Computational challenges in the SKA era

Rossella Cassano

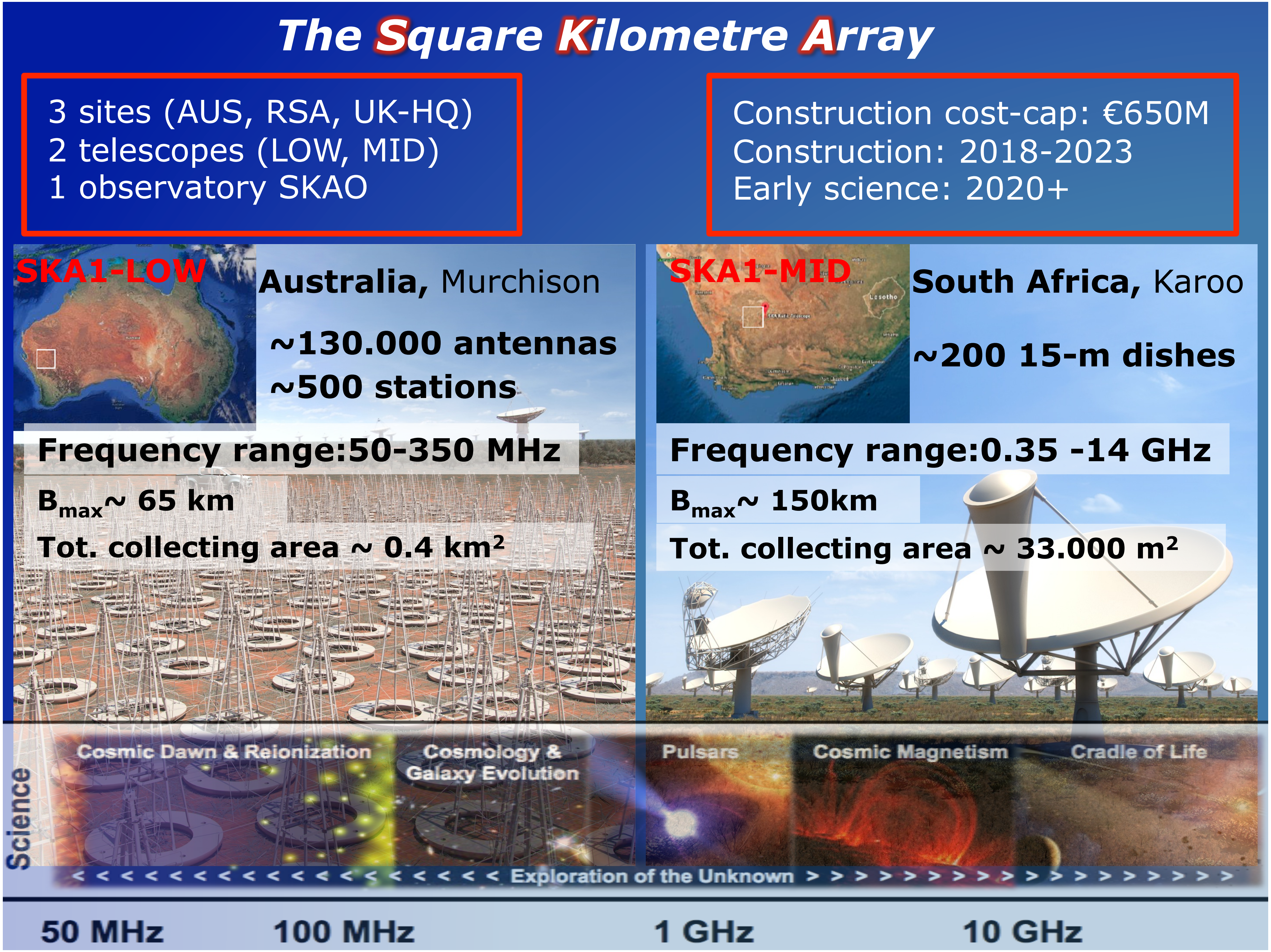
INAF-IRA, Bologna, ITALY

R. Smareglia, S. Tingay, M. Nanni

Credits: SKAO

"What about computing @ INAF ?"

20-21 June 2016, Roma


The Square Kilometre Array
3 sites (AUS, RSA, UK-HQ)
2 telescopes (LOW, MID)
1 observatory SKAO
Construction cost-cap: €650M
Construction: 2018-2023
Early science: 2020+
SKA1-LOW
Australia, Murchison
~130.000 antennas
~500 stations
Frequency range:50-350 MHz
B$_{max}$~ 65 km
Tot. collecting area ~ 0.4 km²
SKA1-MID
South Africa, Karoo
~200 15-m dishes
Frequency range:0.35 -14 GHz
B$_{max}$~ 150km
Tot. collecting area ~ 33.000 m²
Science
Cosmic Dawn & Reionization
Cosmology & Galaxy Evolution
Pulsars
Cosmic Magnetism
Cradle of Life
< < < < < < < < < < < < < < < < < < < Exploration of the Unknown >
50 MHz
100 MHz
1 GHz
10 GHz

The Square Kilometre Array

How will SKA1 be better than today's best radio telescopes?

Astronomers assess a telescope's performance by looking at three factors - **resolution**, **sensitivity**, and **survey speed**. With its sheer size and large number of antennas, the SKA will provide a giant leap in all three compared to existing radio telescopes, enabling it to revolutionise our understanding of the Universe.

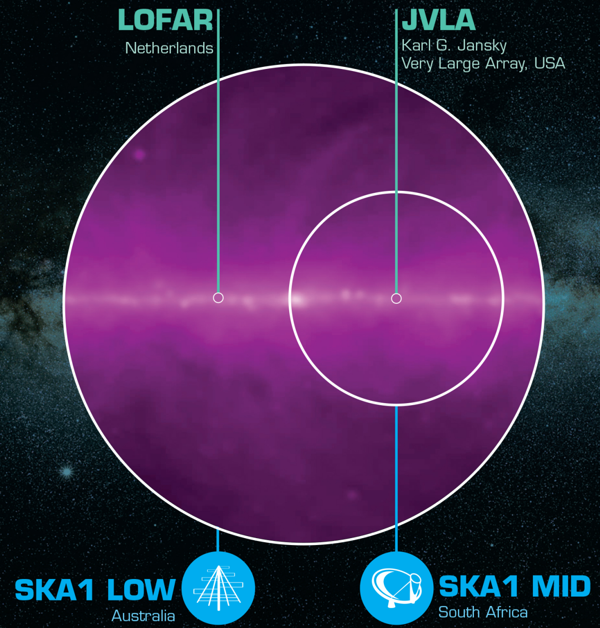
The **Square Kilometre Array** (SKA) will be the world's largest radio telescope. It will be built in two phases - SKA1 and SKA2 - starting in 2018, with SKA1 representing a fraction of the full SKA. SKA1 will include two instruments - **SKA1 MID** and **SKA1 LOW** - observing the Universe at different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

SKA1 LOW **x1.2** LOFAR NL

SKA1 MID **x4** JVLA

RESOLUTION

Thanks to its size, the SKA will see smaller details, making radio images less blurry, like reading glasses help distinguish smaller letters.

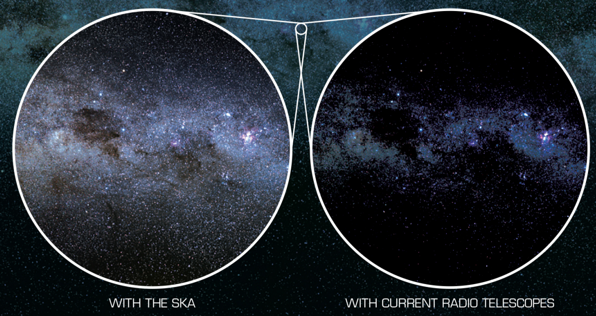


SKA1 LOW **x135** LOFAR NL

SKA1 MID **x60** JVLA

SURVEY SPEED

Thanks to its sensitivity and ability to see a larger area of the sky at once, the SKA will be able to observe more of the sky in a given time and so map the sky faster.



SKA1 LOW **x8** LOFAR NL

SKA1 MID **x5** JVLA

SENSITIVITY

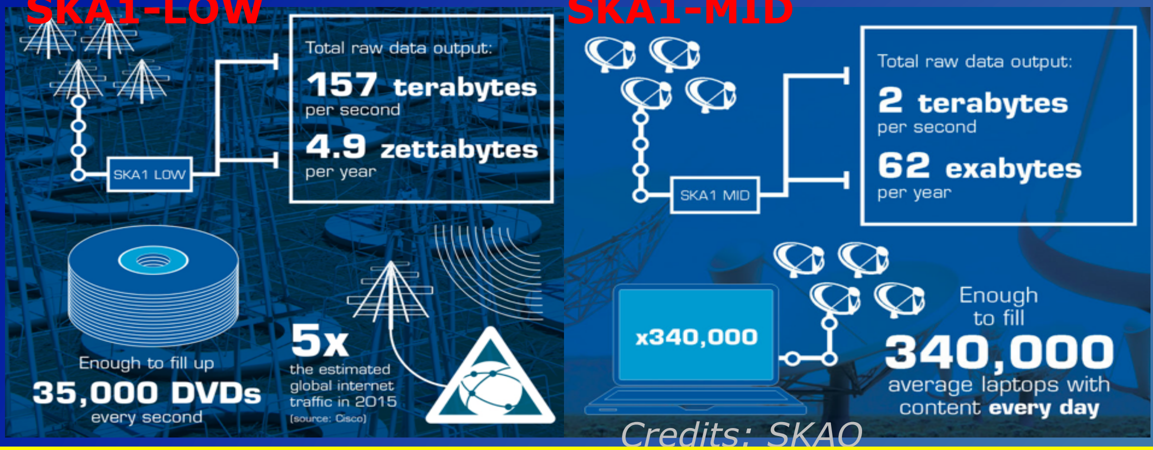
Thanks to its many antennas, the SKA will see fainter details, like a long-exposure photograph at night reveals details the eye can't see.

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

As the SKA isn't operational yet, we use an optical image of the Milky Way to illustrate the concepts of increased sensitivity and resolution.

SKA1 Data Challenges

✓ Estimated total raw data rate



✓ Estimated Power: **300 PetaFlops**
- @ Cineca new supercomputer **Marconi**, will reach **20 PetaFlops** and storage capability of 20 PetaByte within 2017
- the fastest supercomputer in the world (the China's **Tianhe-2)**, runs at **33.86 PetaFlops**,

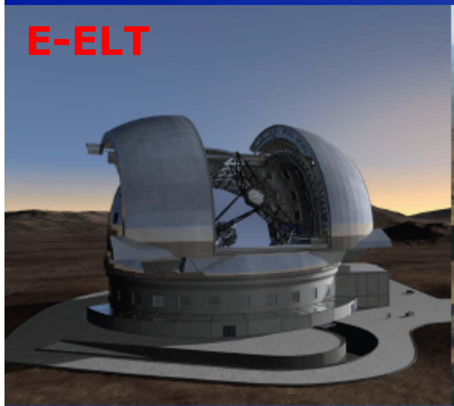
SKA will need a computer between <u>six and 10 times</u> faster than the fastest machine on earth.

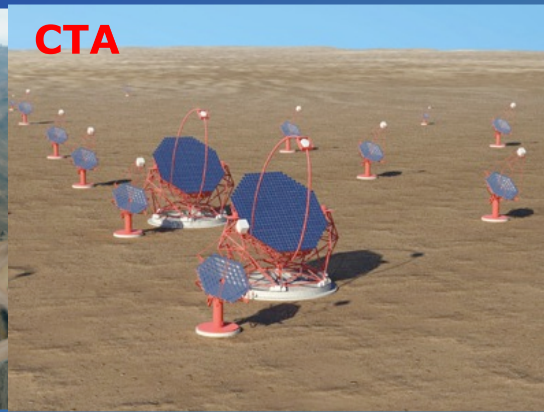
Regional Science and Data Center(?)

✓ Estimated Storage capability
- the SDP consortium estimates **50-300 Petabytes per year** of standard data product to be archived (no raw data)

✓ Signal Transport & Networks
-**160 Gbit/s** to transmit data from each radio dish to a central processor=> the high frequency dishes will produce *ten times the current global internet traffic!*
-**10-100 Gbit/s** to transmit the processed data to the international community

Data Intensive Astrophysics





- ✓ astrophysics is entering in the era of *big-data*
- ✓ massive data collections
- ✓ most science extraction is based on the archived data
- ✓ current instruments already producing petascale datasets

Evolution of Science Extraction

Desktop Systems
(simple tools)

SKA pathfinder
(Europe)

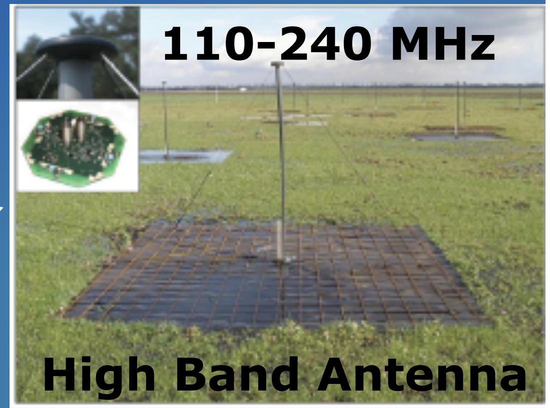
Moderate Cluster
(pipelines)

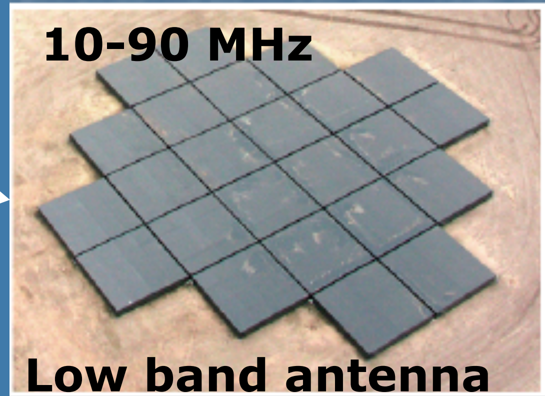
SKA precursor
(Australia)

Science Data Centres
(smart systems)

In the meantime…LOFAR (LOw Frequency Array)

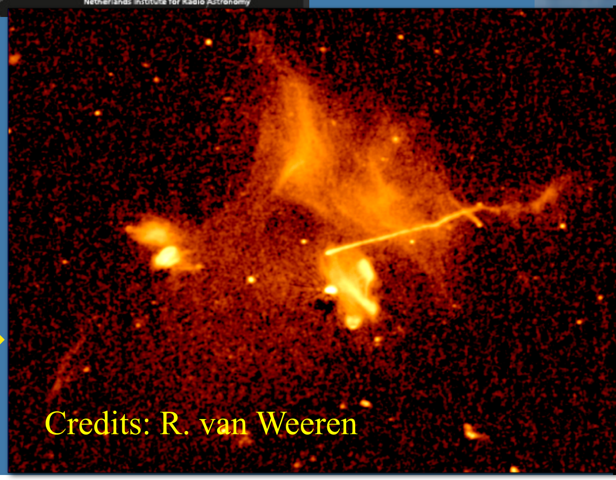
- World's largest radio telescope
- unprecedented resolution and sensitivity at ν~15-250 MHz
- wide field of view => excellent for surveys

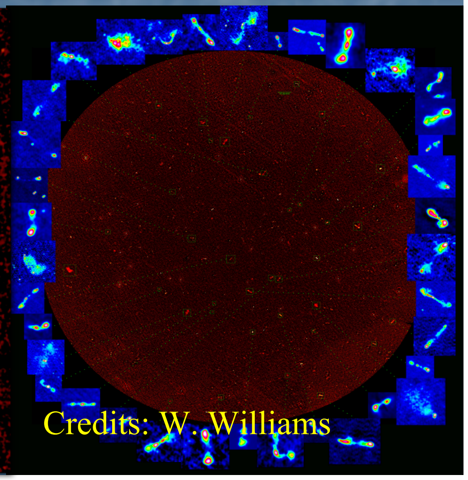




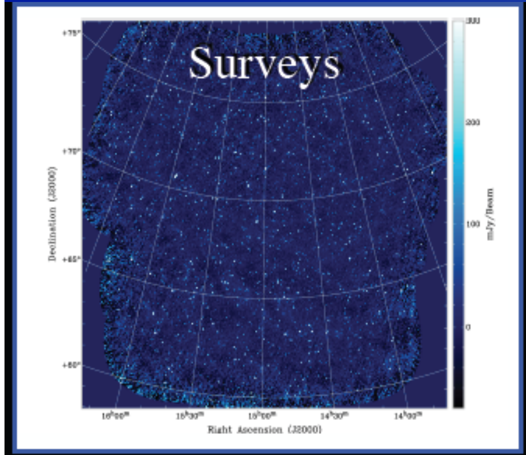
51 stations in European countries
38 in Netherland, **6** in Germany,
3 in Poland and **1** in France, UK,
Sweden and Ireland

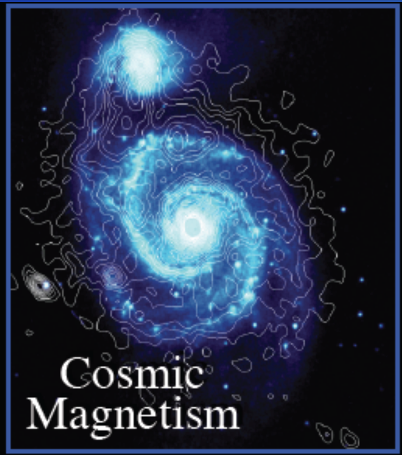
LOFAR first results

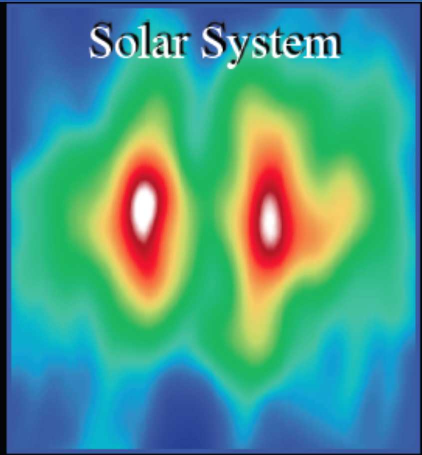


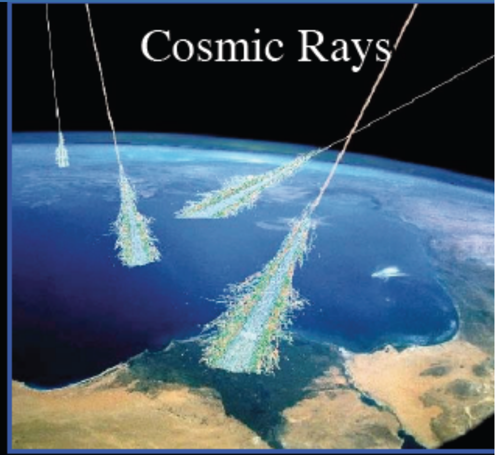


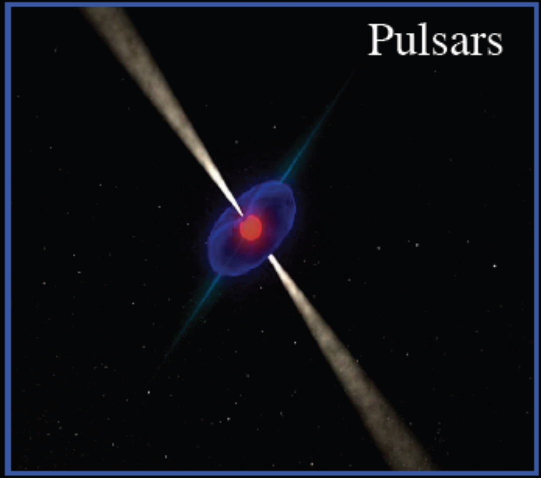
LOFAR Science Key Project

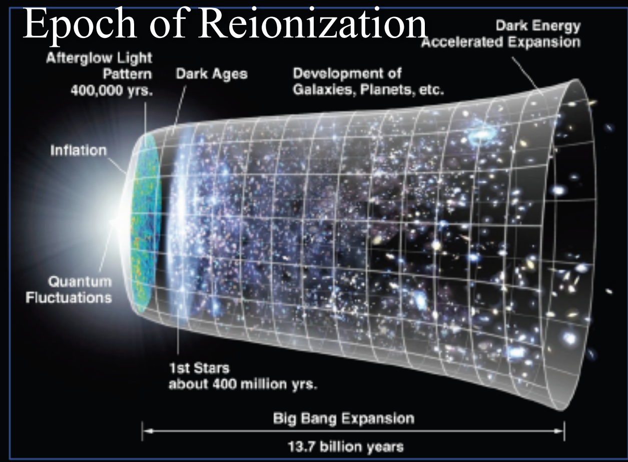






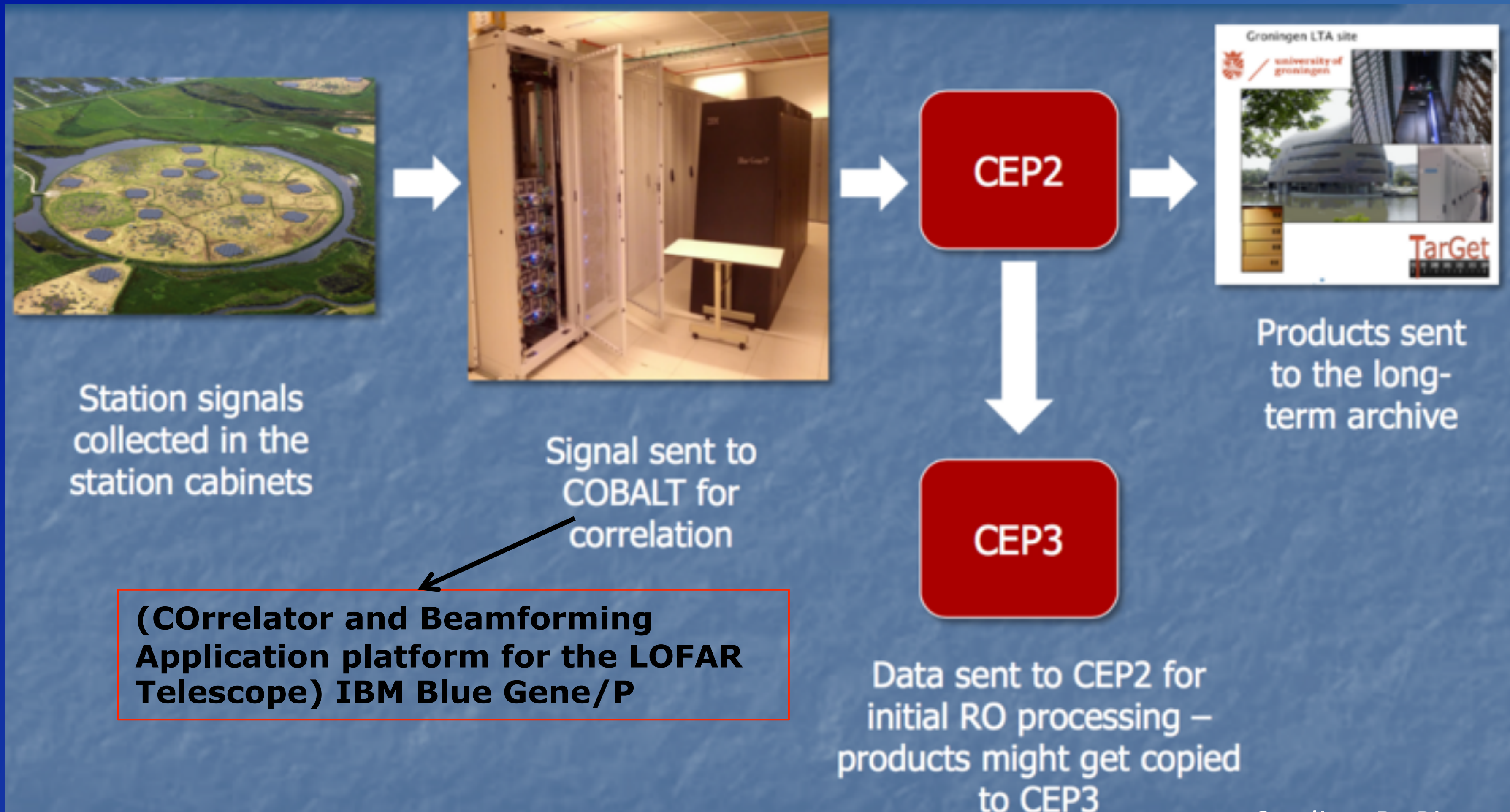






Credits: M. Wise

The LOFAR System: data flow



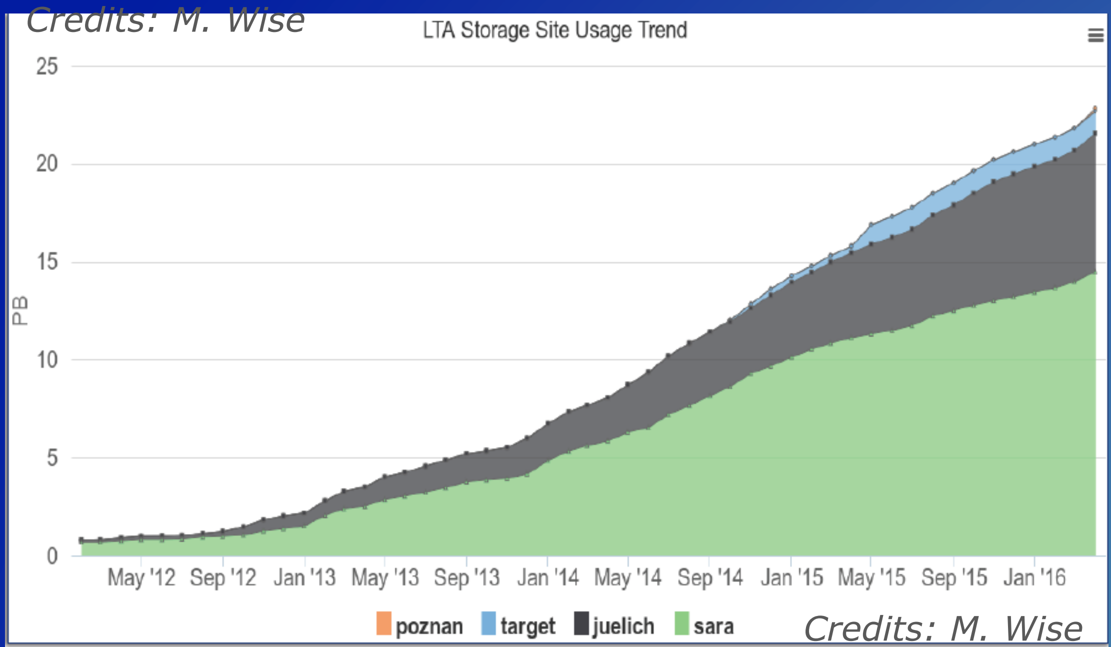
Credits: R. Pizzo

✓ Large data transport rates => data storage challenges (35 TB /h)

✓ LOFAR is the first of a number of new astronomical facilities dealing with the transport, processing and storage of these large amounts of data and therefore represents an important technological pathfinder for the SKA

LOFAR Data Accumulation

LOFAR Long-Term Archive (LTA) is federated over 4 locations, 3 countries



Data Storage

- ✓ 23 Petabytes
- ✓ 3 PB/yr grow
- ✓ 300 TB/month ingest
- ✓ 100 TB/month staged

Contents

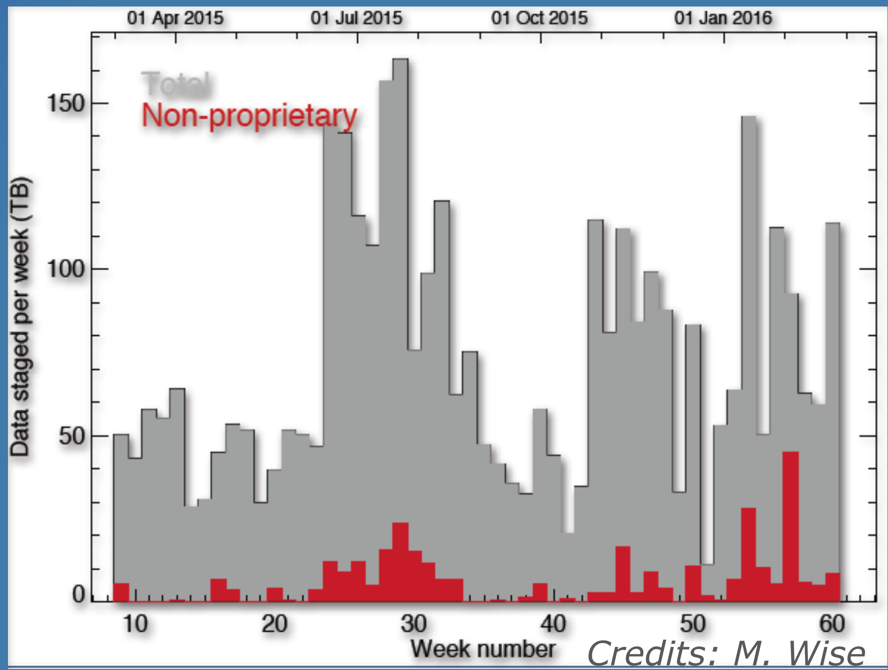
- ✓ 10^7 products
- ✓ visibilities, images, BF data
- ✓ does not include raw visibility

Typical data size is 10-100 Tb

Problematic for many researchers!

→ Data transfer from archive to institutes too slow: ~ 10 Mb/s

→ Current Processing/ Observing for a single observation too high: 10-100



SKA Science Archive

PER YEAR

• 1 Petabyte

Credits: M. Wise

SKA Regional Science and Data Centres

RSDC will likely host the SKA science archive

Provide access and distribute data products to users

Provide access to compute and storage resources for users

Provide analysis capabilities

Multiple national RSDC, locally resourced

Provide user support

RSDC will be the primary interface for astronomers to extract science from SKA data!

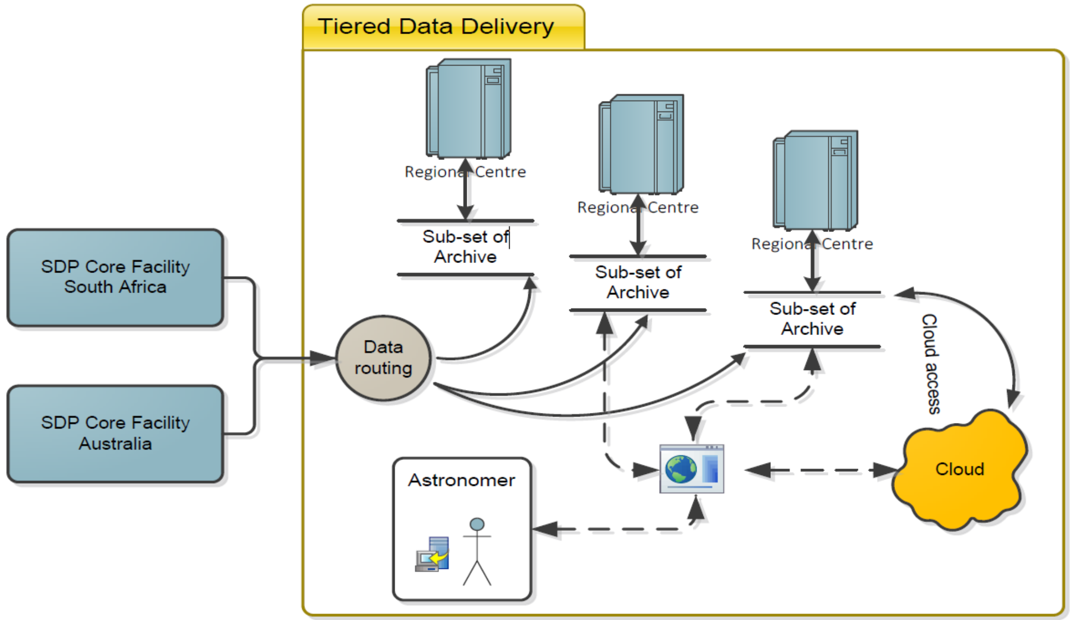


Figure 1: Schematic illustrating the connections between the proposed Regional Science & Data Centres and the SKA primary archive sites. Note the RSDCs are assumed to host subsets of the full SKA science archive.

The **MWA** has produced **>10 PB** of data and distributes from Perth to Canada, India, New Zealand, USA, and Australia.

So, MWA is a good example of what SKA will have to do - distribute data from Australia/South Africa to the SKA partner countries around the world

The AENEAS Project

Aeneas

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

Credits: M. Wise

Design and specification of a distributed, European Science Data Centre (ESDC) to support the pan-European 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)

✓WP2: ESDC Governance Structure and Business Models *(van Haarlem ,ASTRON)*
✓WP3: ESDC Computing + Processing Requirements *(Scaife (Mach.Uni),Bolton (Camb.Uni))*
✓WP4: SKA Data Transport and Optimal European Storage *(Capone,Hugh-Jones (GEANT))*
✓WP5: User Data Access and Knowledge Creation *(**Steven Tingay (INAF**))*

The **ALMA Regional Centres (ARCs)** is an excellent examples of the distributed data centre model working for radio astronomy in Europe. The AENEAS project is inspired by the success of the ARC model and will likely seek to scale the ALMA model up to satisfy SKA user requirements.

Conclusions

- Astronomy is entering in the era of Big Data (LSST, E-ELT...LOFAR, SKA)

- The Square Kilometre Array (SKA) Project is the biggest science project on the Earth and will be an ICT-driven science facility

[raw data output => **5 Exabytes per day**
computational power=> **300 PetaFlops**
data storage=> **50-300 Petabytes per year**]

SKA will be a huge data and computational challenge

LOFAR, MWA are already dealing with the transport, processing and storage of large amounts of data and therefore represent important technological pathfinders for the SKA

The Italian community has a science involvement in the SKA precursors/pathfinders (*MeerKAT, ASKAP, MWA, LOFAR*). We need to understand which kind of resources INAF scientists need to support the use of these 4 facilities in Italy. These will be a big learning step toward SKA/AENEAS etc, putting us in a position to do precursor science now.