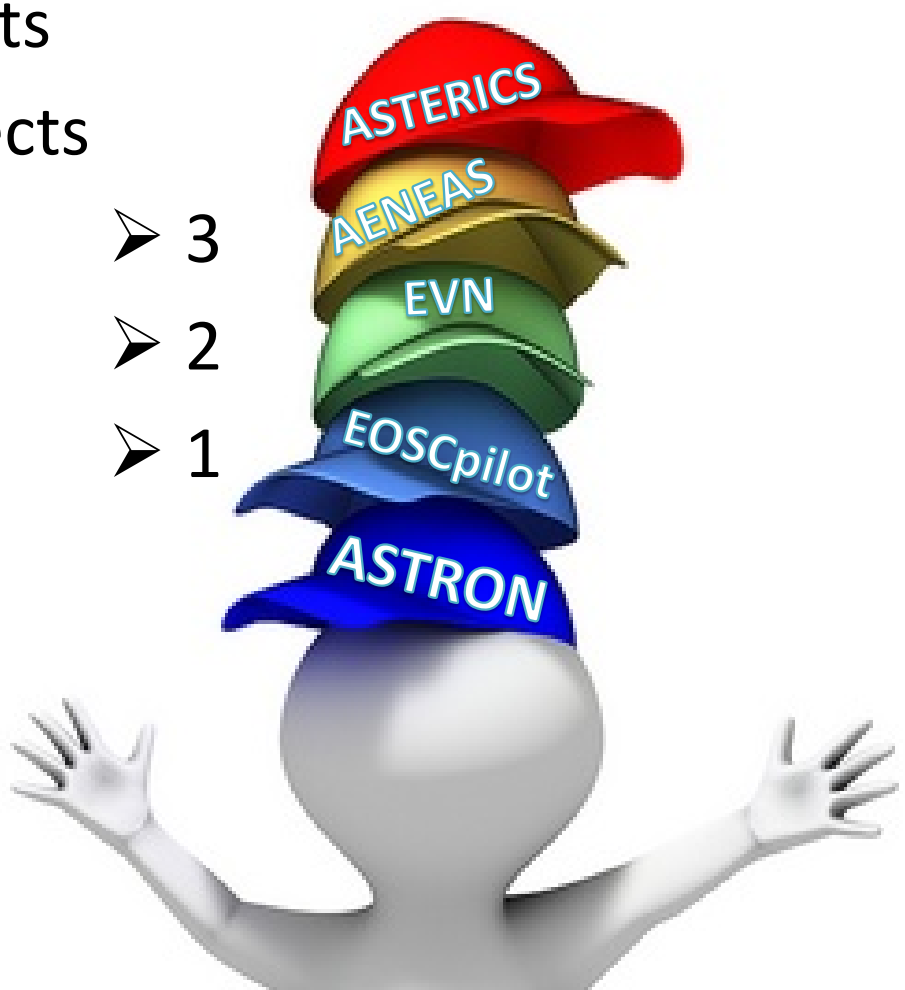
EOSCpilot Science Demonstrator LOFAR

Rob van der Meer - ASTRON

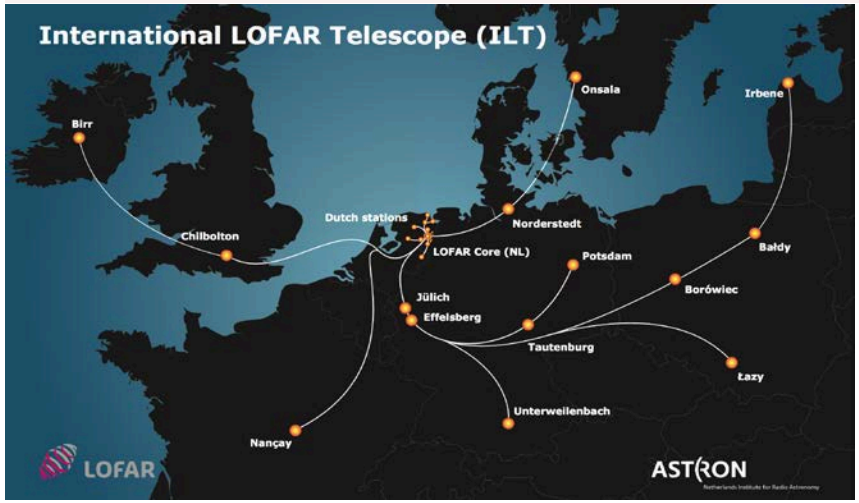
Arpad Szomoru - JIVE

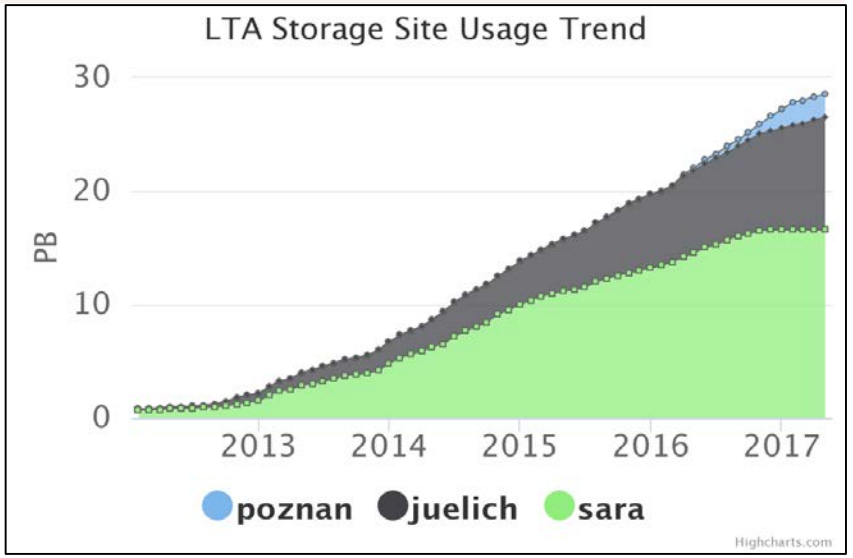
Points of today

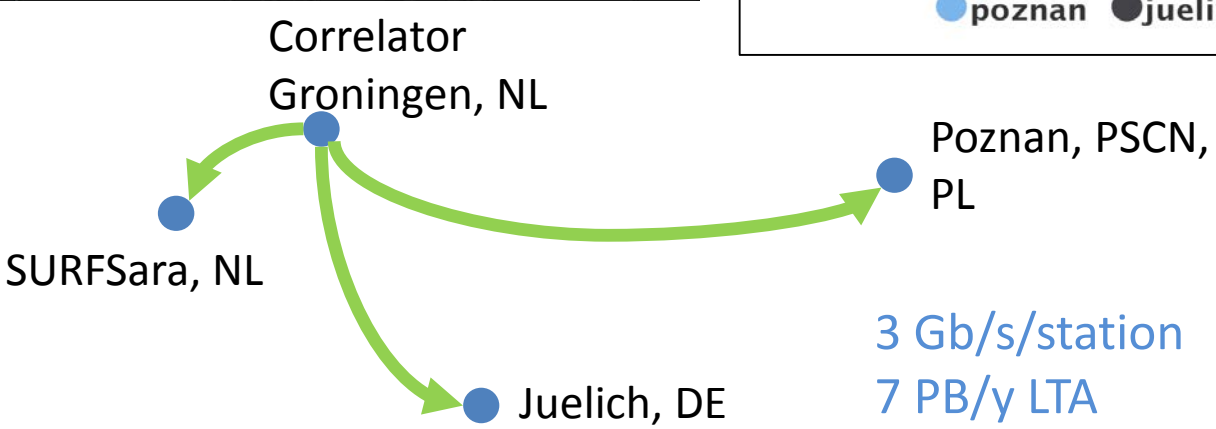
- Multiple hats
- Three subjects
  - 3
  - 2
  - 1



International LOFAR Telescope





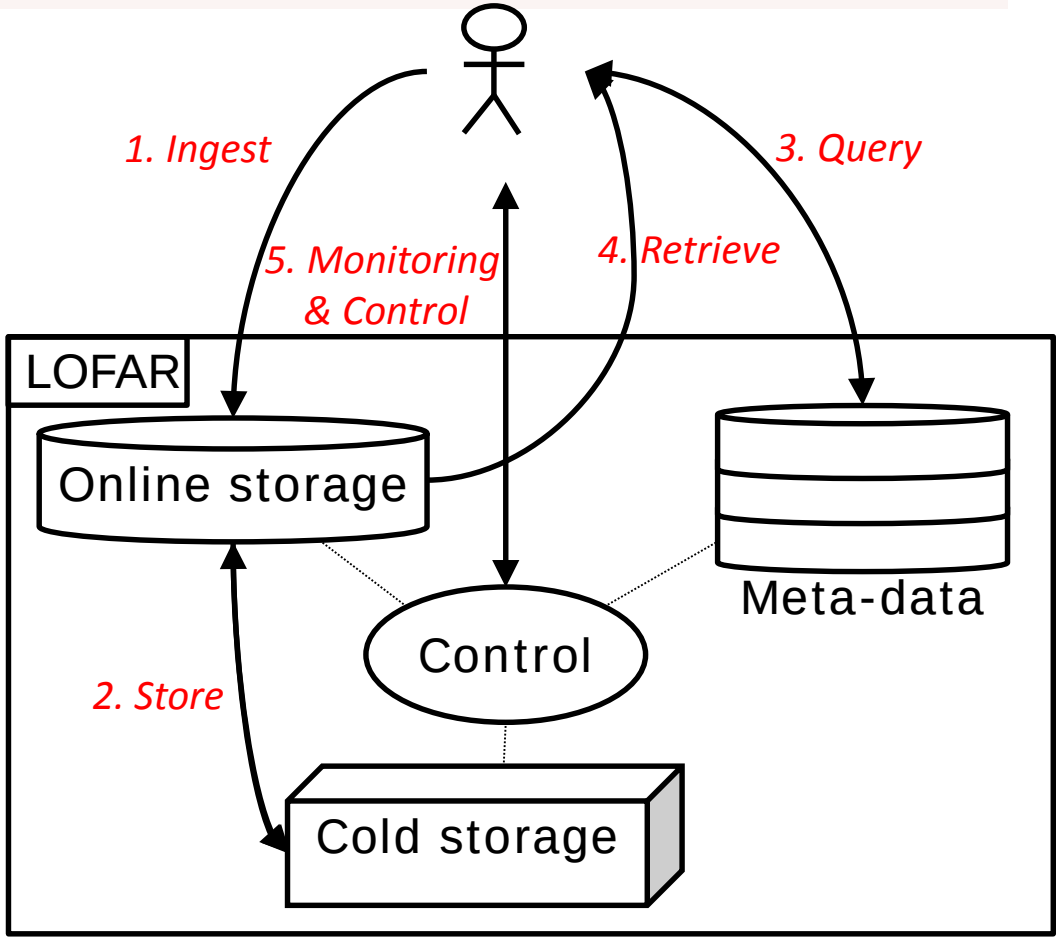


3 Gb/s/station
7 PB/y LTA

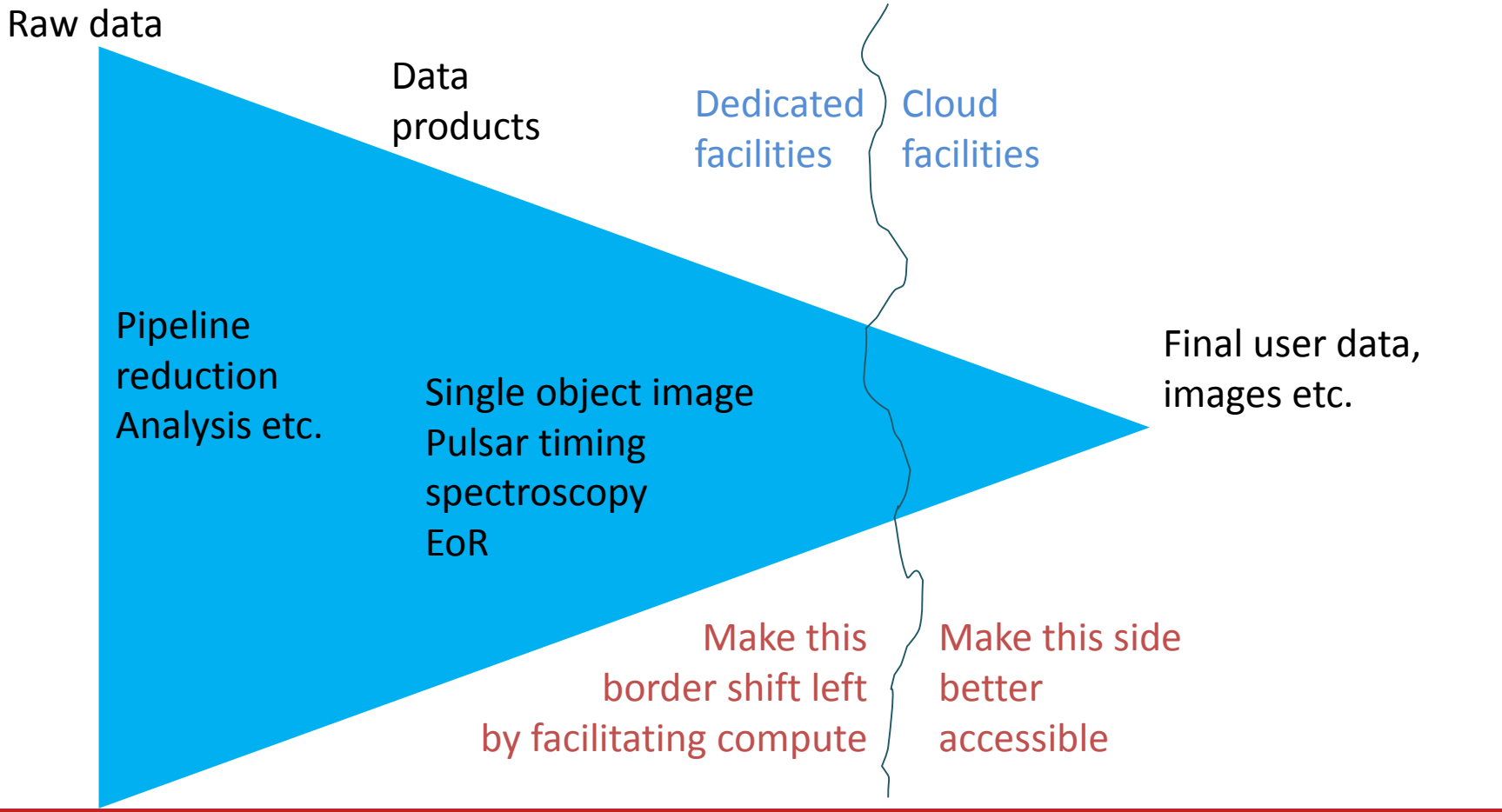
LOFAR Long Term Archive

High level use-case:

1. ingest data
2. store data
3. query meta-data
4. retrieve data
5. monitor & control



EOSC pilot tasks



Challenge & Use cases

Challenges

- Data provenance
- Federated Indentity
- Compute to data
- Multiple LTA sites
- Where → what is my data

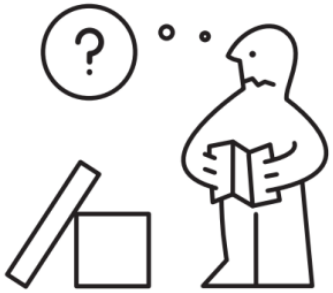
Facilitate

- easy access for power user. Free/sandbox compute with own algorithm, parameters, on small local data set.
  - Then scale up to larger data set on remote cluster
- Make LOFAR LTA accessible to non power users
  - Standard pipeline and GUI for ~10 free parameters.

Plan of attack

1. Define “perfect” environment

2. Existing tools and resources
   immediately start building



3. from there define new projects
   for improving the working system

Use this demonstration to show both possibilities and limitations of current software and e-Infrastructure.

Project plan

- SURFsara → connectivity to others
- Standardize existing pipelines → CWL
- Investigate notebooks
- Build web frontend → User settings
- Existing viewer for resulting workflow
- Use Zenodo/B2Share + Research Object → persistant storage with DOI

Resource & interoperability requirements

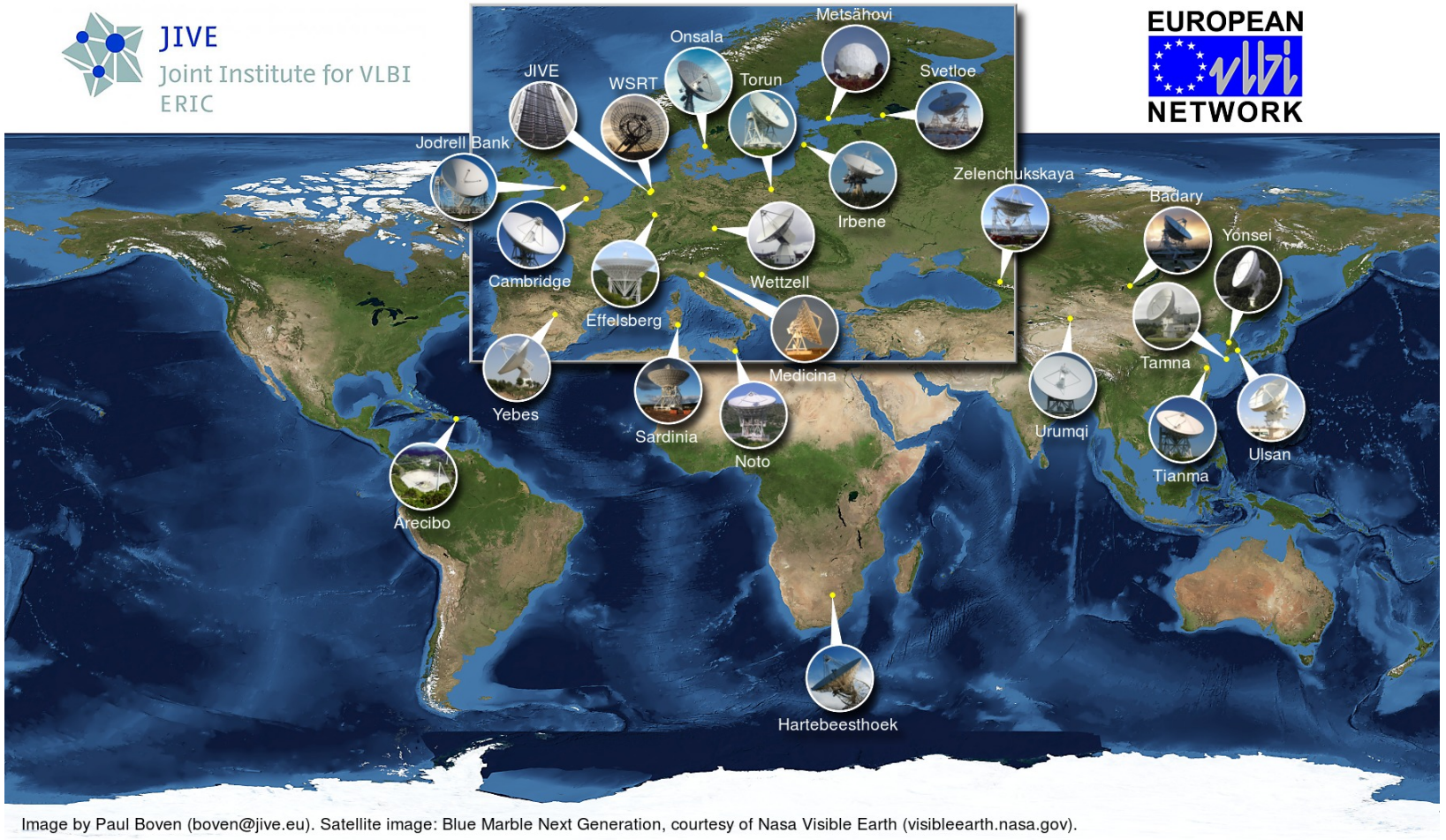
- ✓ Access to data in the LOFAR LTA @ SURFsara
- ✓ Acces to compute facilities @ SURFsara
- ✓ Access to data in the LOFAR LTA @ Juelich
- ✓ Transport of data between LTA sites
  Contact with GÉANT
- ✓ Access to compute facilities at other sites

Succes criteria

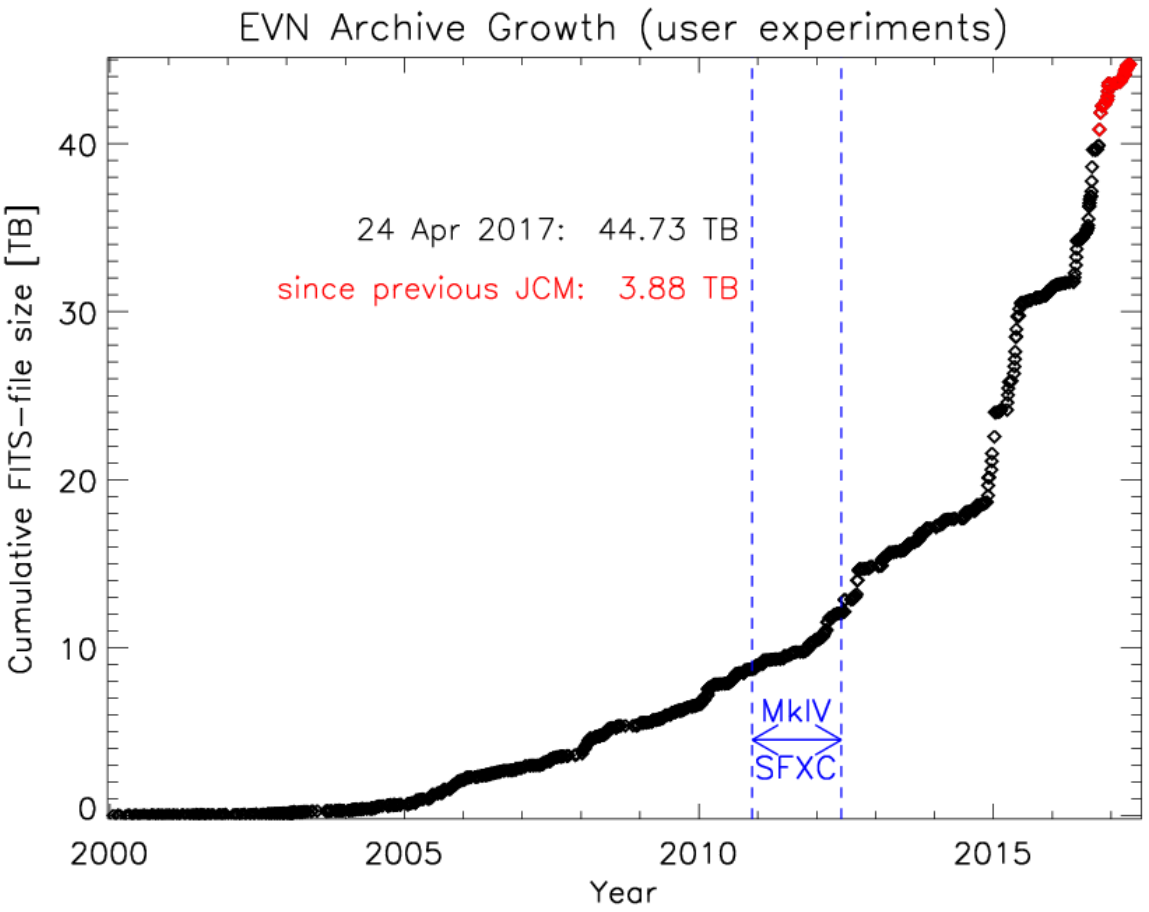
Demonstration that a complete system can be constructed from existing tools.

End of EOSC pilot

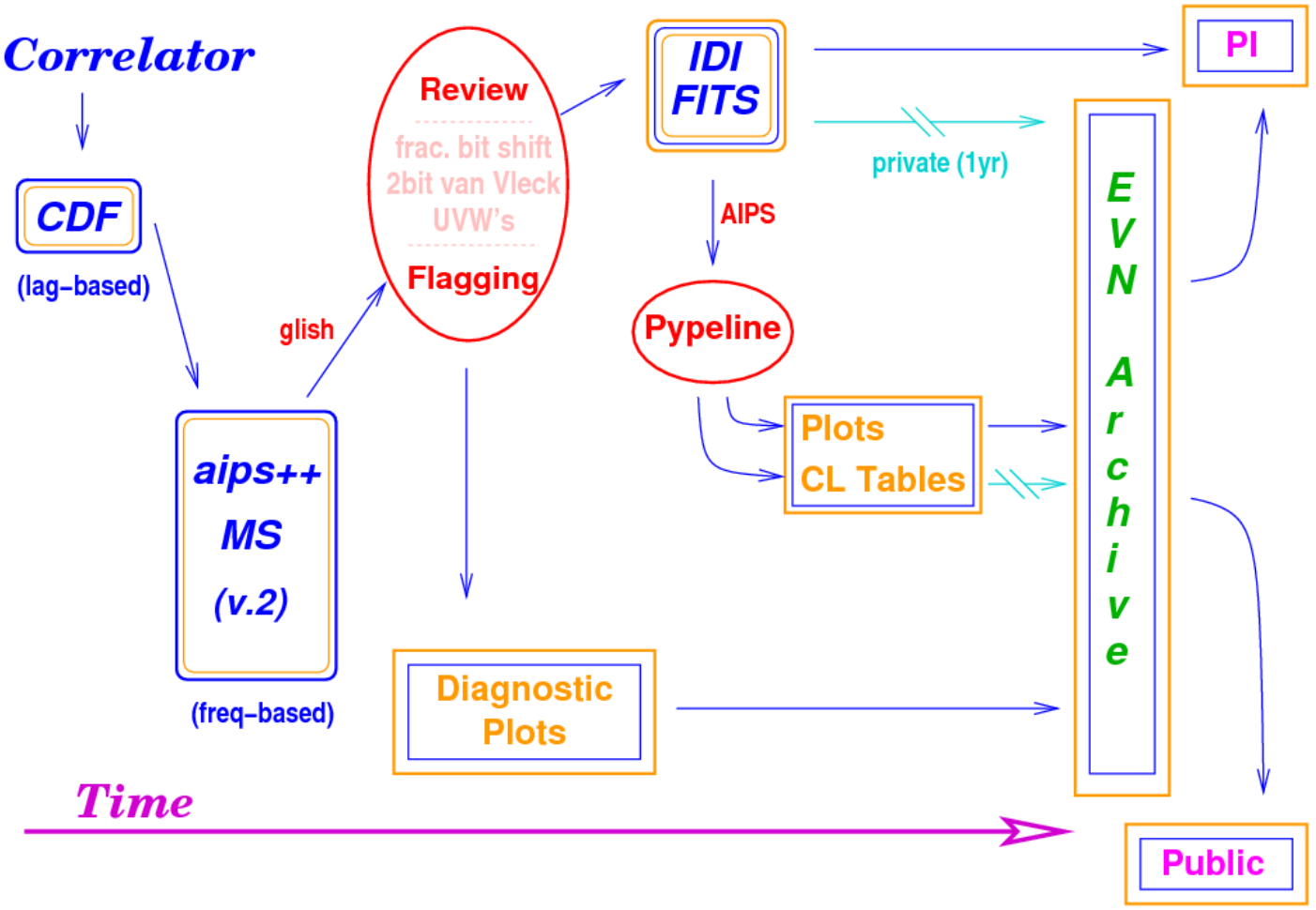
EVN as cloud version



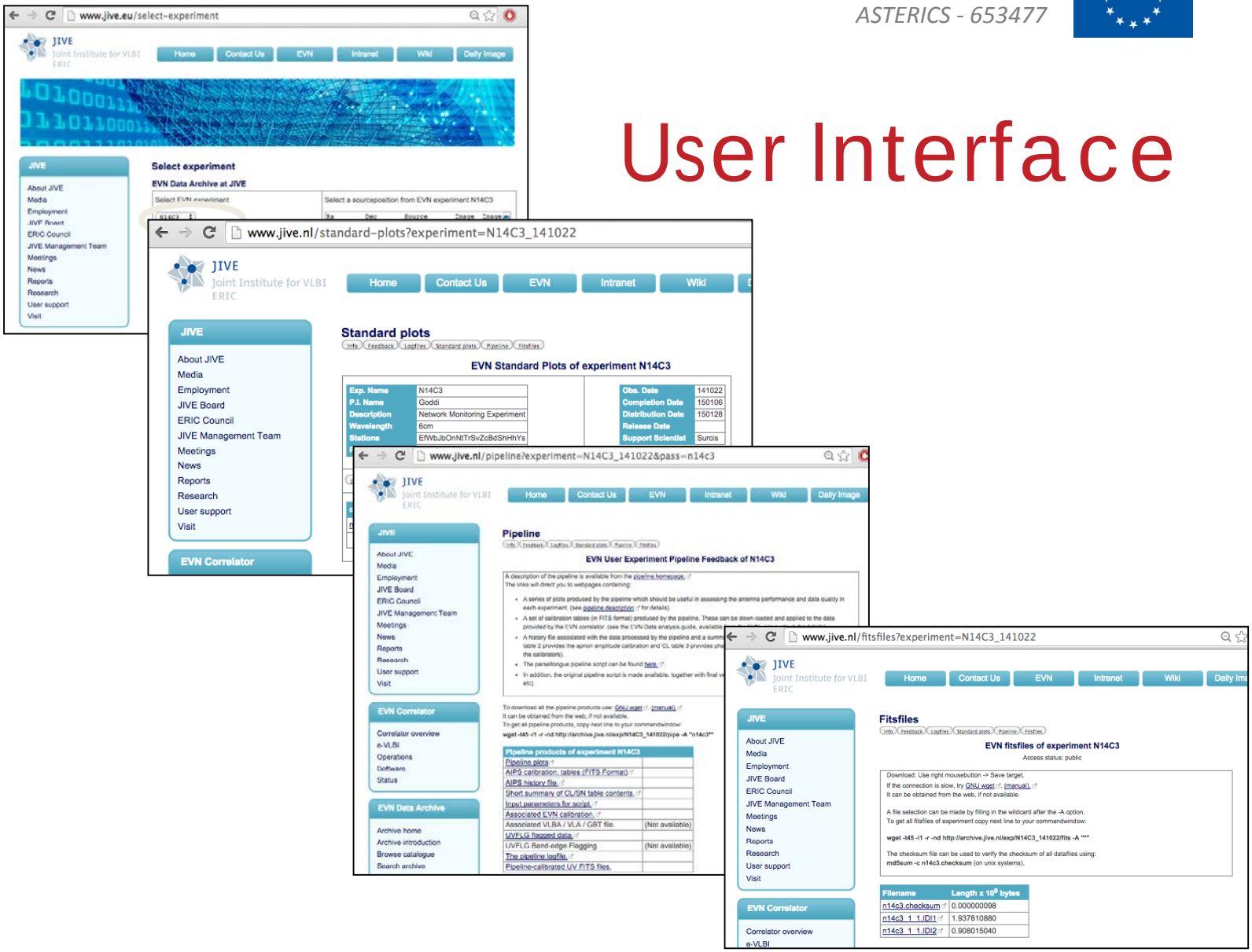
EVN Archive growth



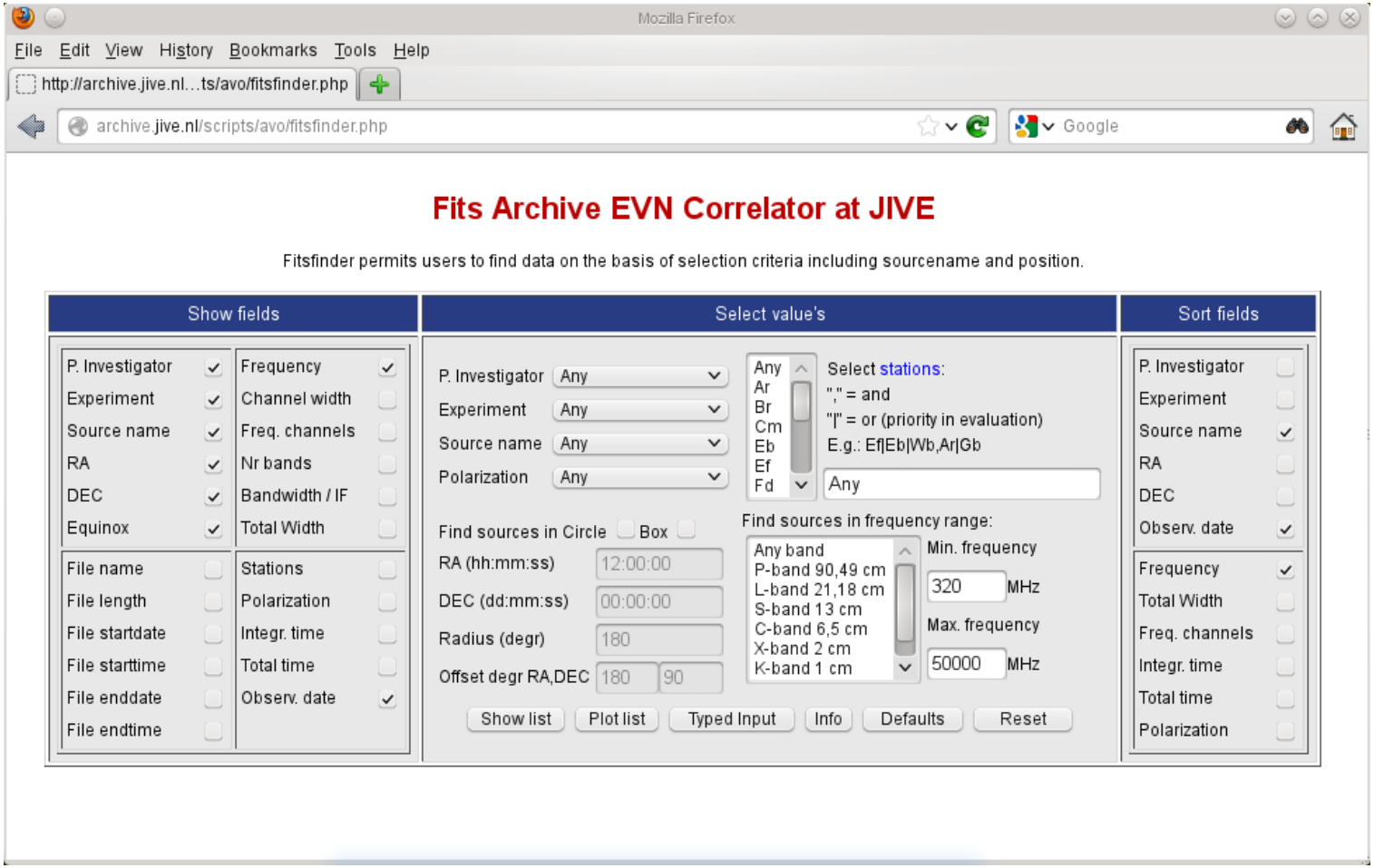
Processes


Correlator
CDF
(lag-based)
aips++
MS
(v.2)
(freq-based)
glish
Review
frac. bit shift
2bit van Vleck
UVW's
Flagging
IDI
FITS
AIPS
Pypeline
Plots
CL Tables
Diagnostic
Plots
PI
private (1yr)
EVN Archive
Public
Time

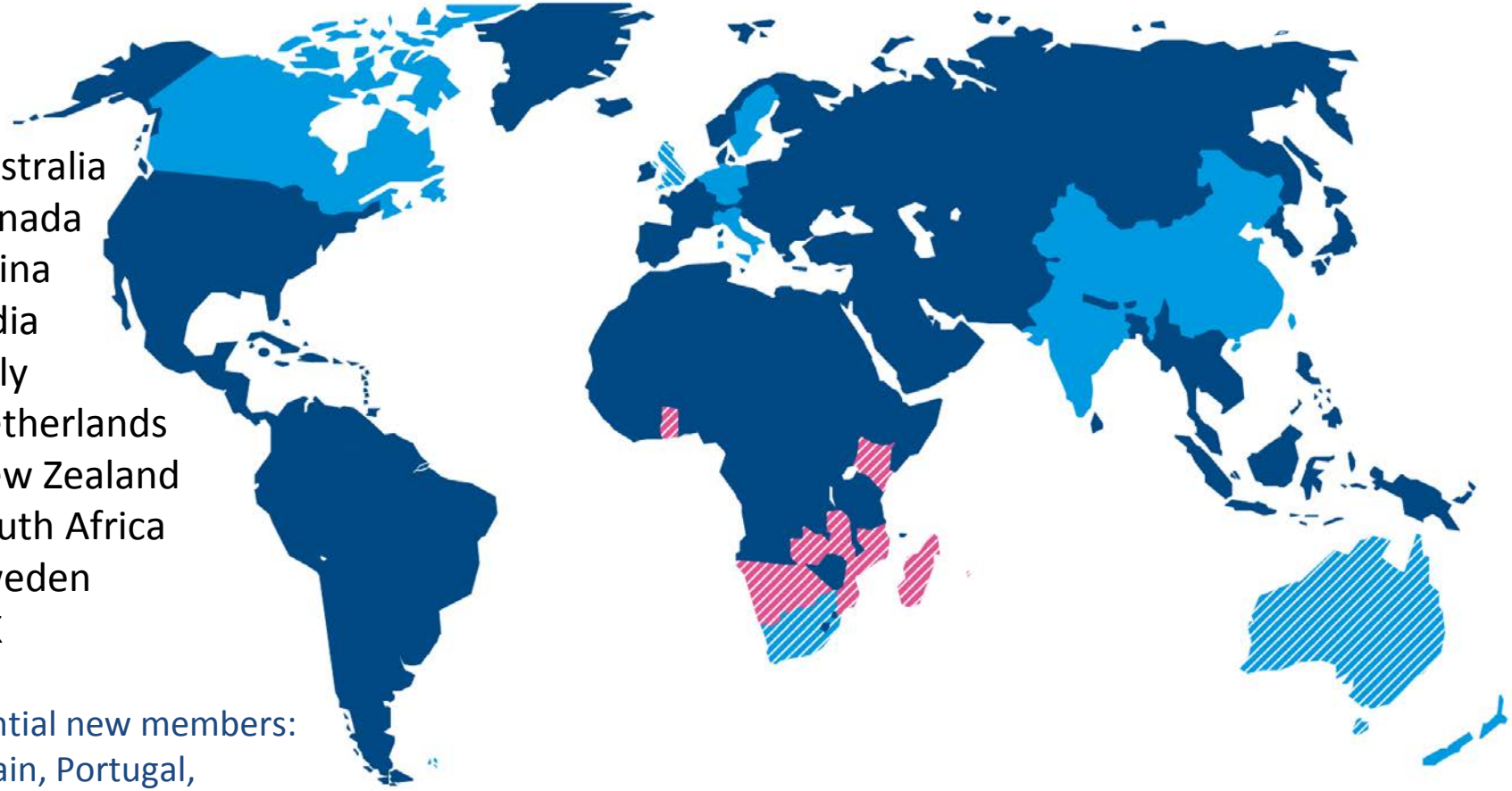
User Interface



Searching the EVN archive



Square Kilometer Array



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

Potential new members:

- Spain, Portugal,
- Germany, France, others…

The AENEAS Project

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

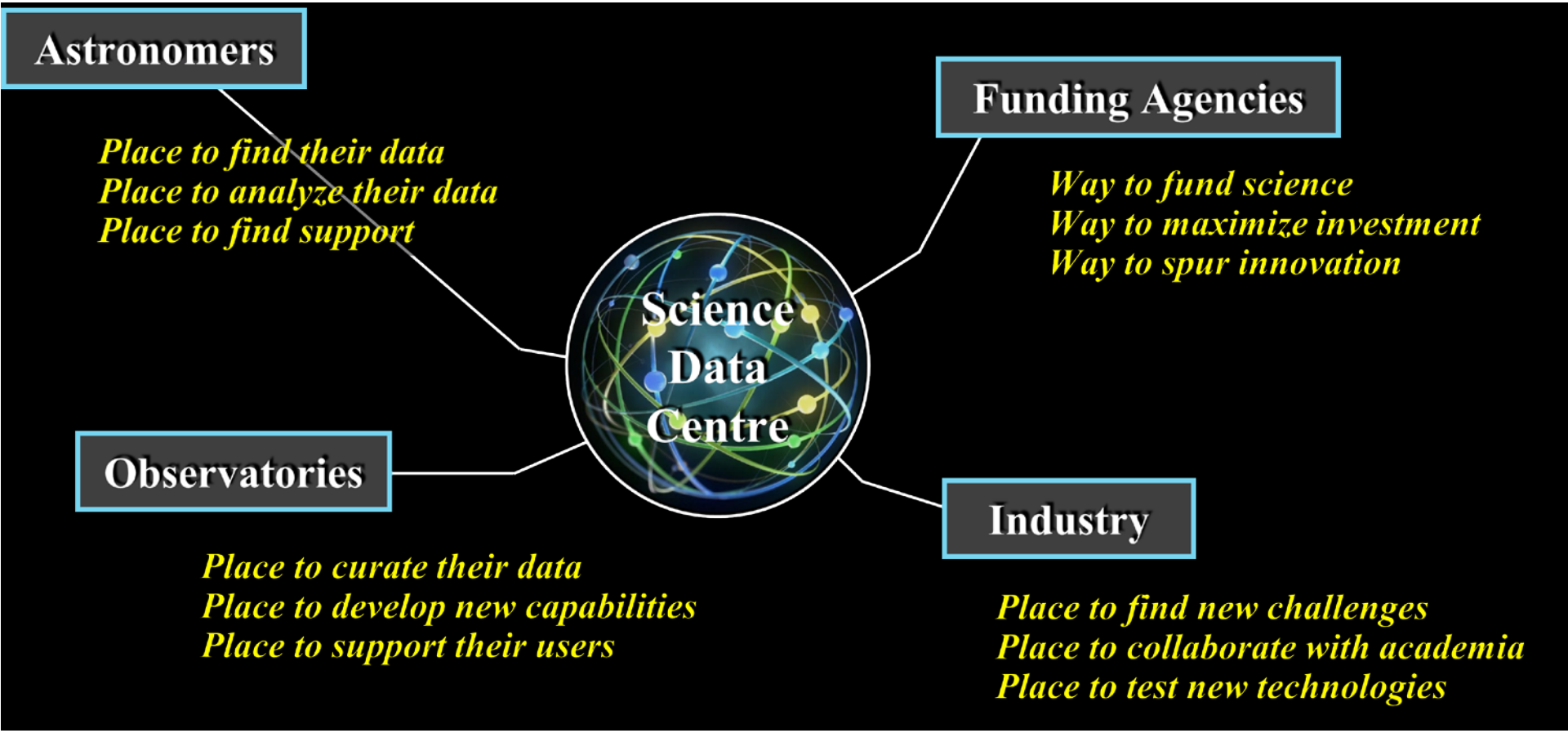
Design and specification of a distributed, European Science Data Centre (ESDC) to support the pan-European astronomical community in achieving the scientific goals of the SKA

EC H2020, 28 partners, three years (2017-2019)

- ❑ Governance Structure and Business Models
- ❑ Computing and Processing Requirements
- ❑ Data Transport and Optimal European Storage Topologies
- ❑ User Data Access and Knowledge Creation

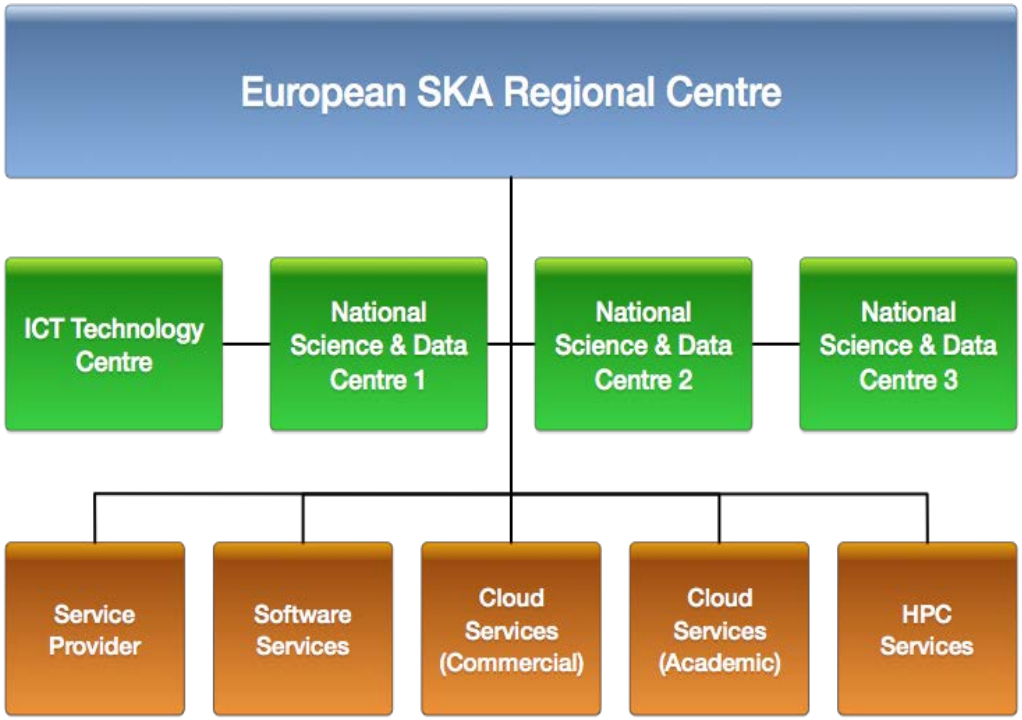
Global network of SD centers

What is a Science Data Centre?



European SKA Regional Centre

- Create a European-scale, federated Regional Centre for the SKA
- Provide resources for SKA science extraction to users
- Coordination with ICT communities, industry, and service providers
- Facilitate shared development, interoperability, and innovation
- European counterpart for engagement with other SRCs internationally



AENEAS & EOSC

- Building on the existing infrastructure
- using knowledge and requirements of current large archives and compute facilities
- scale increase of one to two orders of magnitude will stretch the capacity of any cloud or existing infrastructure to the limit.
- It is therefore very important that the design of the ESDC for SKA runs parallel to the emergence of the EOSC and learning from each other.

Open Questions

- How do we build and sustain large-scale research infrastructures that can support multiple domains?
- How do we deploy science as a service to researchers and maintain the cycle of discovery and innovation?
- How do we commodify public private partnerships?