

ASTRON

Netherlands Institute for Radio Astronomy

Establishing the SKA Regional Centre Network *Mesh Management and Culture Change*

Michael Wise

Head of Astronomy, ASTRON

ADASS 2018

College Park, MD, November 15, 2018

ASTRON is part of the Netherlands Organisation for Scientific Research (NWO)

Today's Astronomy is The History of the Universe



Testing General Relativity
(Strong Gravity, Gravitational Waves)

Cradle of Life
(Planets, Molecules, SETI)

Exploration of the Unknown

Cosmic Dawn
(First Stars & Galaxies)

Galaxy Evolution
(Normal Galaxies at $z \sim 2-3$)

Cosmology
(Dark Matter, Large-scale Structure)

Formation and Evolution of Galaxies • The Dawn of Galaxies: Searching for the Epoch of First Light • 21-cm Emission and Absorption Mechanisms • Preheating the IGM • **SKA Imaging of Cosmological HI** • Large Scale Structure and Galaxy Evolution • A Deep SKA HI Pencil Beam Survey • Large scale structure studies from a shallow, wide area survey • The Ly- α forest seen in the 21-cm HI line • **High Redshift CO** • Deep Continuum Fields • Extragalactic Radio Sources • The SubmicroJansky Sky • Probing Dark Matter with **Gravitational Lensing** • Activity in Galactic Nuclei • The SKA and Active Galactic Nuclei • Sensitivity of the SKA in VLBI Arrays • Circum-nuclear MegaMasers • H₂O megamasers • OH Megamasers • Formaldehyde Megamasers • The **Starburst** Phenomenon • **Interstellar Processes** • HII Regions: High Resolution Imaging of Thermal Emission • Centimetre Wavelength Molecular Probes of the ISM • **Supernova Remnants** • The Origin of Cosmic Rays • Interstellar Plasma Turbulence • Recombination Lines • Magnetic Fields • Rotation Measure Synthesis • Polarization Studies of the Interstellar Medium in the Galaxy and in Nearby External Galaxies • Formation and Evolution of Stars • Continuum Radio Emission from Stars • Imaging the **Surfaces of Stars** • Red Giants and Supergiant Stars • Star Formation • Protostellar Cores • Protostellar Jets • Uncovering the Evolutionary Sequence • Magnetic Fields in Protostellar Objects • Cool Star Astronomy • The Radio **Sun** • Observing Solar Analogs at Radio Wavelengths • Where are the many other Radio Suns? • Flares and Microflares • X-ray Binaries • Relativistic Electrons from X-ray Transients • The Faint Persistent Population • Imaging of Circumstellar Phenomena • Stellar Astrometry • Supernovae • Radio **Supernovae** • The Radio After-Glows of **Gamma-ray Bursts** • Pulsars • Pulsar Searches • **Pulsar Timing** • Radio Pulsar Timing and General Relativity • Solar System Science • Thermal Emission from Small Solar System Bodies • Asteroids • Planetary Satellites • **Kuiper Belt Objects** • Radar Imaging of Near Earth Asteroids • The Atmosphere and Magnetosphere of Jupiter • Comet Studies • Solar Radar • Coronal Scattering • Formation and Evolution of Life • Detection of **Extrasolar Planets** • Pre-Biotic Interstellar Chemistry • The Search for **Extraterrestrial Intelligence**

SKA Science

- Australia
- Canada
- China
- France
- India
- Italy
- Netherlands
- New Zealand
- South Africa
- Spain
- Sweden

Potential new members: Germany, Portugal,
Switzerland, Japan, etc.





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

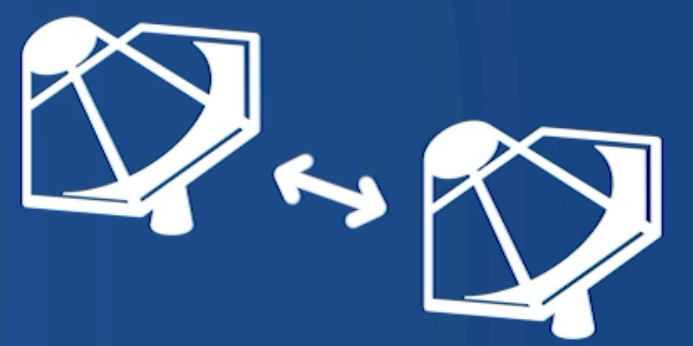


Frequency range:
350 MHz to
14 GHz



~200 dishes
(including 64 MeerKAT dishes)

Total
collecting
area:
33,000m²



Maximum distance
between stations:
150km

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

SKA1 MID
South Africa

33,000m²
~200 dishes



JVLA
Karl G. Jansky Very Large
Array, USA
13,200m²
27 dishes

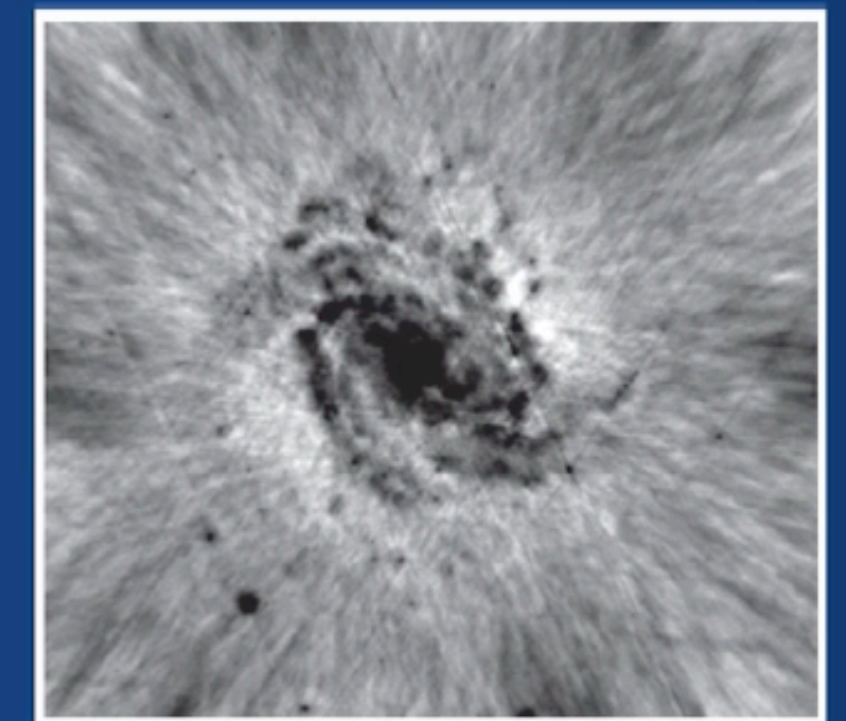
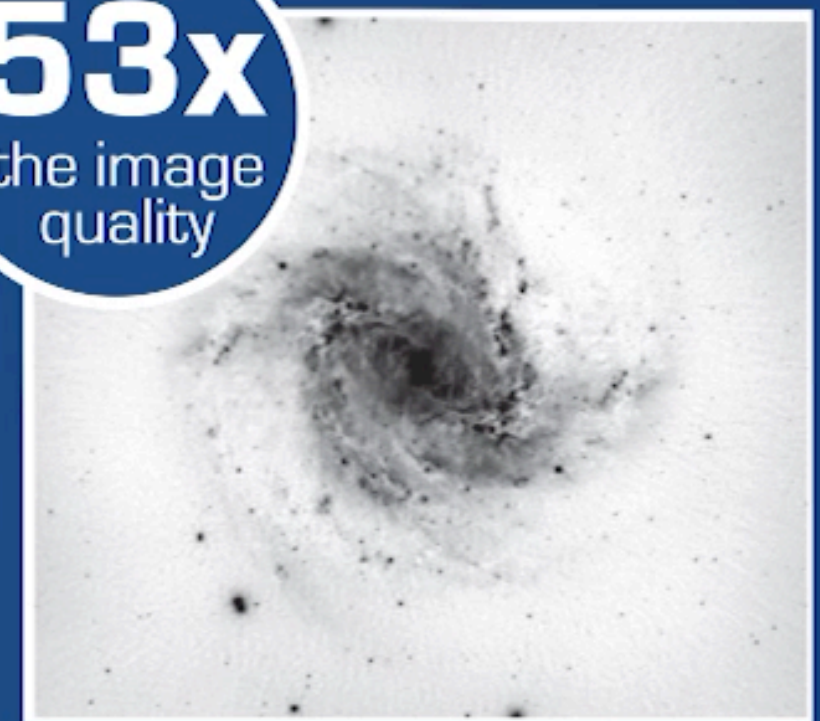


4x
the
resolution

5x
more
sensitive

60x
the survey
speed

53x
the image
quality





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

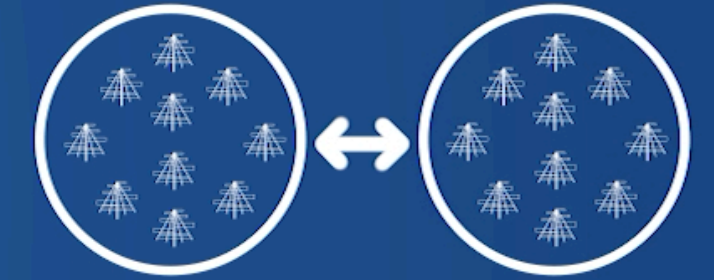


Frequency range:
50 MHz to
350 MHz



~130,000
antennas spread between
500 stations

Total
collecting
area:
0.4km²



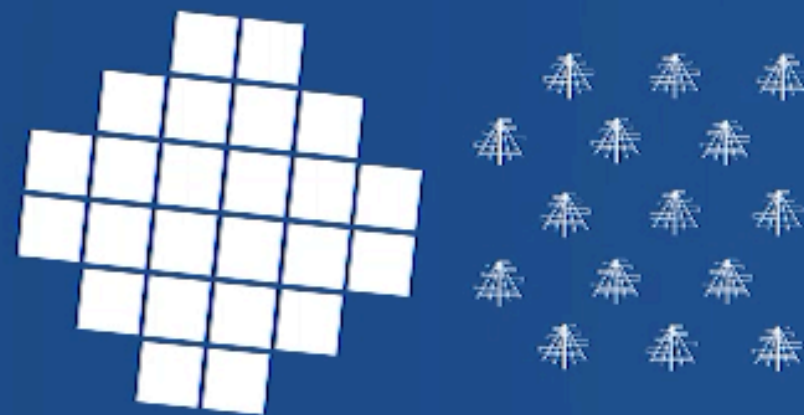
Maximum distance
between stations:
>65km

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

LOFAR
LOw Frequency ARray
52,000m²
34,000 antennas

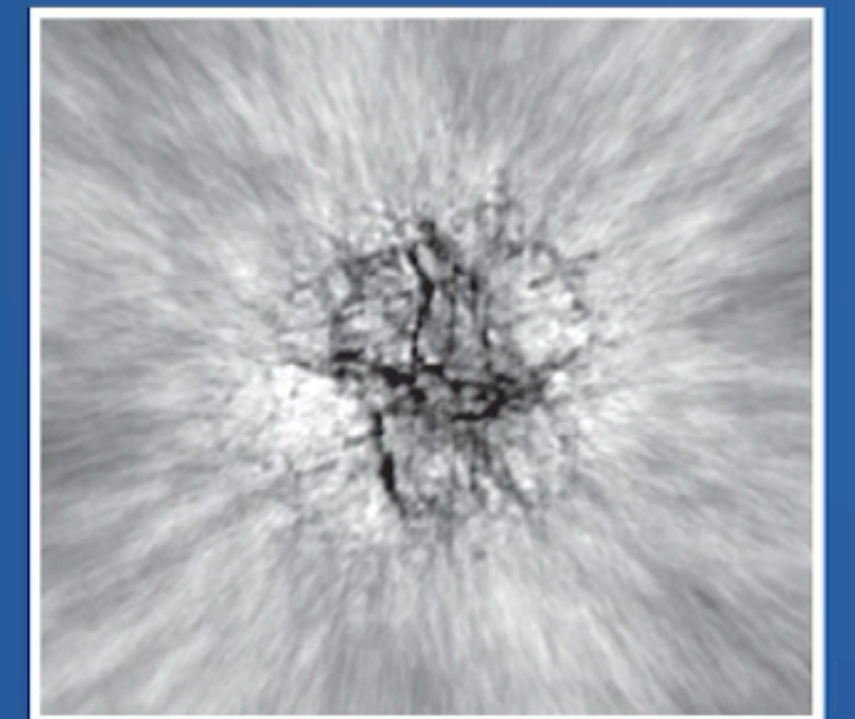


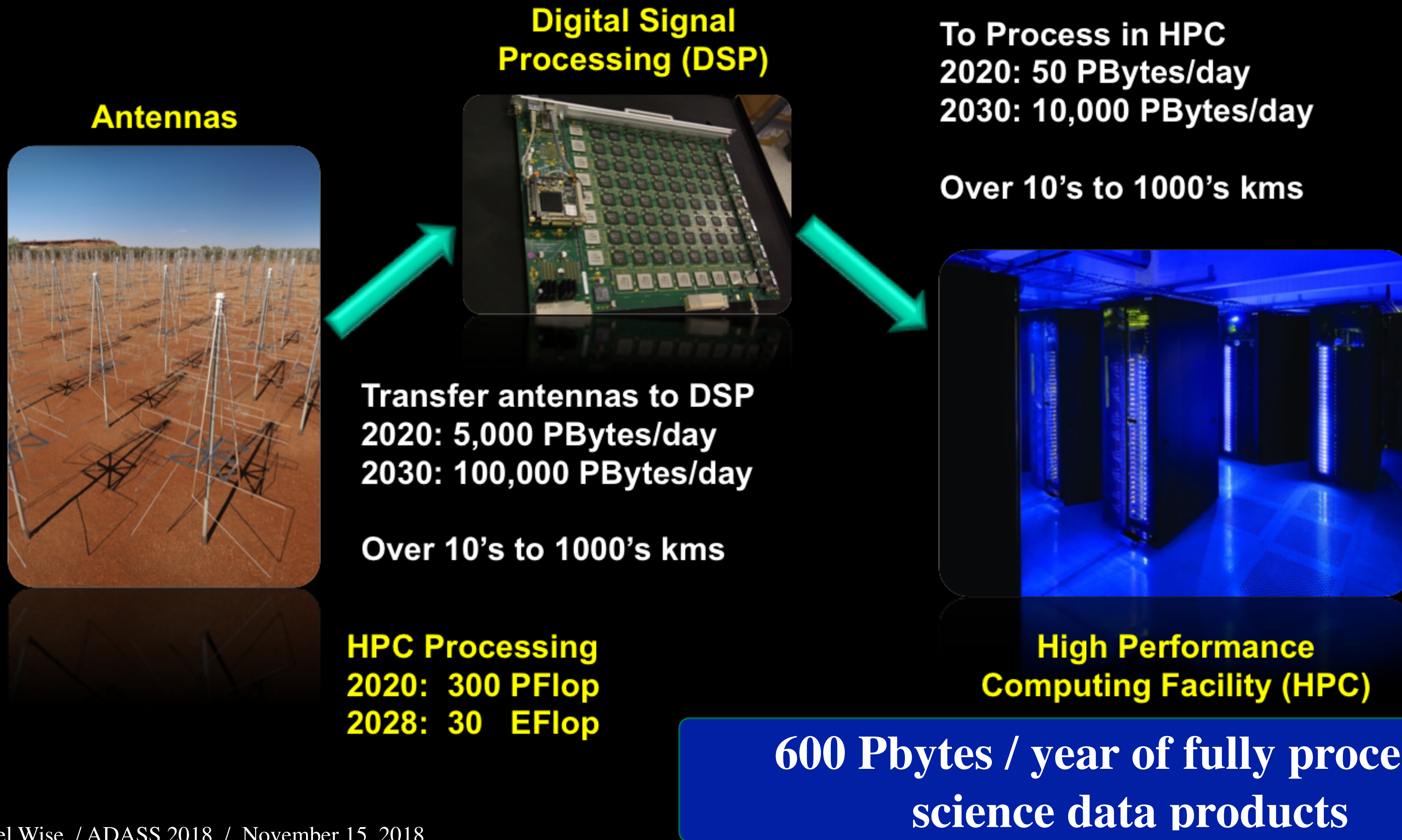
25%
better
resolution

8x
more
sensitive

135x
the survey
speed

68x
the image
quality



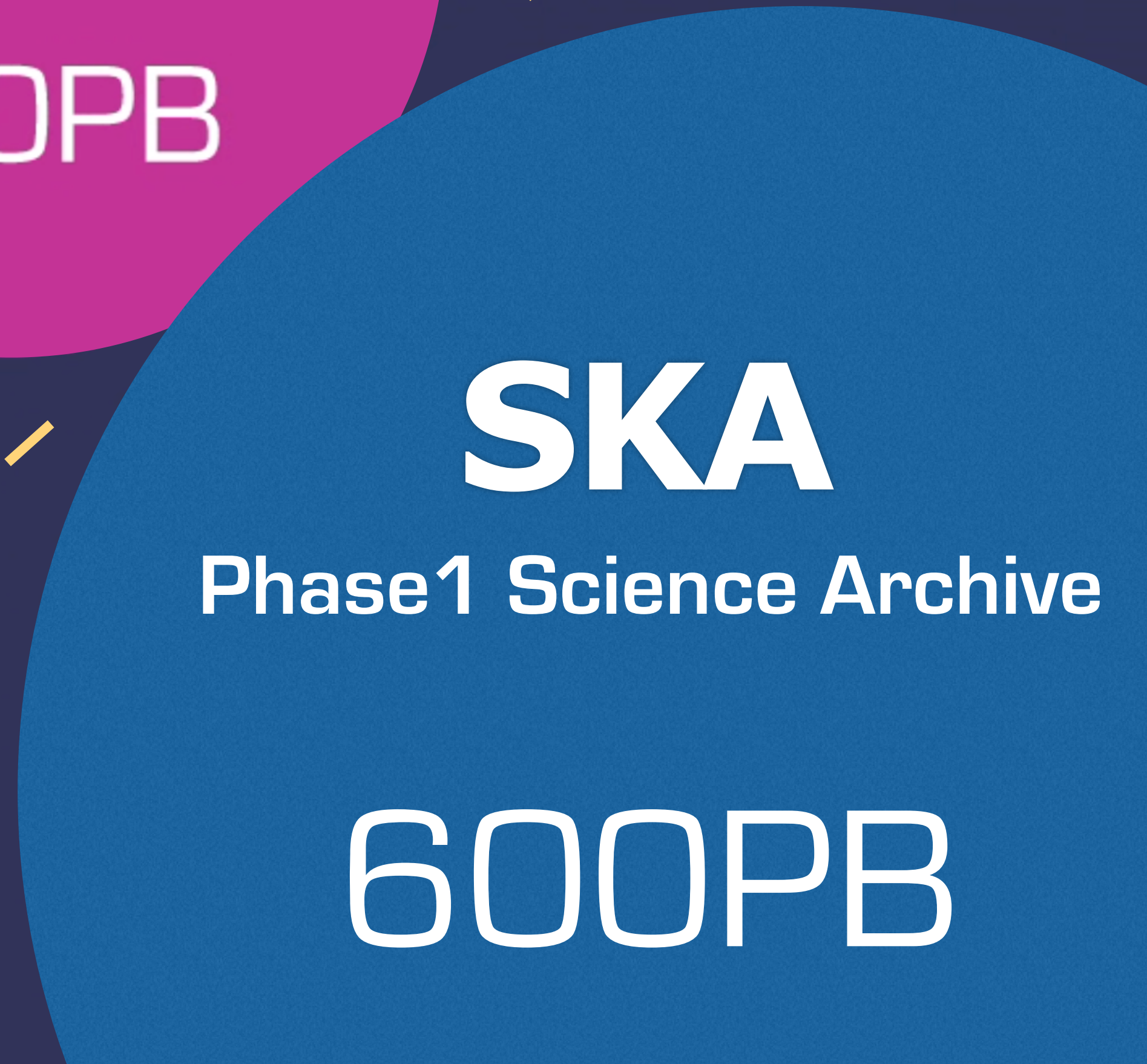


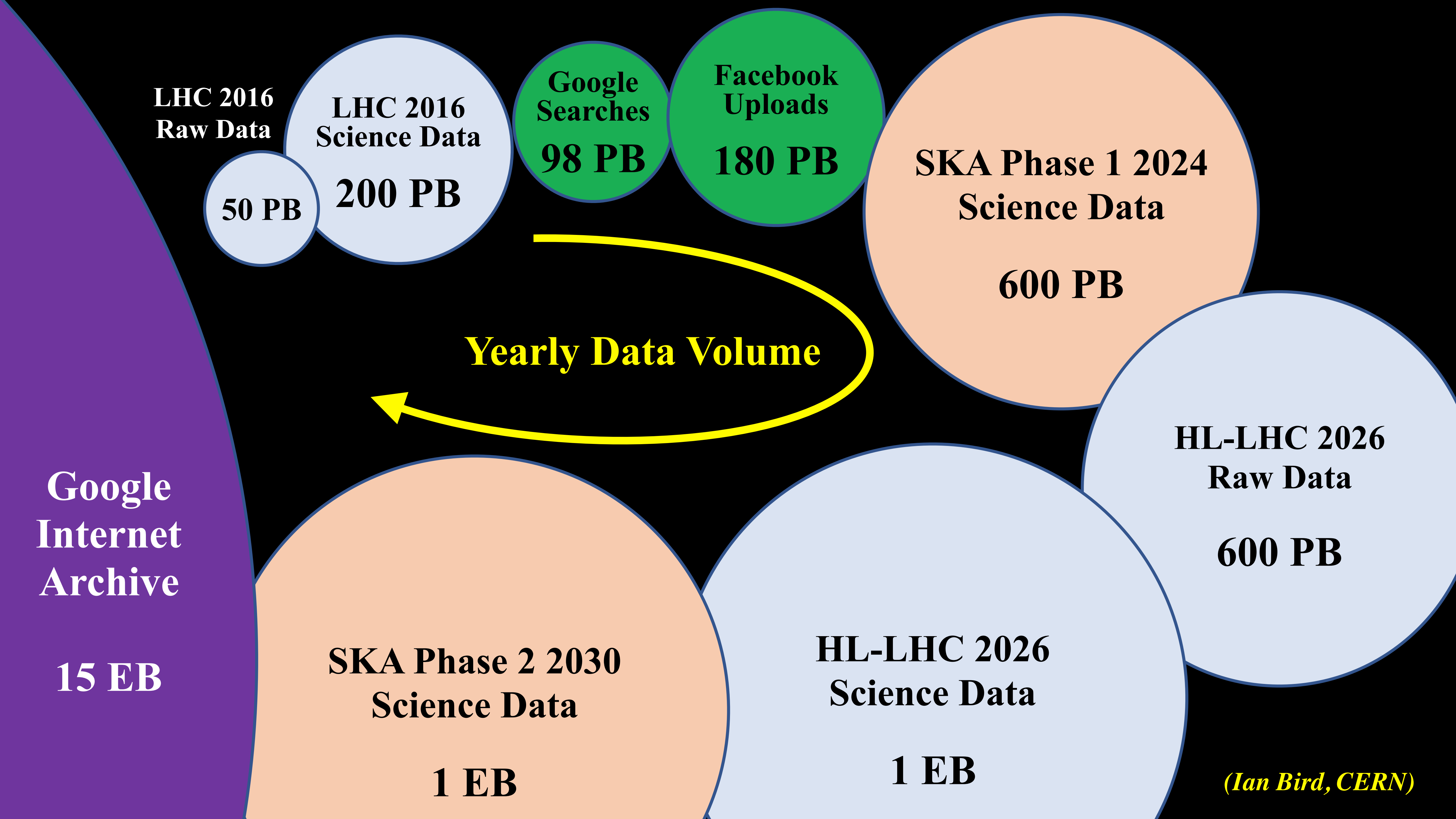
Future SKA Science Archive

2018
—
2024



PER YEAR
1 Petabyte





**LHC 2016
Raw Data**

50 PB

**LHC 2016
Science Data**

200 PB

**Google
Searches**

98 PB

**Facebook
Uploads**

180 PB

**SKA Phase 1 2024
Science Data**

600 PB

Yearly Data Volume

**Google
Internet
Archive**

15 EB

**SKA Phase 2 2030
Science Data**

1 EB

**HL-LHC 2026
Science Data**

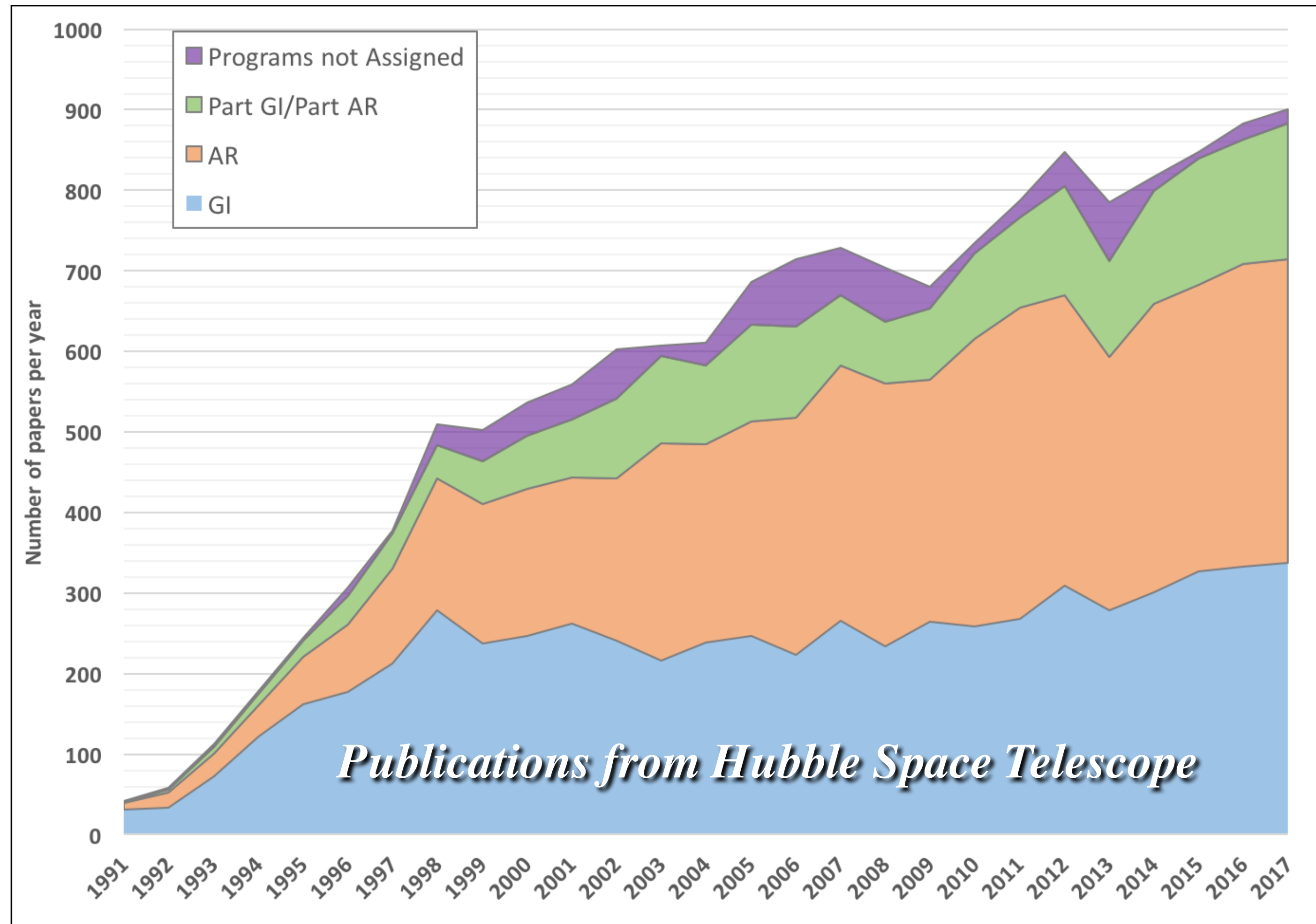
1 EB

**HL-LHC 2026
Raw Data**

600 PB

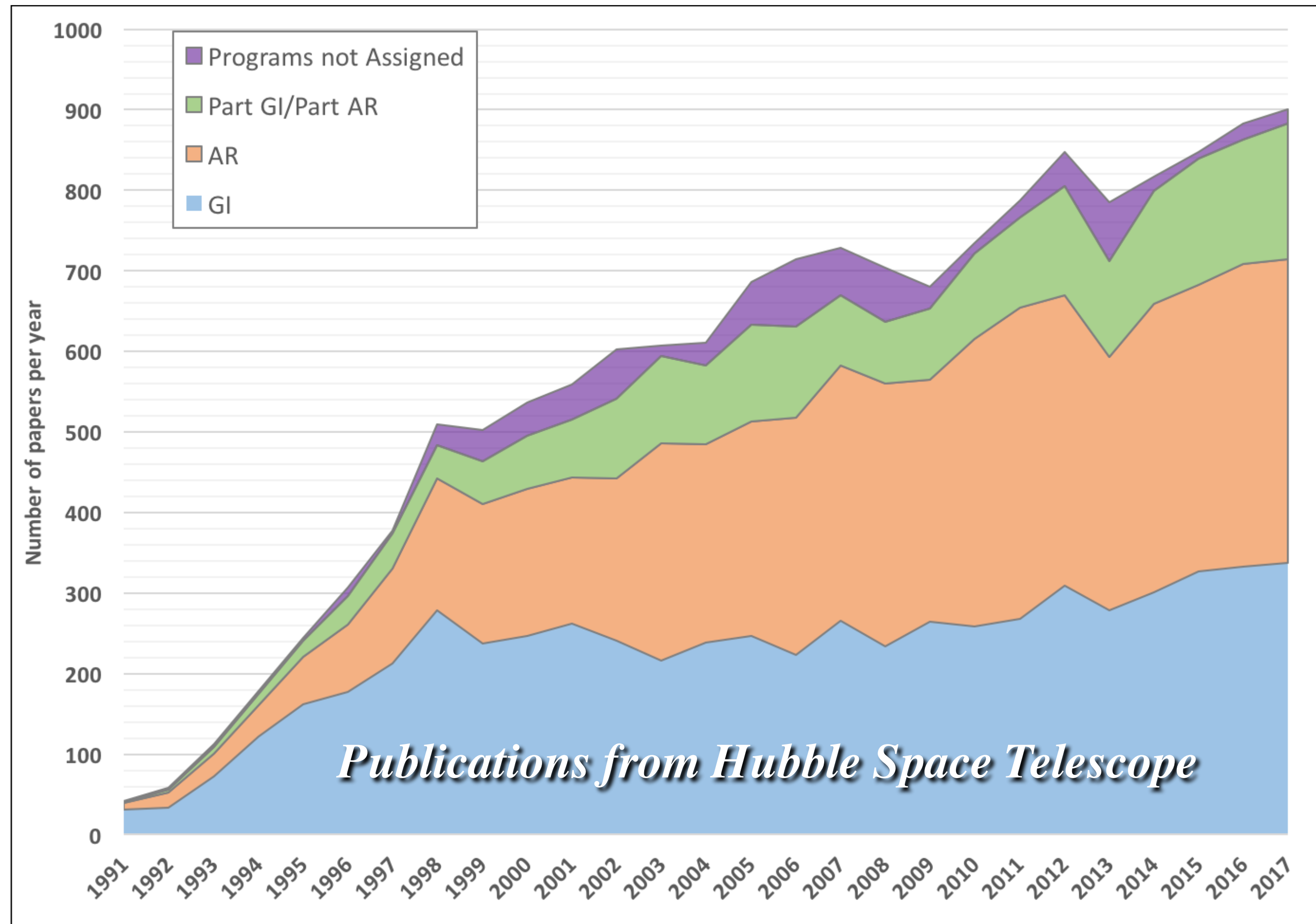
(Ian Bird, CERN)

Science archives are a multiplier for total science output



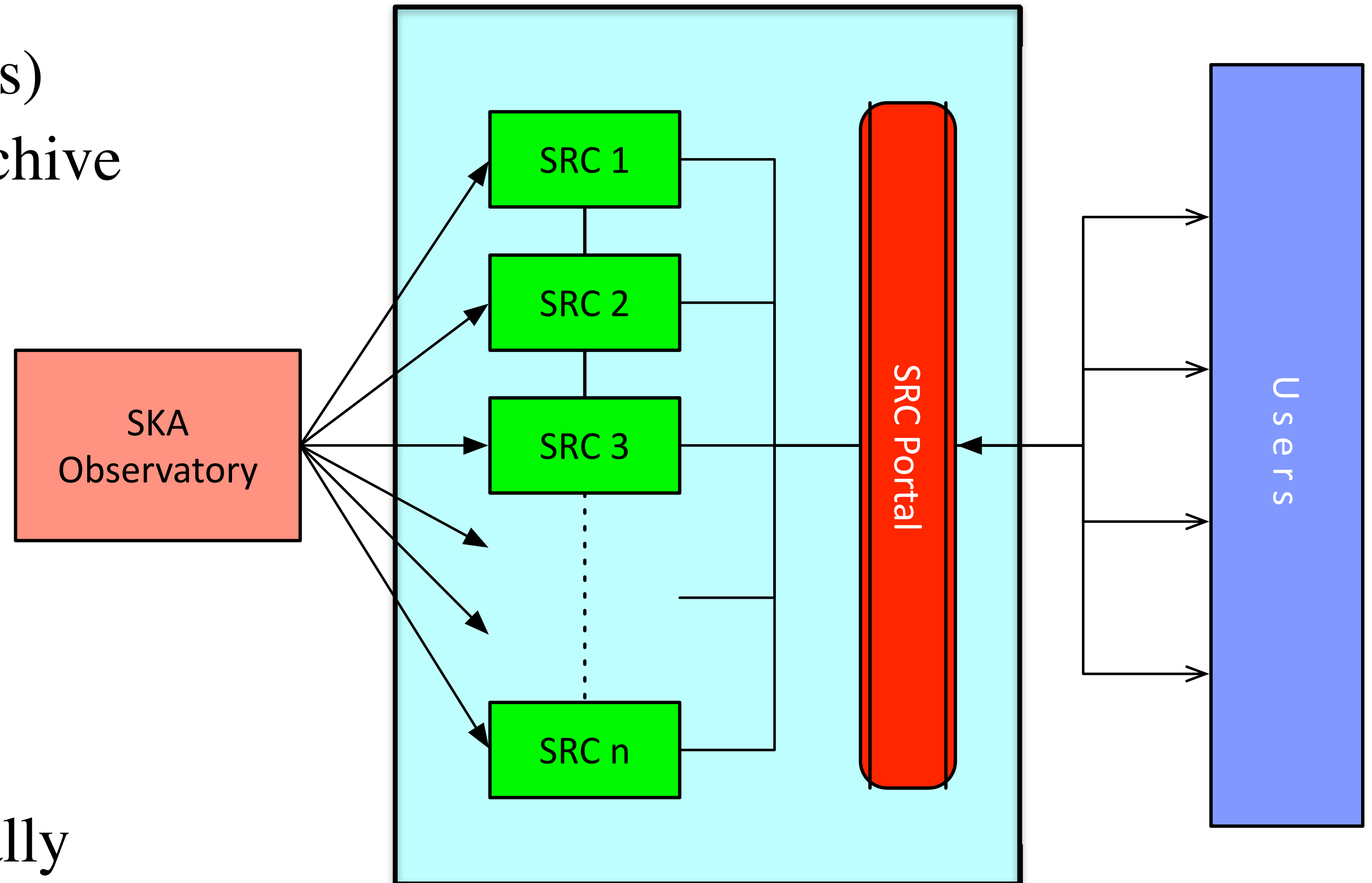
- Assumes the archives are persistent and maintained
- Assumes archival data is open and accessible to users
- Assumes data products stored are appropriate for general use
- Assumes users retrieving data have resources to process to a science result

Science archives are a multiplier for total science output



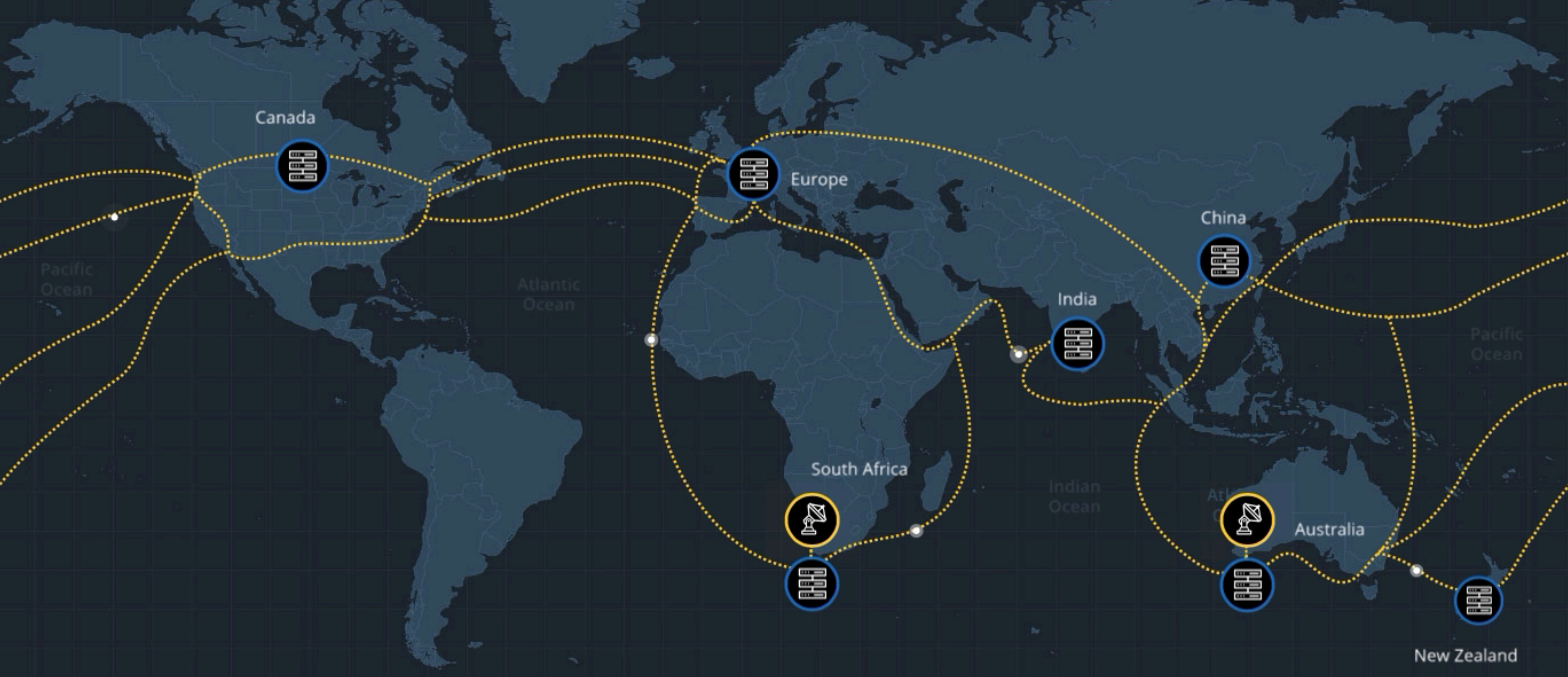
- Assumes the archives are persistent and maintained
- Assumes archival data is open and accessible to users
- Assumes data products stored are appropriate for general use
- Assumes users retrieving data have resources to process to a science result

- SKA Regional Centres (SRCs) will host the SKA science archive
- Provide access and distribute data products to users
- Provide access to compute and storage resources
- Provide analysis capabilities
- Provide user support
- Multiple regional SRCs, locally resourced and staffed



Primary interface for SKA data analysis

Global Network of SKA Regional Centres



Global Network of SKA Regional Centres

Where will the SKA science archive data be hosted?

How will that data be distributed around the SRC network?

How can we take optimal advantage of existing infrastructure?

What are the processing requirements and technologies to consider?

What interfaces, tools, and techniques will users need for analysis?

How do we setup and operate an international network of SRCs?



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)

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

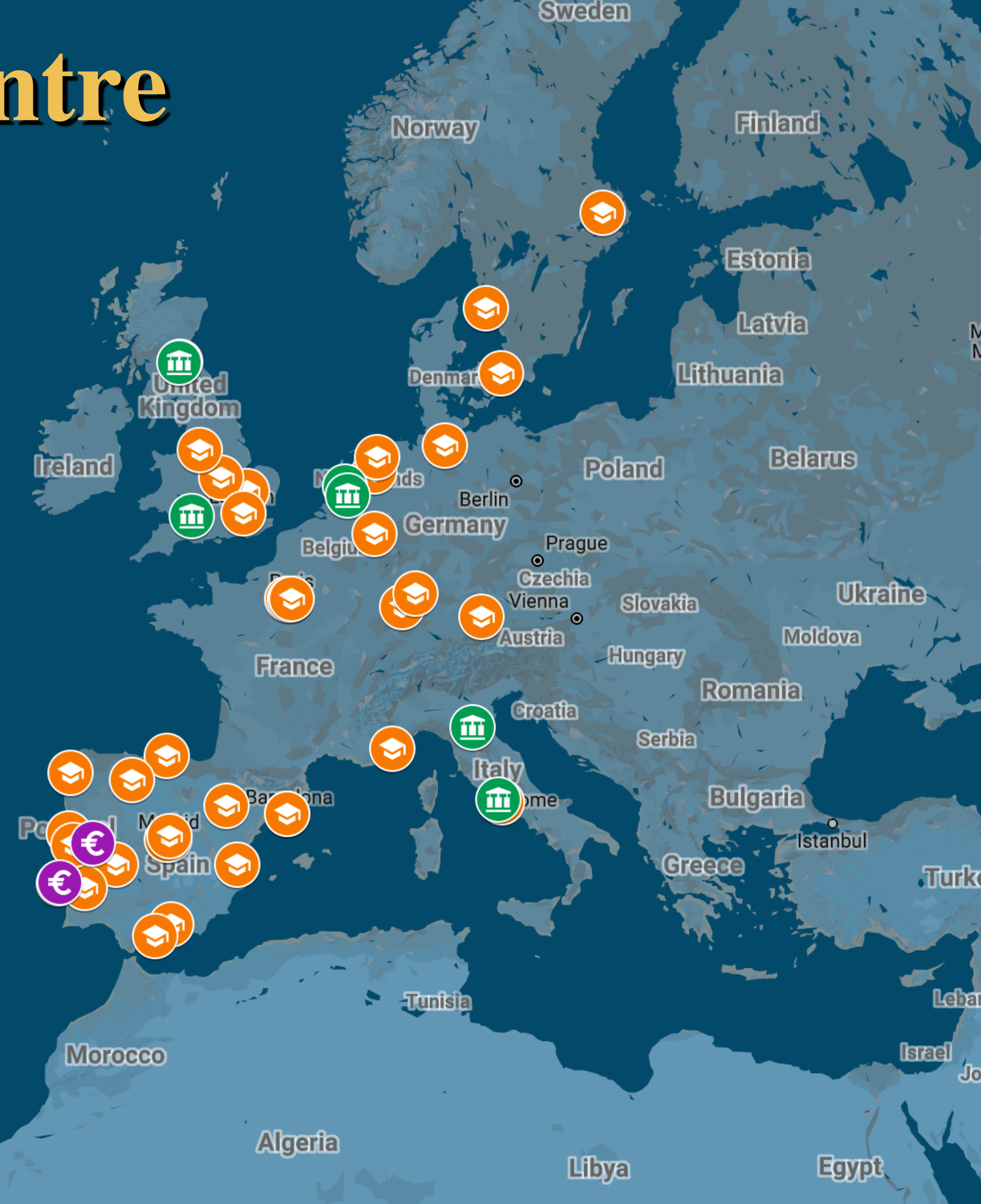
- WP1: Project Management
- WP2: Governance Structure and Business Models
- WP3: Computing and Processing Requirements
- WP4: Data Transport and Optimal European Storage Topologies
- WP5: Data Access and Knowledge Creation
- WP6: User Services



European SKA Regional Centre

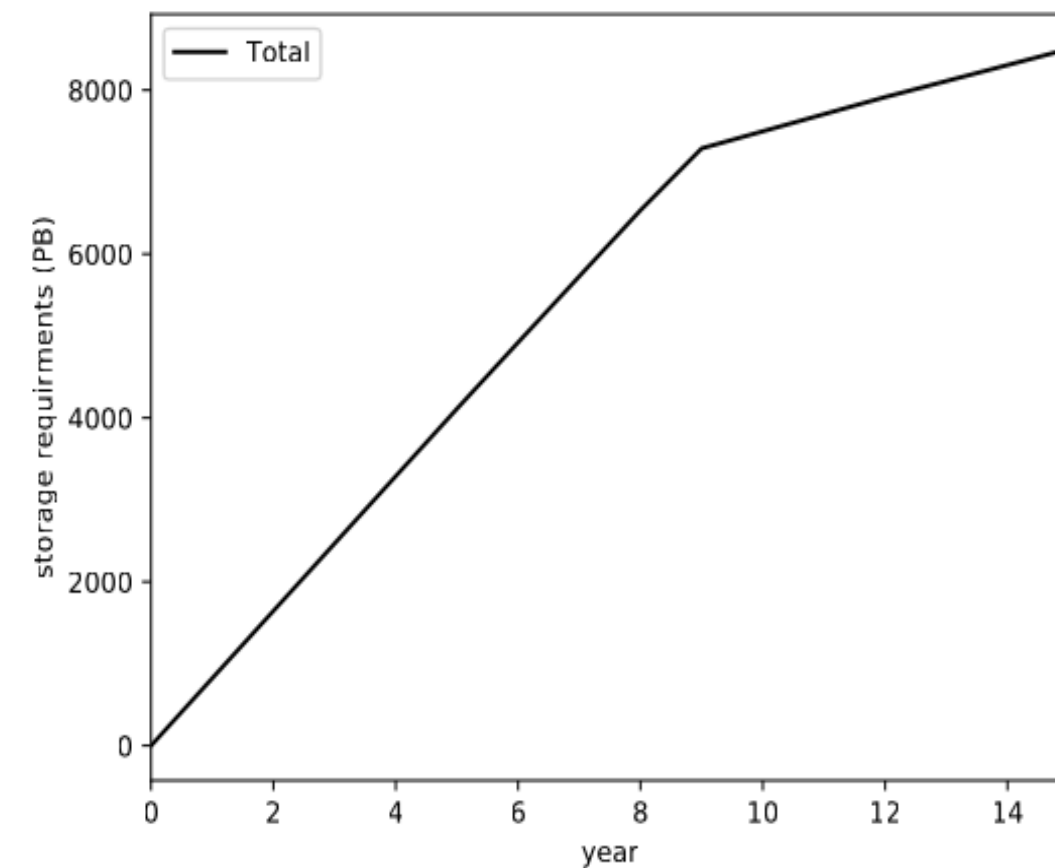
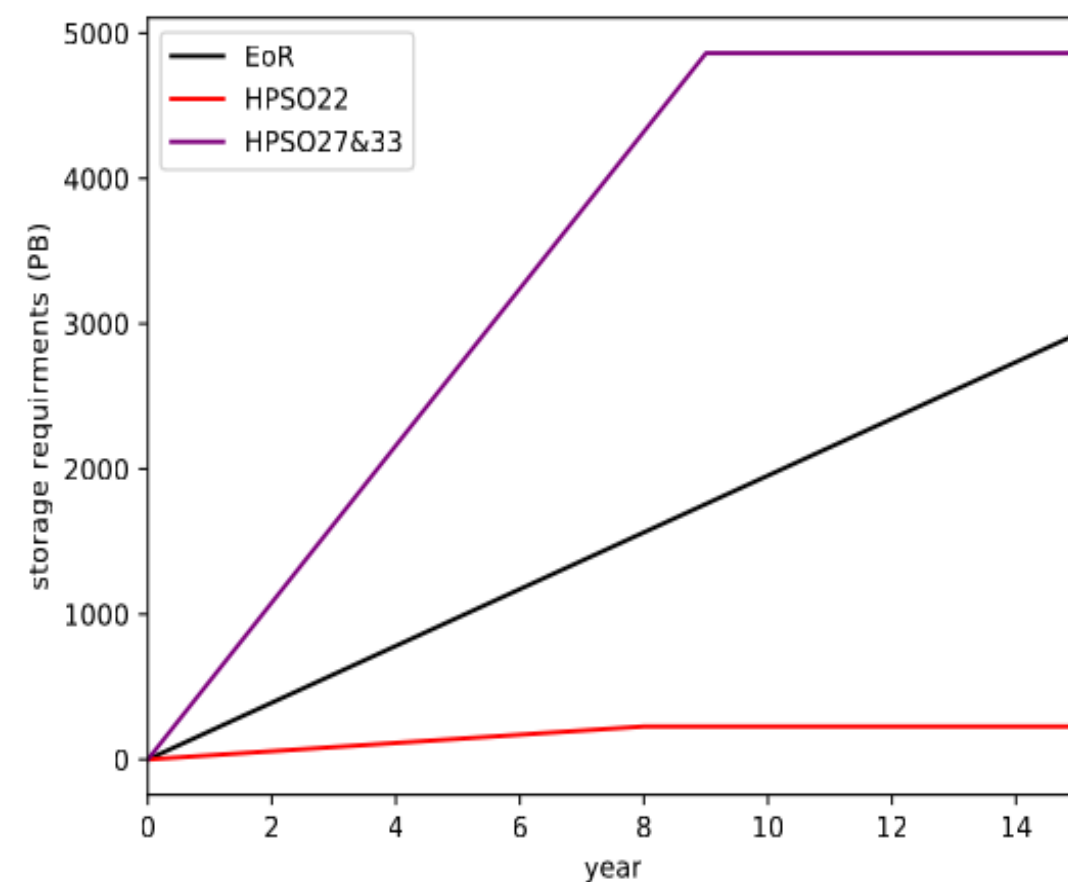
Possible Contributing Nodes

- *Over 52 expressions of interest*
- *Mixture of scientific institutes, infrastructure providers, and industrial partners*
- *Shared development and operations*
- *Combination of national and EC funding*

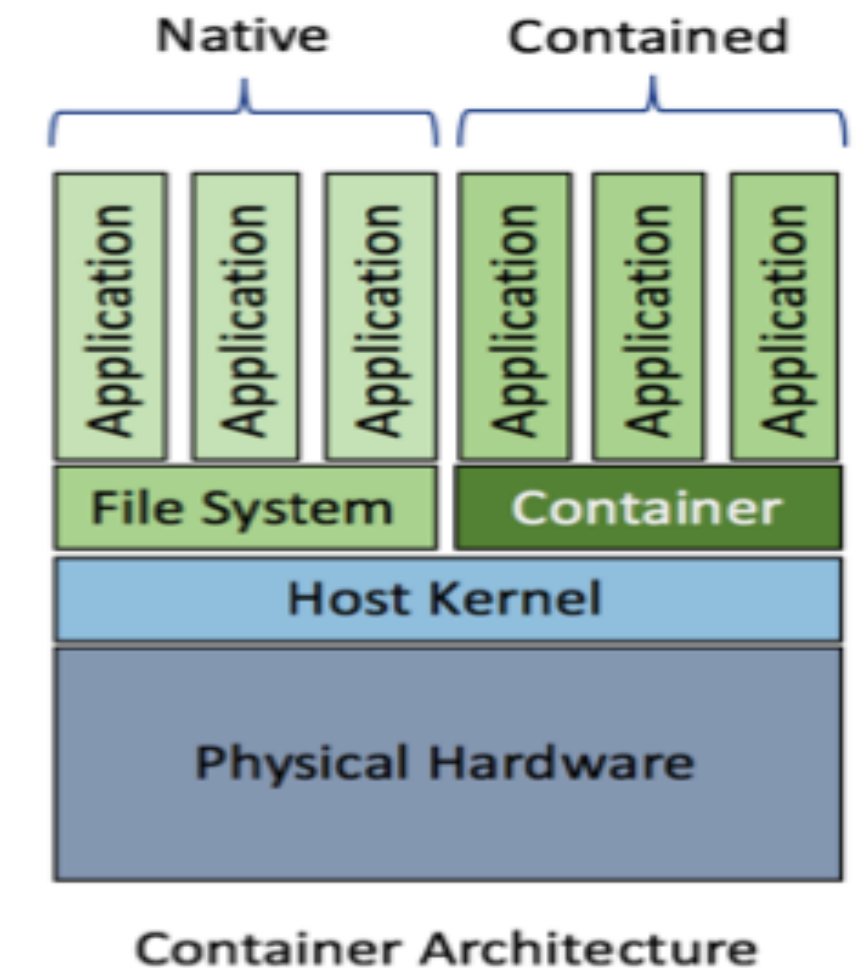
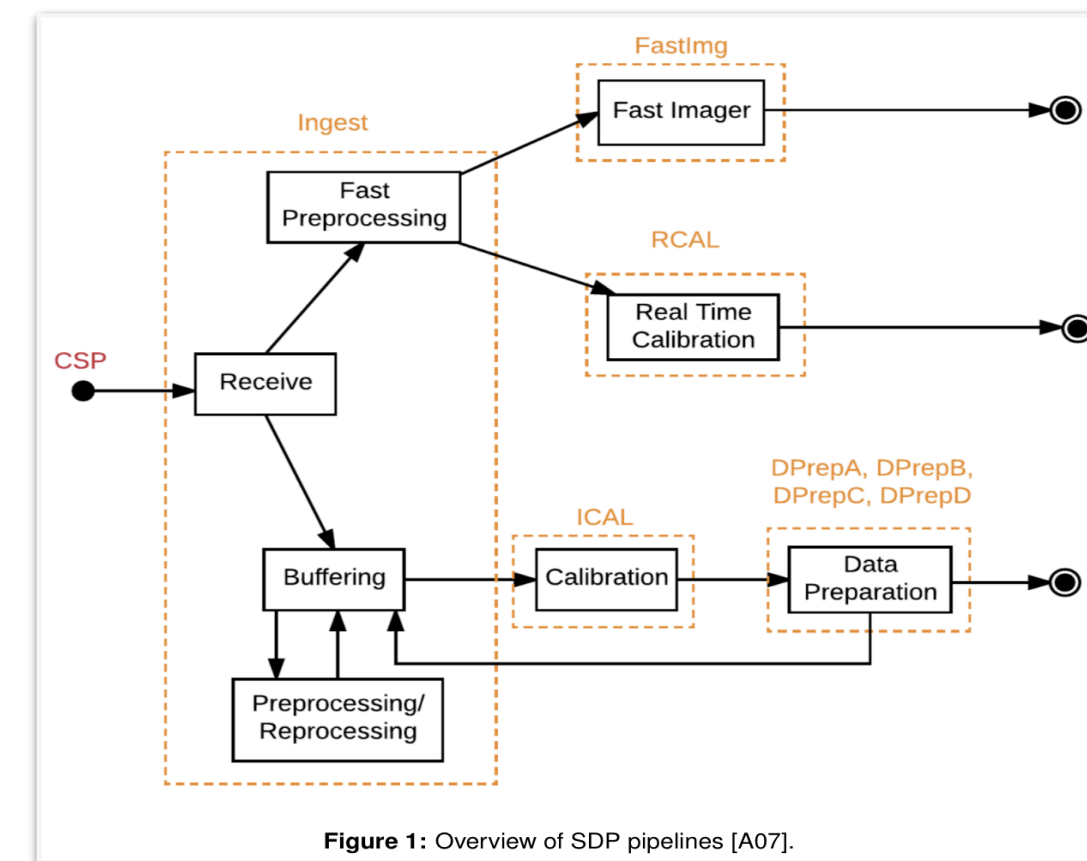


Storage and Computing Models

Storage estimates for HPSOs



Reprocessing and post-processing



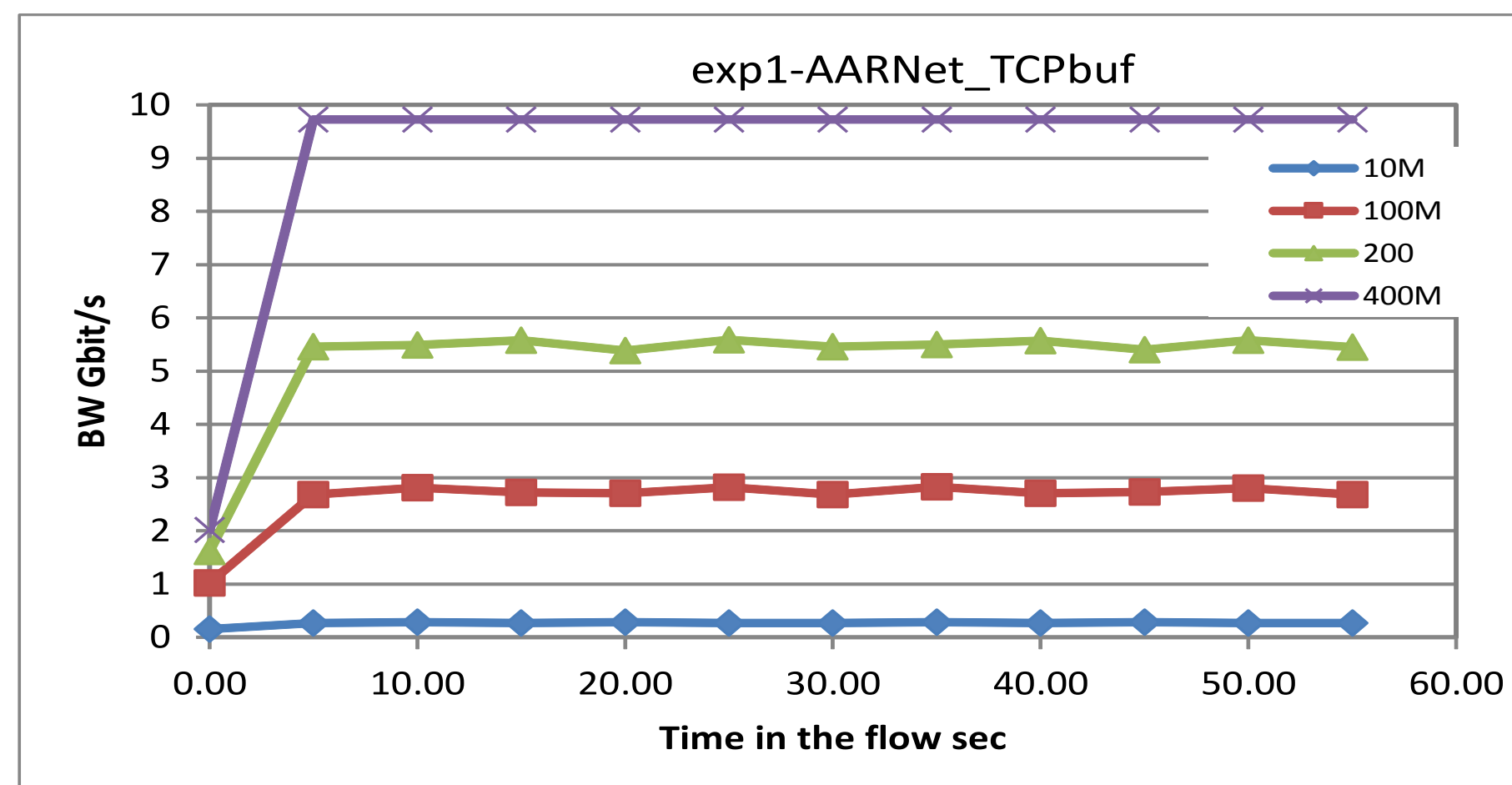
10 ExaBytes over first 15 years of SKA operations

Minimum for HPSOs ~23 PFlops

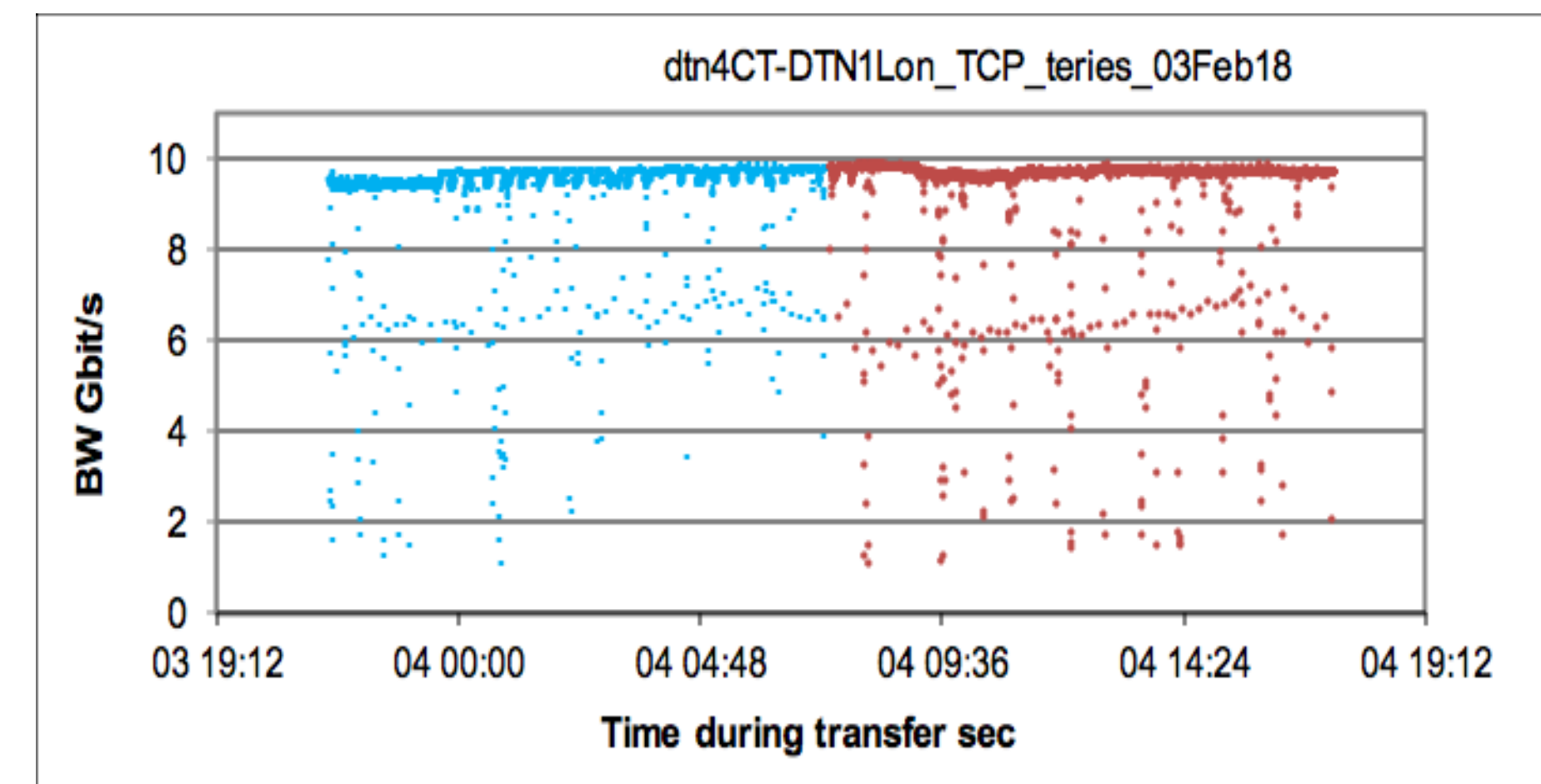
What are the assumptions and the cost implications?

Distributing SKA Data Globally and in Europe

- With AARNet & SANReN test inter-continental performance
- Best practices, protocol tests, and long haul effects (10 & 100 Gigabit)



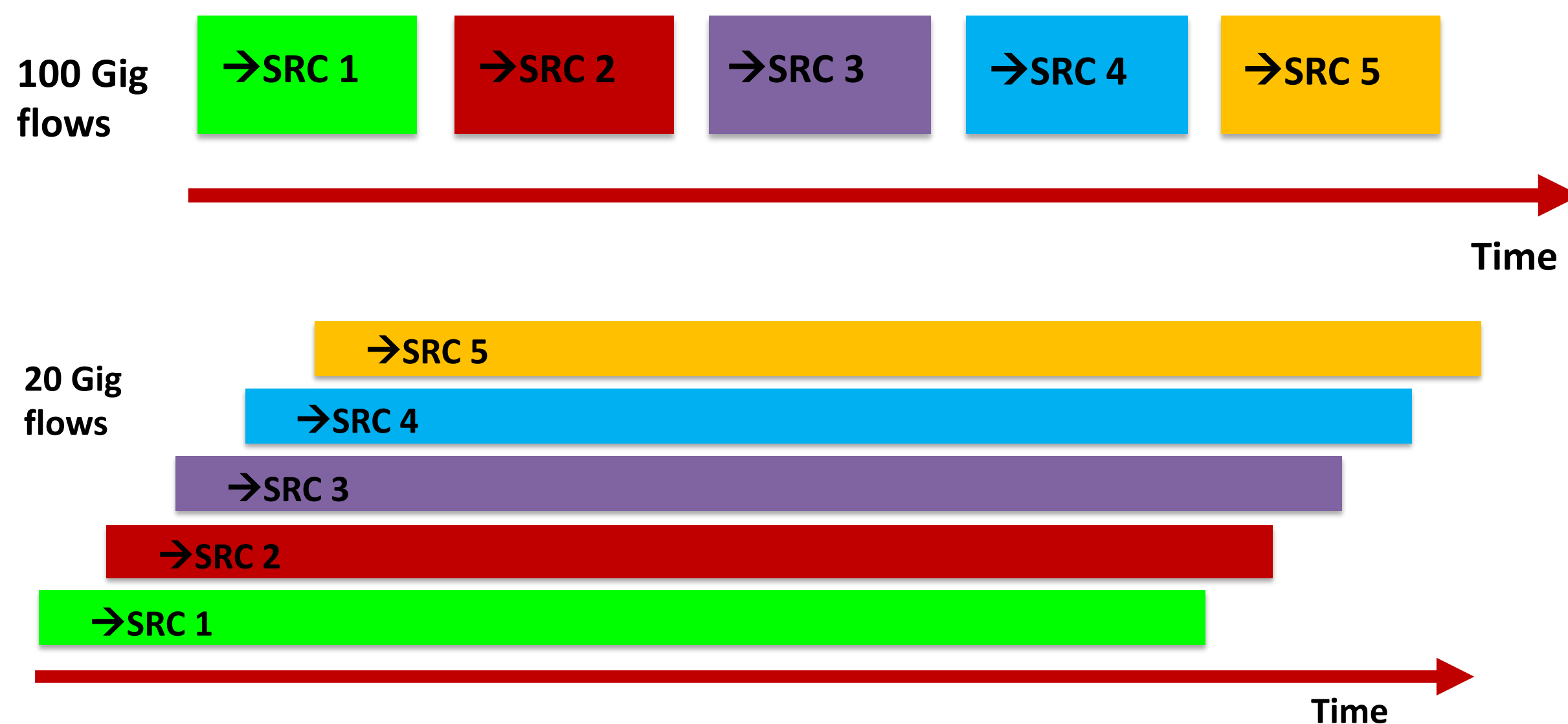
GÉANT London to AARNet Canberra



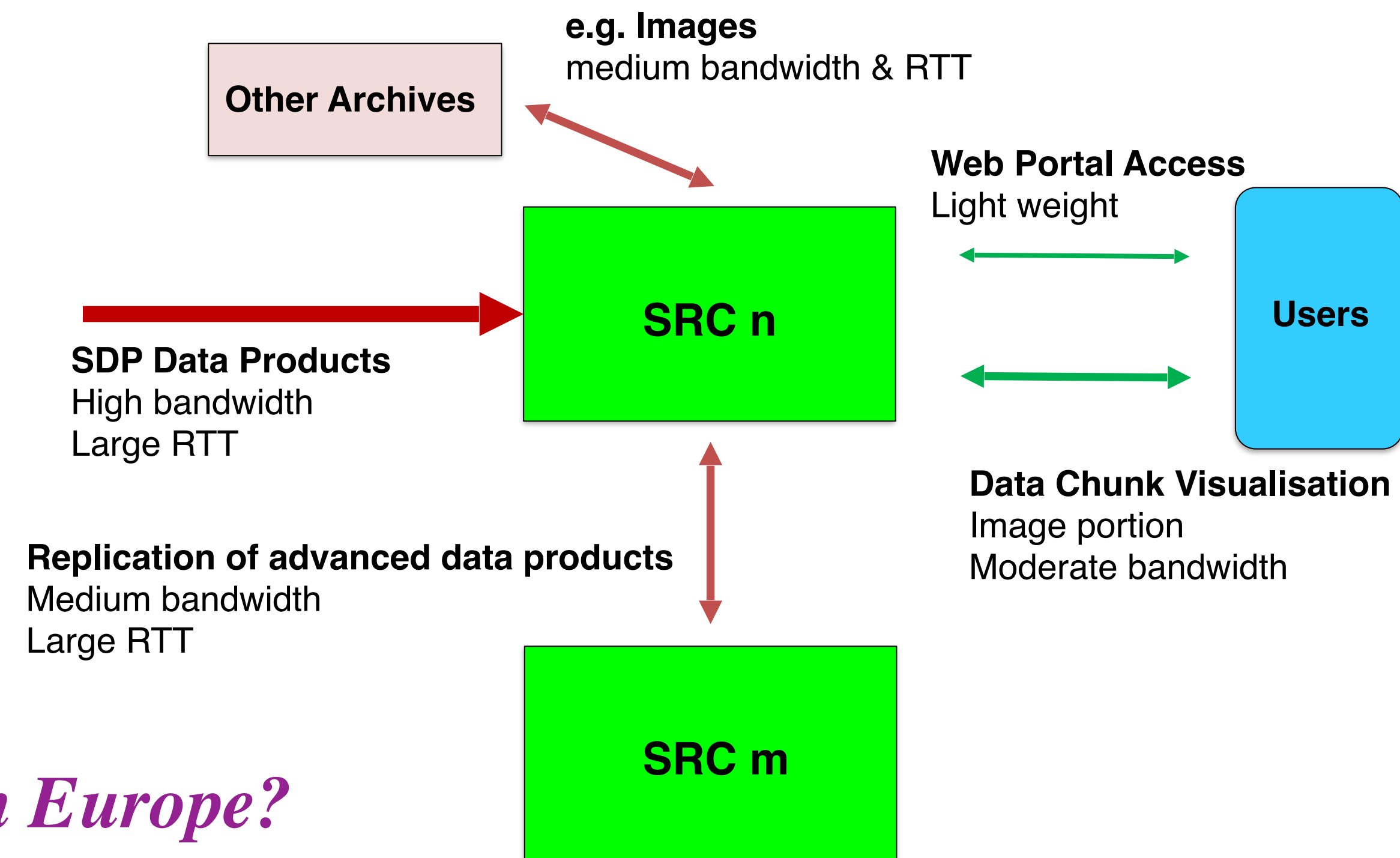
SANReN Cape Town to GÉANT London

Models of Global SKA Data Flow

Distribution from Sites to SRC network

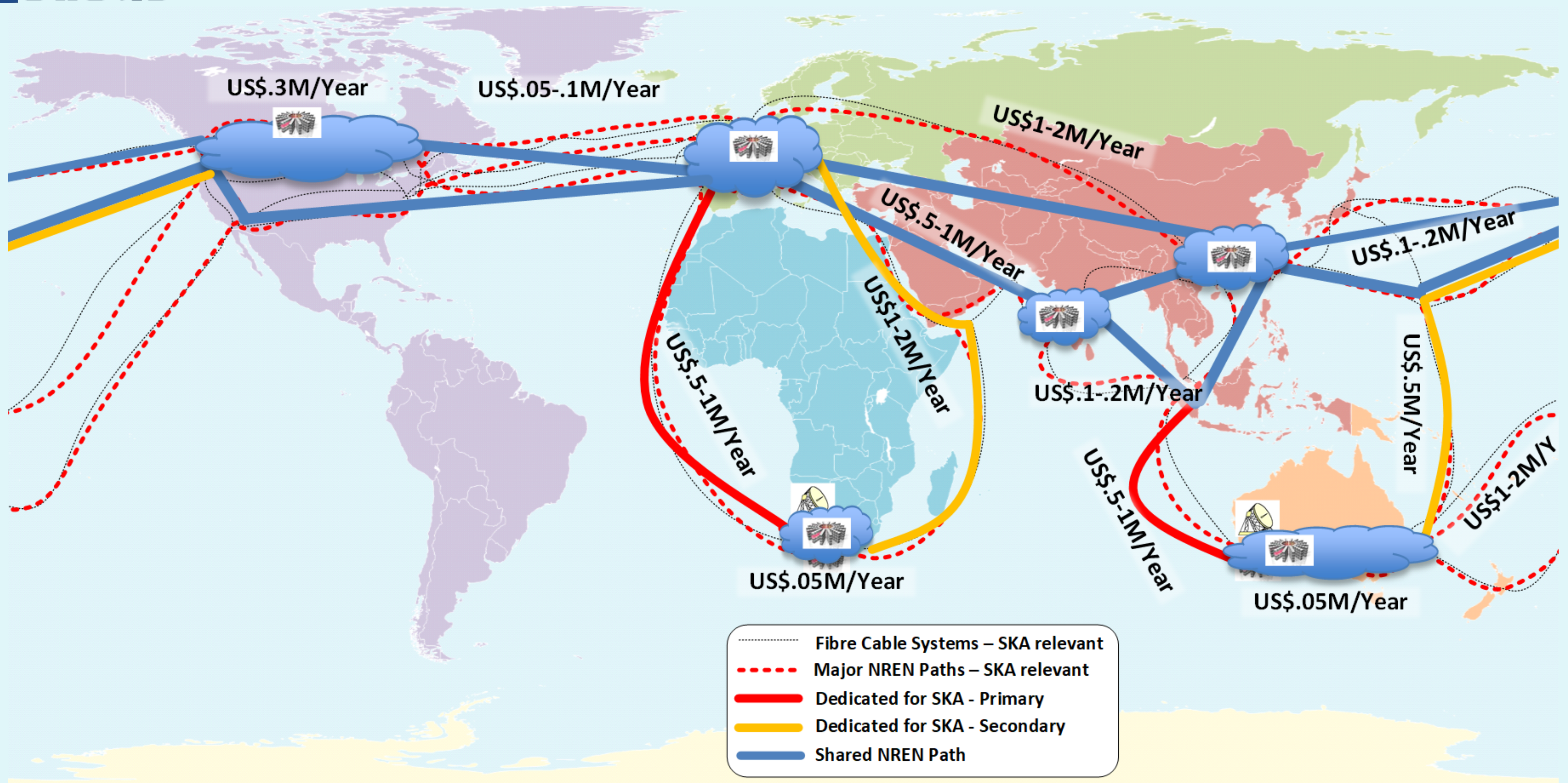


Interactions between SRCs and nodes



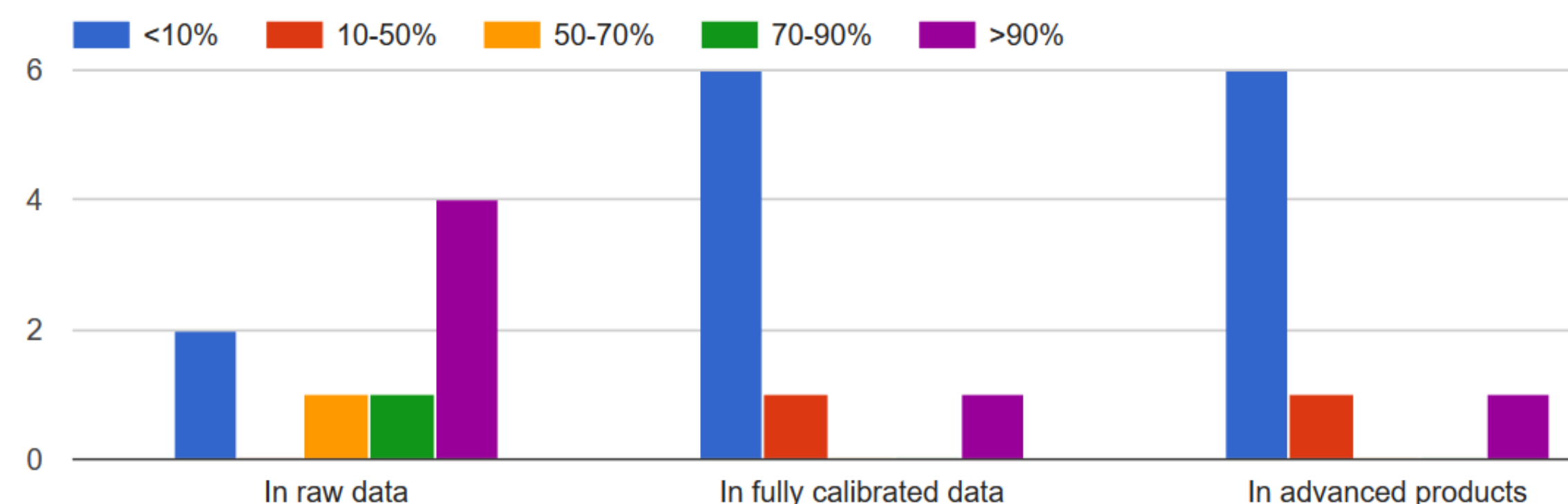
How fraction of the SKA archive will we host in Europe?

Possible Global Network Architecture

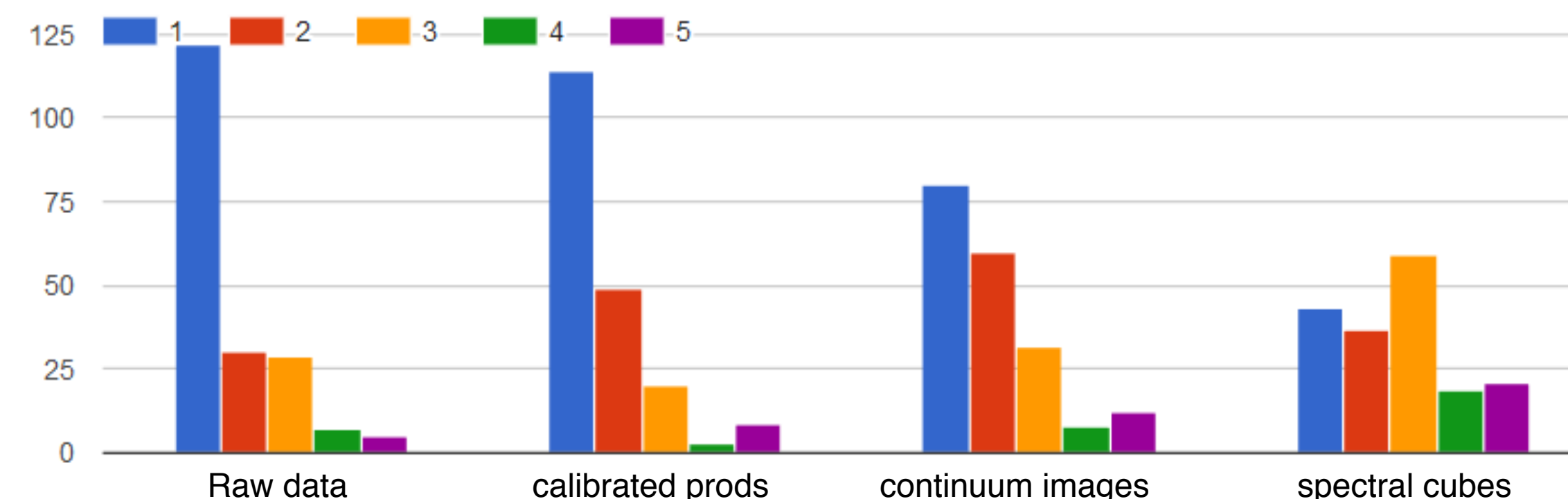


Changing the Way Astronomers Work

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

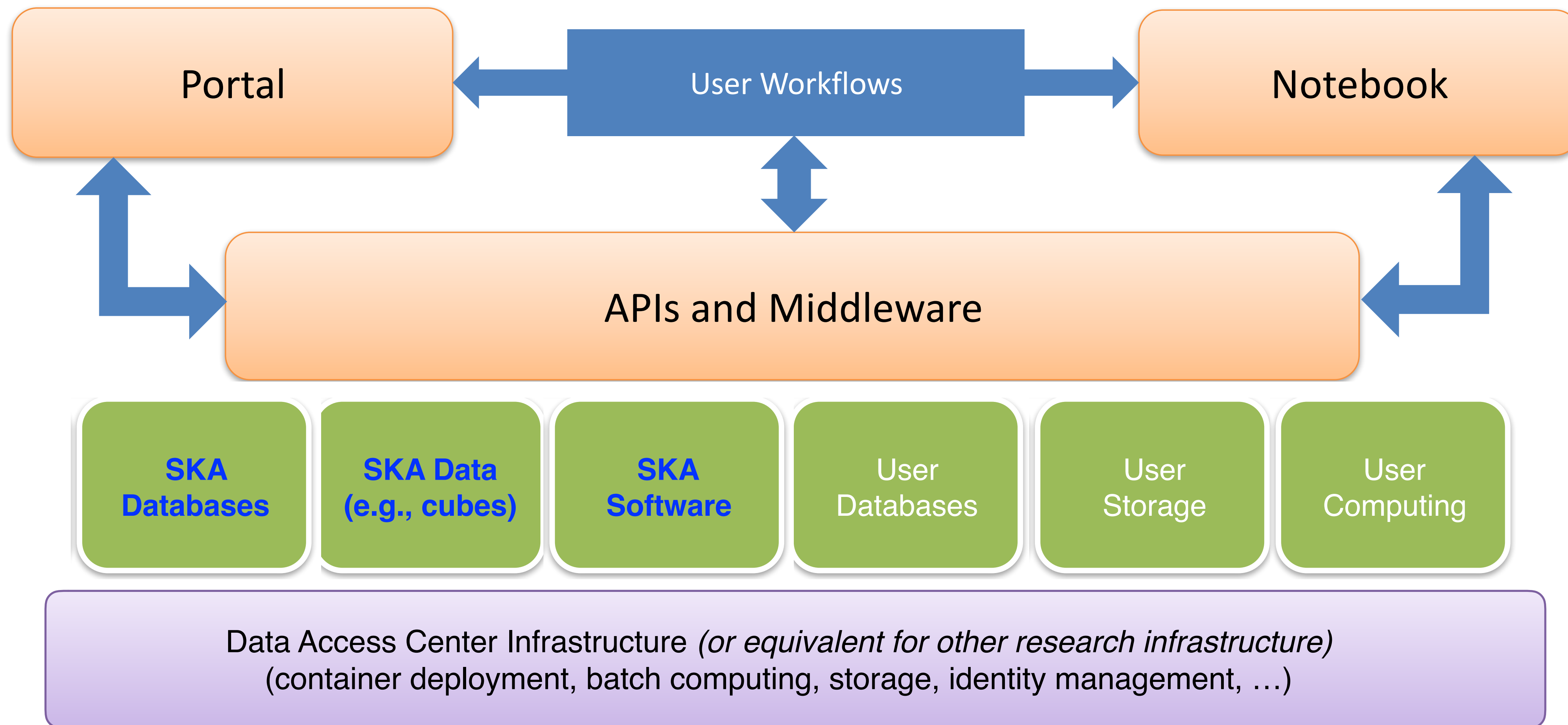


What data products would you like to find in a facility archive (1=necessary, 5=useless)?



⇒ Platform must support creation of advanced products
Users must trust in quality of advanced data products

Science Analysis Platform



(image "borrowed" from LSST Development team)

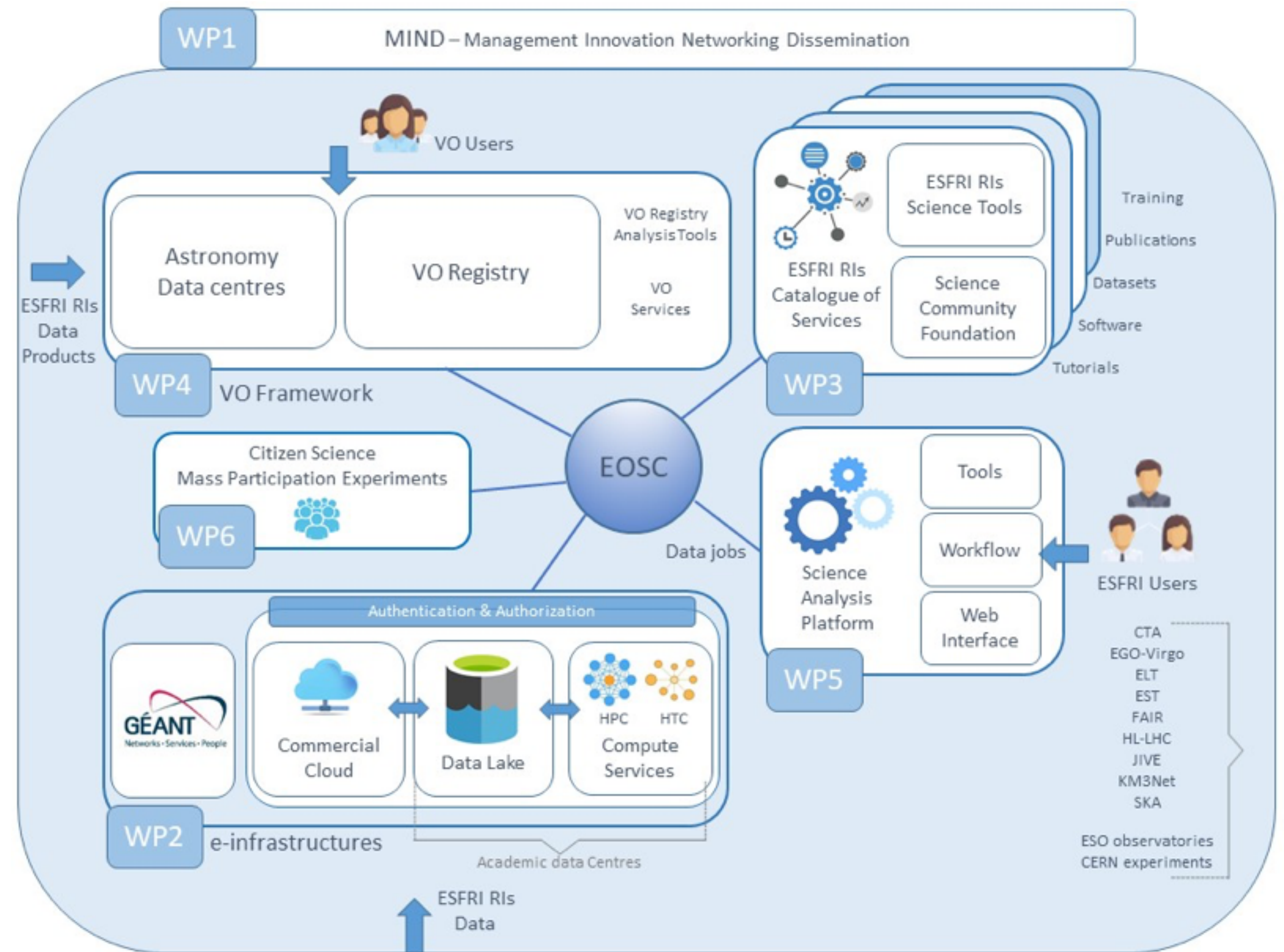


ESCAPE

European Science Cluster of Astronomy & Particle physics
ESFRI research infrastructures

- 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
- CDS leading VO Framework WP
- Work kicks off in January 2019

ESCAPE aims to address the Open Science challenges shared by major European research infrastructures in astronomy and particle physics





ERIDANUS: Asia-Pacific Regional Infrastructure Development

1st Australia-China SKA Big Data Workshop
April 2017, Shanghai



2nd Australia-China SKA Big Data Workshop
April 2018, Shanghai, China



<https://eridanus.net.au>



Exascale Research
Infrastructure
For Data In Asia-Pacific
Astronomy
Using The SKA

Asia-Pacific SRC Prototyping for ASKAP and MWA

PI: Peter Quinn \$4M seed funding received

WORK PACKAGES

1. Data Transfer
2. Pipelines & Algorithms
3. Execution Framework
4. Global Data Management
5. Hardware/Software Co-design
6. System Integration

(slide courtesy of P. Quinn, ICRAR)



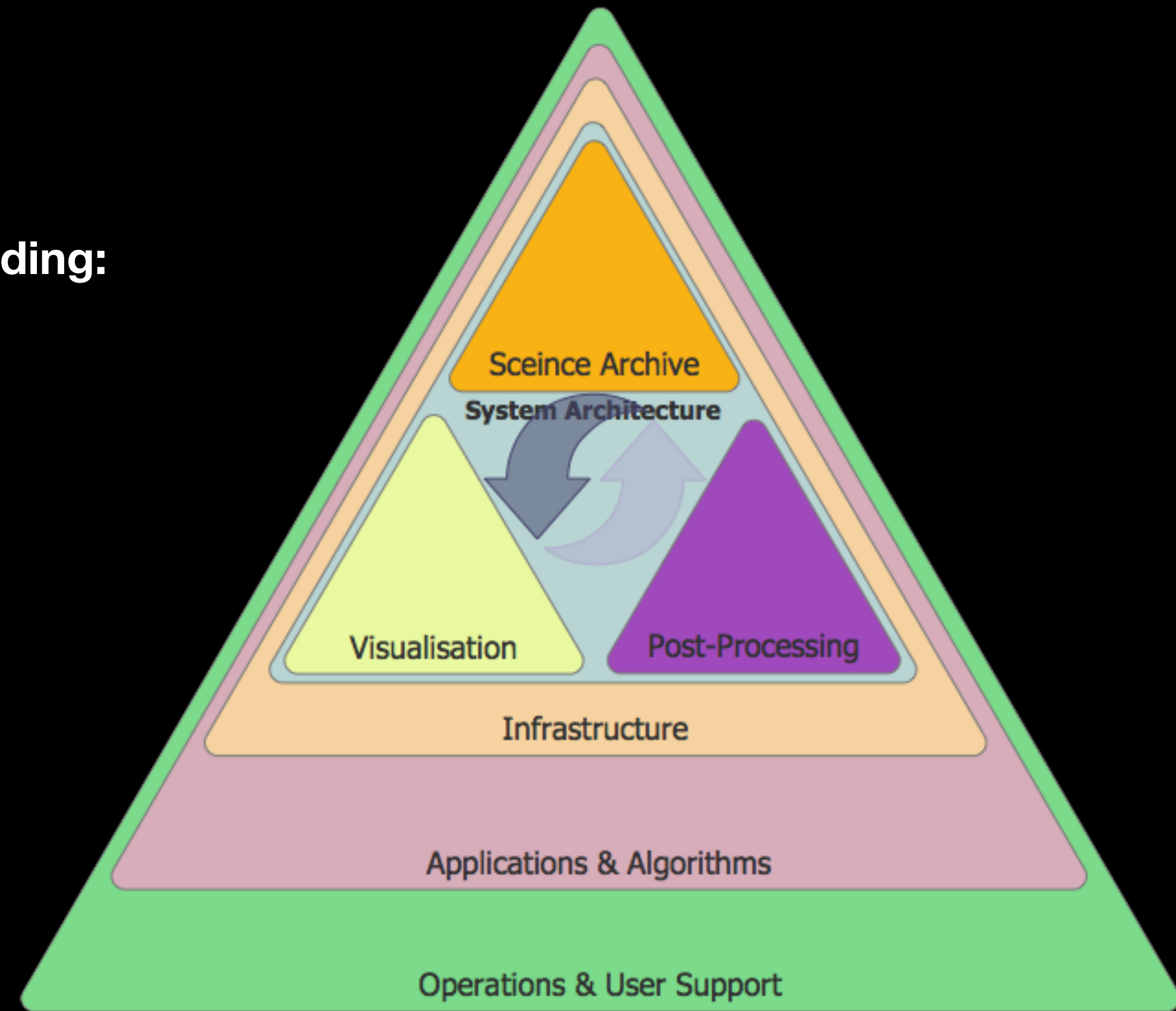
AusSRC Development

Prototype architecture between Pawsey and SHAO deployed including:

- Data transfer and data management tools
- 3 imaging pipelines (ASKAP and MWA) using DaLiuGE
- System monitoring and analysis tools
- Integration & Acceptance framework based on SCRUM agile development
- VO interface to Pawsey & SHAO data archives

Achieved:

- 3 Gbps data transfer Perth (Pawsey) to Shanghai (SHAO)
- Automatic MWA and ASKAP data transfer (using NGAS)
- Automatic pipeline deployment and execution on data arrival
- Comprehensive system monitoring and reports



AusSRC Initiation Project Structure

(slide courtesy of P. Quinn, ICRAR)

Chinese SKA Regional Center Precursor



Shanghai Sheshan Campus (2012-2020)



Current capacity and major equipments:

1. 2x Intel Xeon Phi 7250, each 4nodes, 272cores, 7Tflops, 768GB mem
2. Intel Purley, 4 nodes, 5.6 Tflops, 768GB mem
3. 2x Nvidia GPU, Tesla V100

Prepare (by 2018 Dec) SKA Data Center lab, 200 m²
Phase 0 (by 2020 Dec) SKA China Data center, 1000m²
Phase 2 (by 2025) SKA1 Data Center, 20000 m², according to Shanghai government's arrangement

(slide courtesy of Tao AN, SHAO)

Canadian Initiative for Radio Astronomy Data Analysis

> **CIRADA** (cirada.ca)

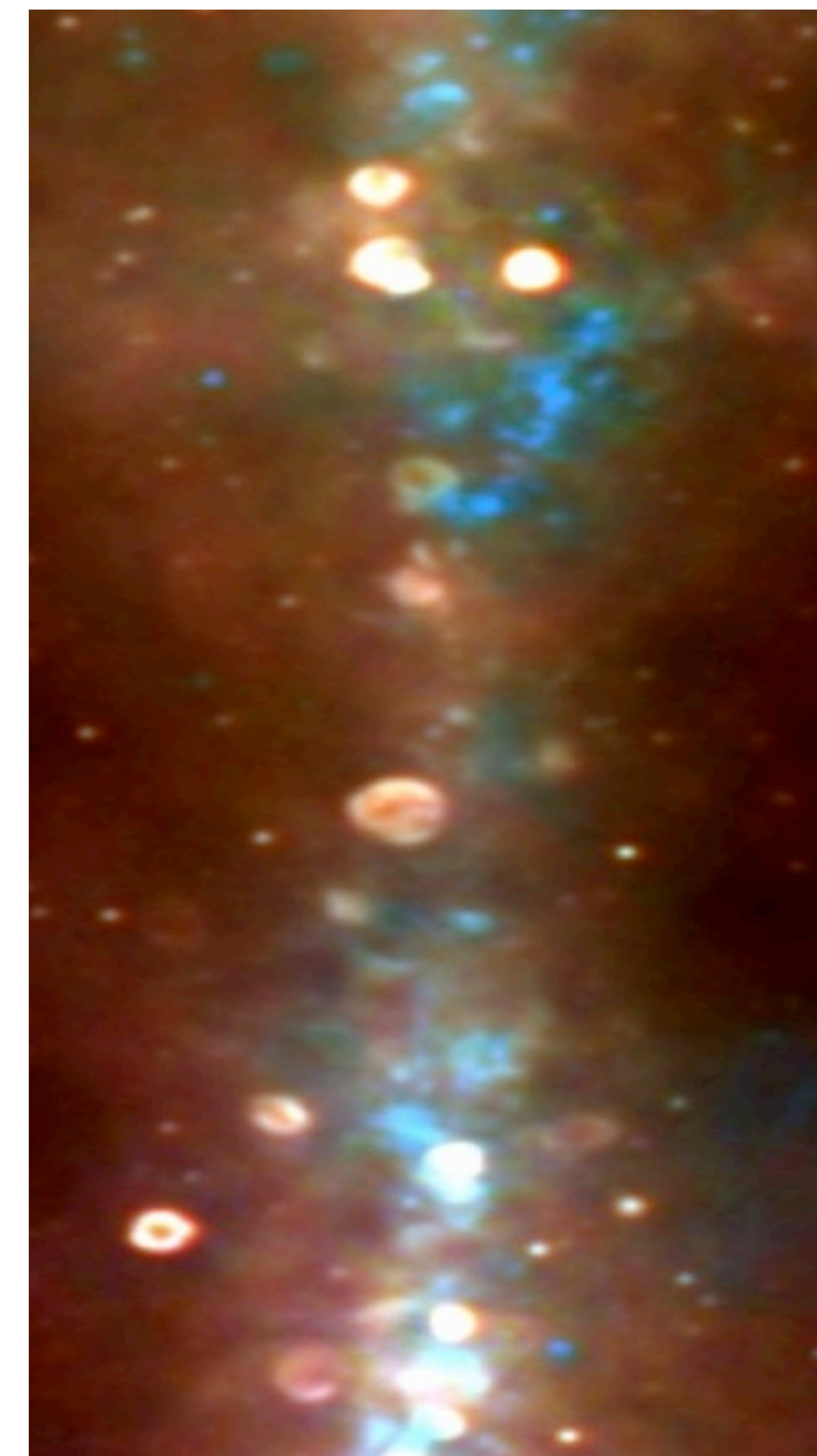
- enhanced data products (EDPs) for VLASS, CHIME and ASKAP surveys
- advanced re-processing of raw data
- unified processing software stack
- cross-matches, advanced analytics, visualisation
- long-term archiving and data access
- ***enables full science return from major Canadian science & instrumentation programs***

> Administrative structure

- CFI Innovation Fund 2017: \$10.6M
- PI: Bryan Gaensler ; Deputy PI: Erik Rosolowsky
- five year program, commenced April 2018
- six Canadian universities: Toronto, Alberta, McGill, Queen's, UBC, Manitoba
- plus NRC/CADC, Compute Canada, NRAO, ASTRON, IDIA, Cornell, Berkeley, Minnesota



(slide courtesy of S. Gaudet, NRC-CNRC)

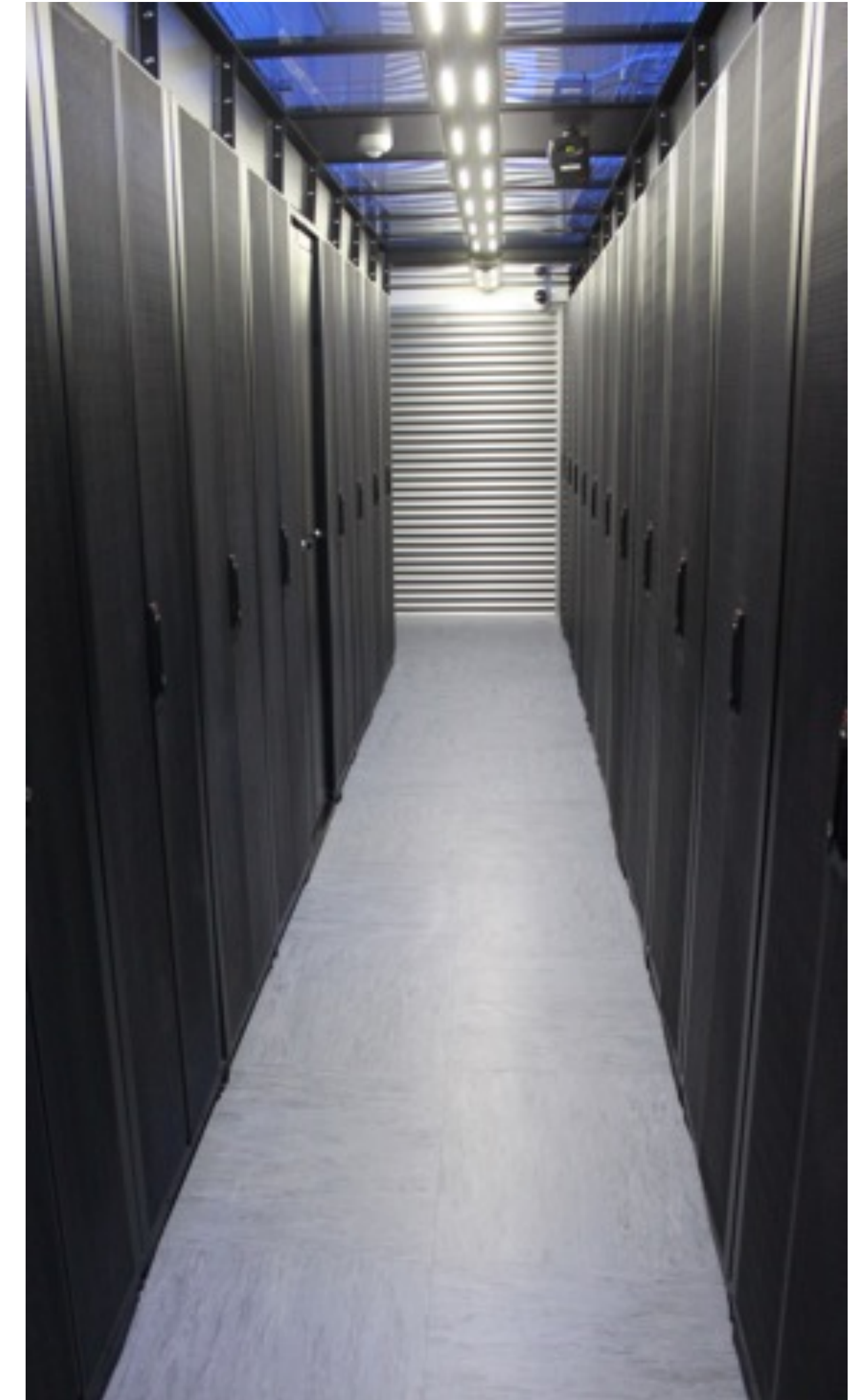
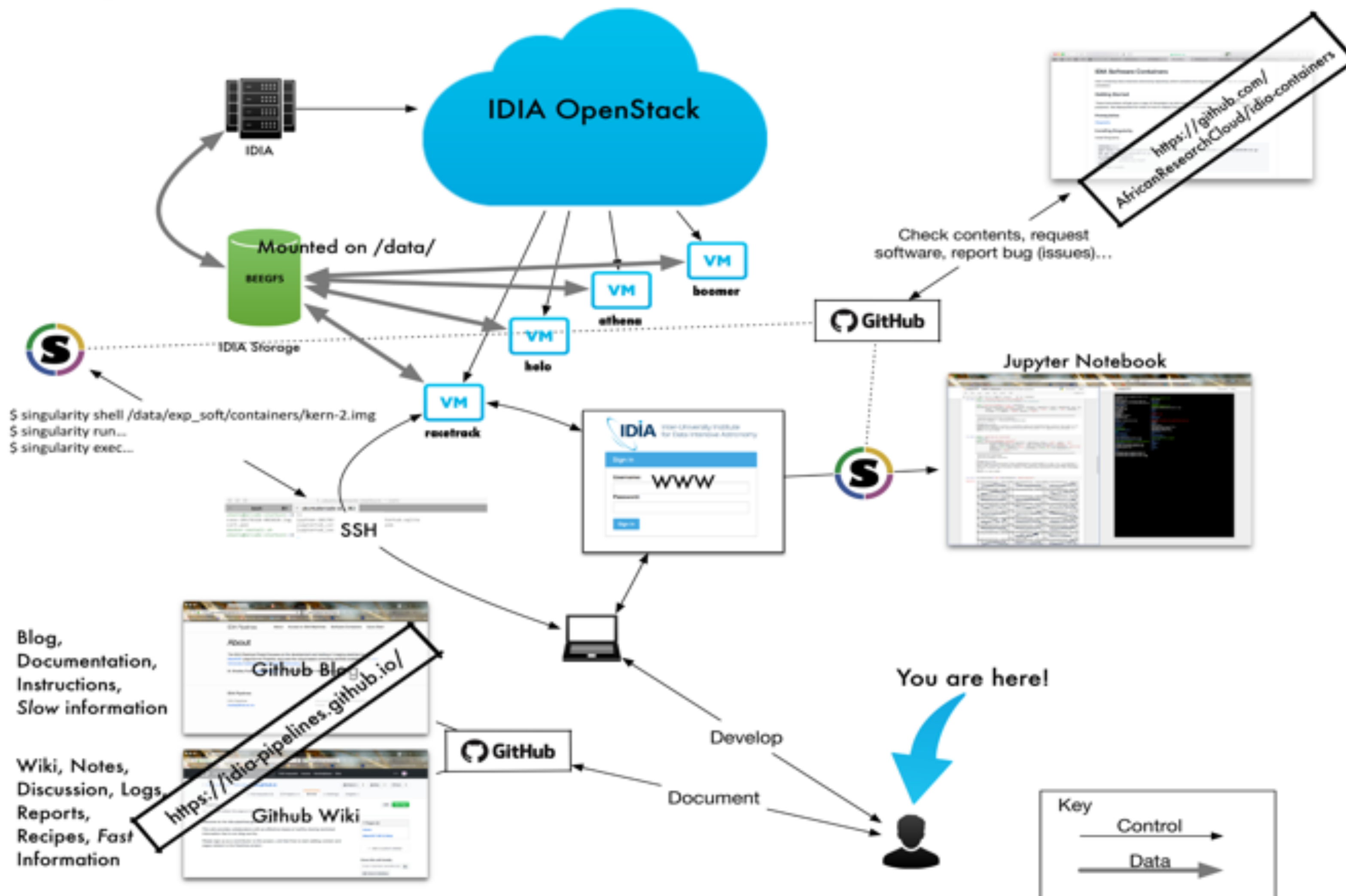


Hindson et al. & Johnston-Hollitt et al.



Institute for Data-Intensive Astronomy

Data-Intensive Cloud: R11M investment

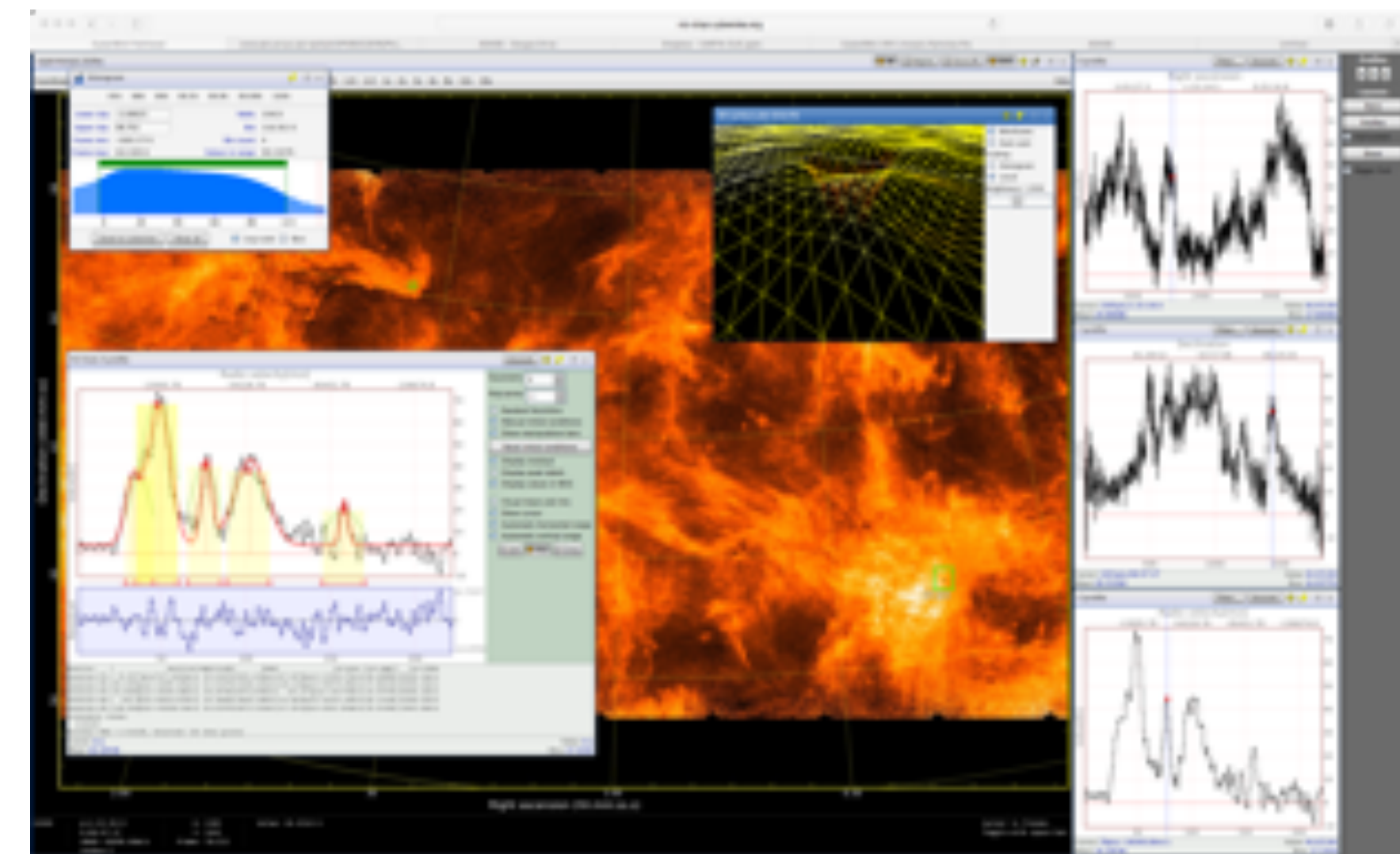
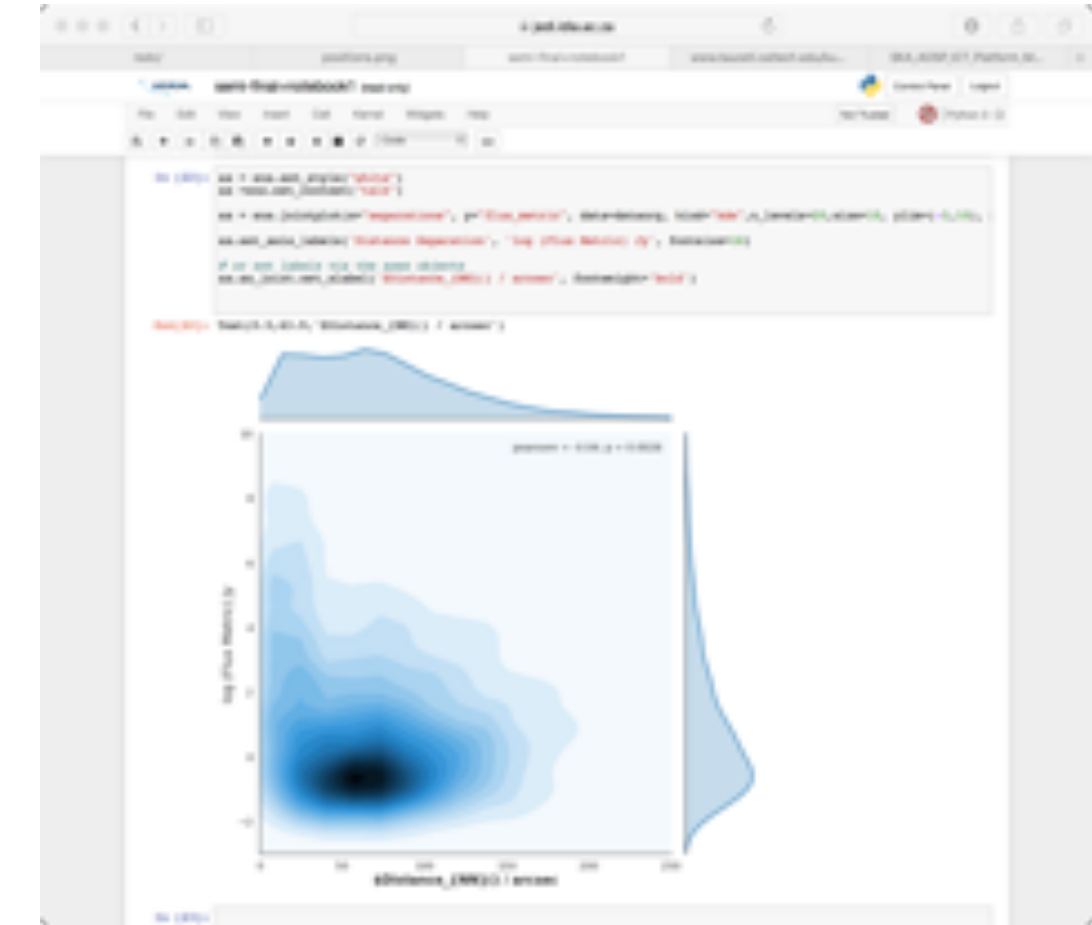


(slide courtesy of R. Taylor, IDIA)



IDIA Cloud-based MeerKAT Large Projects

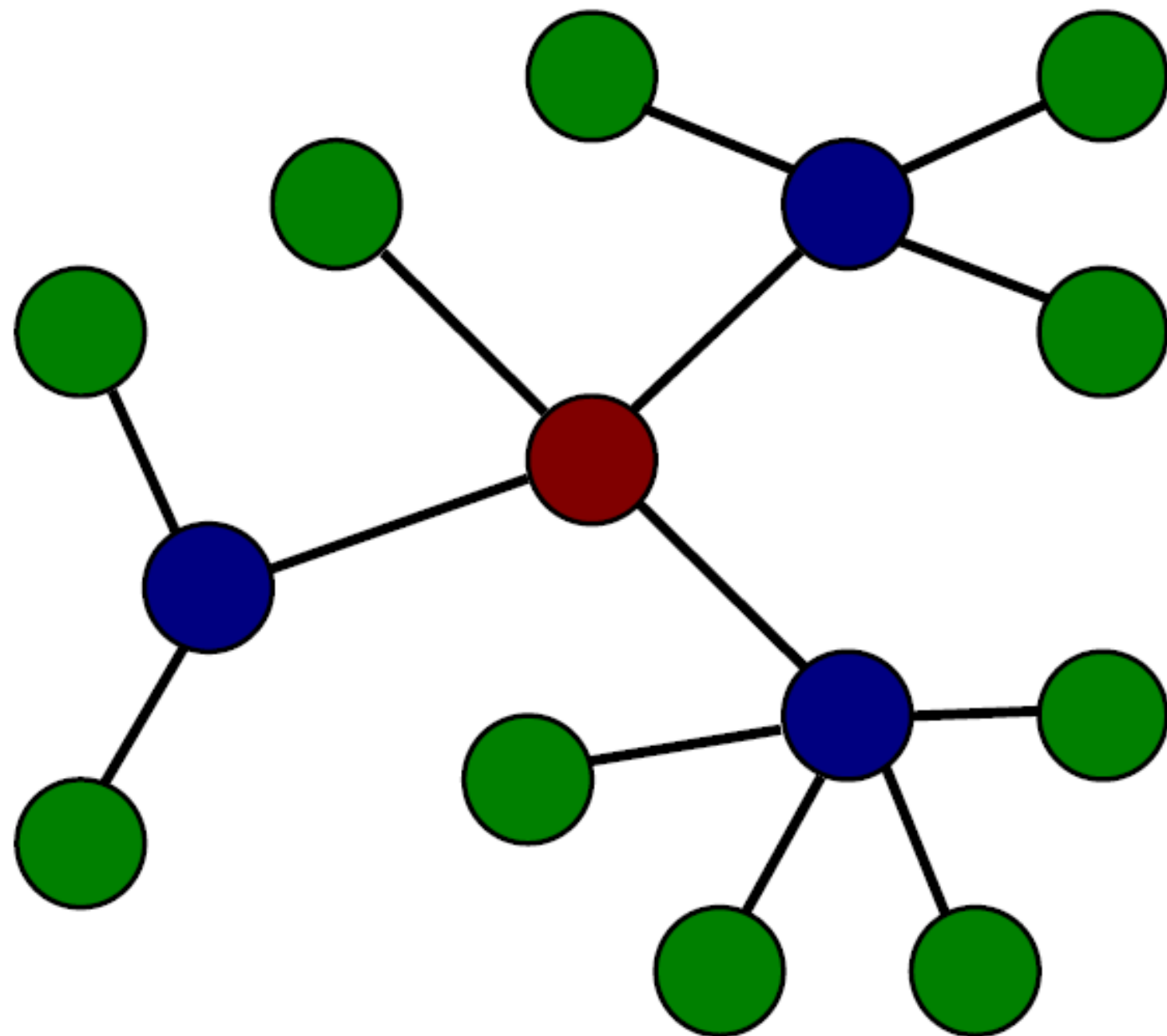
- **Transient IDIA**
 - Processing and joint post-processing analytics for **ThunderKAT** radio and **MeerLICHT** optical observations
- **Pipeline Development for the MeerKAT Imaging Large Survey Projects**
 - Collaboration among **5 MeerKAT LSPs** to jointly develop pipeline processing of MeerKAT data
- **IDIA Visualization Toolkit: Converting Data Into Discoveries**
 - Development of visualization and visual analytics for **MeerKAT big image data sets** and use cases.
- **HIPPO: HELP-IDIA Panchromatic Project**
 - Multi-wavelength data fusion and analysis
 - Machine learning for classification from multi-wavelength data
- **Data Intensive Astronomy with LADUMA**
 - analytics and simulations for **LADUMA** HI science
- **How do Galaxies Form and Evolve**
 - Analytics and simulations for **MONGHOOSE** study of nearby galaxies
- **HI Intensity Mapping**
 - **MeerKLASS** preparatory studies
- **Very Long Baseline Interferometry**
 - Calibration, imaging and analytics of VLBI data sets
- **Open time science projects**
 - E.g. **MHISHAPS, VELA**,...

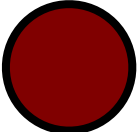
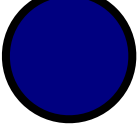
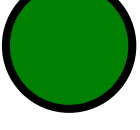


(slide courtesy of R. Taylor, IDIA)

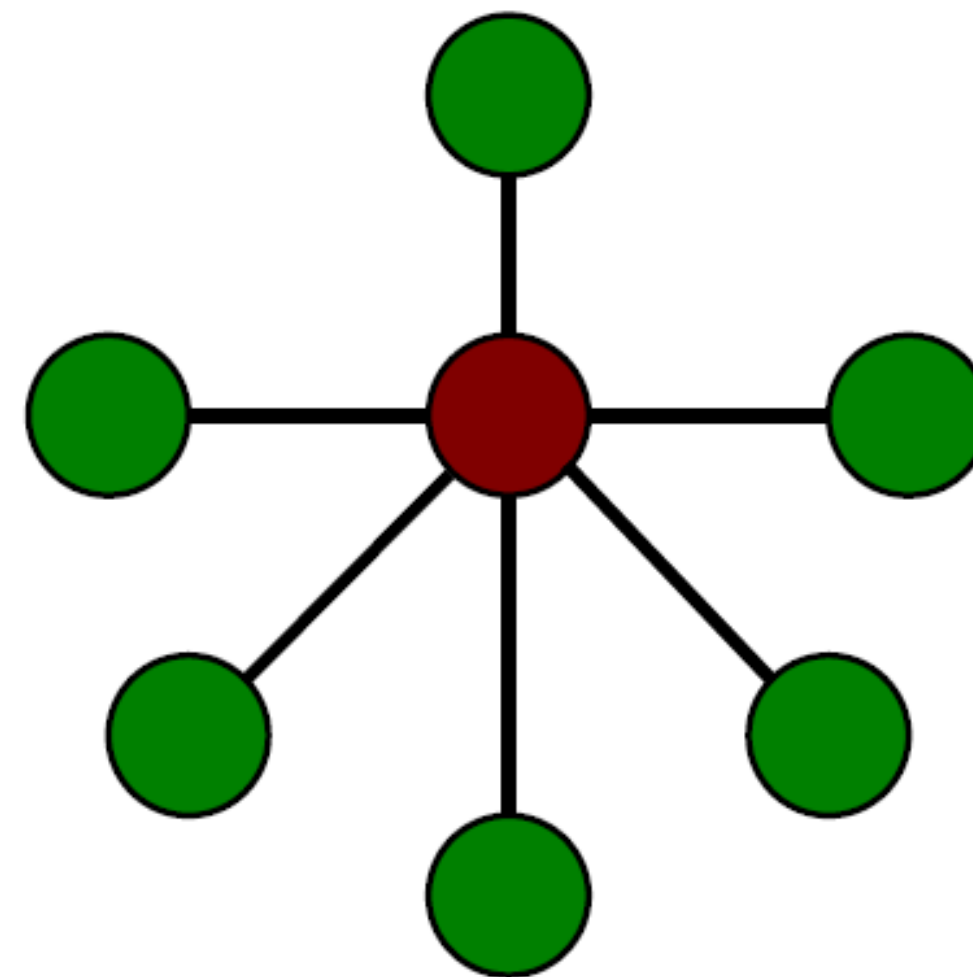
Common Network Topologies

Cluster-Tree Network

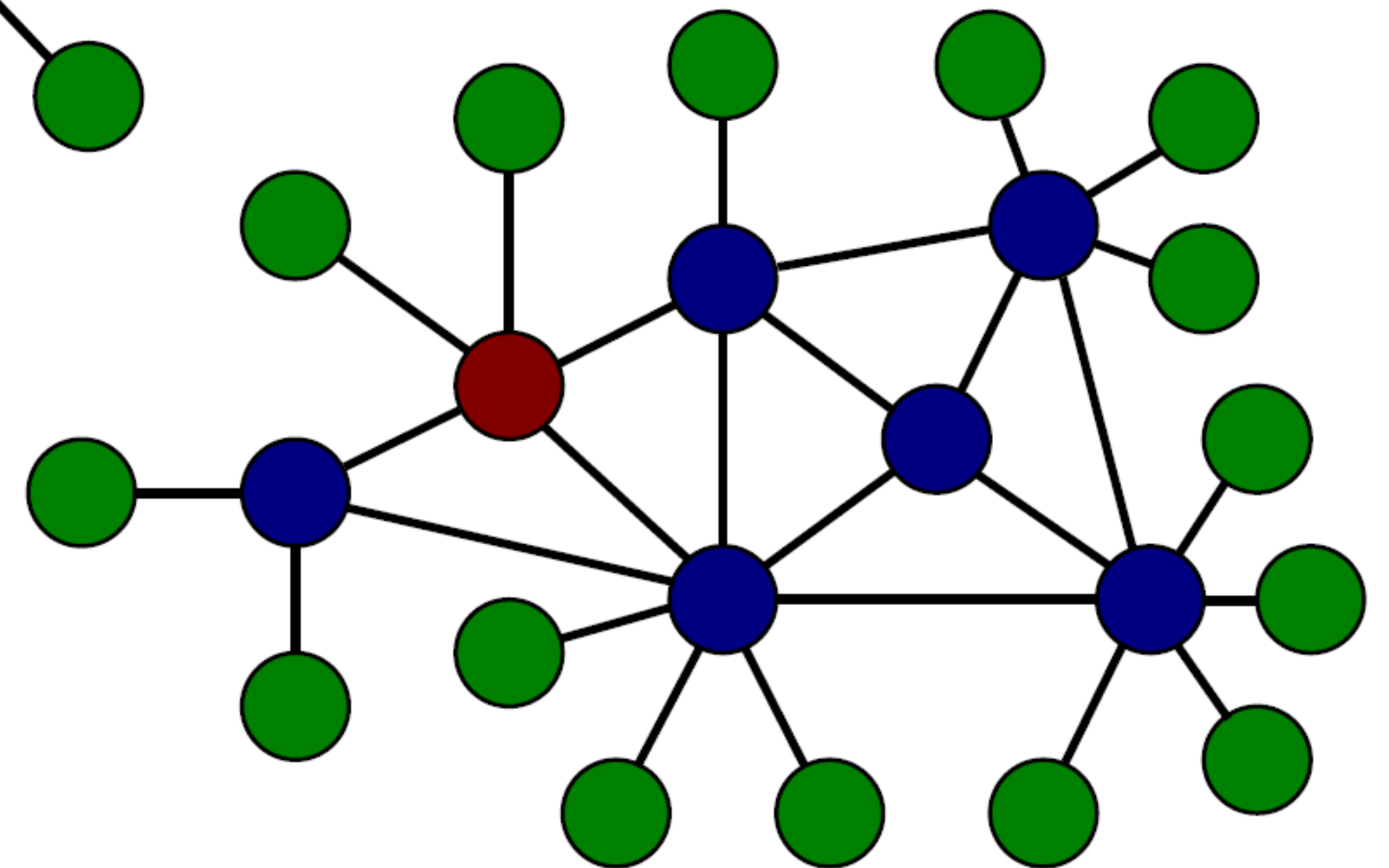


-  Coordinator
-  Router
-  Device

Star Network

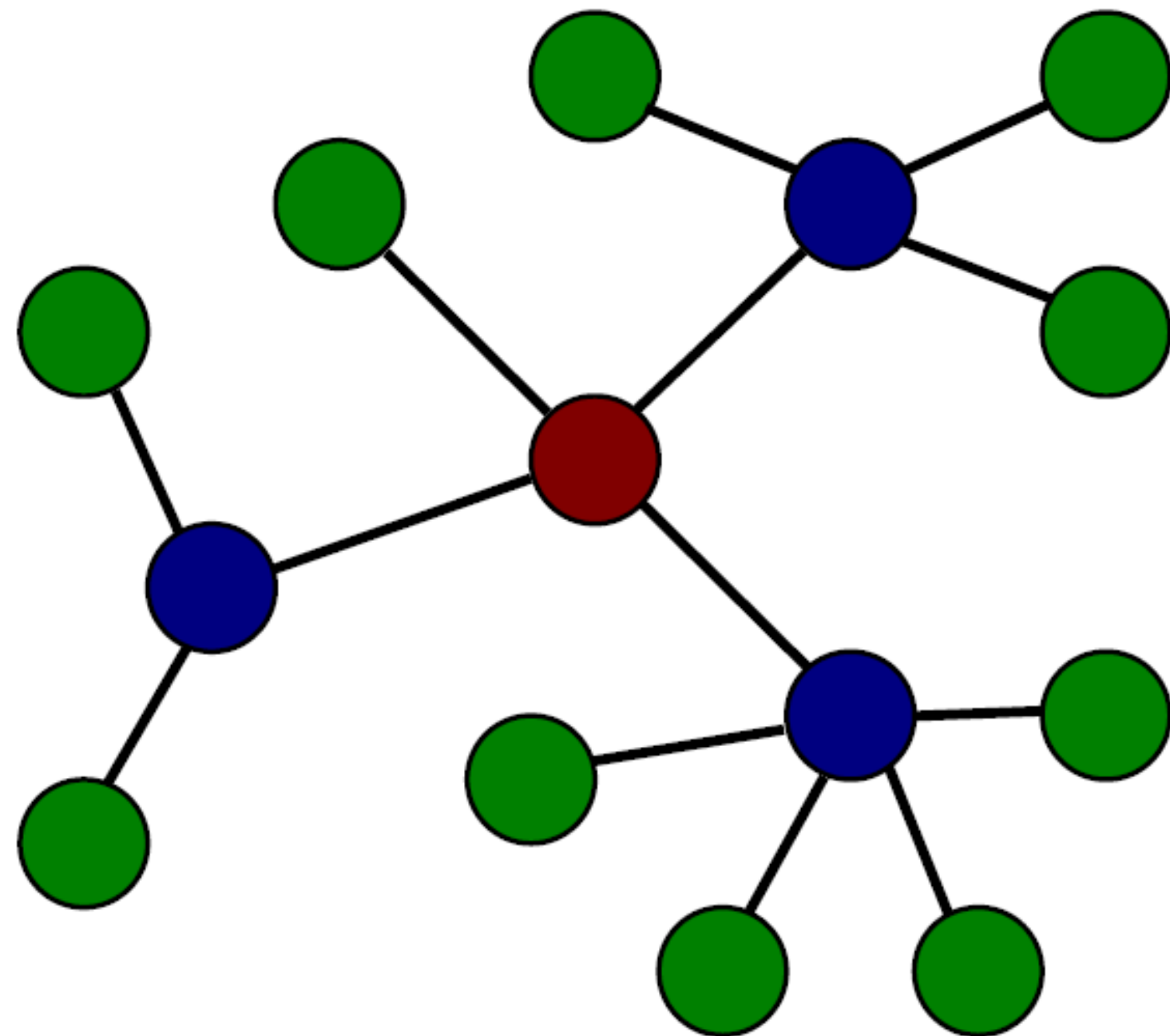


Mesh Network

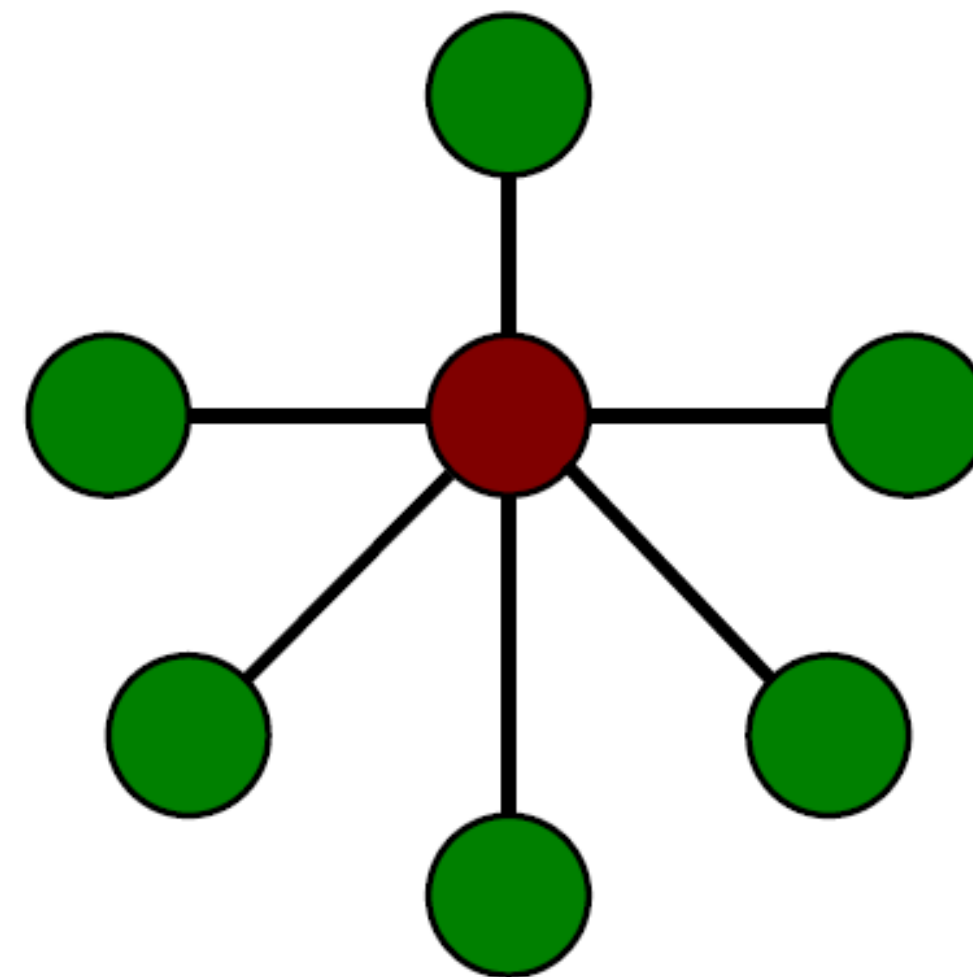


Mesh Management Model

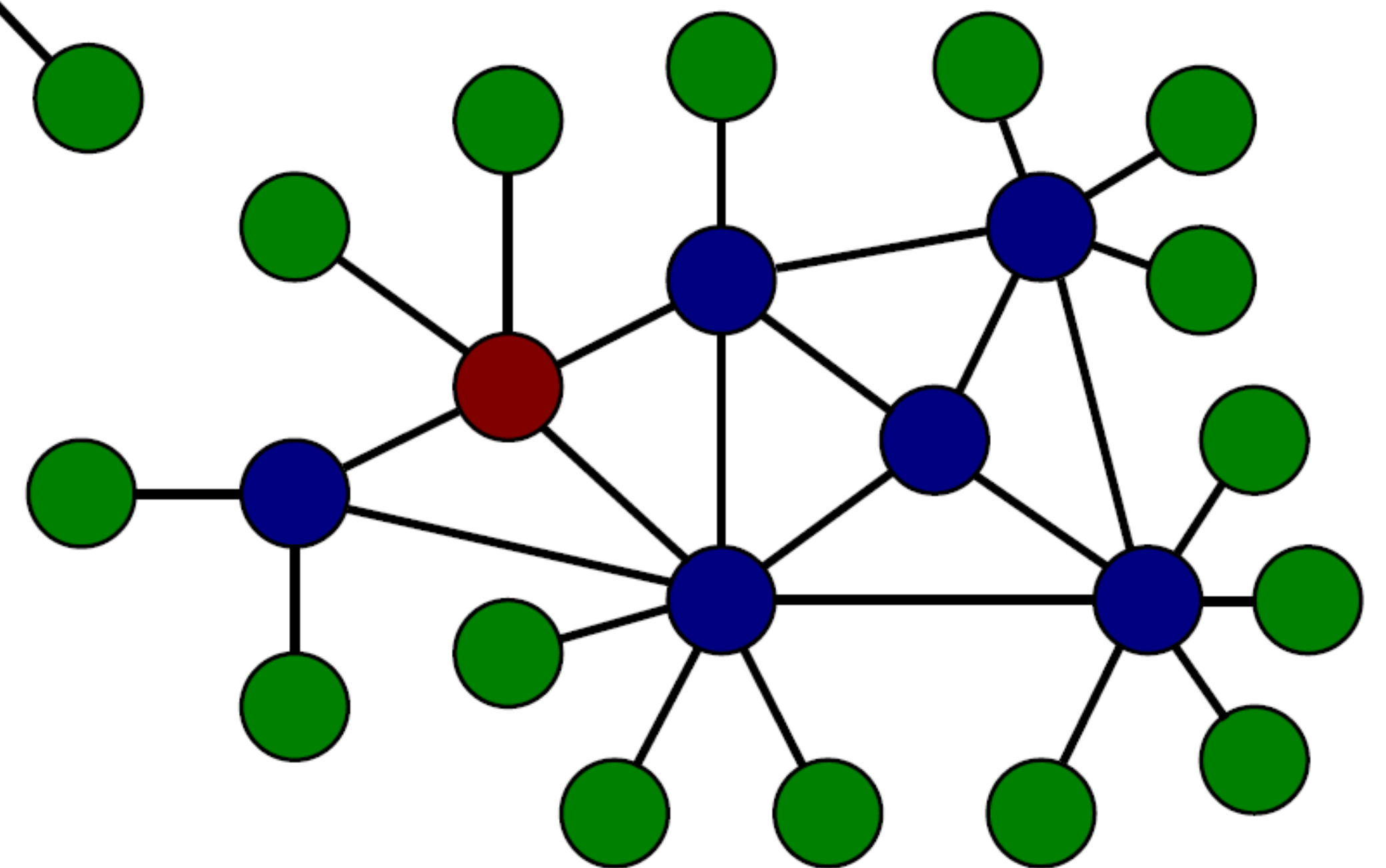
Distributed Support



Direct Support



Fully Connected



-  Observatory
-  Regional Nodes
-  Users

Culture Change

Scientists

- *Limited or no access to raw data*
- *Processing and analysis at the data*
- *Compute becomes part of the proposal*
- *Impact on interactive analysis*

Observatories

- *Increased responsibility for science products*
- *Provide access to HPC and HTC systems*
- *Reliance on external infrastructure*
- *Increased support for custom user analysis*
- *Evolving profiles for support staff*

ADASS

- *Increased focus on distributed, large-scale data analysis*
- *Beyond discovery to staging and large-scale compute*
- *Interoperability of workflows and science platforms*
- *Machine learning, deep learning, and AI (or BlockChain?)*



SQUARE KILOMETRE ARRAY

SQUARE KILOMETRE ARRAY

Thanks for your attention!