full operations, with a future outlook
  - Should also be open to the other scientific disciplines in order to influence an open scientific data processing infrastructure
  - At a global level
- Organise workshops to explore specific topics of common interest, e.g. data infrastructures, networking technology, reproducibility, etc.
  - Propose prototyping projects, demonstrations of key functionality with existing components, ...

# Prototypes and data challenges

- The idea is to try and build value for existing science activities through pragmatic prototypes
    - E.g. use existing data management tools to benefit LOFAR/HERA/MeerKAT/MWA/FAST activities
    - Feed results back to existing WLCG production systems
    - Use prototypes for data challenges for SKA and HL-LHC preparation
- Jointly explore concepts for SKA & LHC Regional Centres
    - E.g. a "data lake": a federated data infrastructure with associated tools
    - Understand how to plug hybrid compute capacity into it

- Thank you

- Any questions?