
ASTRON
Netherlands Institute for Radio Astronomy

Data-Intensive Radio Astronomy with LOFAR and SKA

John Swinbank
swinbank@astron.nl

ASTRON
Netherlands Institute for Radio Astronomy

What do (radio) astronomers want?

- More sensitivity
  - Bigger collecting area


CAMRAS / Wim de Vries

CEphoto, Uwe Aranas


Šarūnas Burdulis, https://www.flickr.com/photos/sarunas_b/49190234907/

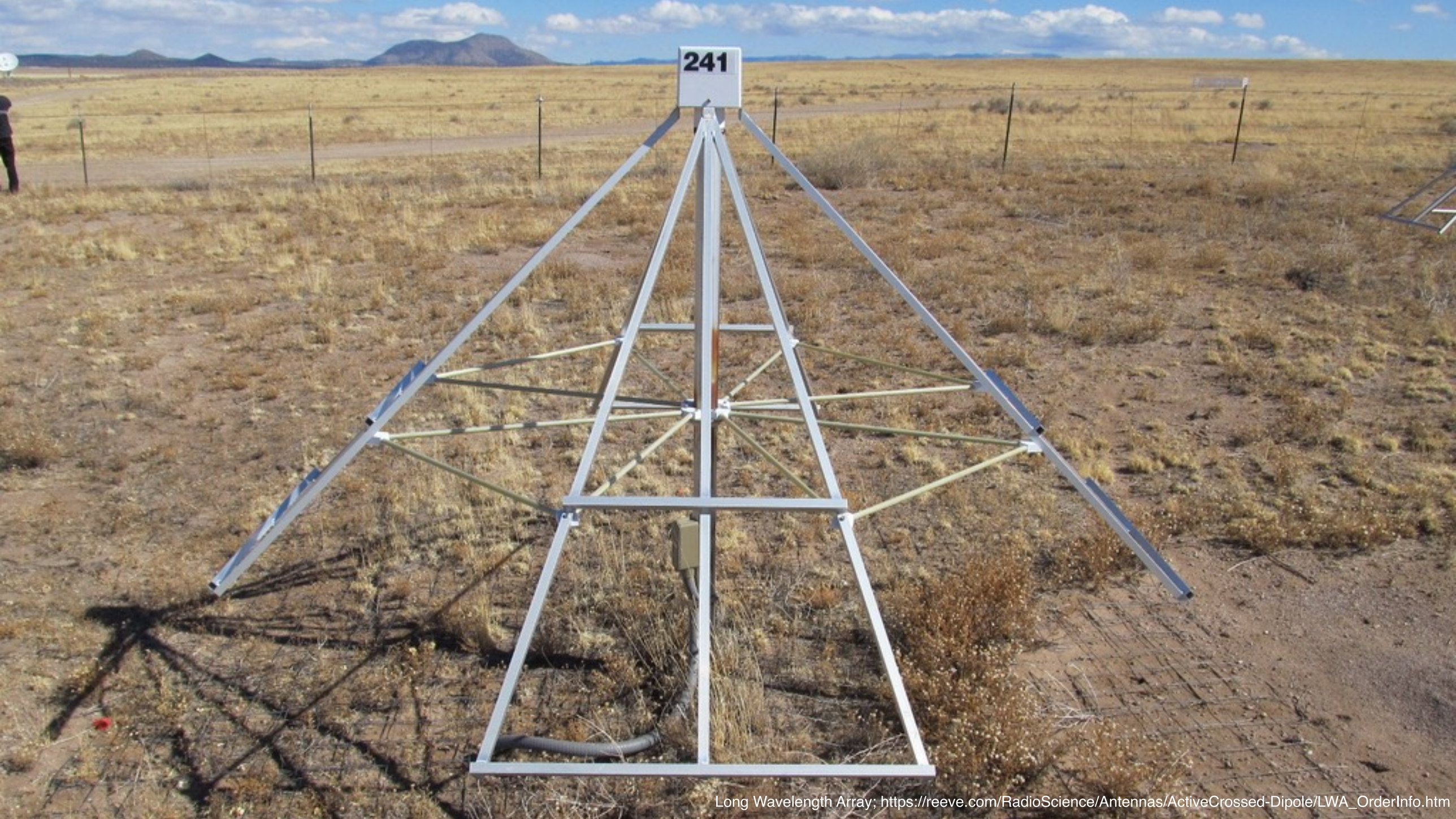
What do (radio) astronomers want?

- More sensitivity
  - Bigger collecting area
- Higher (angular) resolution; $\theta = \lambda / D$
  - Larger telescopes


National Radio
Astronomy
Observatory

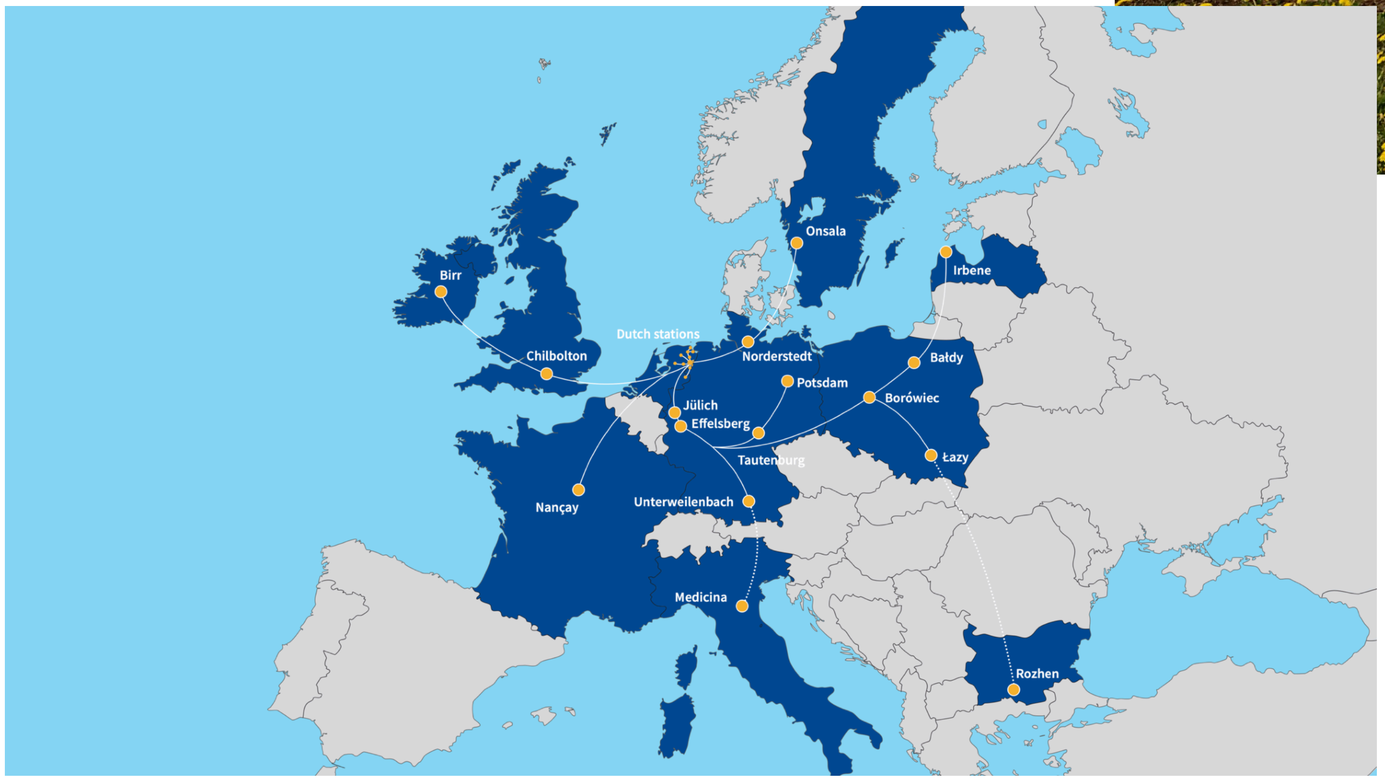
What do (radio) astronomers want?

- More sensitivity
  - Bigger collecting area
- Higher (angular) resolution; $\theta = \lambda / D$
  - Larger telescopes
- Wider field of view
  - All-sky sensitivity



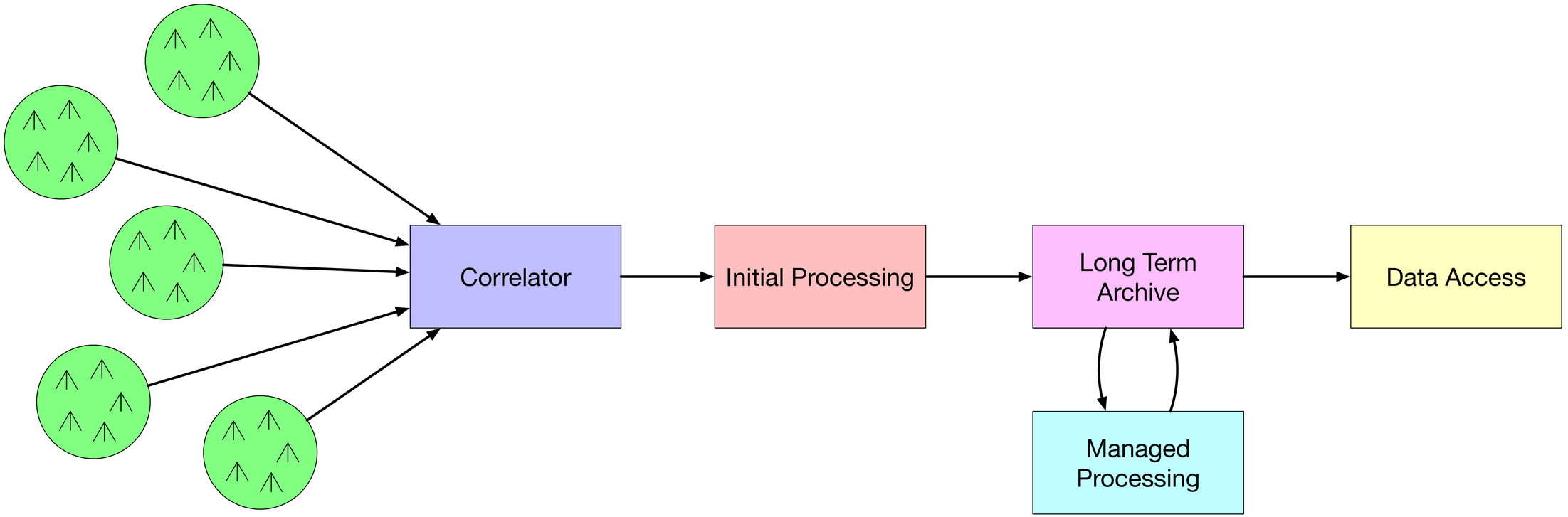
Long Wavelength Array; https://reeve.com/RadioScience/Antennas/ActiveCrossed-Dipole/LWA_OrderInfo.htm

LOFAR ERIC

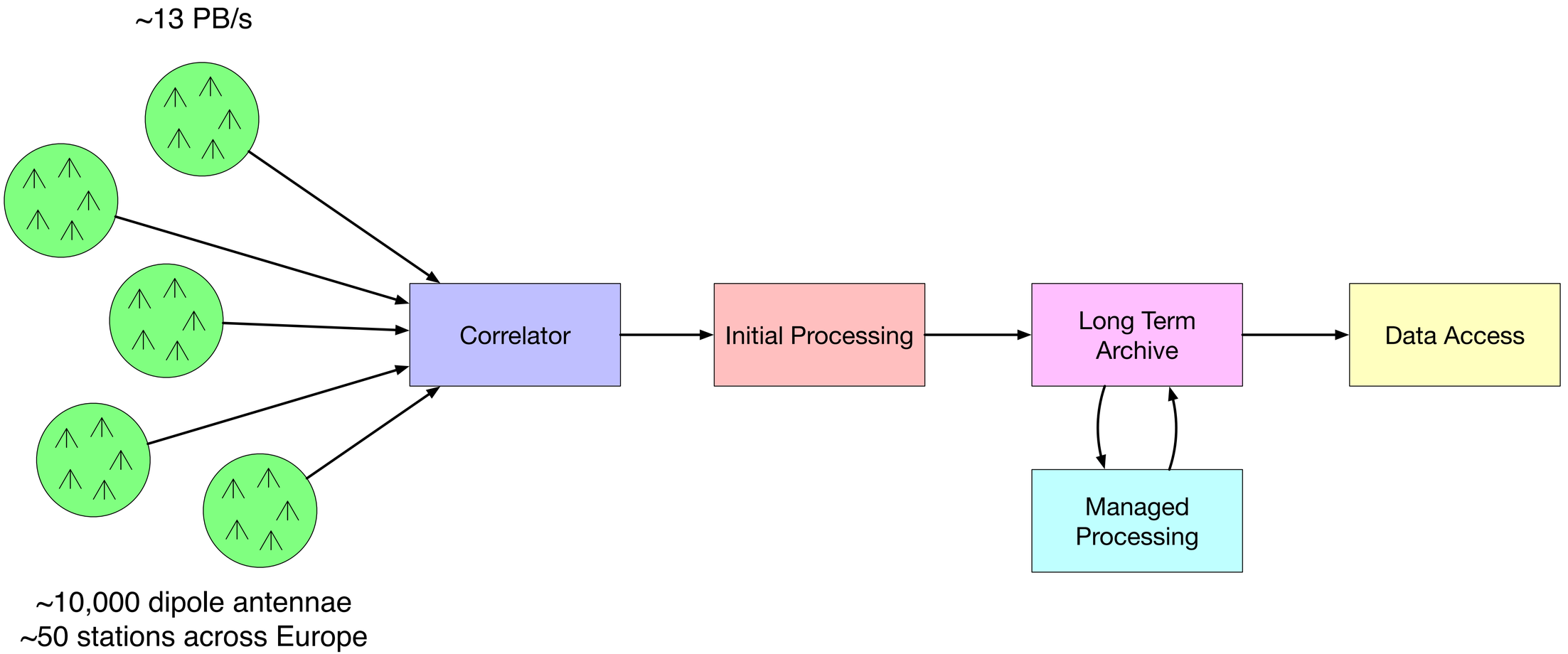


https://www.lofar.eu/

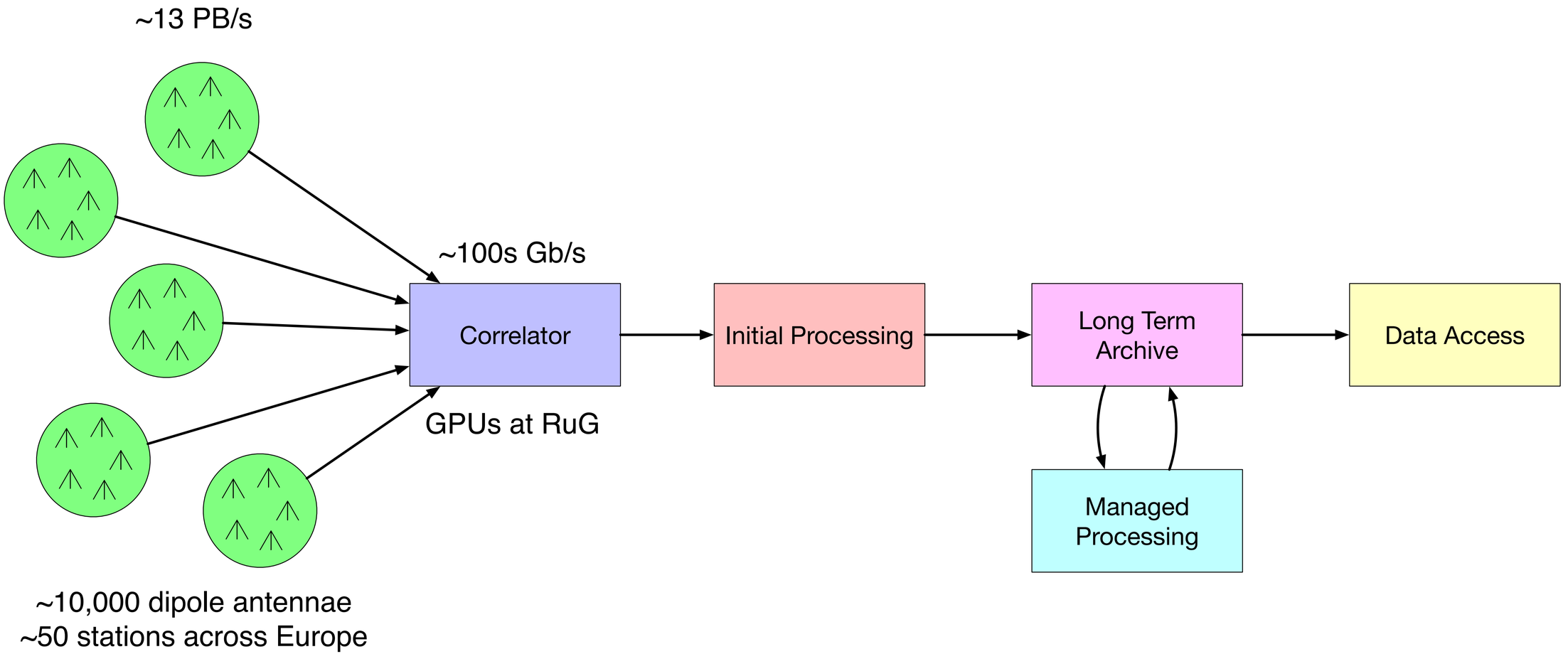
Now


Correlator
Initial Processing
Long Term Archive
Data Access
Managed Processing

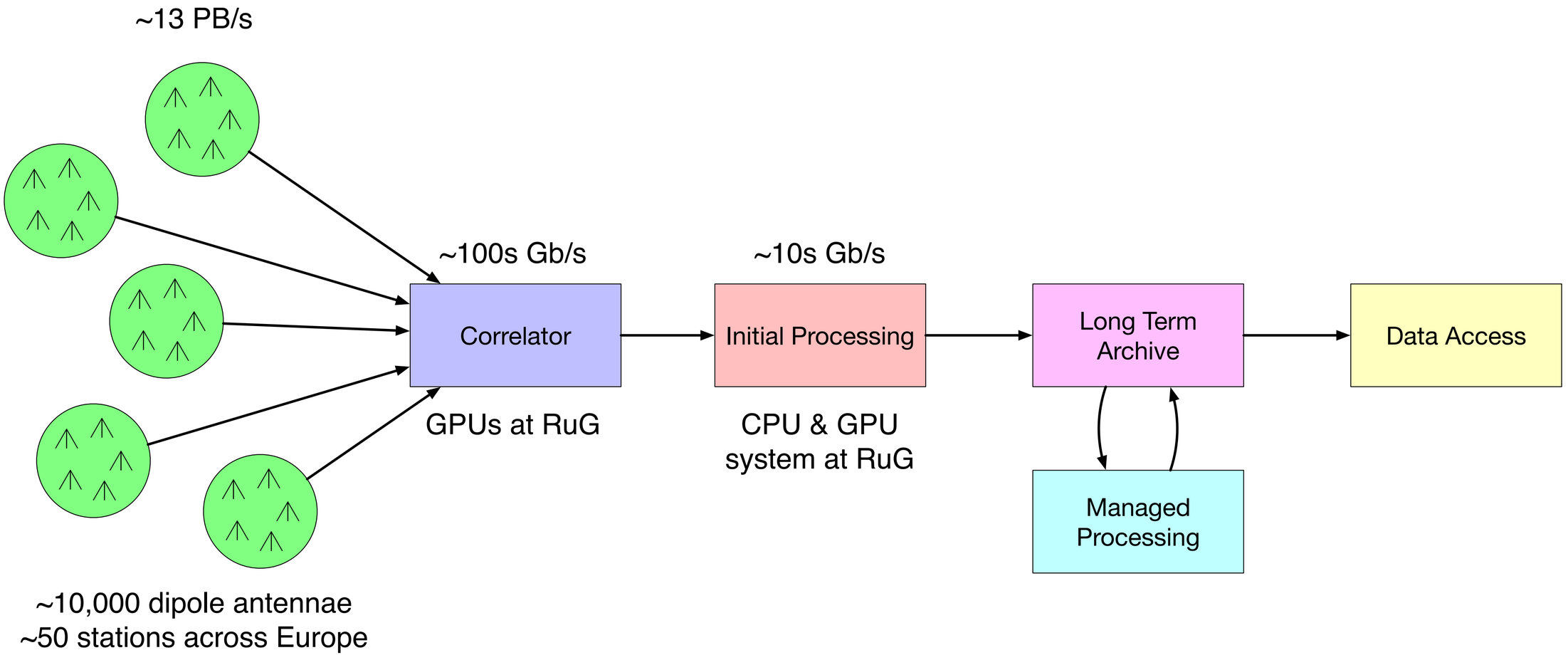
Now


~13 PB/s
Correlator
Initial Processing
Long Term Archive
Data Access
Managed Processing
~10,000 dipole antennae
~50 stations across Europe

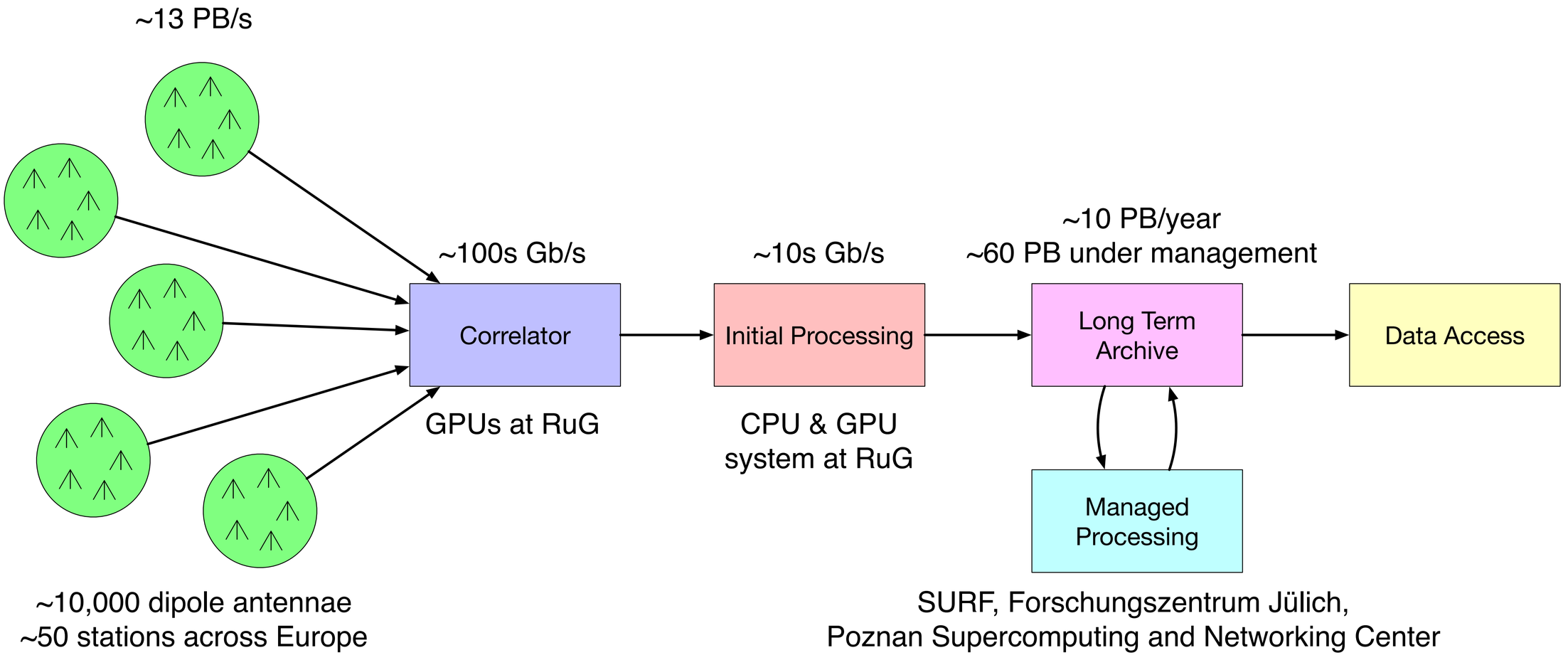
Now


~13 PB/s
~100s Gb/s
Correlator
GPUs at RuG
Initial Processing
Long Term Archive
Data Access
Managed Processing
~10,000 dipole antennae
~50 stations across Europe

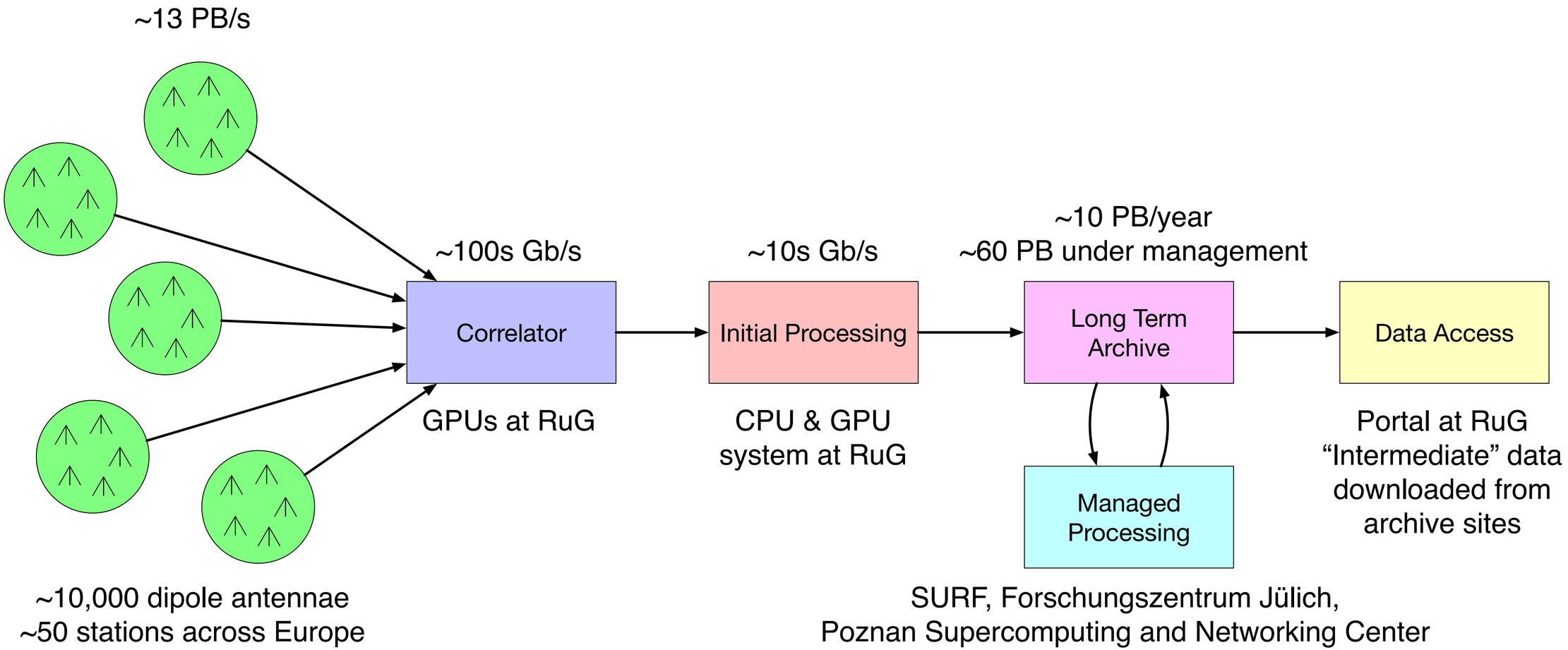
Now


~13 PB/s
~100s Gb/s
~10s Gb/s
Correlator
Initial Processing
Long Term Archive
Data Access
GPUs at RuG
CPU & GPU system at RuG
Managed Processing
~10,000 dipole antennae
~50 stations across Europe

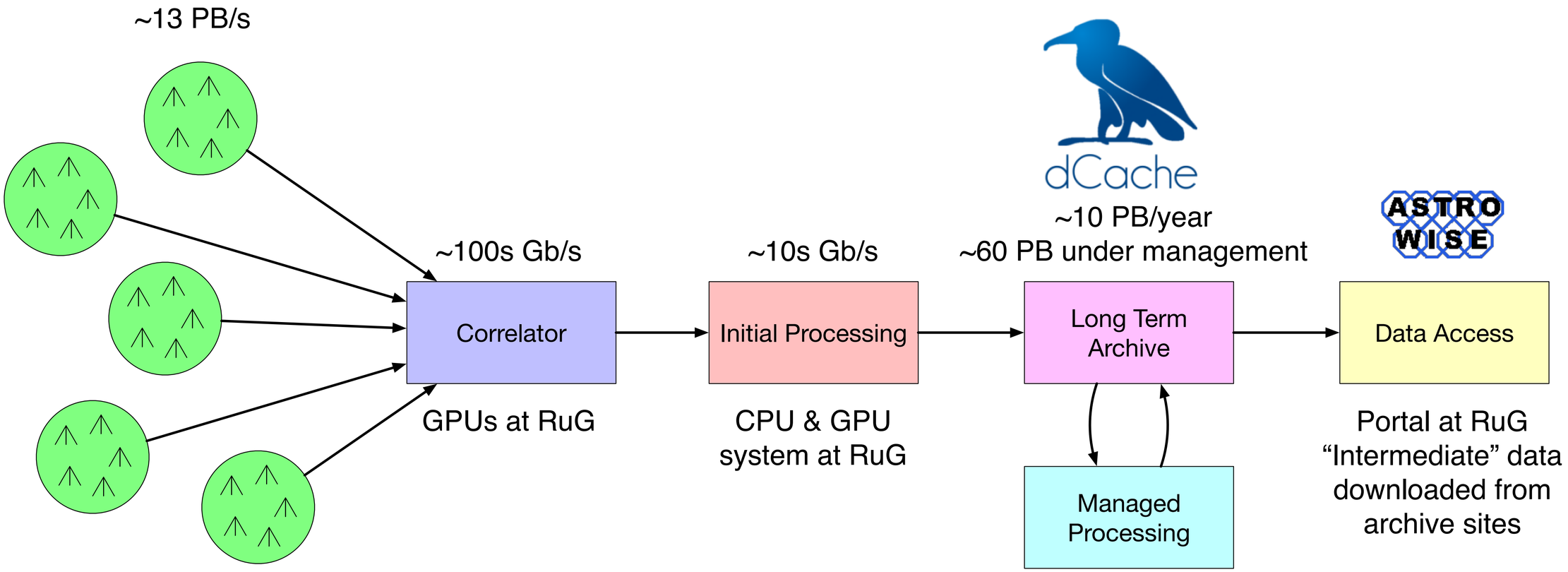
Now


~13 PB/s
~100s Gb/s
~10s Gb/s
~10 PB/year
~60 PB under management
Correlator
Initial Processing
Long Term Archive
Data Access
GPUs at RuG
CPU & GPU system at RuG
Managed Processing
~10,000 dipole antennae
~50 stations across Europe
SURF, Forschungszentrum Jülich,
Poznan Supercomputing and Networking Center

Now

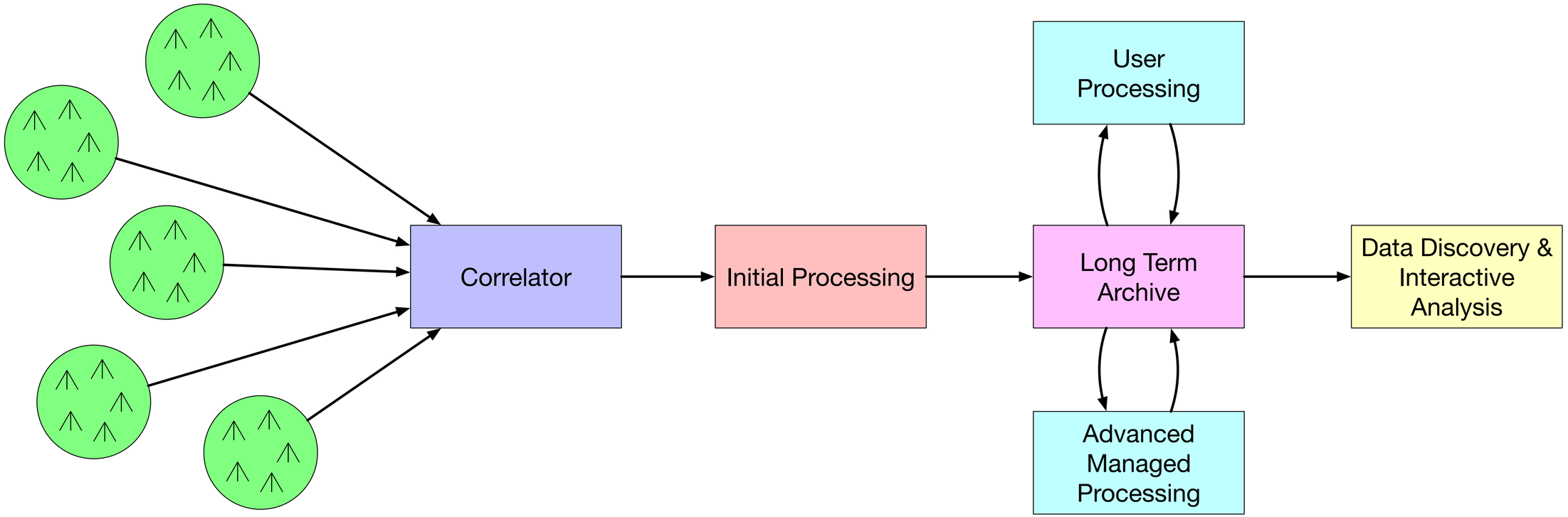

~13 PB/s
~100s Gb/s
~10s Gb/s
~10 PB/year
~60 PB under management
Correlator
Initial Processing
Long Term Archive
Data Access
GPUs at RuG
CPU & GPU system at RuG
Managed Processing
Portal at RuG
“Intermediate” data downloaded from archive sites
~10,000 dipole antennae
~50 stations across Europe
SURF, Forschungszentrum Jülich,
Poznan Supercomputing and Networking Center

Now



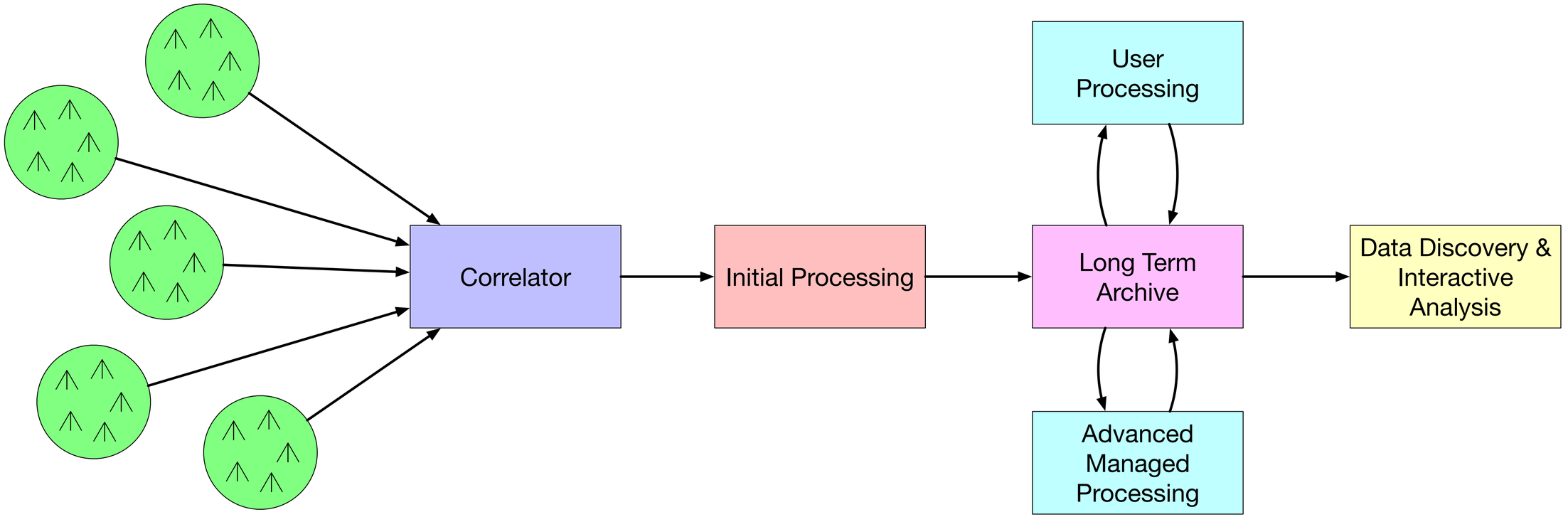
COMMON WORKFLOW LANGUAGE

Next


User
Processing
Correlator
Initial Processing
Long Term
Archive
Data Discovery &
Interactive
Analysis
Advanced
Managed
Processing

Next

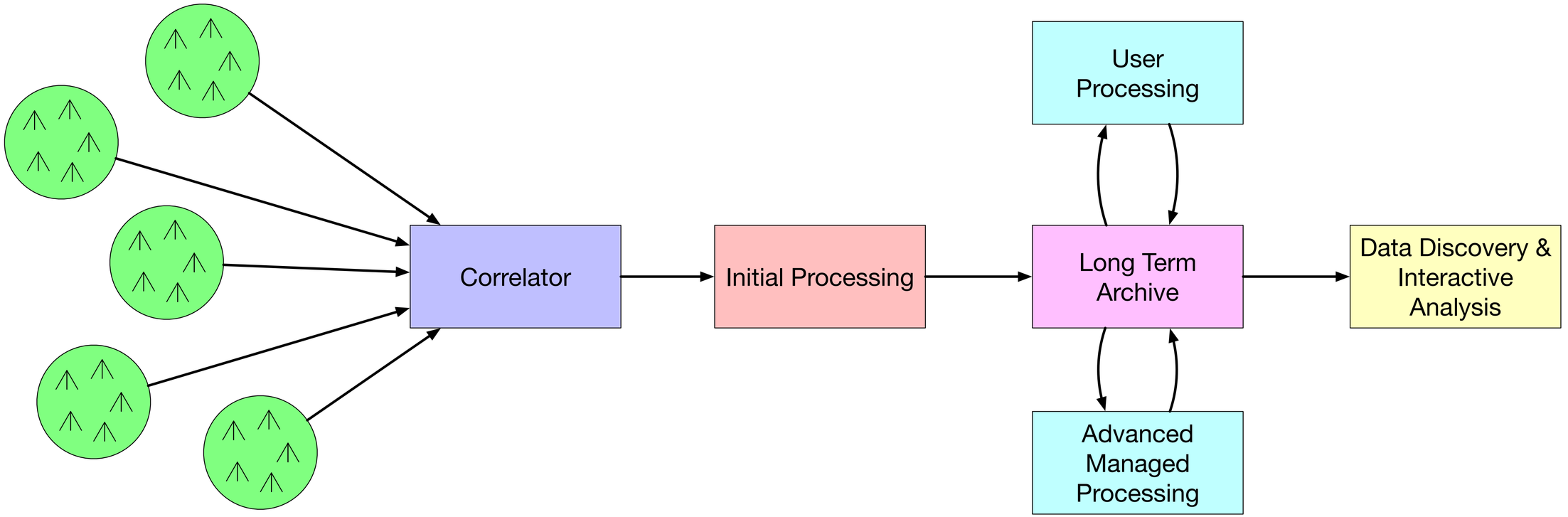
LOFAR2.0 Upgrade (2025)



More flexible observing system;
new modes, more data

Next

LOFAR2.0 Upgrade (2025)

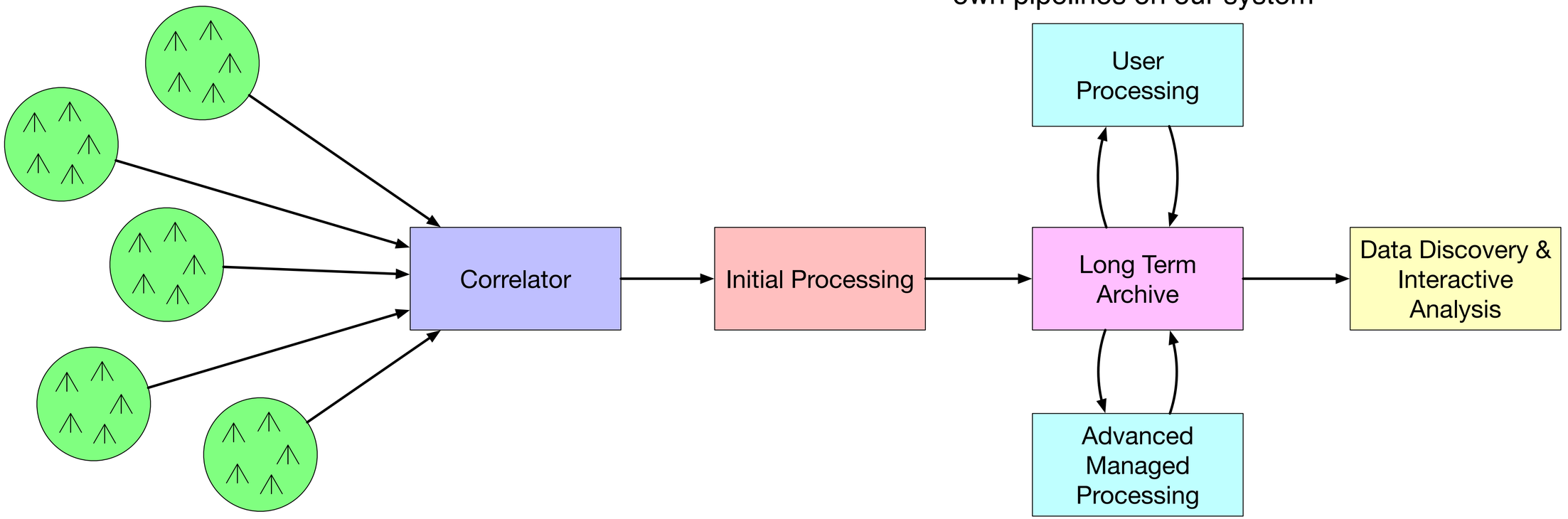


More flexible observing system;
new modes, more data

New & upgraded pipelines,
producing science-ready data

Next

LOFAR2.0 Upgrade (2025)



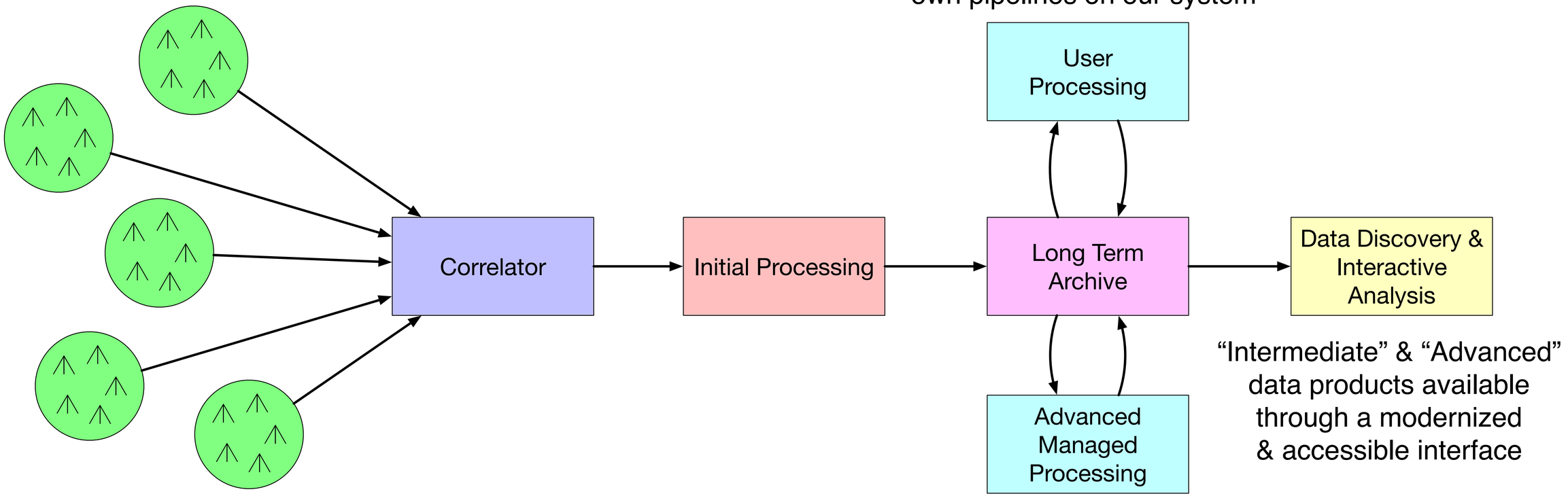
More flexible observing system;
new modes, more data

Users can define & run their own pipelines on our system

New & upgraded pipelines,
producing science-ready data

Next

LOFAR2.0 Upgrade (2025)



More flexible observing system;
new modes, more data

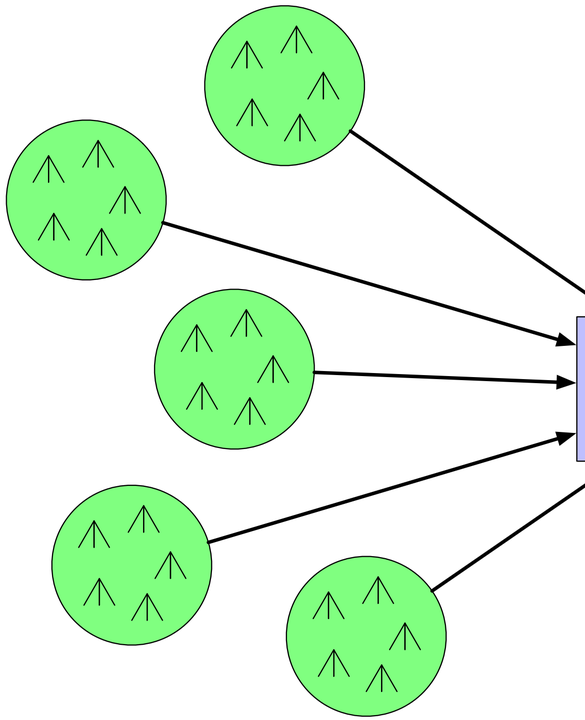
Users can define & run their own pipelines on our system

New & upgraded pipelines,
producing science-ready data

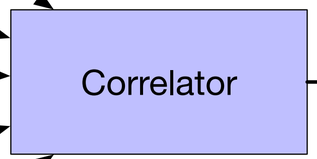
“Intermediate” & “Advanced” data products available through a modernized & accessible interface

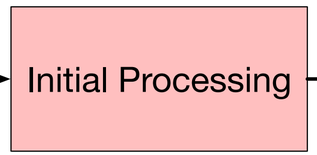
Next

LOFAR2.0 Upgrade (2025)

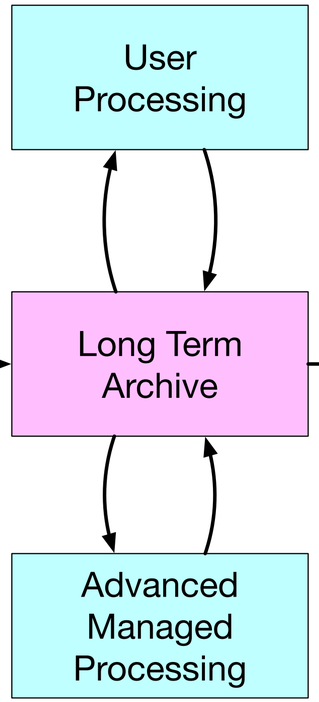


More flexible observing system;
new modes, more data

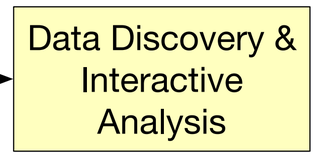




Users can define & run their
own pipelines on our system



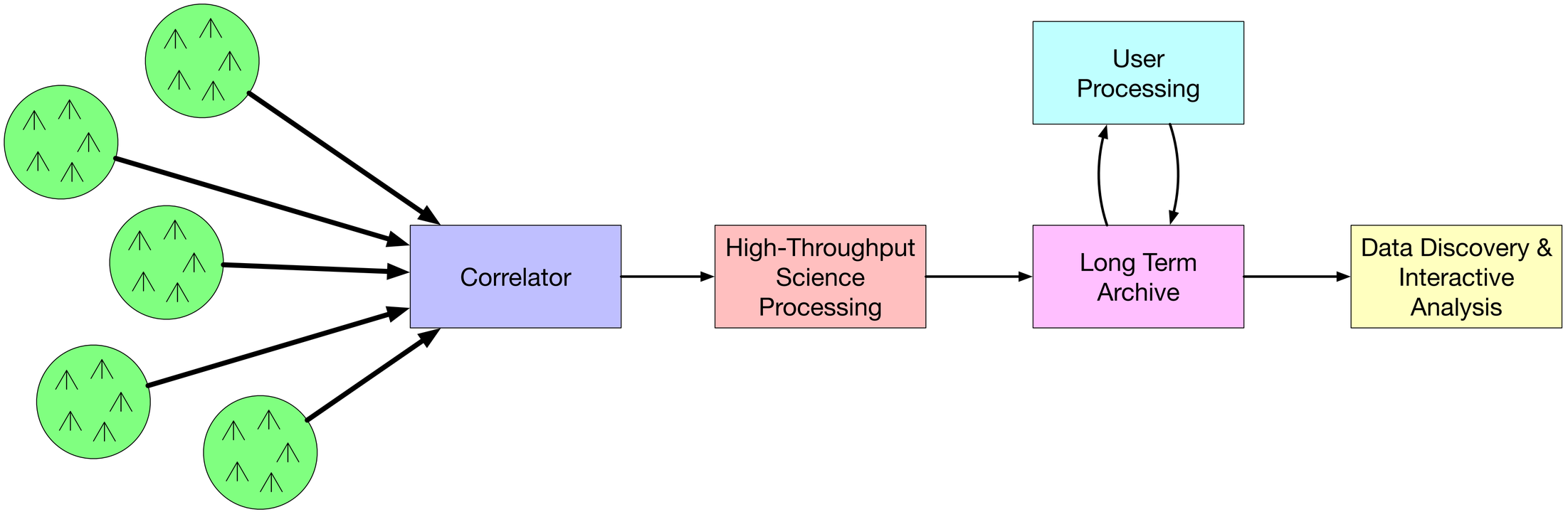
New & upgraded pipelines,
producing science-ready data



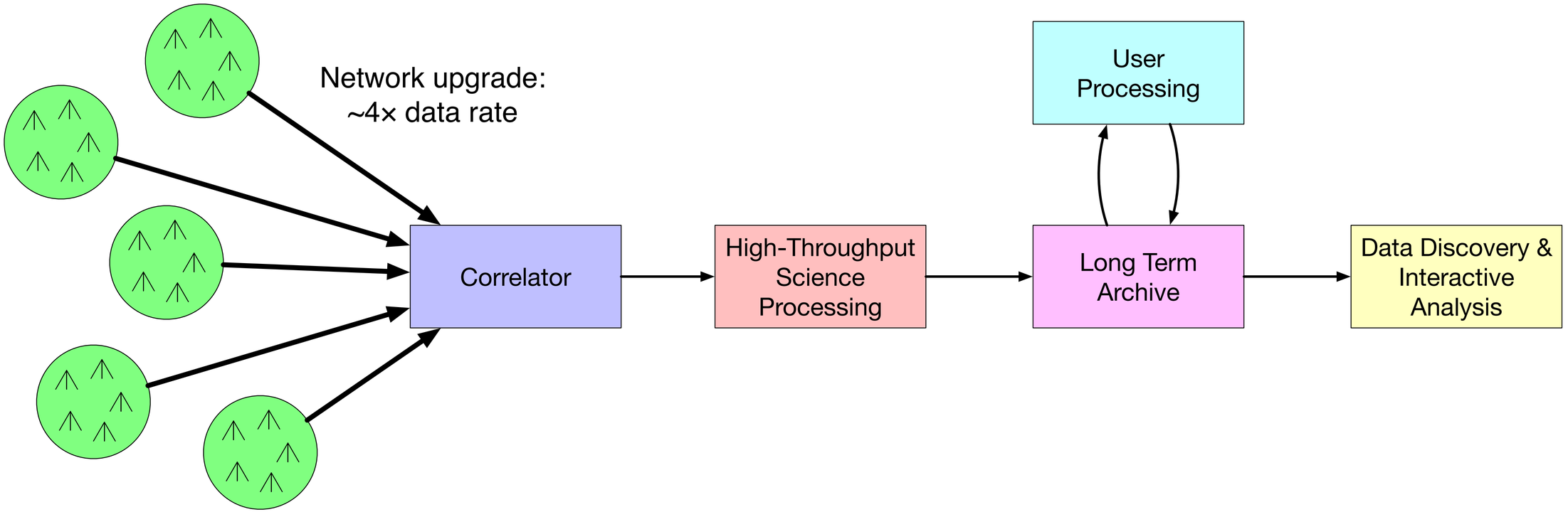
Online analysis tools
running next to data
(e.g. Jupyter)

“Intermediate” & “Advanced”
data products available
through a modernized
& accessible interface

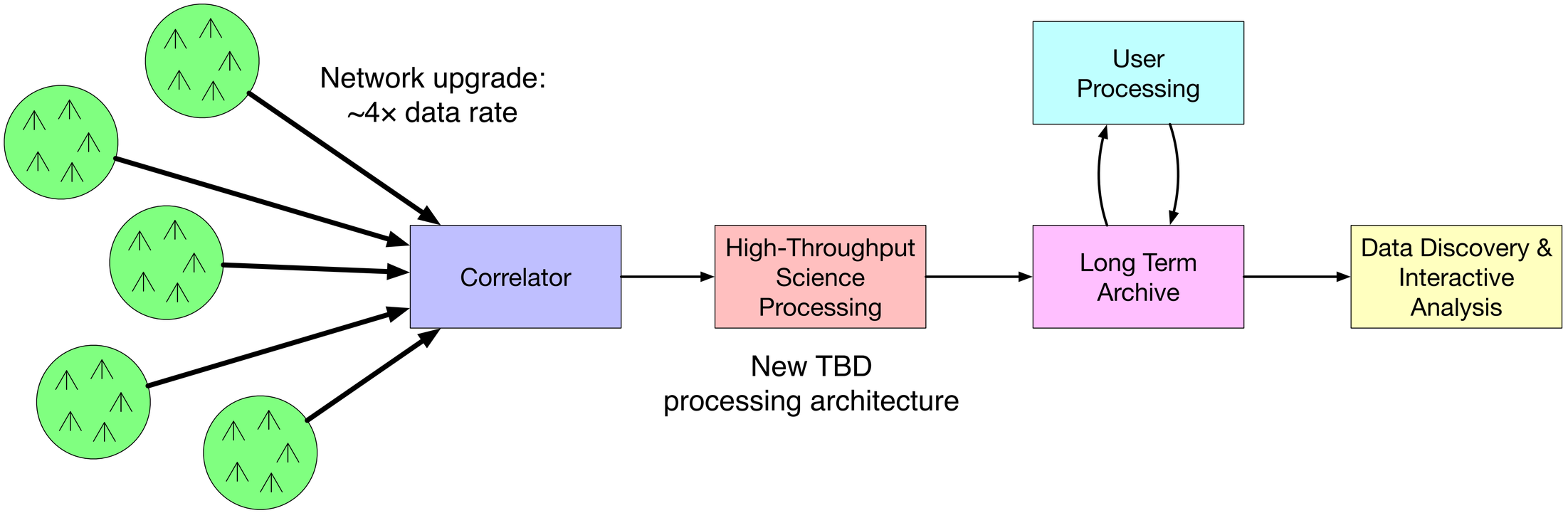
Future


User
Processing
Correlator
High-Throughput
Science
Processing
Long Term
Archive
Data Discovery &
Interactive
Analysis

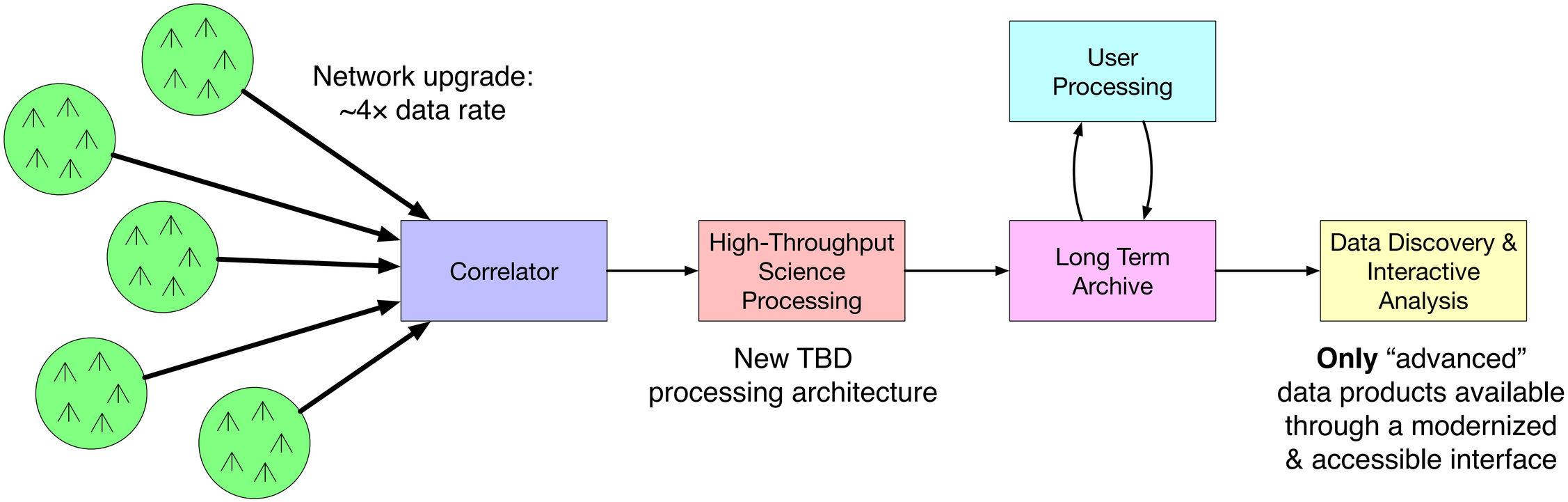
Future


Network upgrade:
~4× data rate
Correlator
High-Throughput
Science
Processing
User
Processing
Long Term
Archive
Data Discovery &
Interactive
Analysis

Future


Network upgrade:
~4× data rate
Correlator
High-Throughput
Science
Processing
New TBD
processing architecture
User
Processing
Long Term
Archive
Data Discovery &
Interactive
Analysis

Future

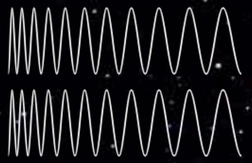

Network upgrade:
~4× data rate
Correlator
High-Throughput
Science
Processing
User
Processing
Long Term
Archive
Data Discovery &
Interactive
Analysis
New TBD
processing architecture
Only “advanced”
data products available
through a modernized
& accessible interface

Coming later this decade

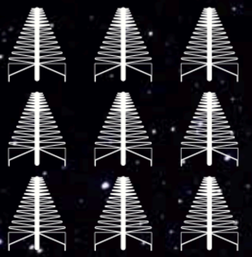
SKA-LOW

THE SKA'S LOW-FREQUENCY TELESCOPE

LOCATION:
AUSTRALIA



FREQUENCY RANGE:
50 MHz–
350 MHz



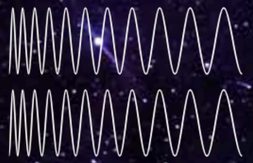
131,072
ANTENNAS
SPREAD ACROSS **512** STATIONS

MAXIMUM BASELINE:
~65km

SKA-MID

THE SKA'S MID-FREQUENCY TELESCOPE

LOCATION:
SOUTH AFRICA

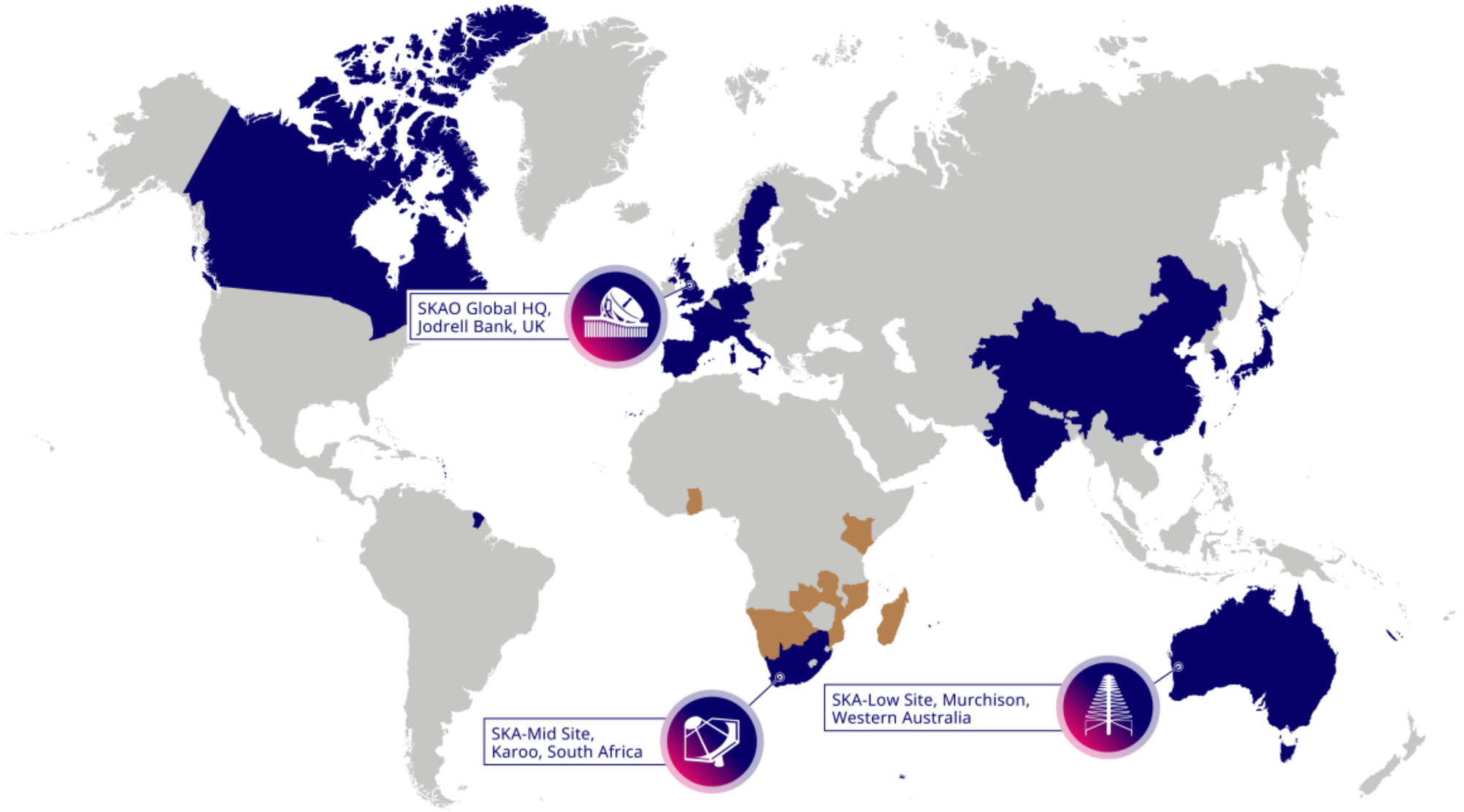


FREQUENCY RANGE:
350 MHz–
15.4 GHz
WITH A GOAL OF 24 GHz

197 DISHES
(INCLUDING 64 MEERKAT DISHES)

MAXIMUM BASELINE:
150km

https://www.skao.int/

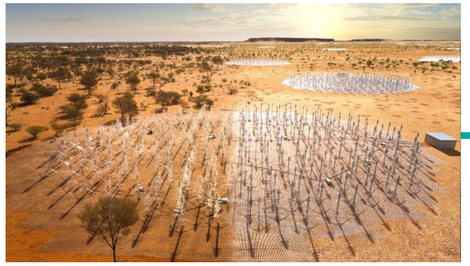


SKAO Partnership - includes SKAO Member States* and SKAO Observers (as of July 2023)

African Partner Countries

* * * * * * * * *

Regional Centre Concept



SKA Central Signal Processor

SKA Central Signal Processor

SKA Science Data Processor

SKA Science Data Processor

350 PB/year

Observatory Data Products

350 PB/year

SKA Regional Centre

SKA Regional Centre

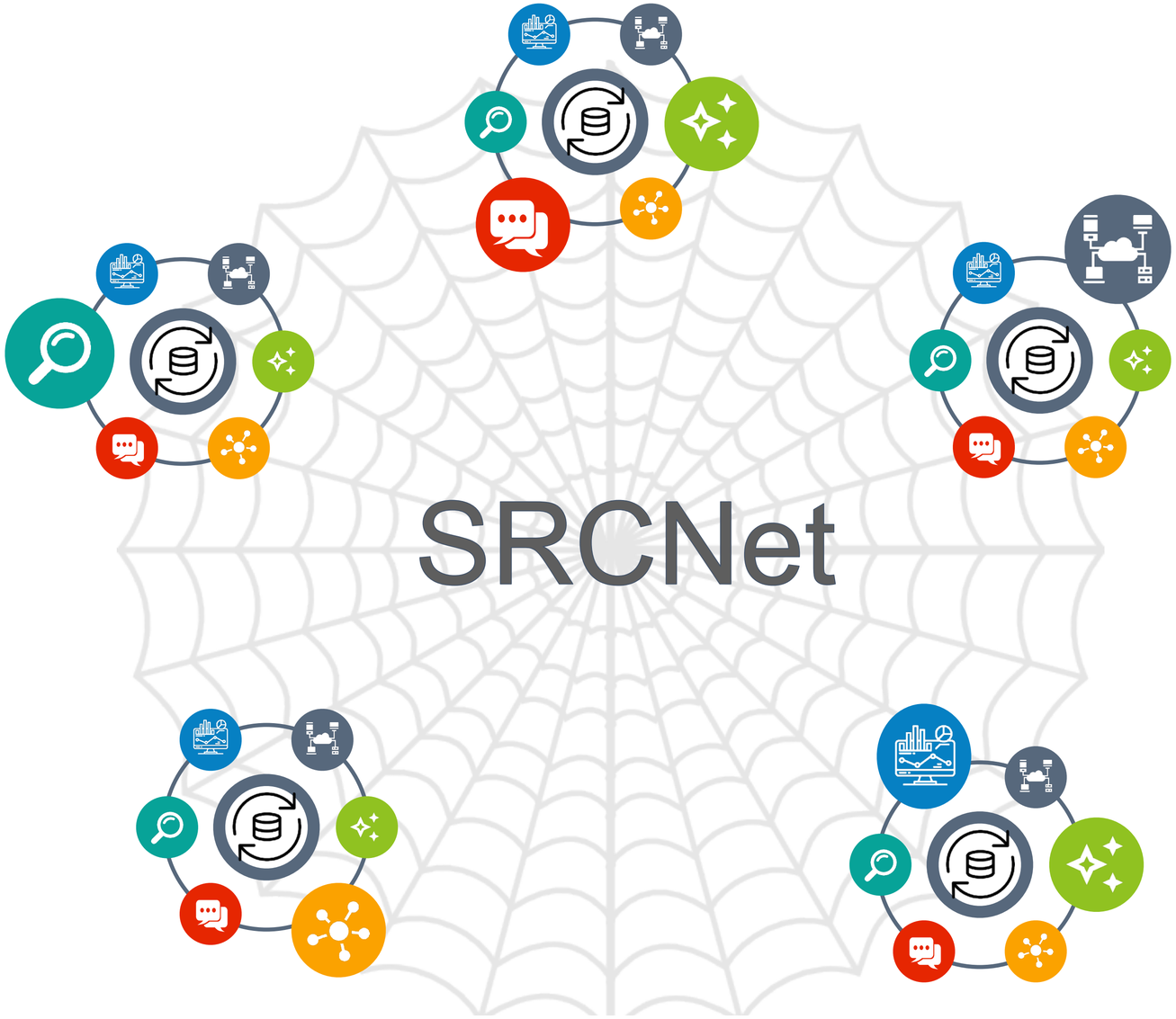
SKA Regional Centre

SKA Regional Centre

...

- The SKA Observatory will generate “science ready” data (images, etc).
  - Science users have no direct access to the processing system or data products; it is operated by SKAO staff only.
- Data products are supplied by SKAO to a network of *SKA Regional Centres.* SRCs are responsible for:
  - Archiving and curating science data;
  - Making that data available to end users;
  - Enabling scientific discovery.

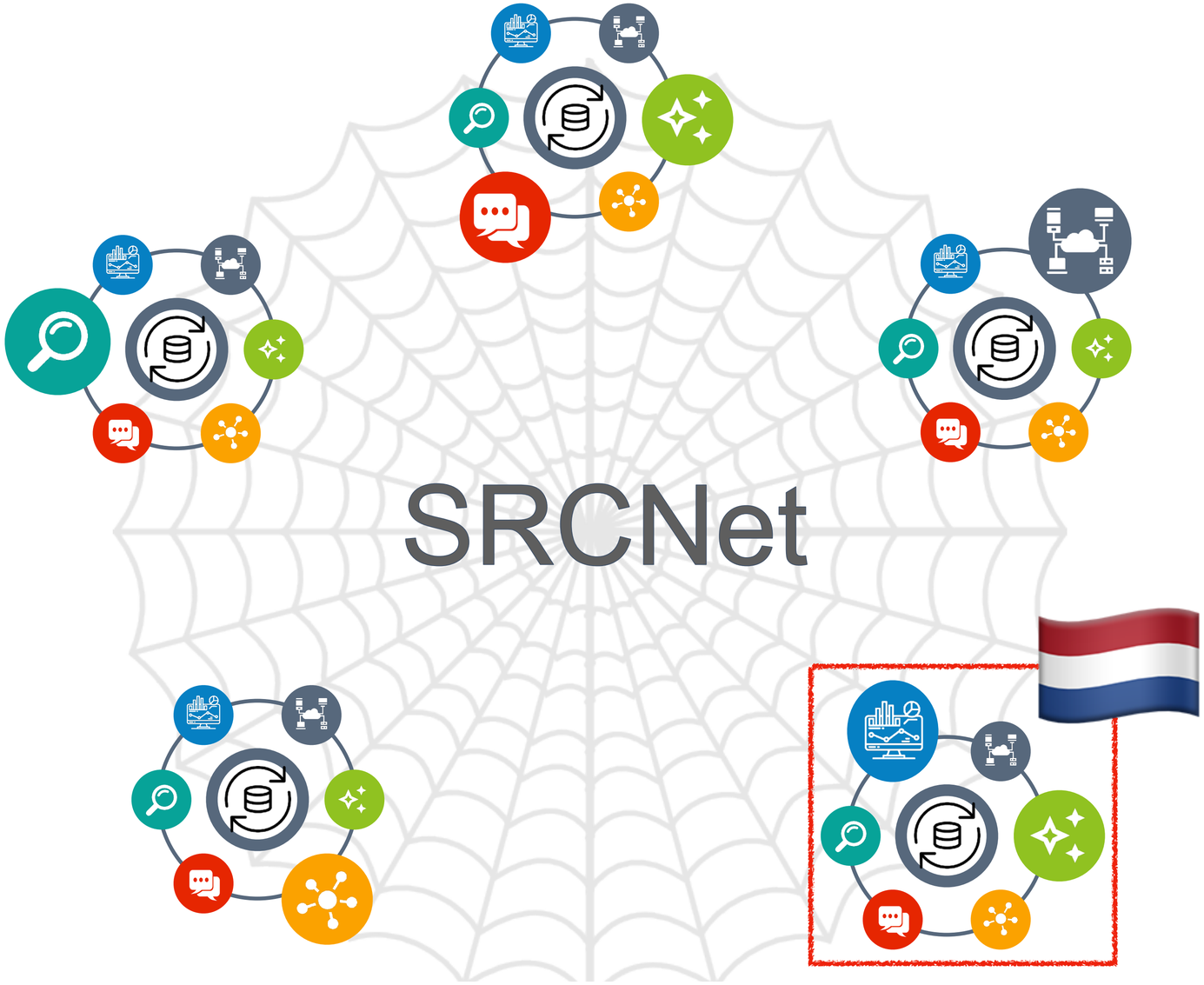
The Global SRC Network



Collectively meet the needs of the global community of SKA users.

Heterogeneous SRCs, with different strengths, working together in a federated infrastructure.

The Global SRC Network



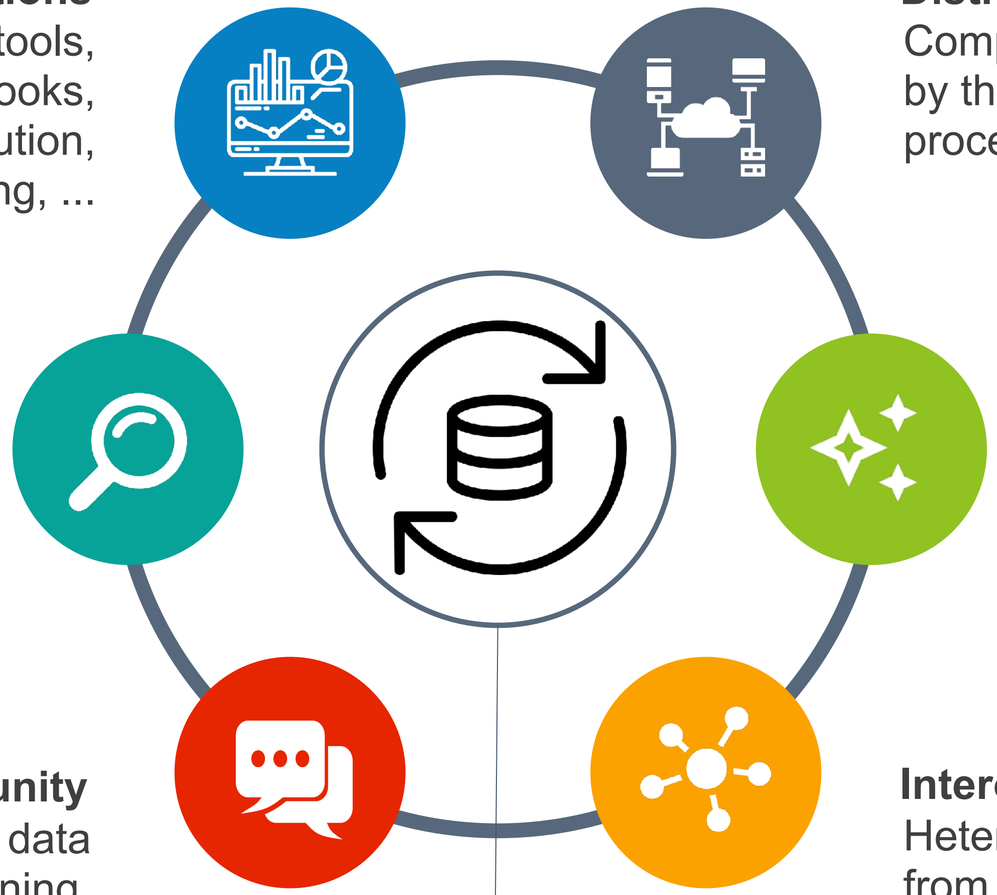
Collectively meet the needs of the global community of SKA users.

Heterogeneous SRCs, with different strengths, working together in a federated infrastructure.

SRC Capabilities

Scientific Applications
Analysis tools, Jupyter notebooks, workflow execution, machine learning, ...

Distributed Data Processing
Computing capabilities provided by the SRCNet to allow data processing



Data Discovery
Discovery of SKA data from the SRCNet, local or remote, transparently to the user

Visualization
Advanced visualization tools for both SKA data and data from other facilities

Support to Science Community
Support community on SKA data use, SRC services use, training, project impact dissemination

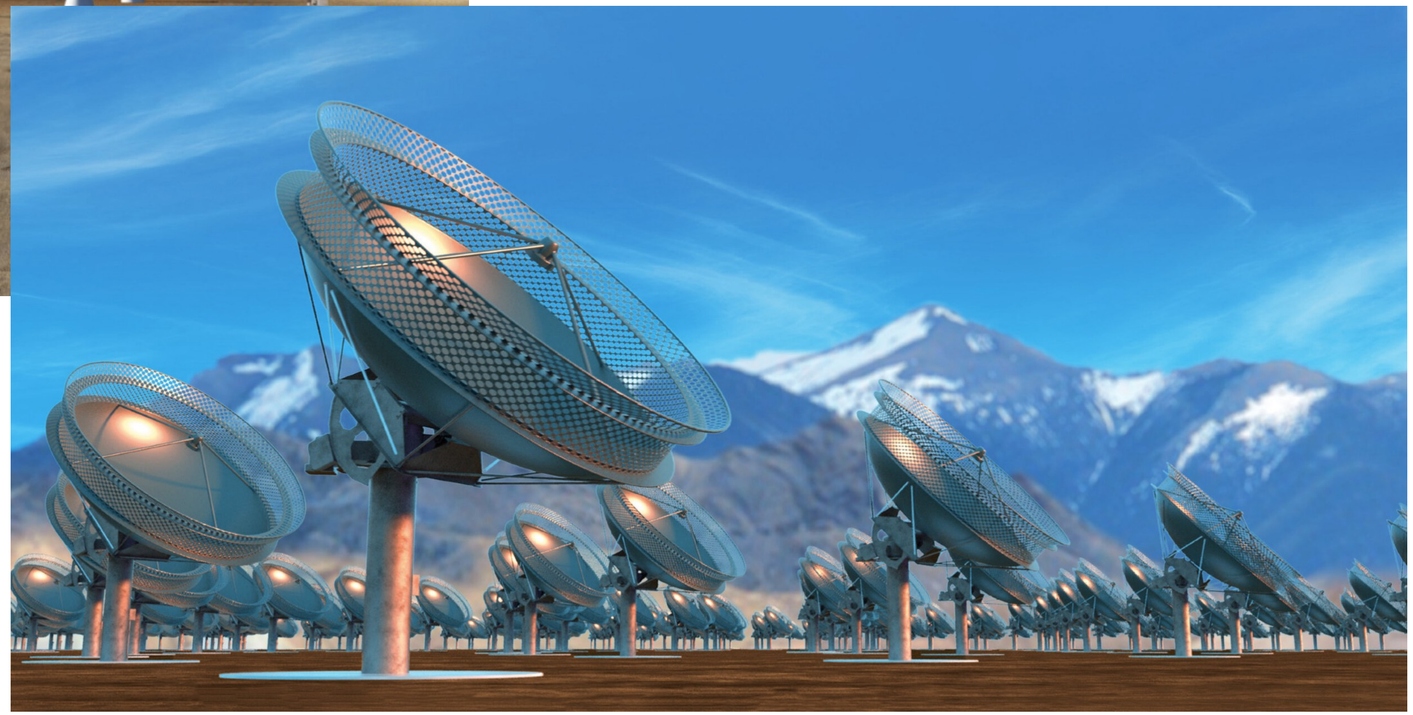
Interoperability
Heterogeneous data and services from various SRCs and other facilities

Data Management
Dissemination of data to SRCs and distributed data storage

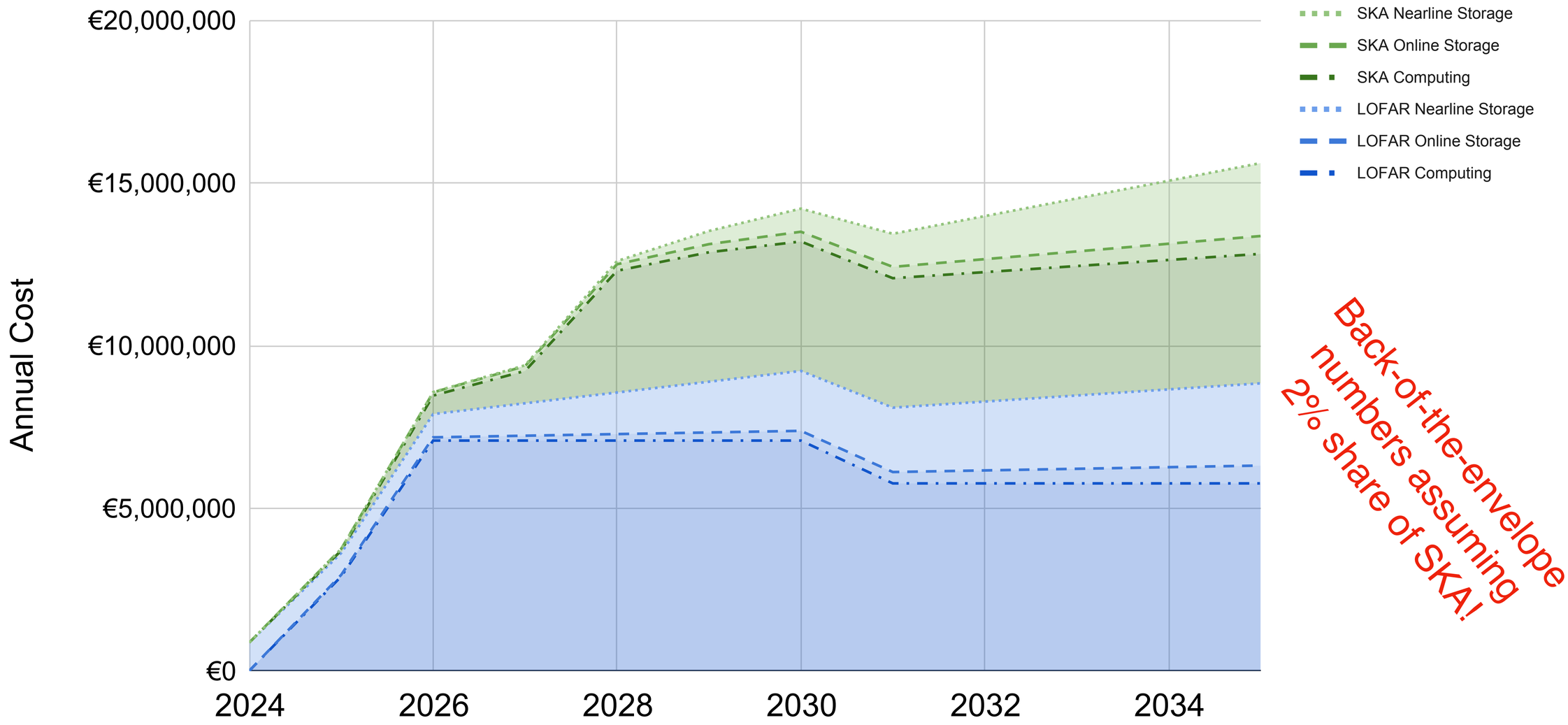
The Radio Data Explosion

http://ngvla.nrao.edu

https://www.deepsynoptic.org



Cost$


€20,000,000
€15,000,000
€10,000,000
€5,000,000
€0
Annual Cost
2024
2026
2028
2030
2032
2034
SKA Nearline Storage
SKA Online Storage
SKA Computing
LOFAR Nearline Storage
LOFAR Online Storage
LOFAR Computing
Back-of-the-envelope numbers assuming 2% share of SKA!

Challenges & Opportunities

- Environmental & financial sustainability; look to new technologies, innovative approaches
- A limited talent pool; an explosion of competition
- International collaborations & the dangers of megaprojects

Conclusions

- For the last decade+, LOFAR has been a pioneer for a new generation of radio telescopes.
- Only now do we have the data infrastructure to make the most of it.
- Coming LOFAR upgrades, as well as the SKA and a range of other facilities, present unprecedented challenges in data archiving and processing.
- We have ideas & opportunities, but there's lot of work to be done and a lot of opportunity for collaboration.