ASTRON

NWO

Netherlands Institute for Radio Astronomy

Challenges and plans of SKA Regional Centres

Zheng Meyer-Zhao

ASTRON Science Data Centre Development Lead

Email: meyer@astron.nl

Square Kilometre Array

Headquarters @ Jodrell Bank Observatory, UK

Two sites: South Africa & Australia

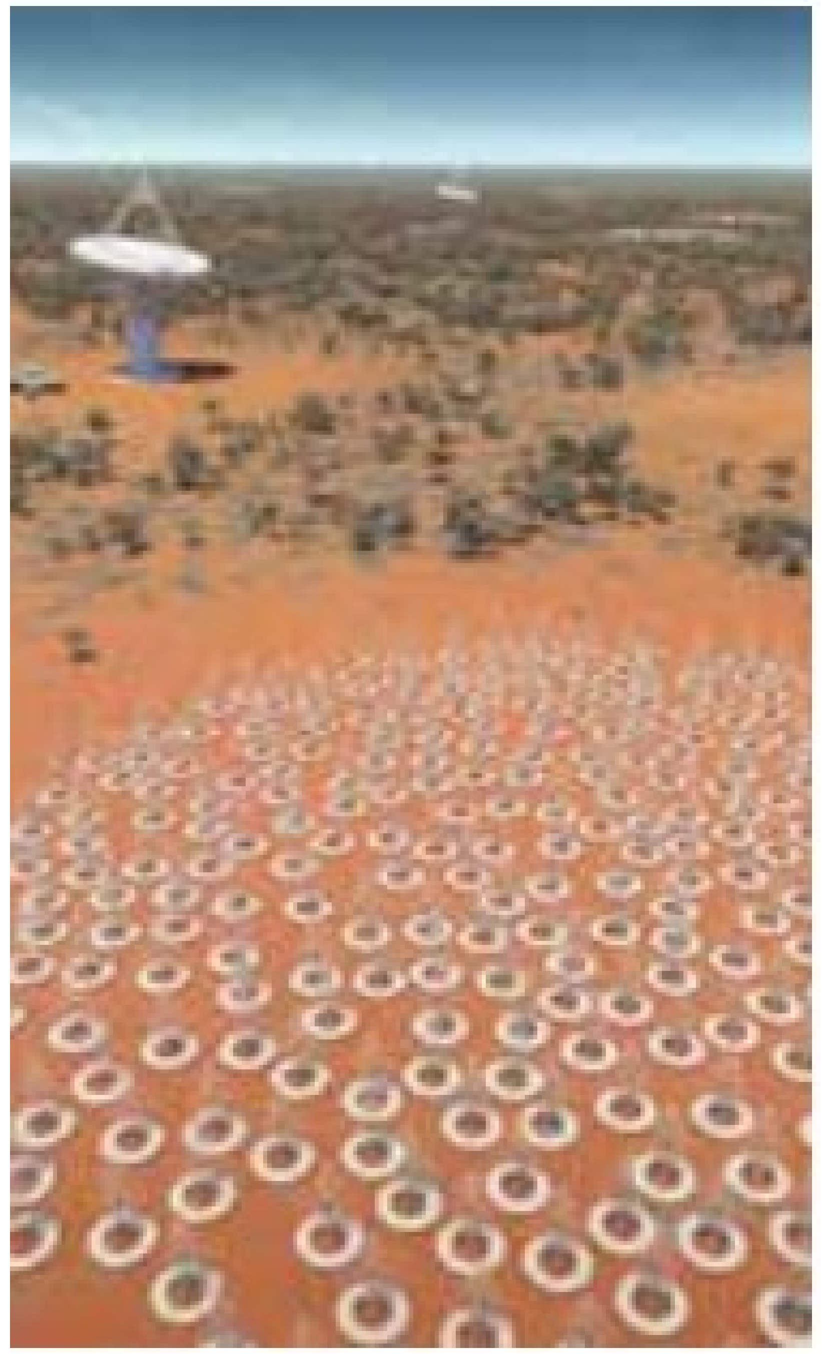
South Africa (Credit: SKA SA)

Australia (Credit: Rob Millenaar (ASTRON)

Two antenna types:

- Dishes
- Low frequency aperture array

Provide continuous frequency coverage from 50 MHz to 14 GHz






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

Location:
South Africa

Frequency range:
350 MHz to **14 GHz**

~200 dishes
(including 64 MeerKAT dishes)

Total collecting area:
33,000m²

Maximum distance between stations:
150km

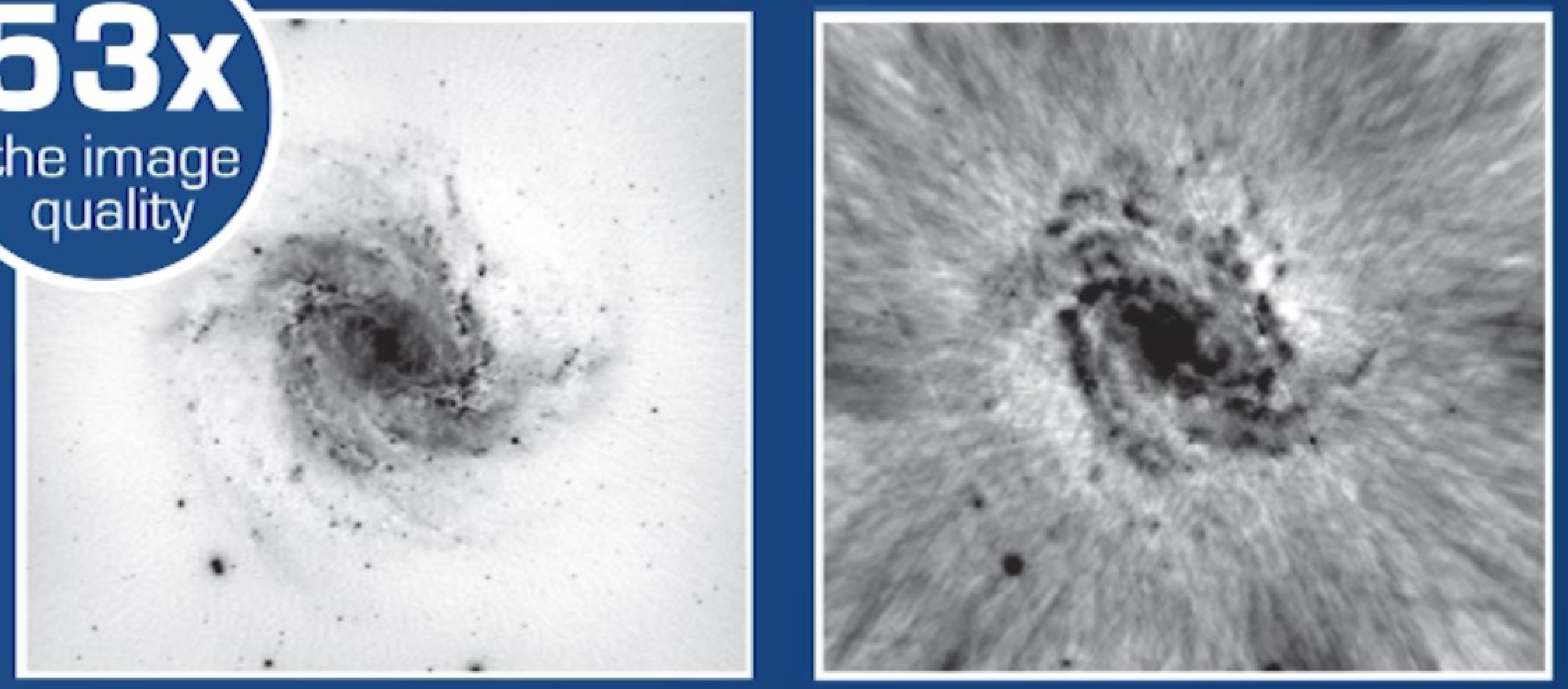
How SKA1-mid compares with the Janksy Very Large Array (JVLA), the current best similar instrument in the world.

4x the resolution

5x more sensitive

60x the survey speed

53x the image quality



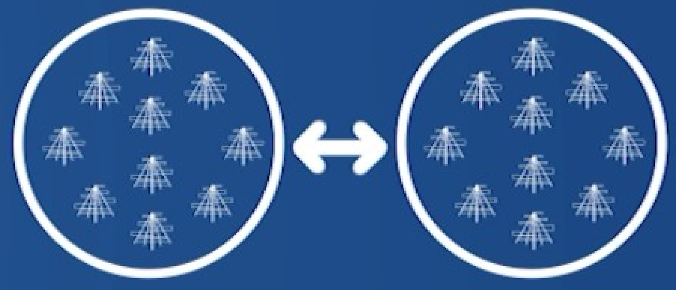
SKA1-low - the SKA's low-frequency instrument

Location: Australia

Frequency range:
50 MHz to 350 MHz

~130,000
antennas spread between
500 stations

Total collecting area:
0.4km²

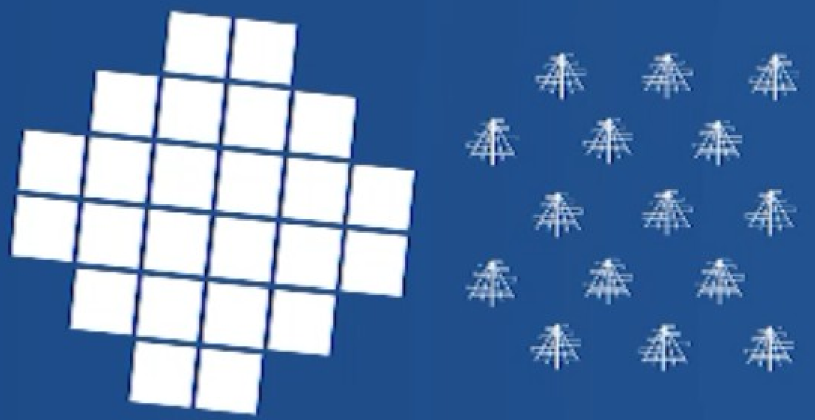


Maximum distance between stations:
>65km

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

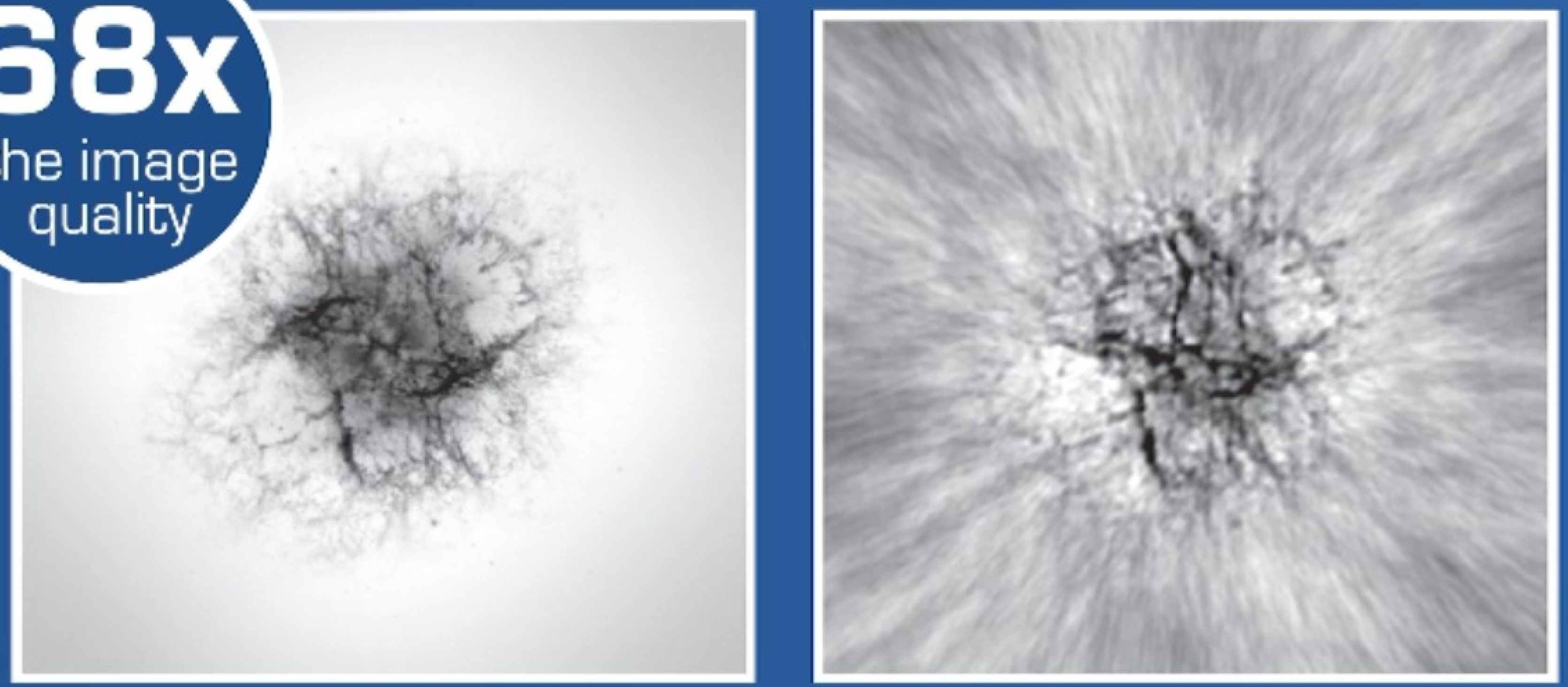


25% better resolution

8x more sensitive

135x the survey speed

68x the image quality



Today's Astronomy is The History of the Universe

Testing General Relativity
(Strong Gravity, Gravitational Waves)

Cosmic Dawn
(First Stars & Galaxies)

Galaxy Evolution
(Normal Galaxies at z~2-3)

Cradle of Life
(Planets, Molecules, SETI)

Cosmology
(Dark Matter, Large-scale Structure)

Exploration of the Unknown

Members of the SKA Organisation
Host Countries: Australia, South Africa, United Kingdom

African Partner Countries

SKA Observatory will be established as an Intergovernmental Organisation in 2020, taking over from the SKA Organisation. It will undertake the construction and operation of the telescope.

As of March 2019, confirmed SKA Observatory members are

SKA Observatory Convention signed on 12 March 2019

Members: AU, UK, ZA, CA, CN, NZ, IN, NL, IT, SE, DE, ES, FR

Non full members will not have voting rights.

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
HL-LHC 2026
Raw Data
600 PB
Google
Internet
Archive
15 EB
SKA Phase 2 2030
Science Data
1 EB
HL-LHC 2026
Science Data
1 EB
(Ian Bird, CERN)

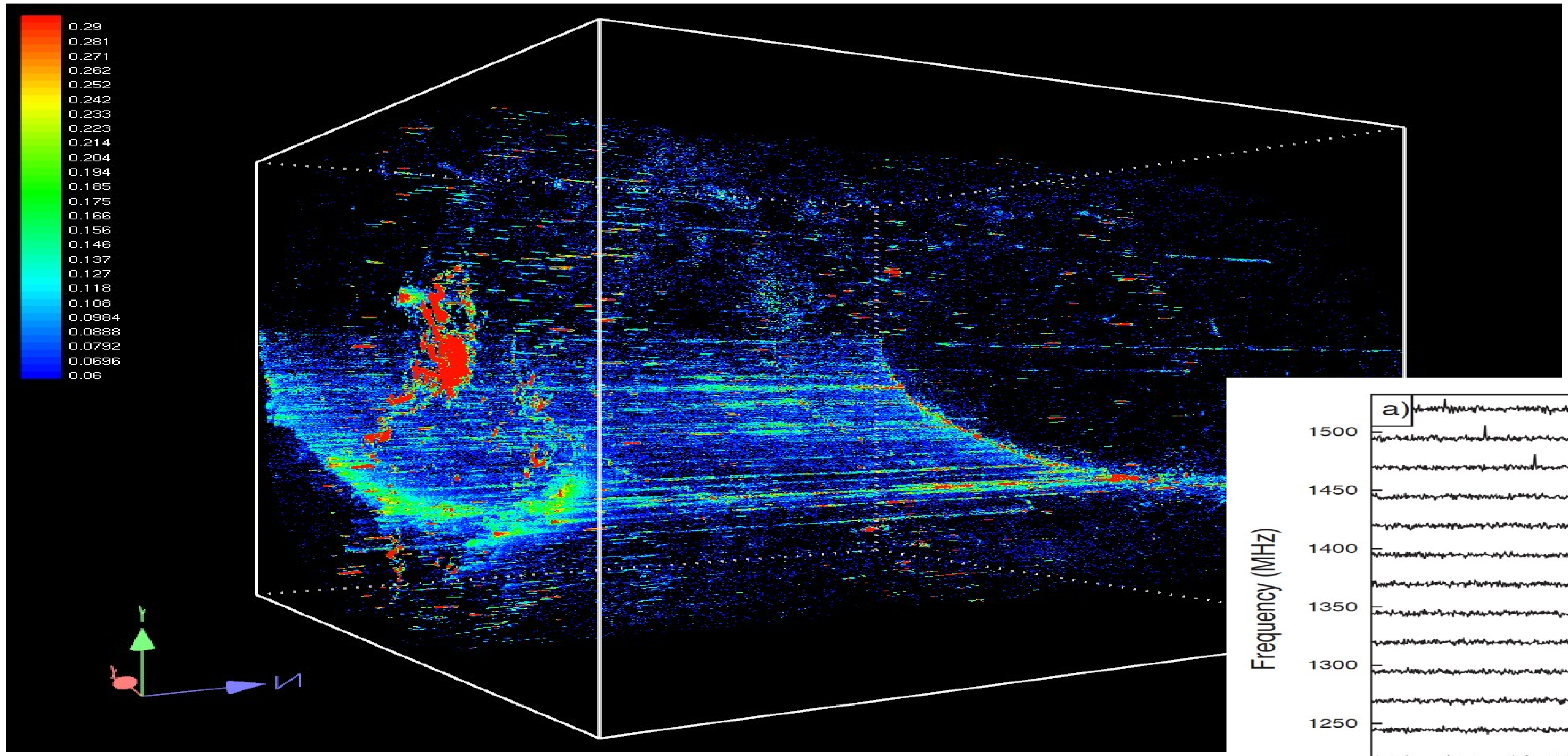
SKA Precusors and Pathfinders

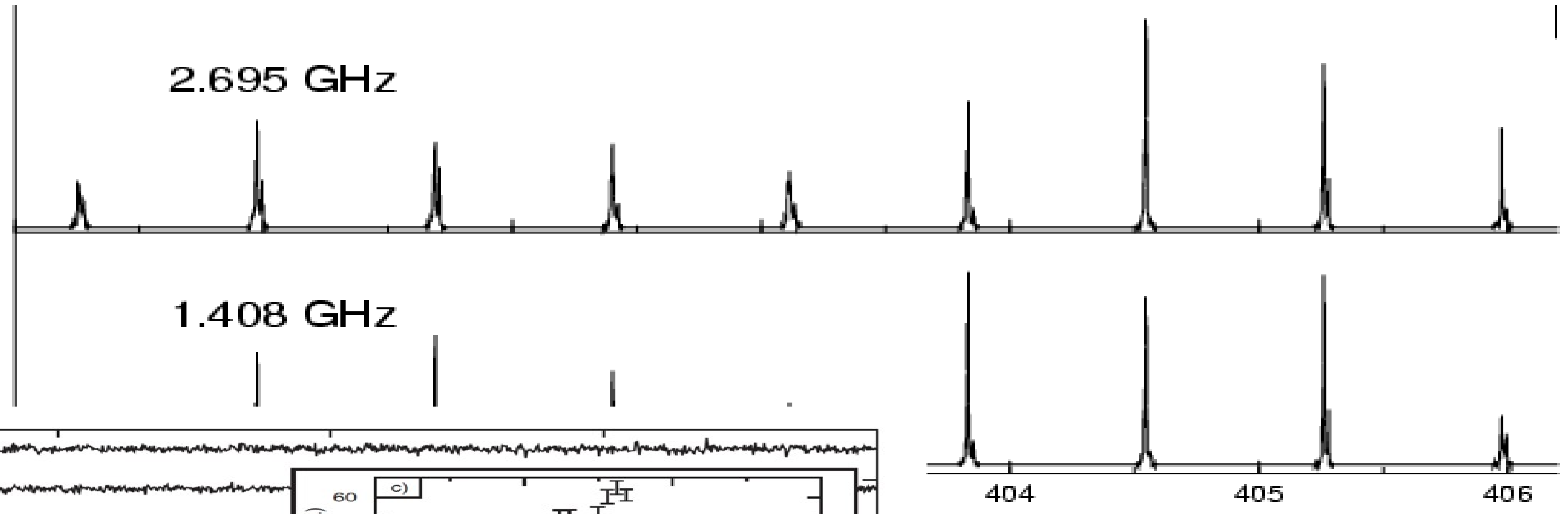
Data flow challenges

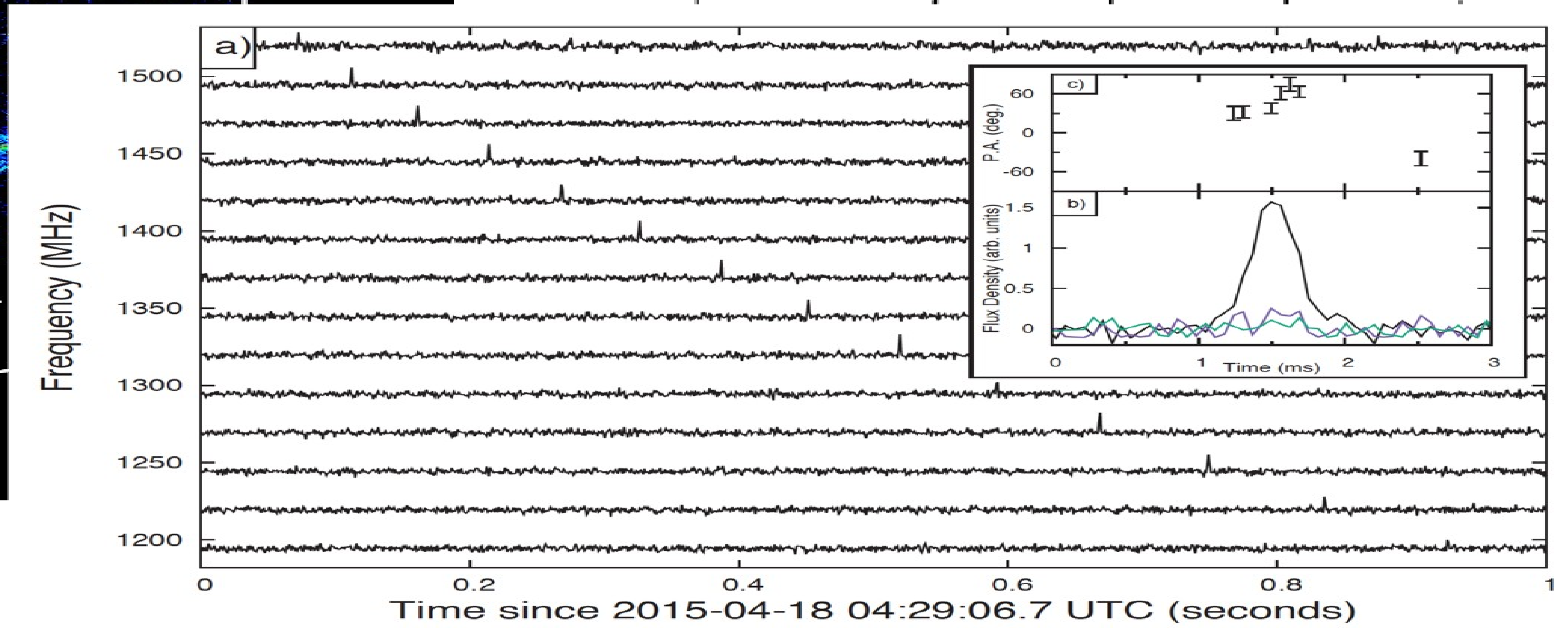
10 – 50 x data rate reduction by SDP in observatory.
600 PBytes/year output to users.

SKA Observatory data products – not raw data

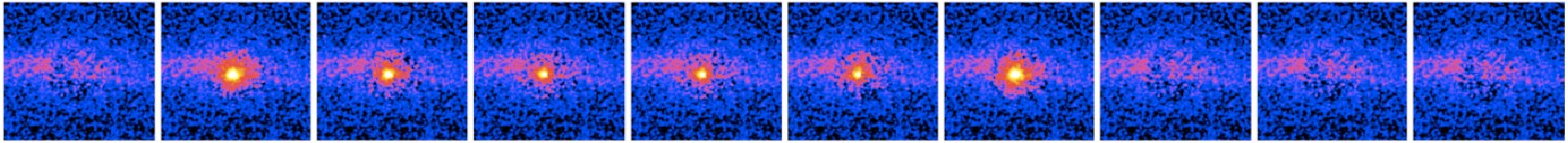
Image cubes up to 1PByte…







Keane et al., 2016, Nature, 530, 453



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 transported from the sites to Europe?

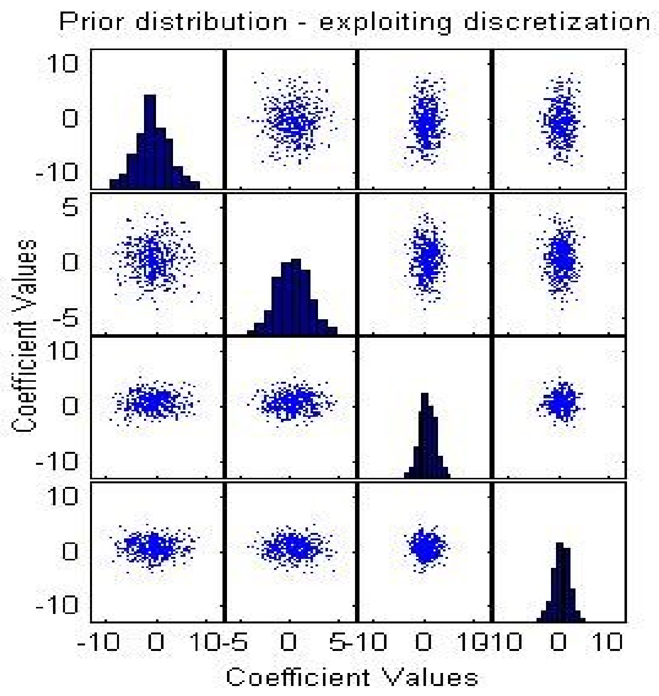
How can we take optimal advantage of existing infrastructure?

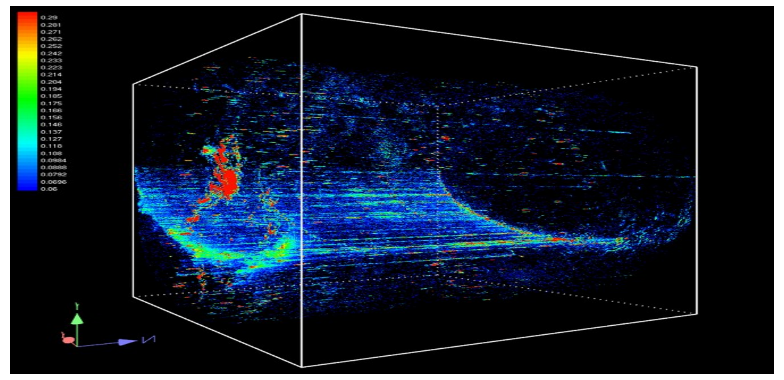
What are the processing requirements and technologies to consider?

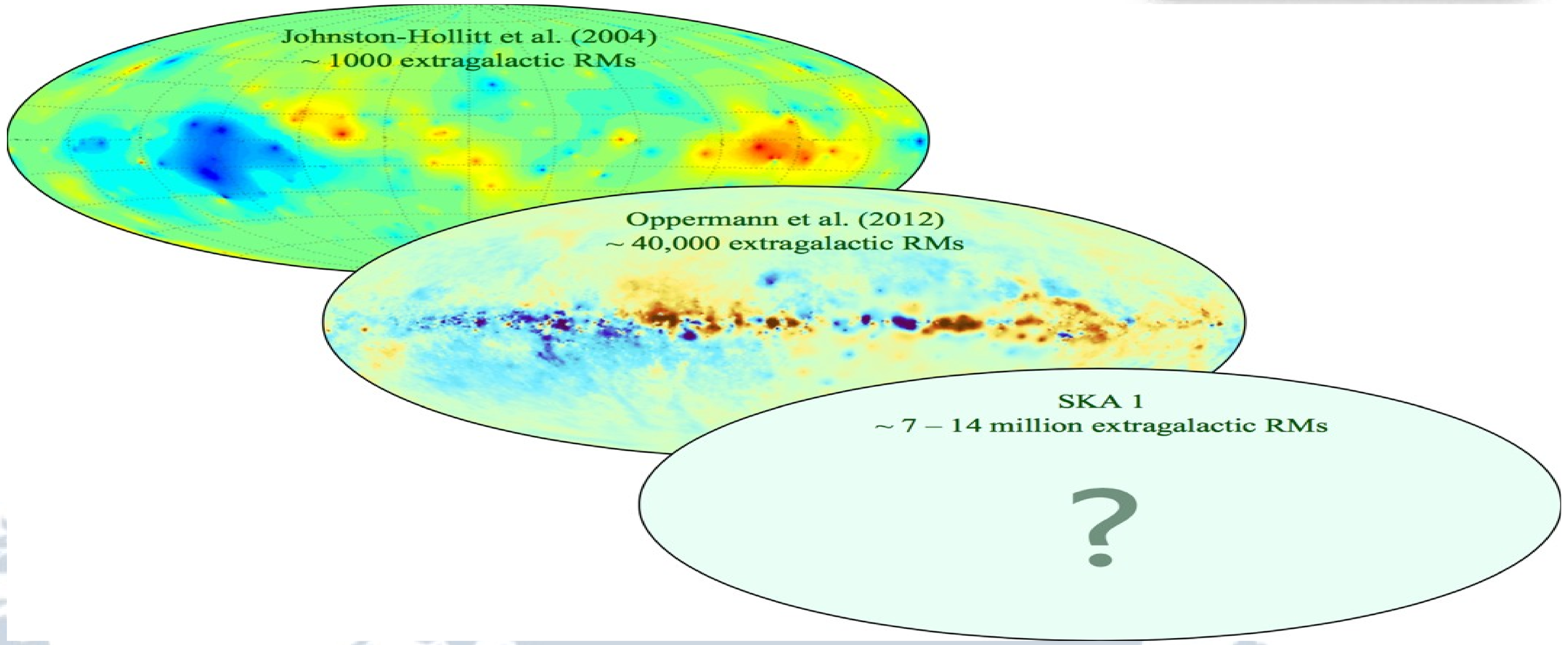
What interfaces, tools, and techniques will users need for analysis?

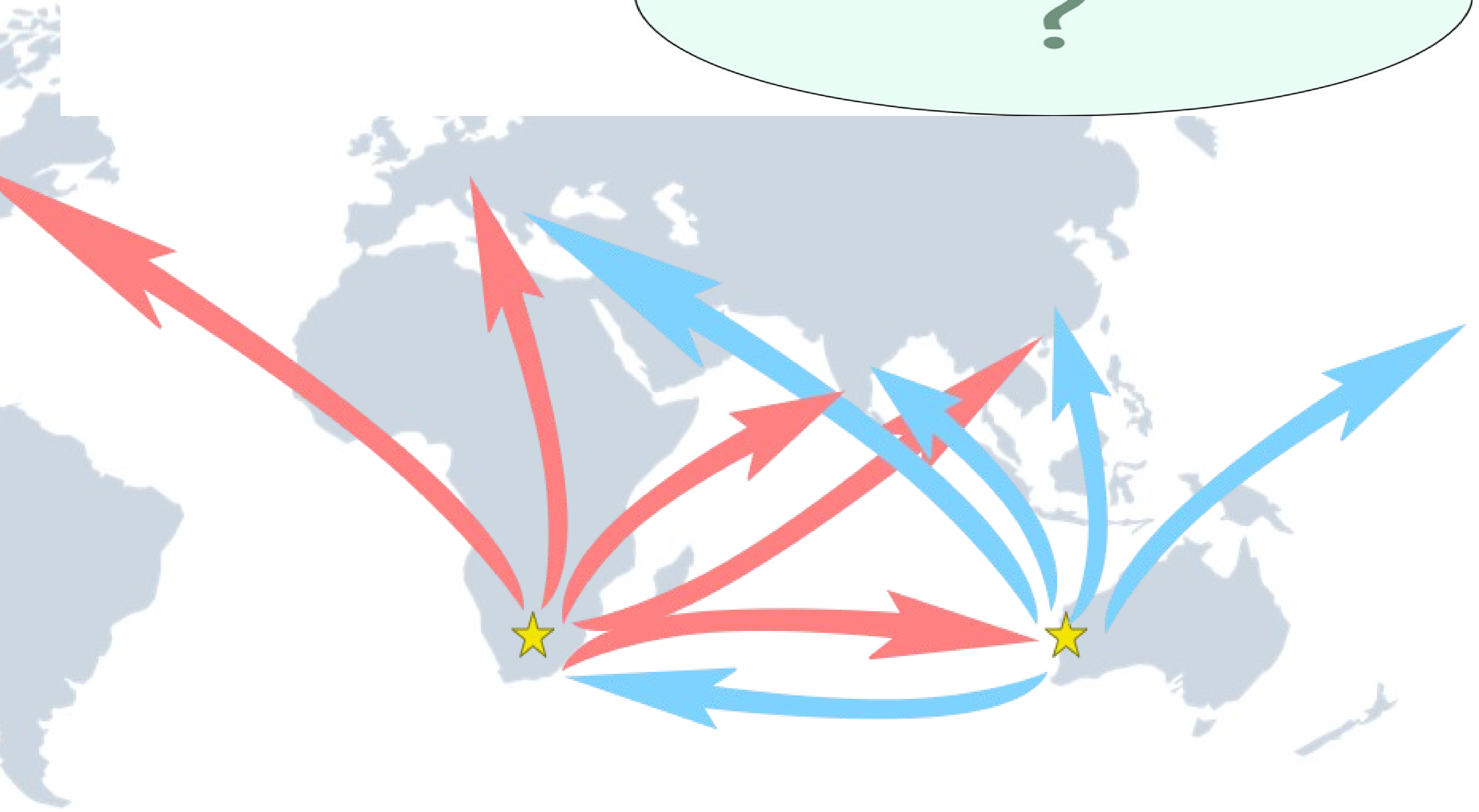
How do we setup and operate an international network of SRCs?

What are SRCs for?









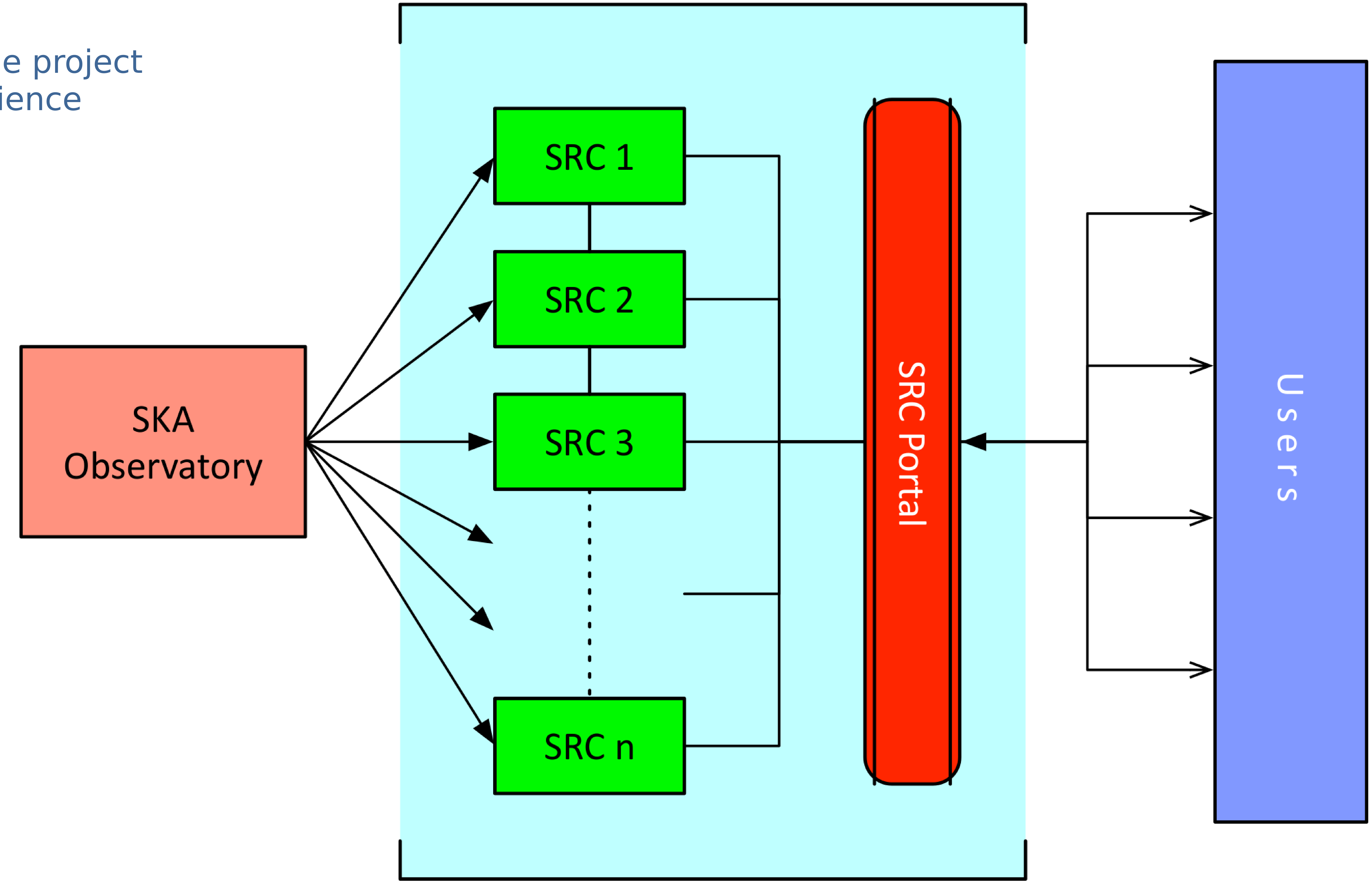
- A collaborative network for collaborative science
- Transparent and location-agnostic interface for users
- All SKA users access their project data via SRCs
- A forum for development of software tools: analysis, modelling, visualisation
- Local user support functions

SKA Regional Centres

Outside the cost cap for the project
Essential for delivery of science

Ideals:

- Resources pledged into system
- Access given by user-linked data privileges
- Accounting to track resource use
- Users do not need to know where “their” data are



Strategic Partnerships and Opportunities

CIRADA
CANADIAN INITIATIVE FOR RADIO ASTRONOMY DATA ANALYSIS

ERIDANUS

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

CERN-SKA MoU

ESCAPE
European Science Cluster of Astronomy & Particle physics
ESFRI research infrastructures

EOSC pilot
The European Open Science
Cloud for Research Pilot Project

EOSC-hub

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

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)

- Computing and Processing Requirements
- Data Transport and Optimal European Storage Topologies
- Data Access and Knowledge Creation
- User Services

Final deliverable: preliminary ESDC Design and Implementation Plan

https://www.aeneas2020.eu/

AENEAS Workpackages

WP1: Project Management – lead: ASTRON (NL)

WP2: ESDC Design & Governance - lead: ASTRON (NL), Onsala (SE)

WP3: Computing and Processing Requirements - lead: Cambridge & Manchester (UK)

WP4: Data Transport and Optimal European Storage Topologies - lead: GEANT (EU)

WP5: Data Access and Knowledge Creation – lead: INAF (IT)

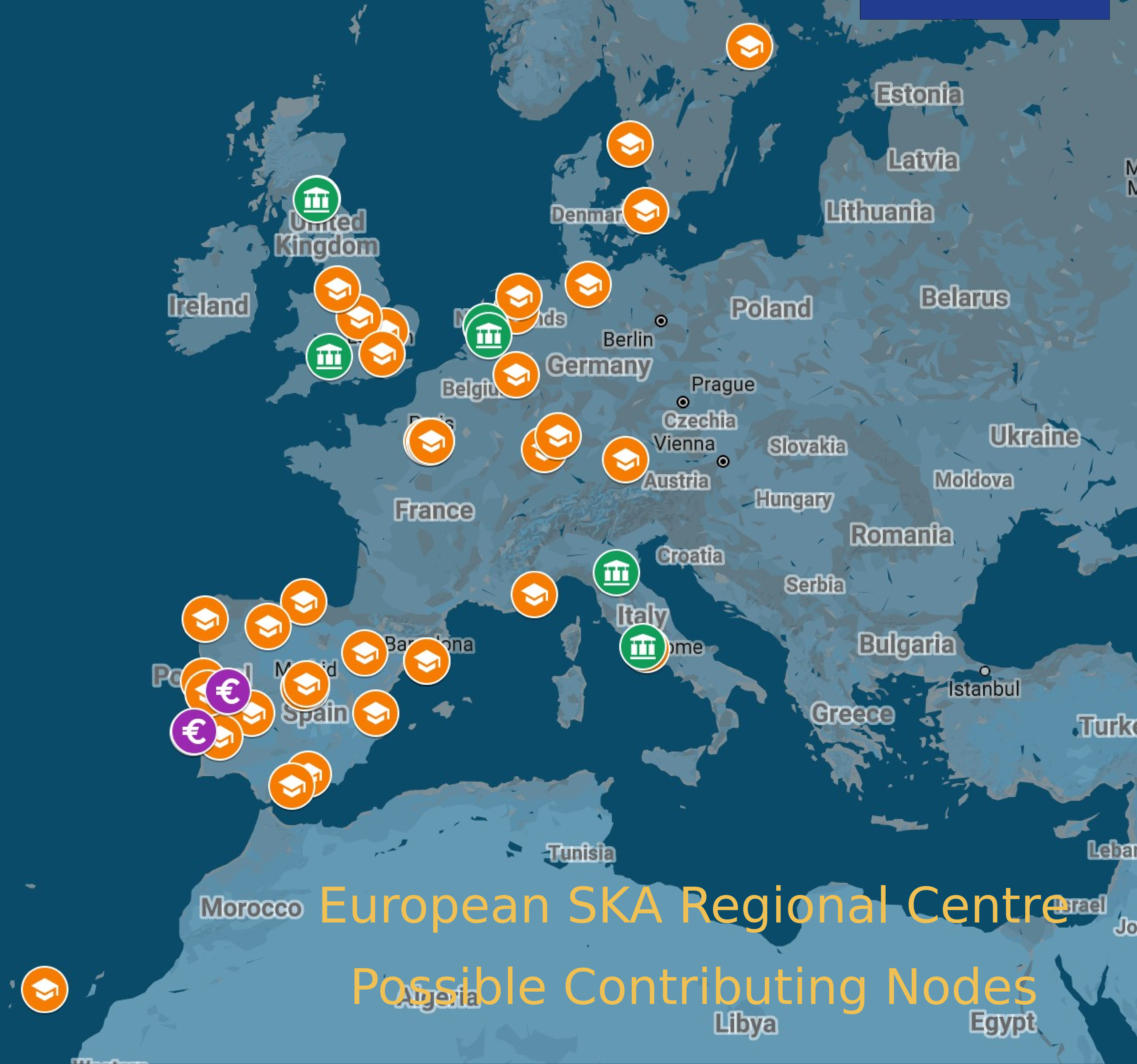
WP6: User Services – lead: EGI (EU)

WP2: ESDC Design & Governance

Survey of Potential Providers

- *Over 50 expressions of interest*
- *Mixture of scientific institutes, infrastructure providers, and industrial partners*

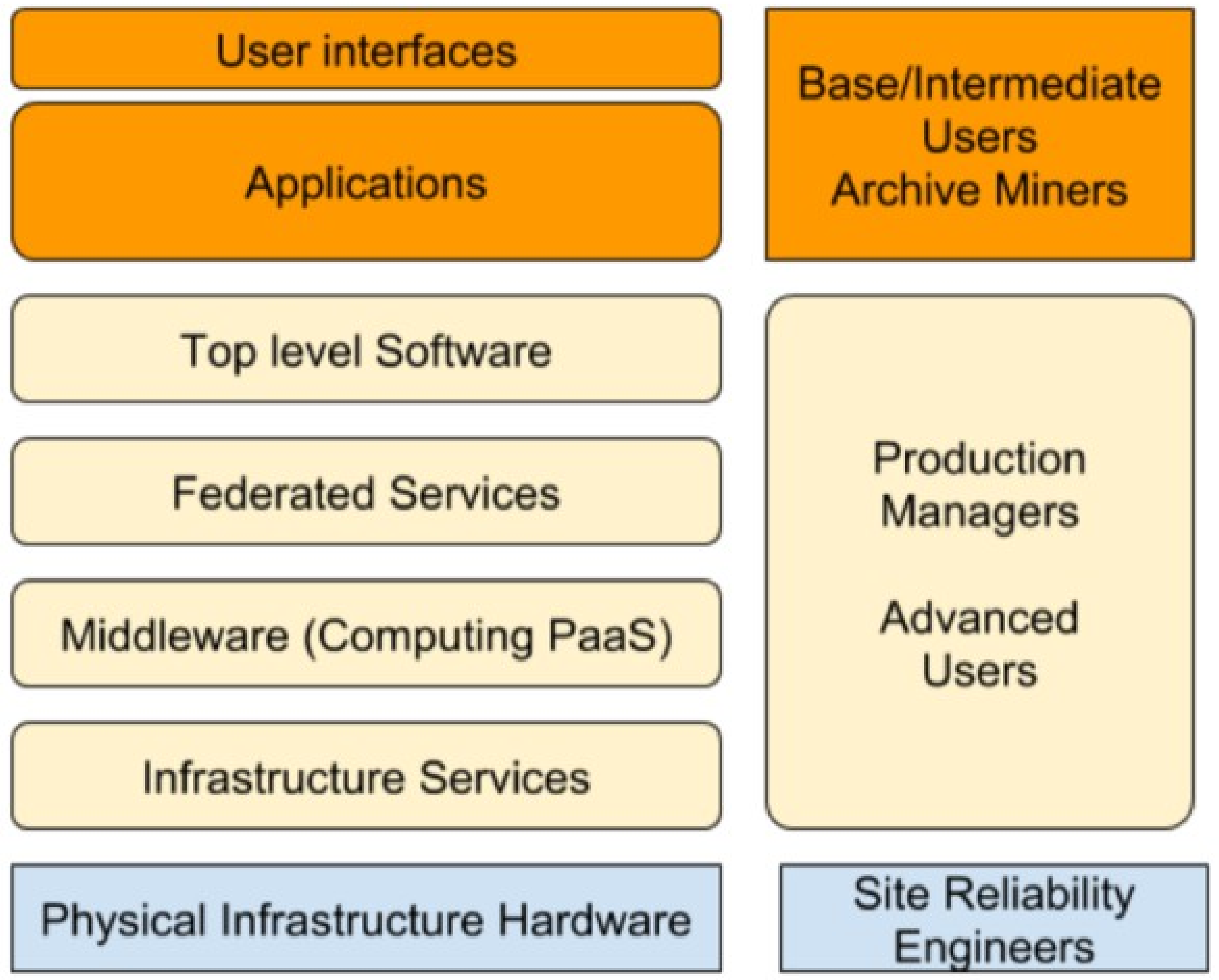
- *ESDC Requirements based on those developed by SRCCG*
- *Final deliverable: preliminary ESDC Design and Implementation Plan*
- *User input needed!*



European SKA Regional Centre
Possible Contributing Nodes

WP3: Computing and Processing Requirements

- Analysis of compute load, data transfer and data storage anticipated as required for SKA Key science
- Suggested solutions to address each of the key software areas associated with running a distributed ESDC
- Initial System Sizing



Reprocessing and post-processing

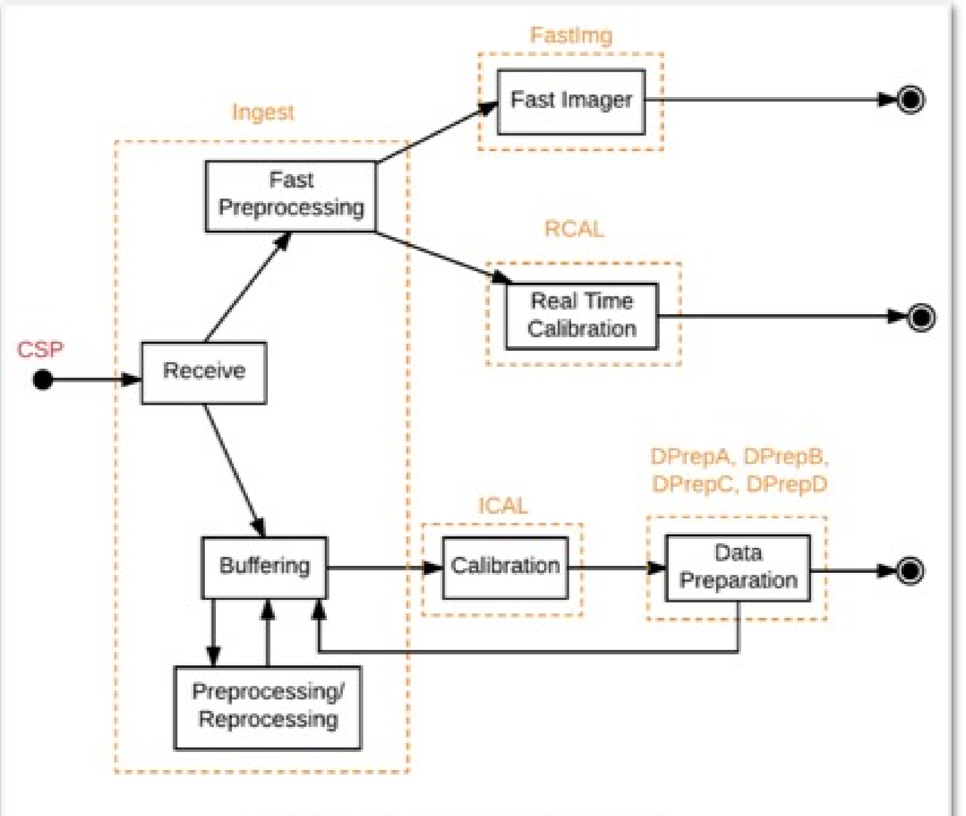


Figure 1: Overview of SDP pipelines [A07].

Processing is being examined in terms of (1) compute load; (2) memory requirements; (3) potential for distribution; (4) suitability of platform.

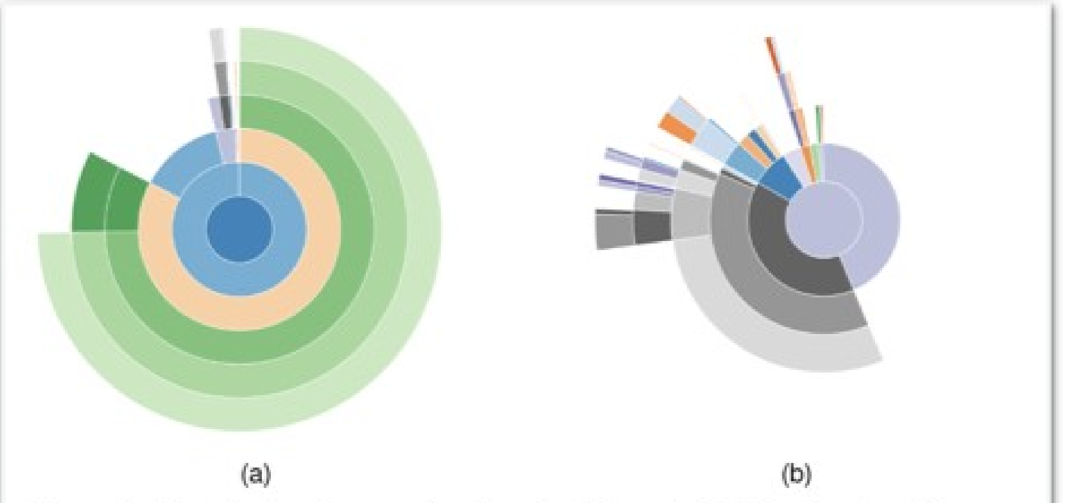
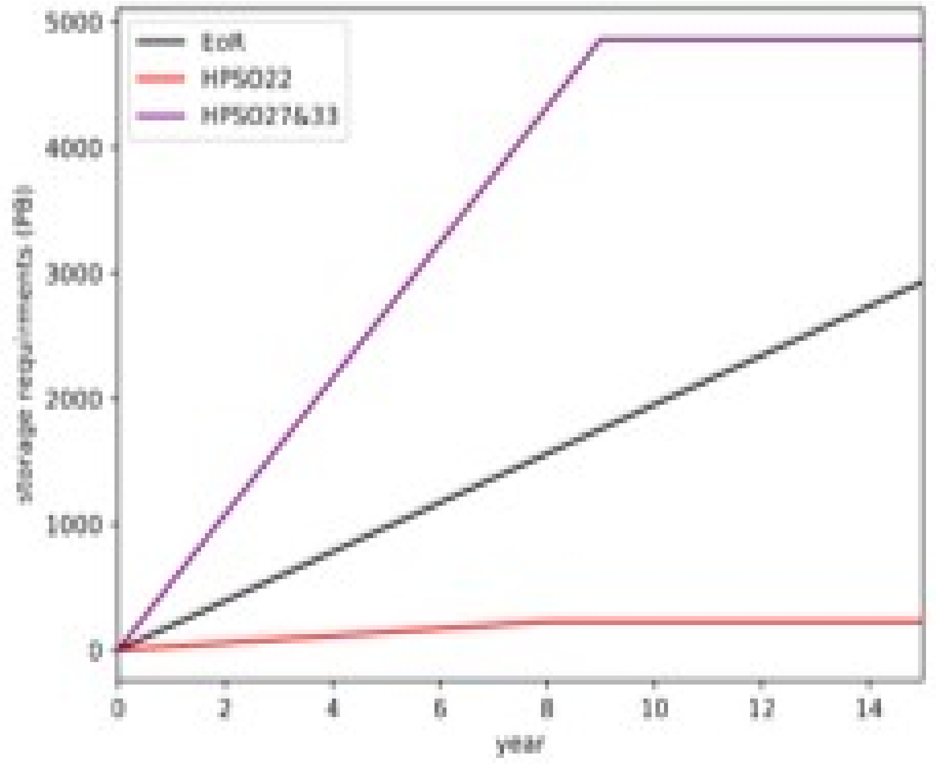
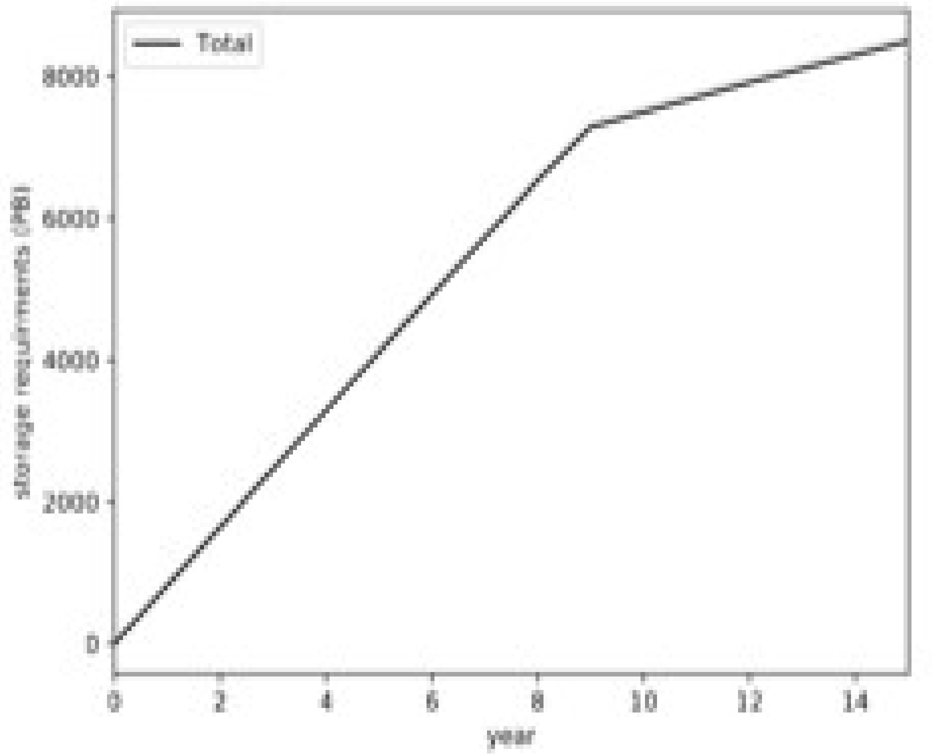


Figure 9: Visualisation of processing time breakdown for PyRMSynth. (a) The overall breakdown, dominated by `write_output_files` (green); (b) the breakdown beneath `compute_dirty_image`, dominated by the FFT (grey). See text for details.

Minimum for HPSOs ~13 PFlops

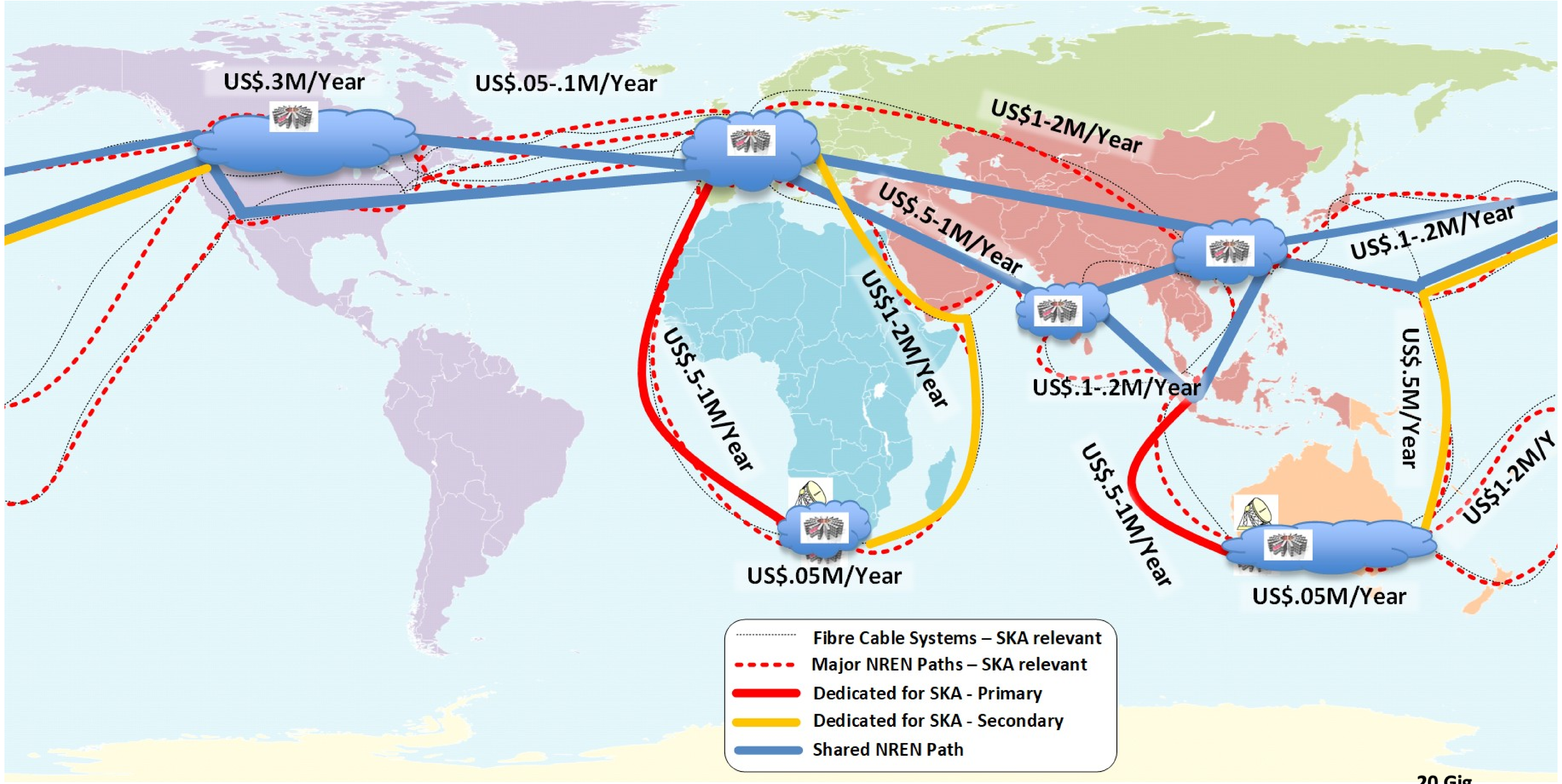
Storage estimates for HPSOs

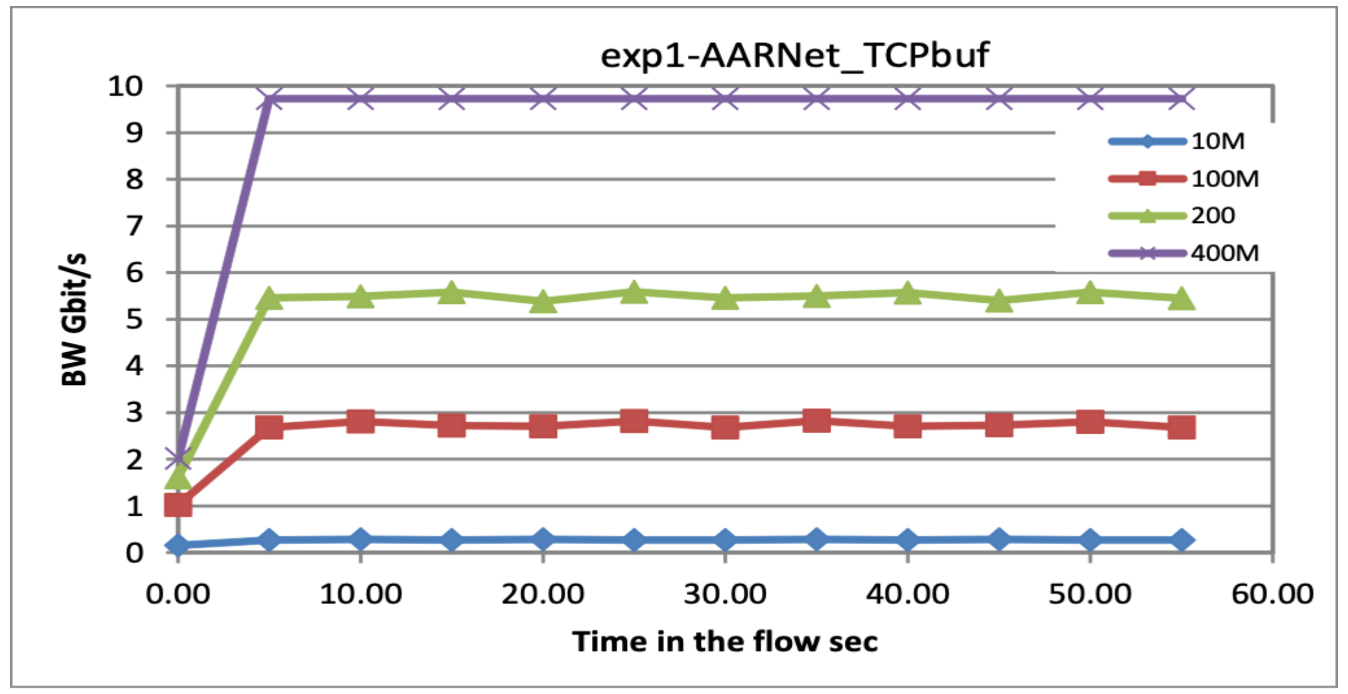




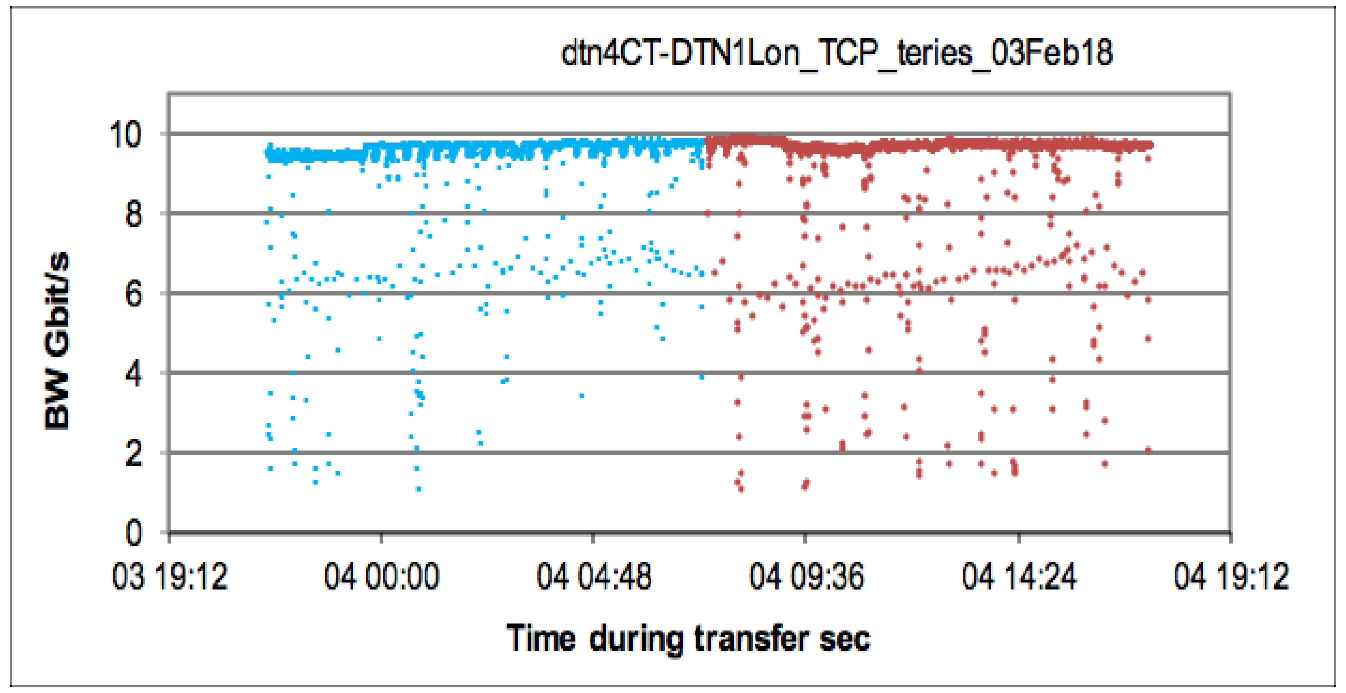
10 ExaBytes over first 15 years of SKA operations

WP4: Data Transport and Optimal Storage Topologies



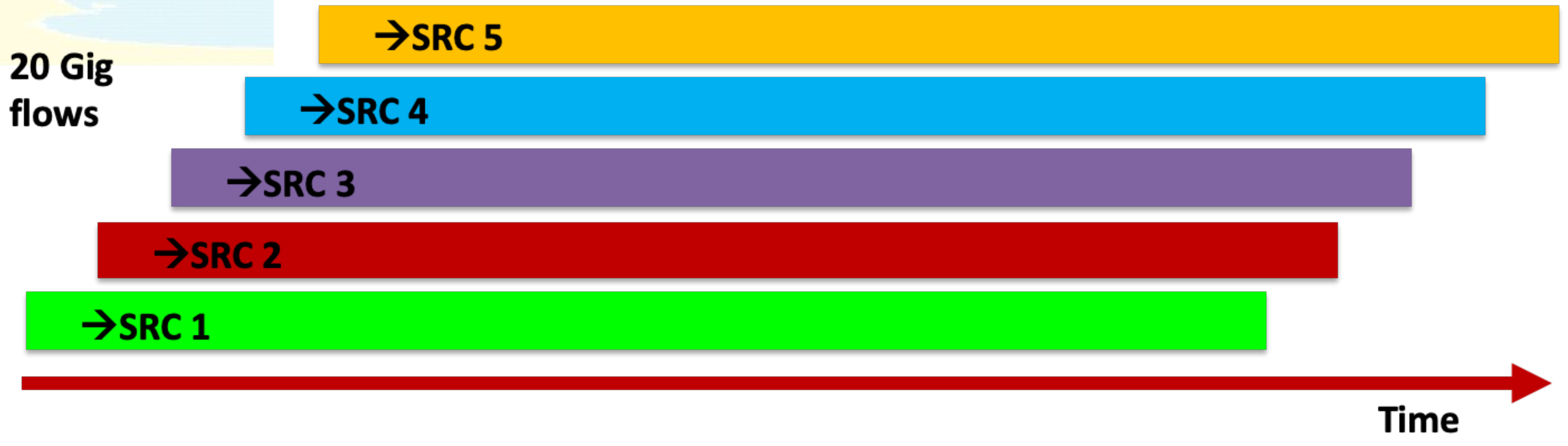


GÉANT London to AARNet Canberra



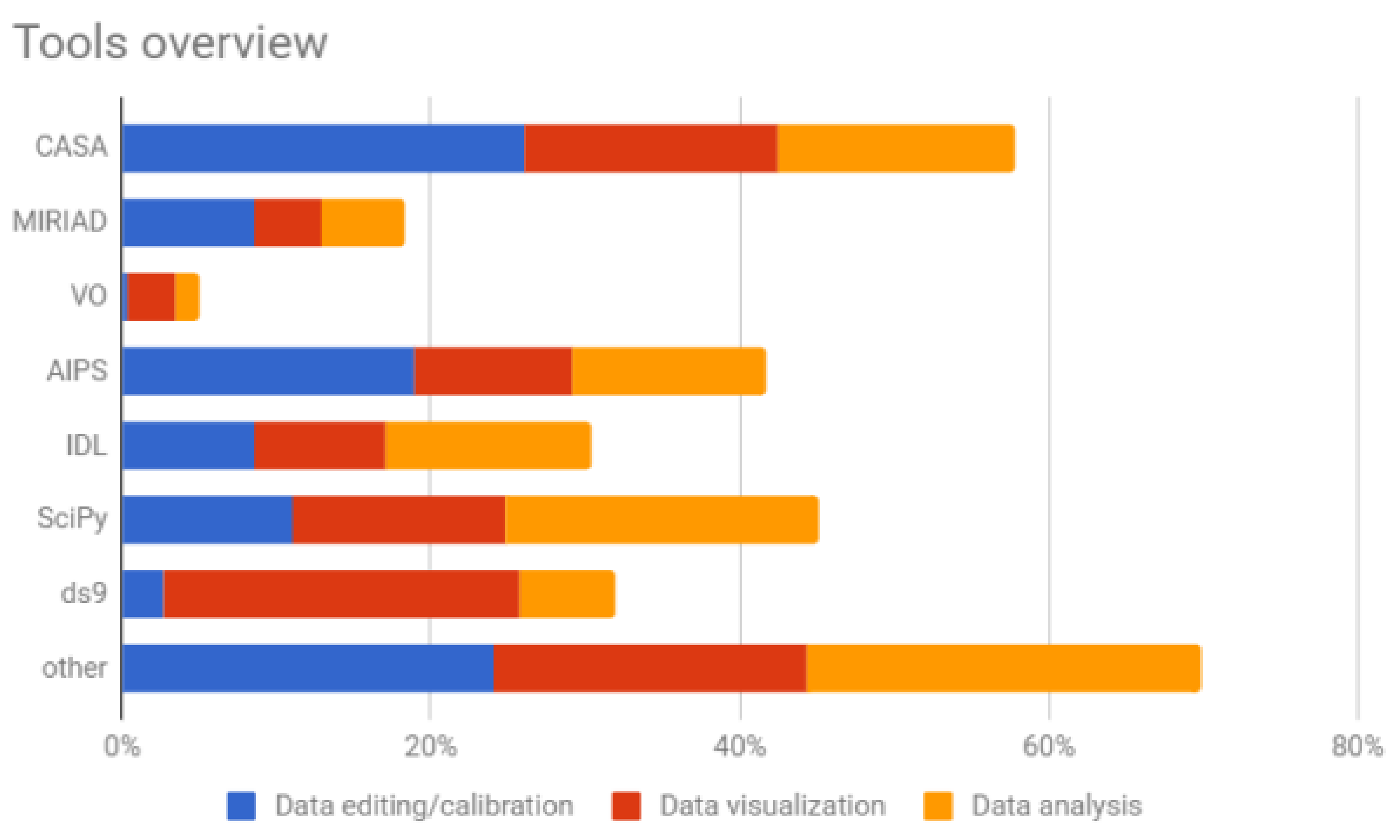
SANReN Cape Town to GÉANT London

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

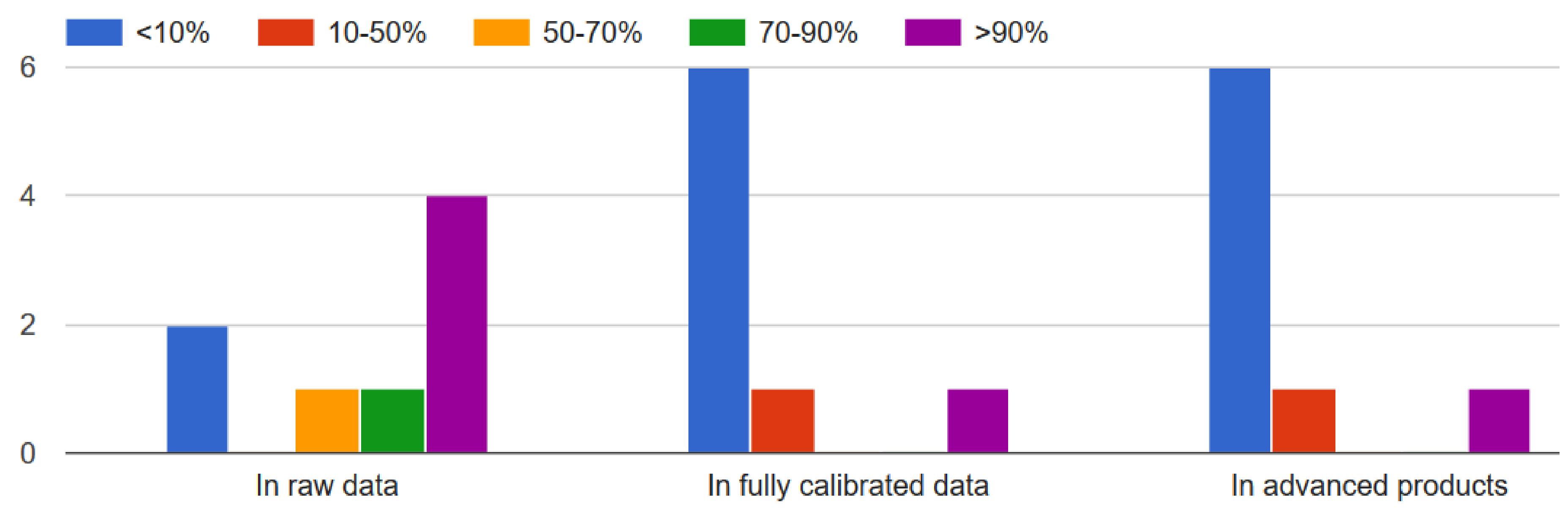


WP5: Data Access and Knowledge Creation

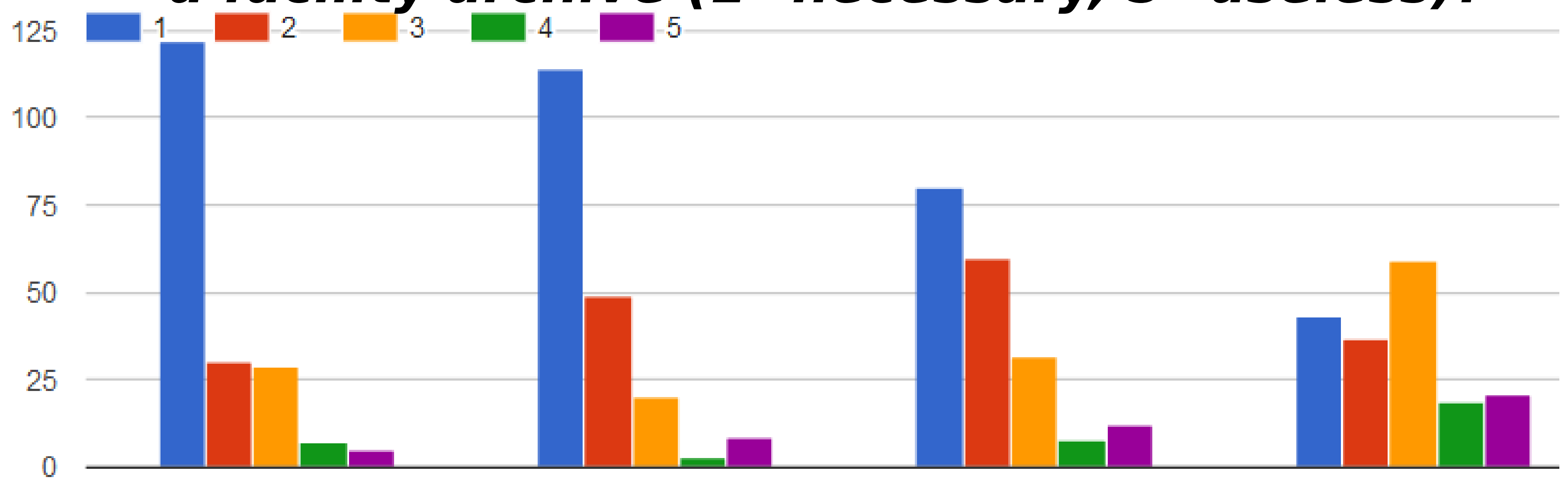
- Surveys of Astronomical Facilities and of their User Communities
- Gap analysis
- Recommendations on the design of user interfaces
  - for data discovery, access, and retrieval
  - data processing, re-processing, analysis and visualization



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



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



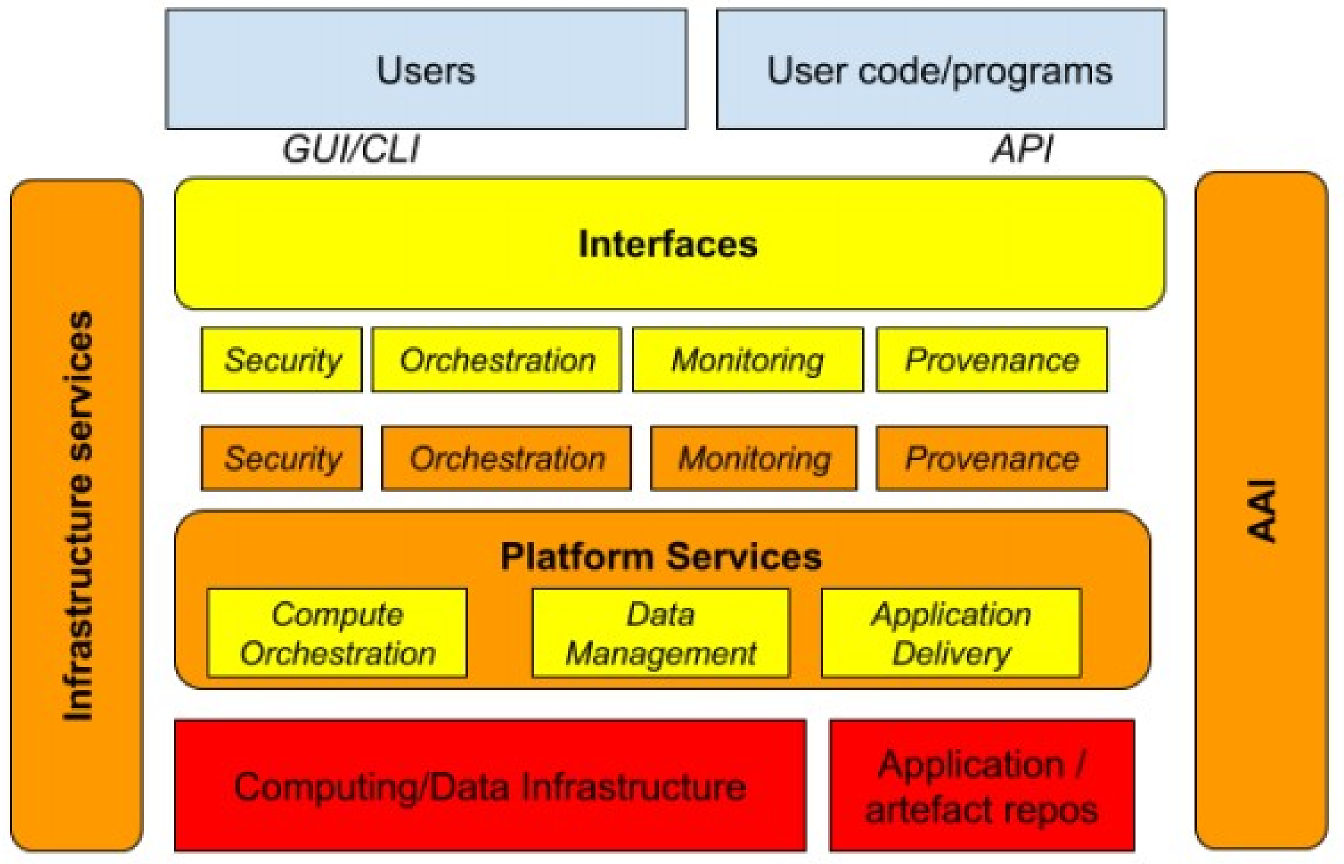
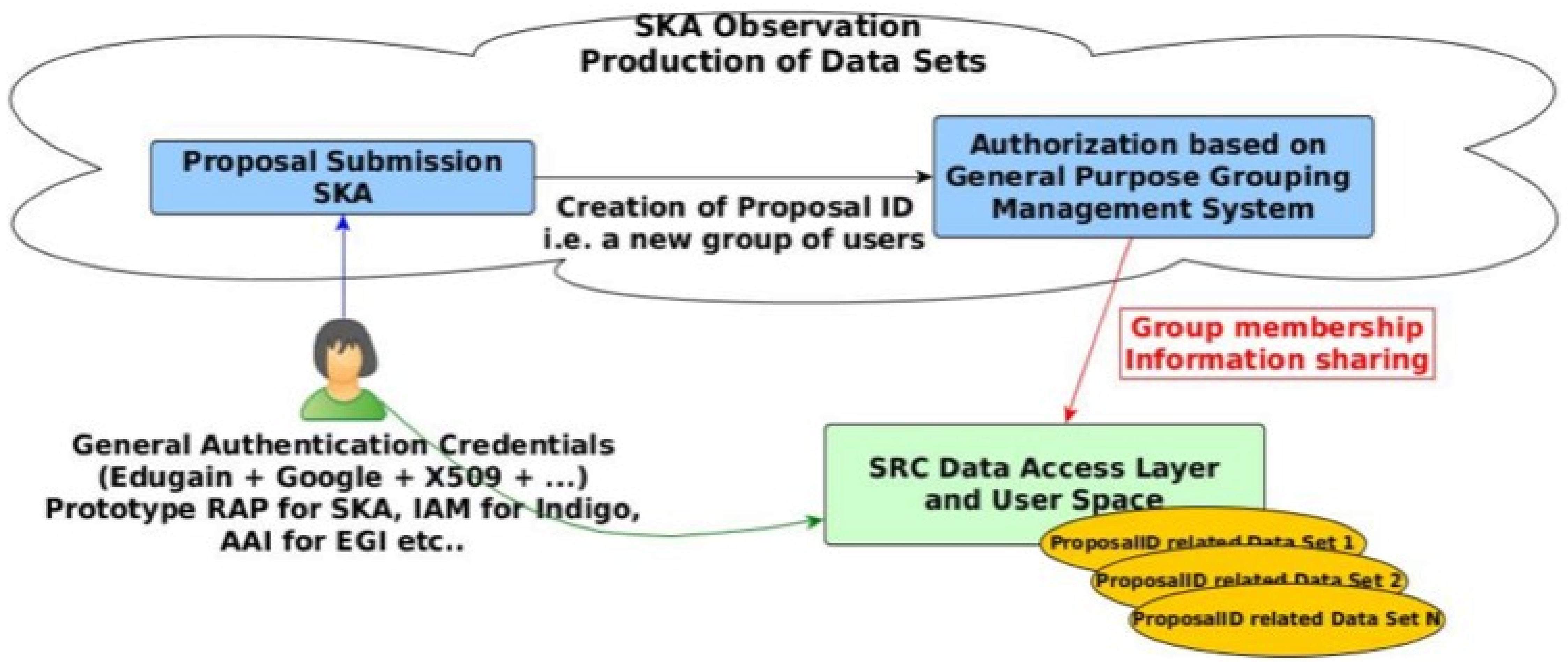


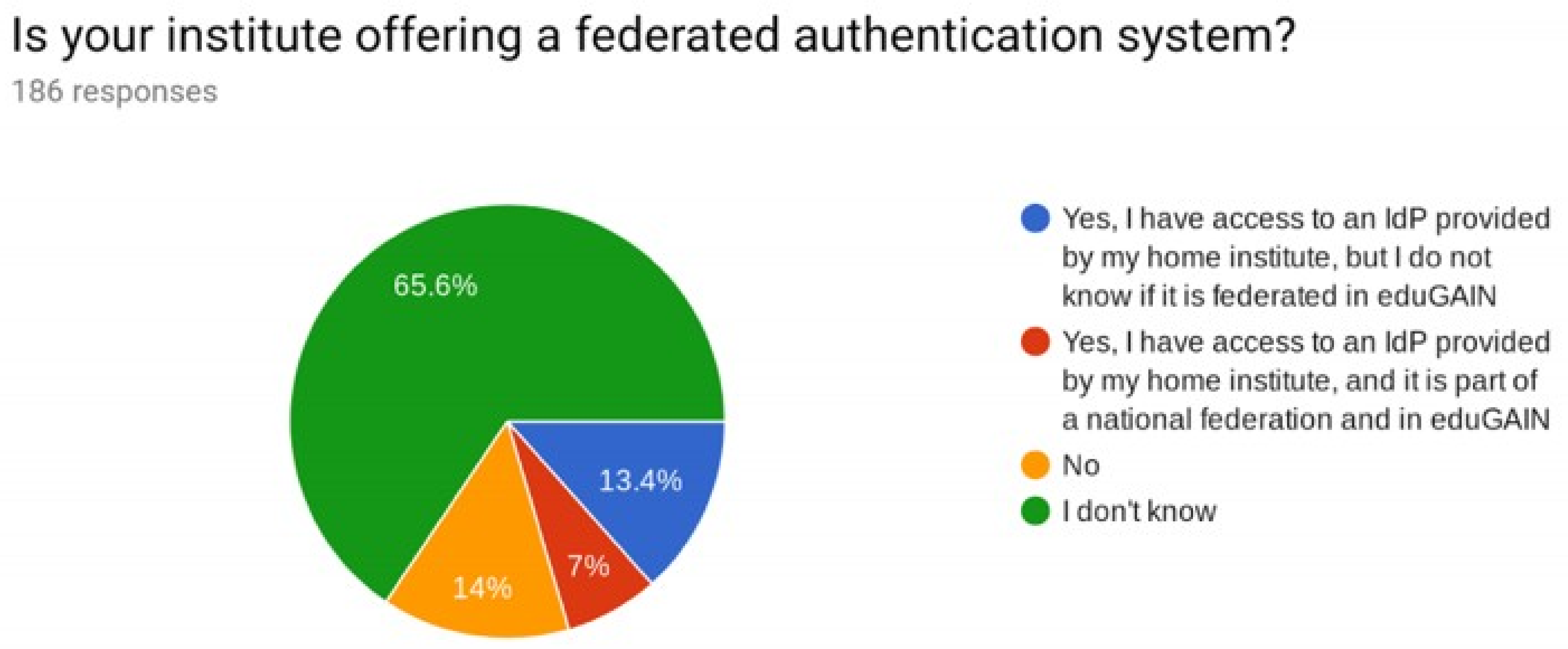
Figure 5.2.1: SG modular components to manage the exploitation of computing/data resources by the user community and user applications.

WP6: User Services

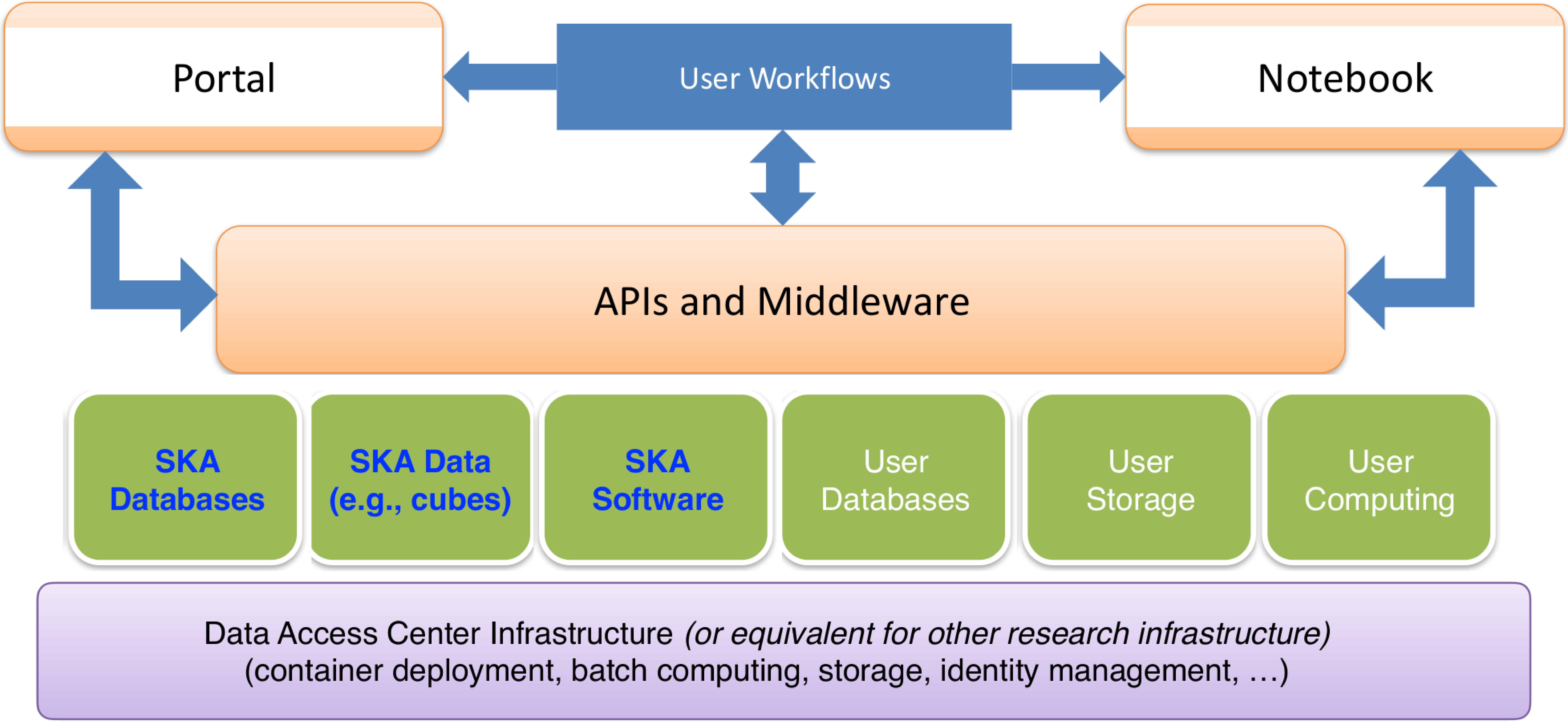
- Authentication and Authorisation Infrastructure
- Federated Access for Research
- Exploration of Technologies
- Proposed AAI Architecture

- Framework for designing and implementing a Service Portfolio for the ESDC and SKA
  - validate users' requests for data access;
  - keep accounts of computing and storage resources for each user or user group;
  - minimize data movement between sites.





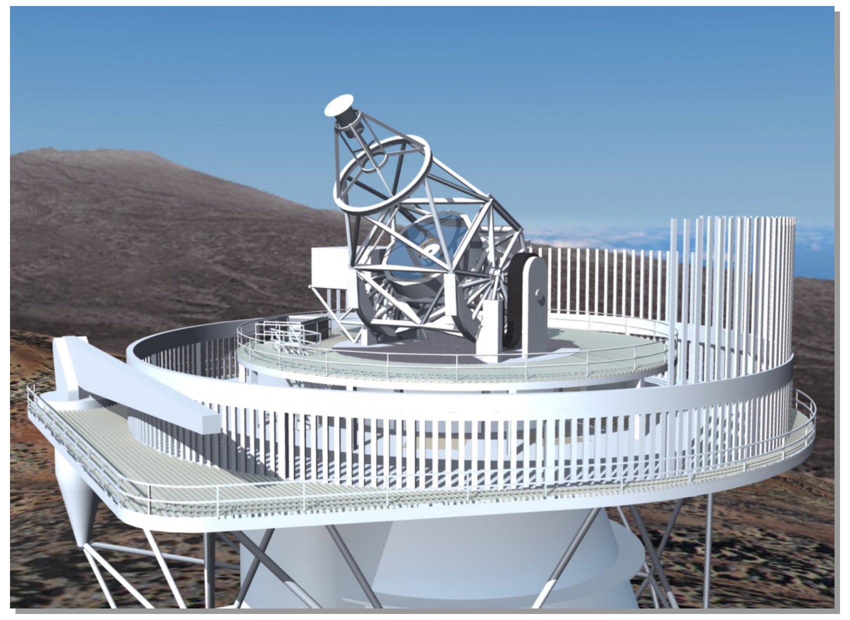
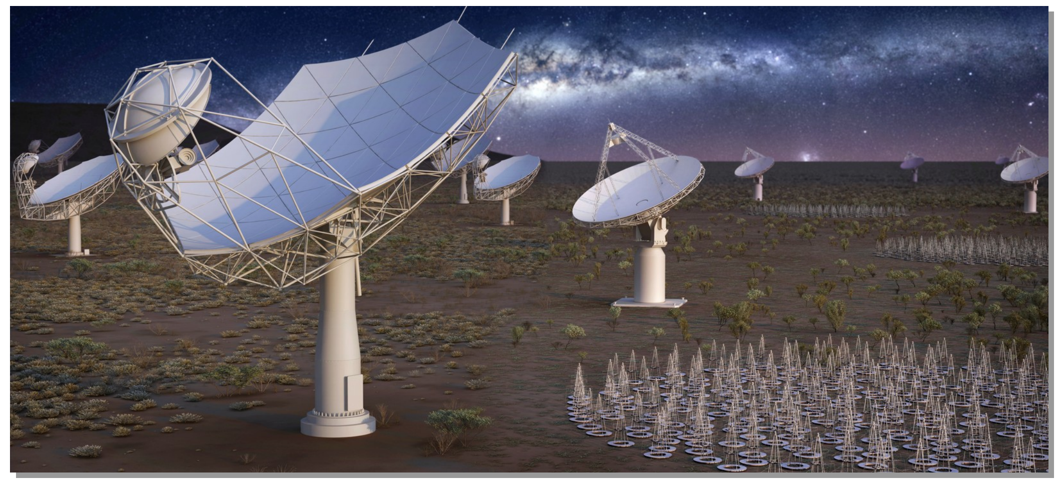
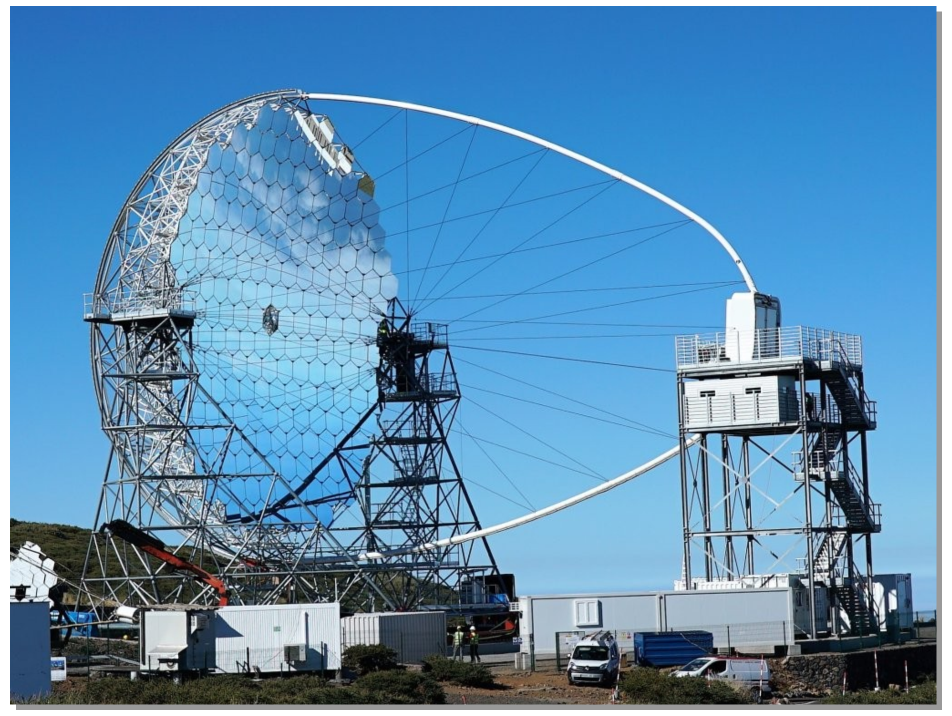
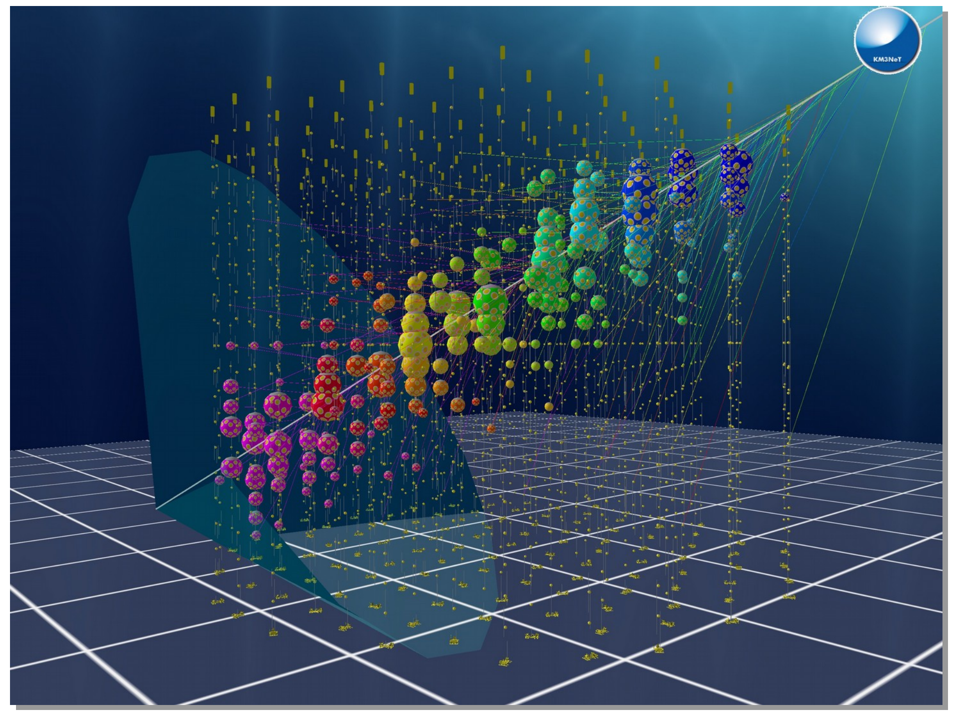
Science Analysis Platform



ESCAPE

European Science Cluster of Astronomy & Particle physics
ESFRI research infrastructures

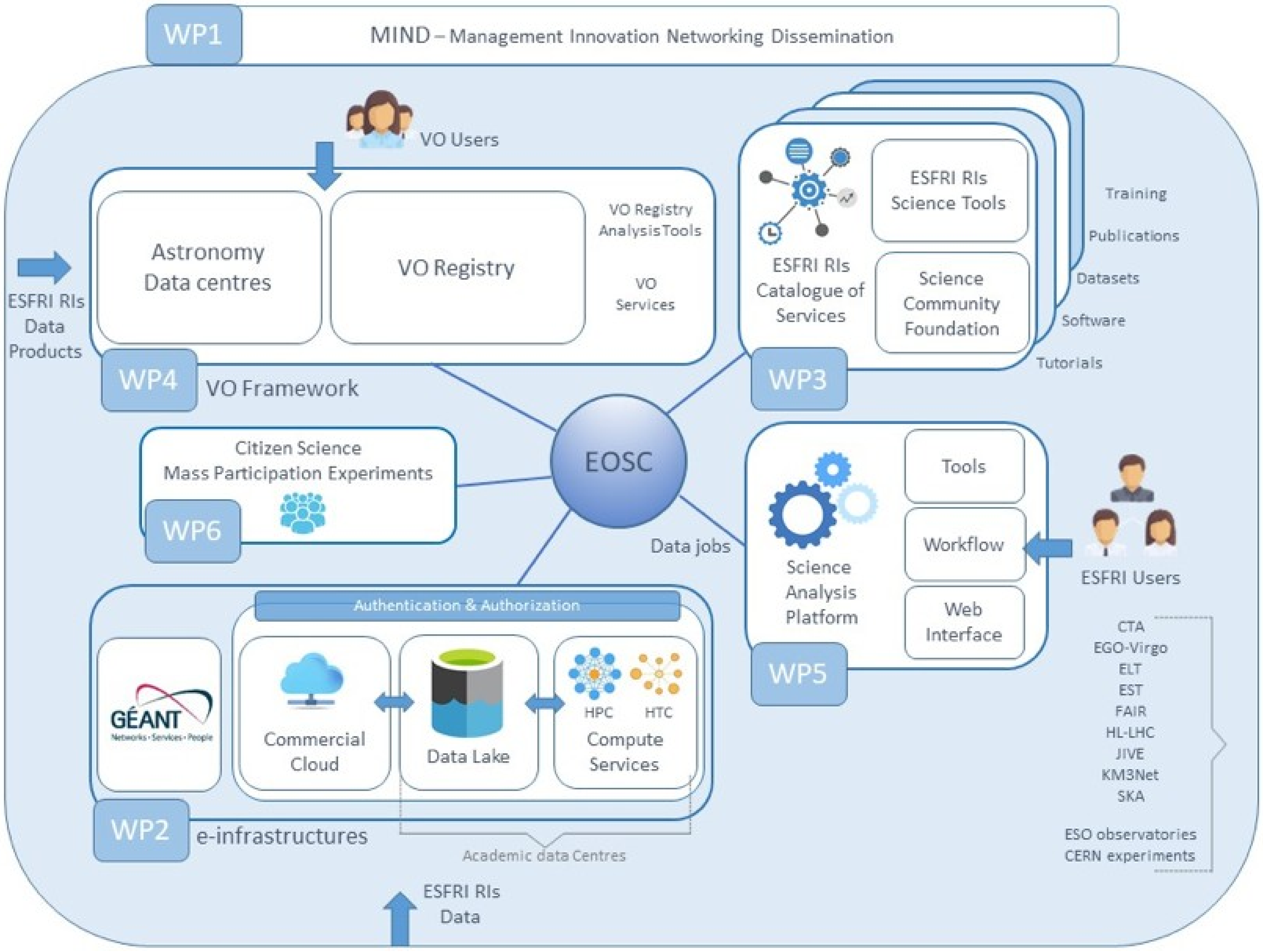
ESCAPE aims to address the Open Science challenges shared by ESFRI facilities as well as other pan-European research infrastructures in astronomy and particle physics


- 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 Analysis Platform Work Package
- Work kicked off in February 2019

EOSC-hub

EOSC pilot
The European Open Science Cloud for Research Pilot Project



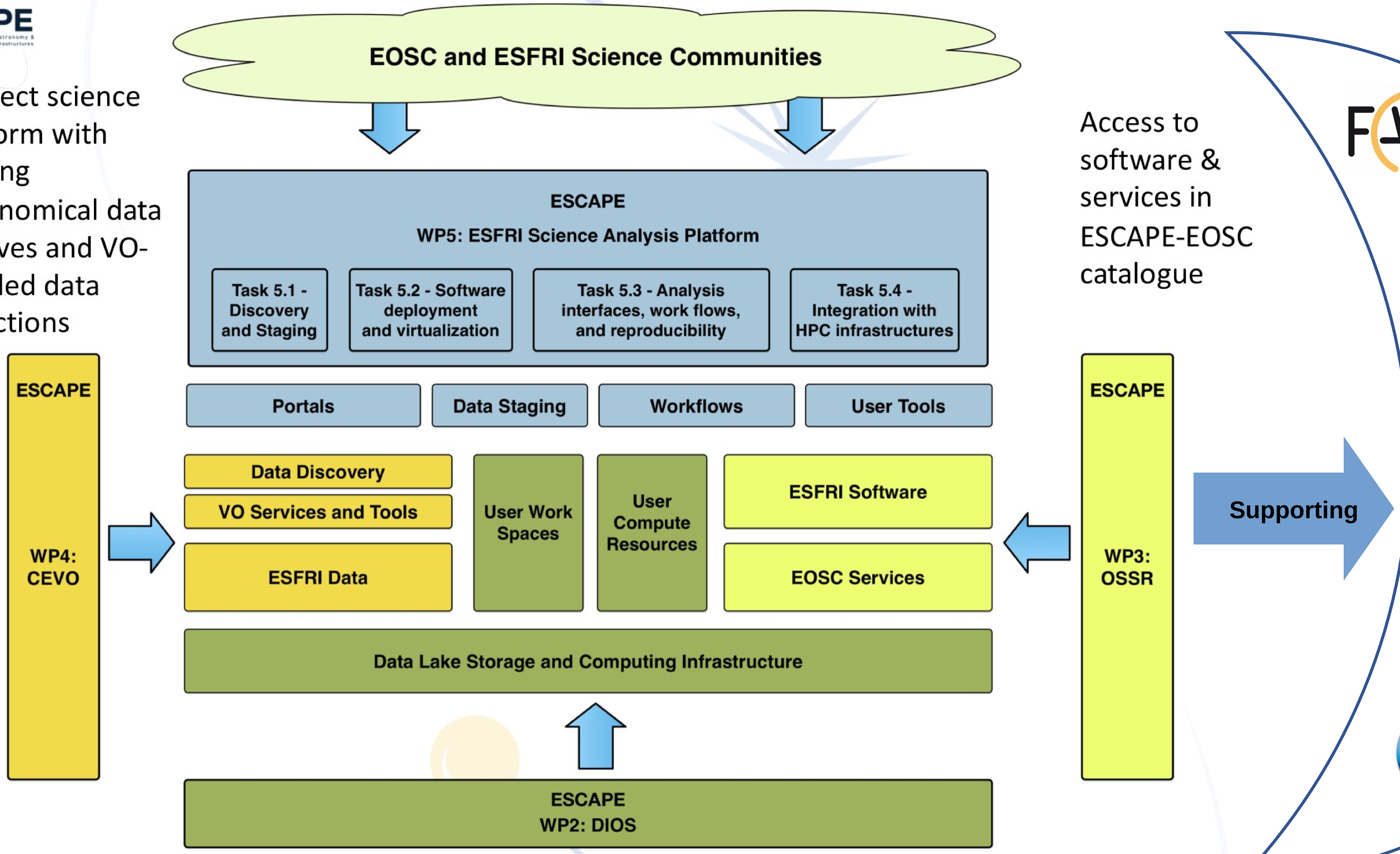
ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures

- WP2: Data Infrastructure for Open Science
  (federating, data lake, networking, AAI)
- WP3: Open-source scientific software and service repositories
- WP4: Connecting ESFRI Projects to EOSC through VO framework
  (integration of astronomy data, FAIR principles, adding value to trusted content in archives)
- WP5: ESFRI Science Analysis Platform

ESCAPE

Connect science platform with existing astronomical data archives and VO-enabled data collections

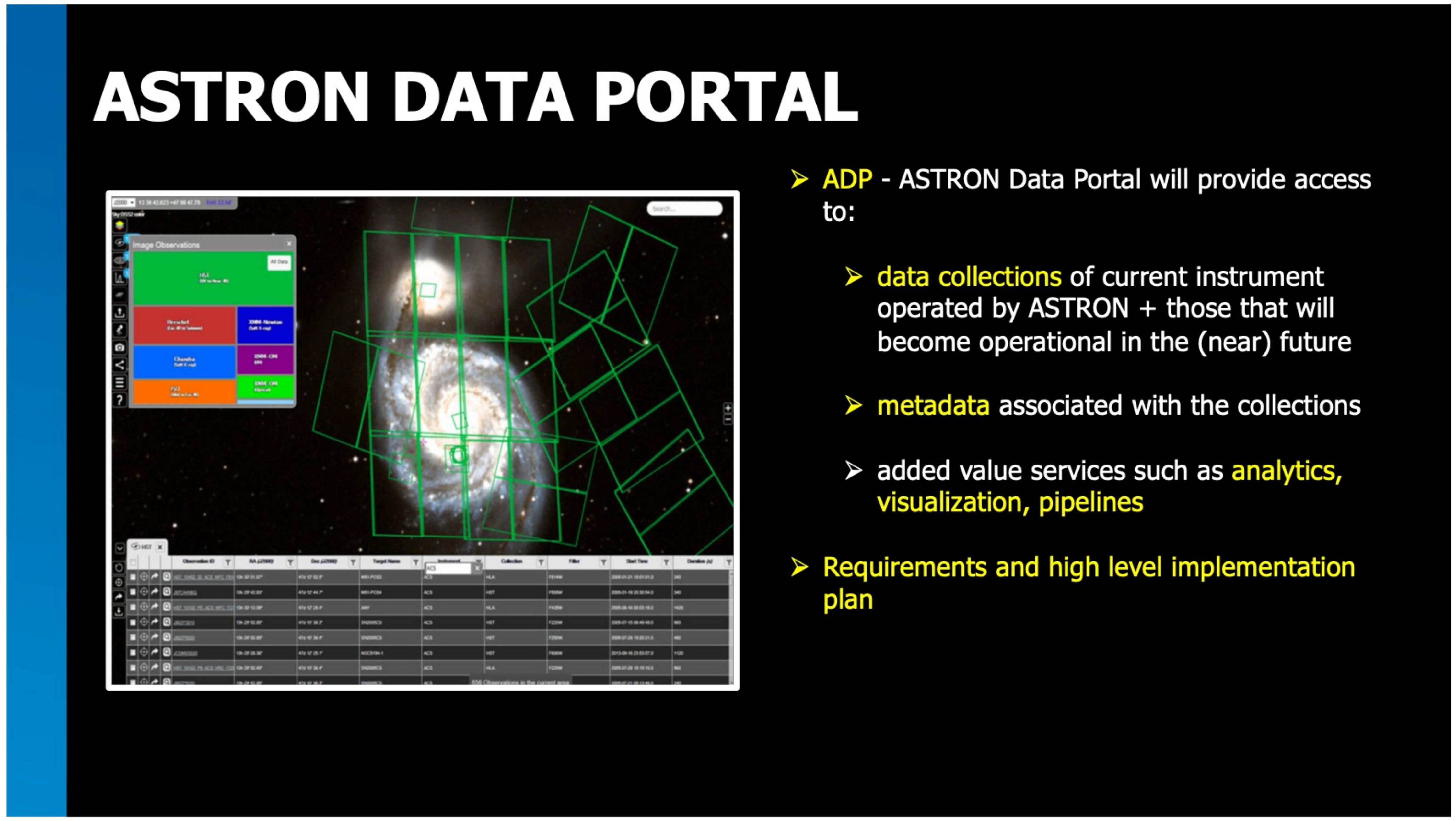
Access to software & services in ESCAPE-EOSC catalogue

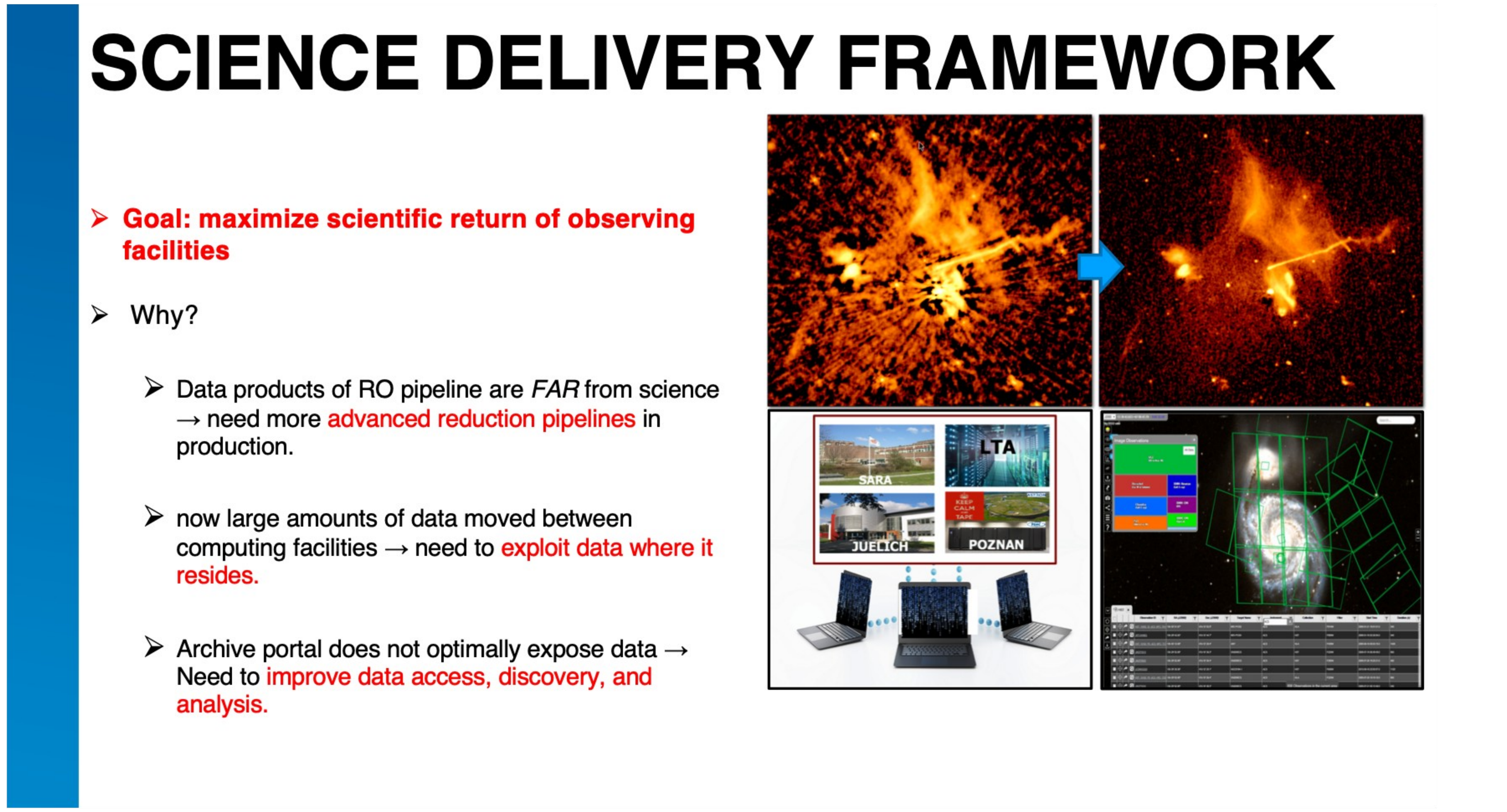


Integration with Data Lake - distributed computing and storage

SRC Related Activities at ASTRON

- ASTRON operates LOFAR/ILT and
- Westerbork/Apertif
- Ongoing projects:
  - ASTRON Data Portal
  - Science Delivery Framework
- EC H2020 Projects
  - ASTERICS
  - AENEAS
  - ESCAPE
  - EOSCpilot & EOSC Hub
- ASTRON Science Data Centre





Thank you!