ASTRON

Netherlands Institute for Radio Astronomy

The Square Kilometre Array

Big Science for Global Astronomy

Michael Wise

Head, ASTRON Astronomy Group

Symposium on the Connecting Strength of Big Science
ILO-net, Neth-ER, Brussels, June 8, 2017

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

LOFAR

ASTRON

Netherlands Institute for Radio Astronomy

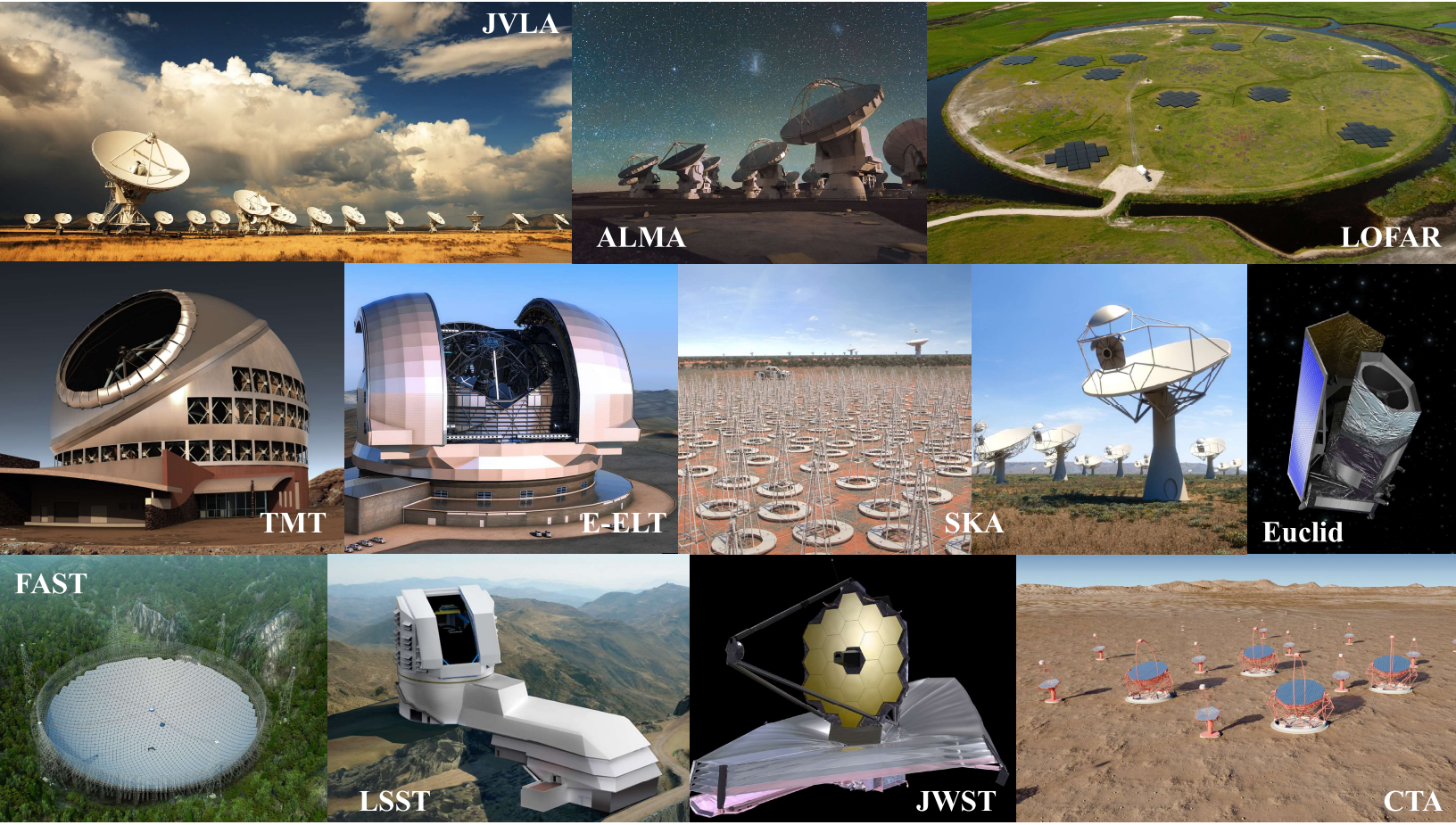
The Square Kilometre Array

Talk Outline

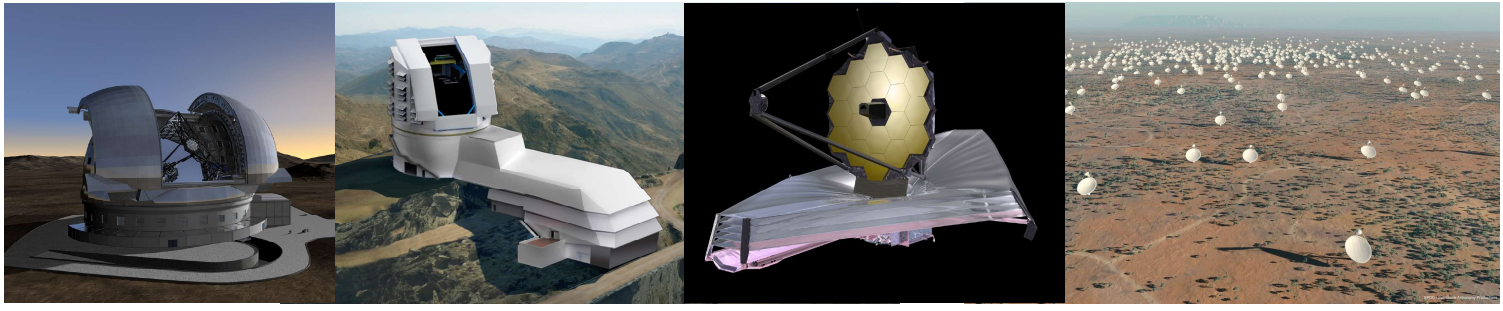
SKA and Data Intensive Astronomy
Global Analysis Infrastructure for SKA Science
Challenges and Boundary Conditions
Experience, Observations, and Open Questions

LOFAR

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


JVLA
ALMA
LOFAR
TMT
E-ELT
SKA
Euclid
FAST
LSST
JWST
CTA

Data Intensive Astronomy



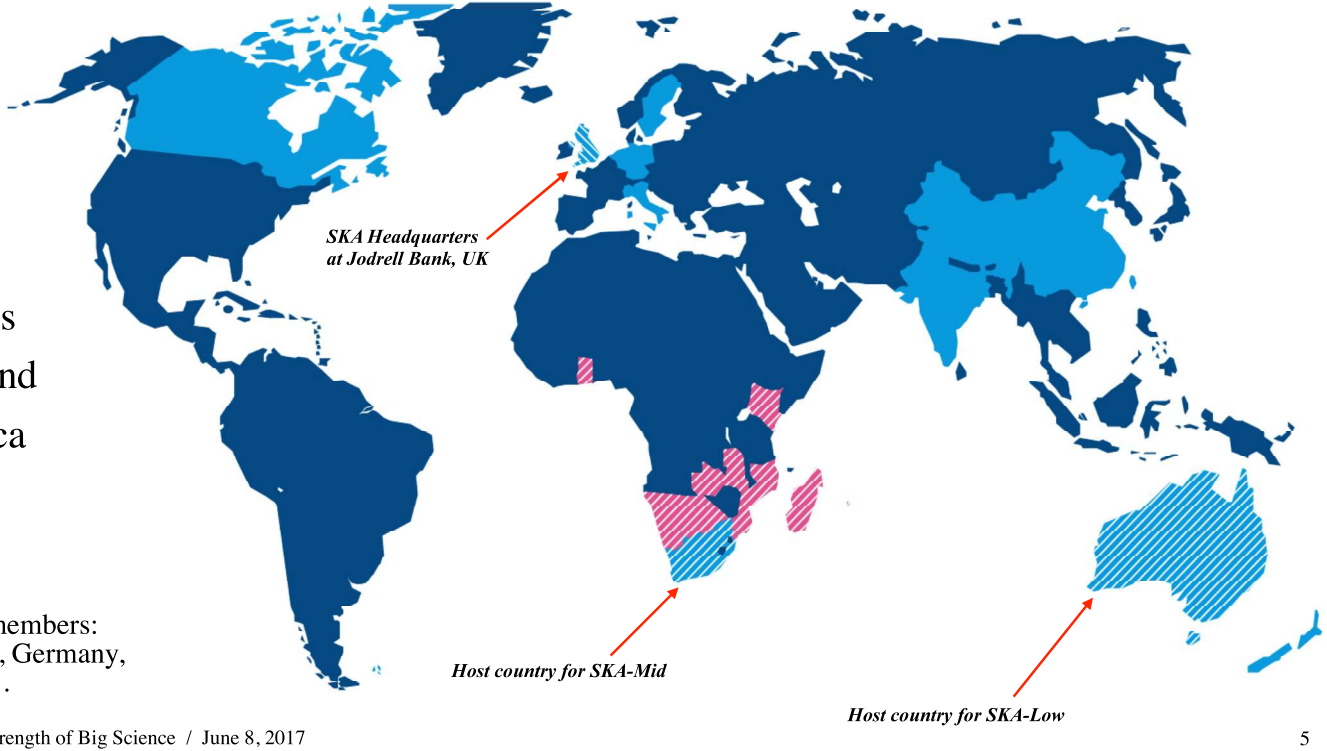
- Science is increasingly driven by large data sets
- Massive data collections and large scientific collaborations
- Most science extraction is based on the archived data
- Current instruments already producing petascale datasets

New science infrastructures will produce exascale data!

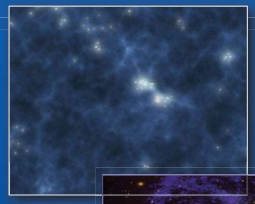
The Square Kilometre Array

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

Potential new members:
Spain, Portugal, Germany, France, others…



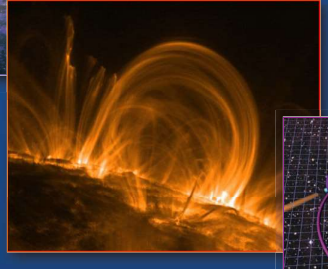
SKA Science Drivers



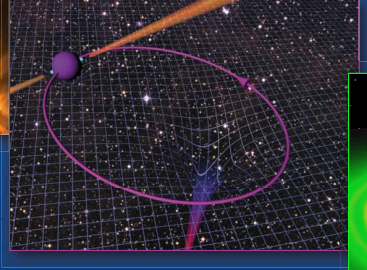
The First Stars

Cosmic Evolution

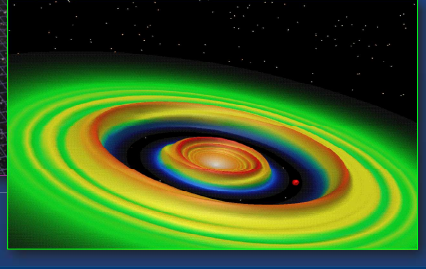
Cosmic Magnetism



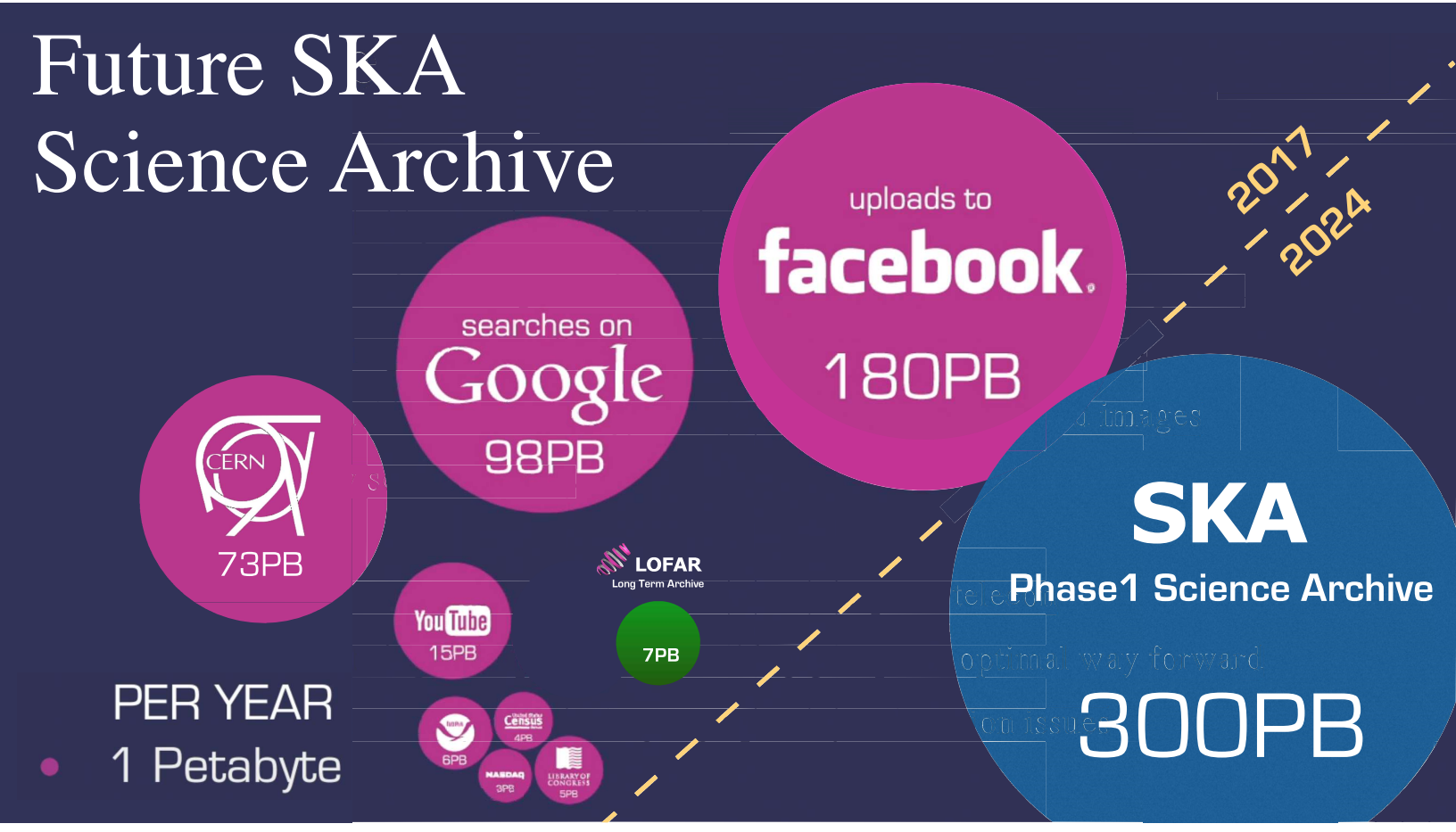
Gravitational Physics



Origins of Life

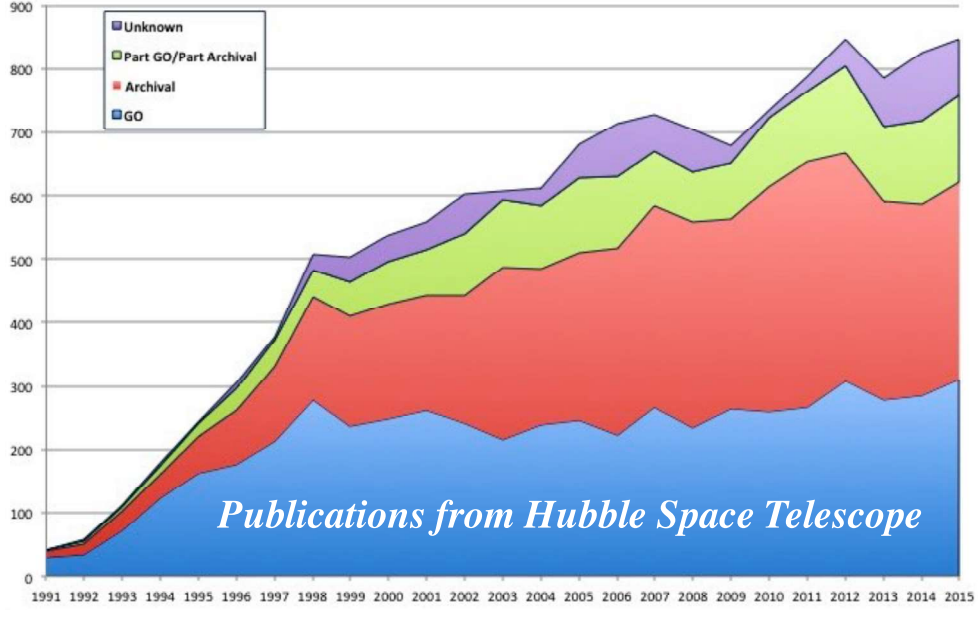


Future SKA Science Archive



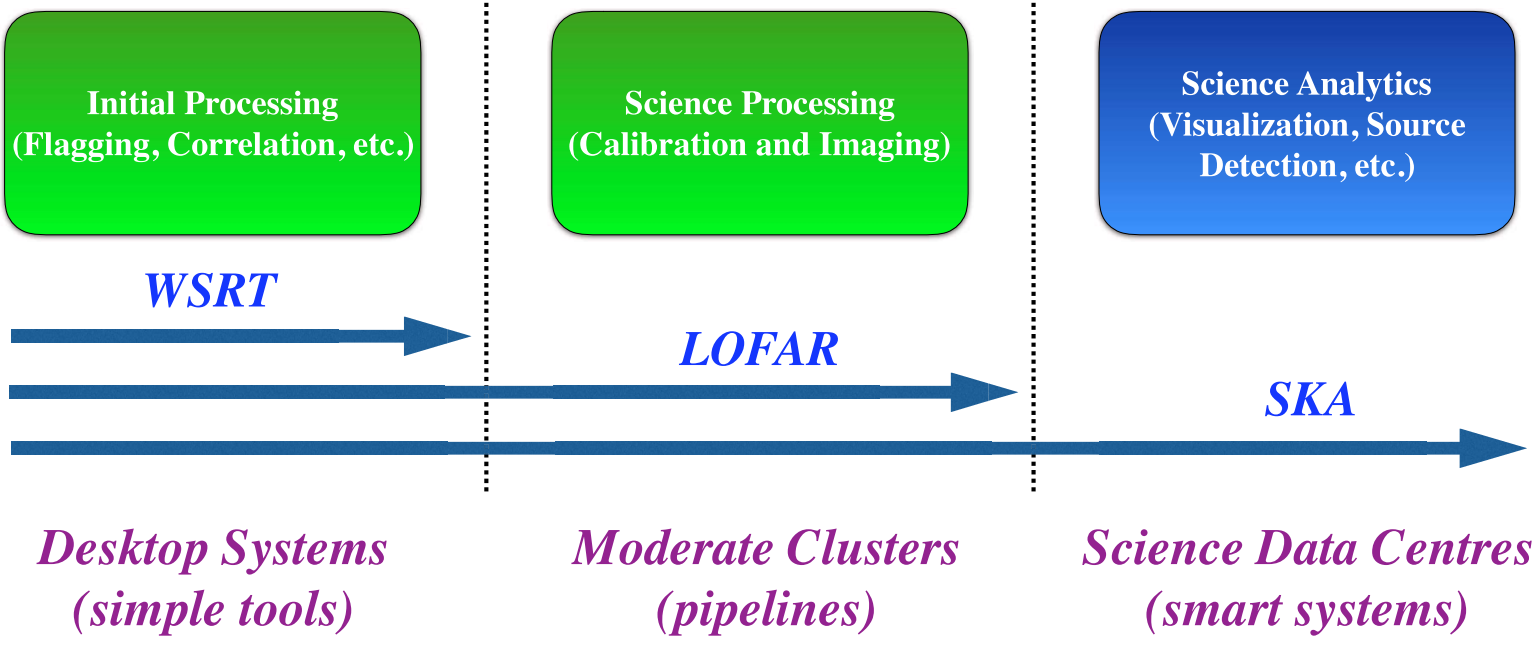
Impact of Science Archives

Science archives are a multiplier for total science output

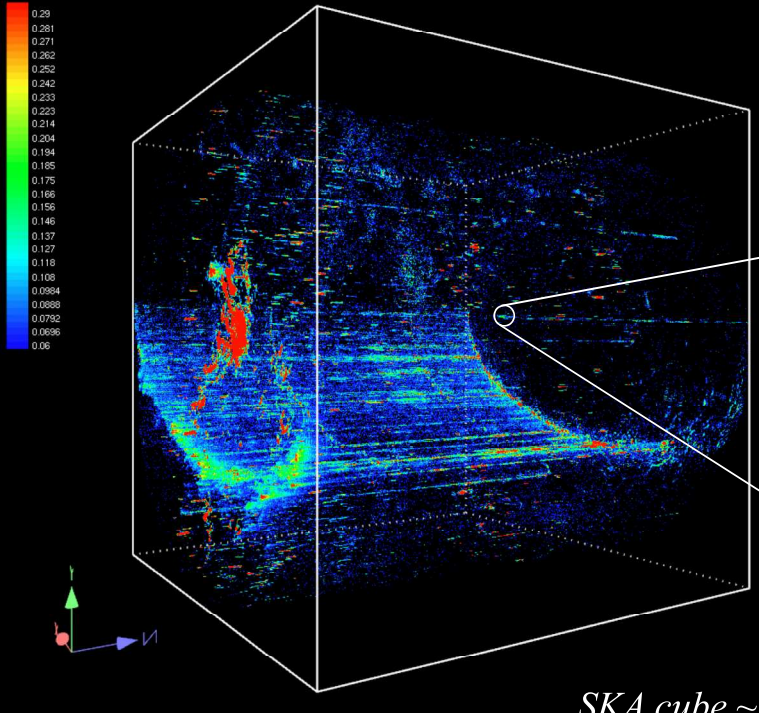


- Assumes the archives are persistent and maintained
- Assumes archival data is open and accessible to users
- Assumes data products stored are appropriate for general use
- Assumes users retrieving data have resources to process to a science result

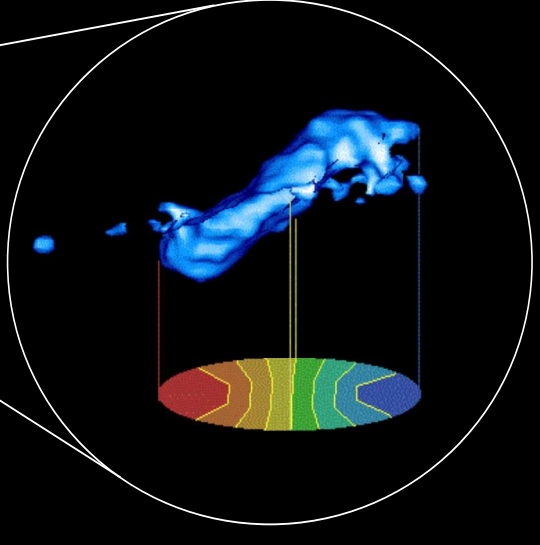
Full Cost of Science Extraction



Detection, Classification, Inference



Automated detection and calculation of galaxy rotation curves in HI surveys

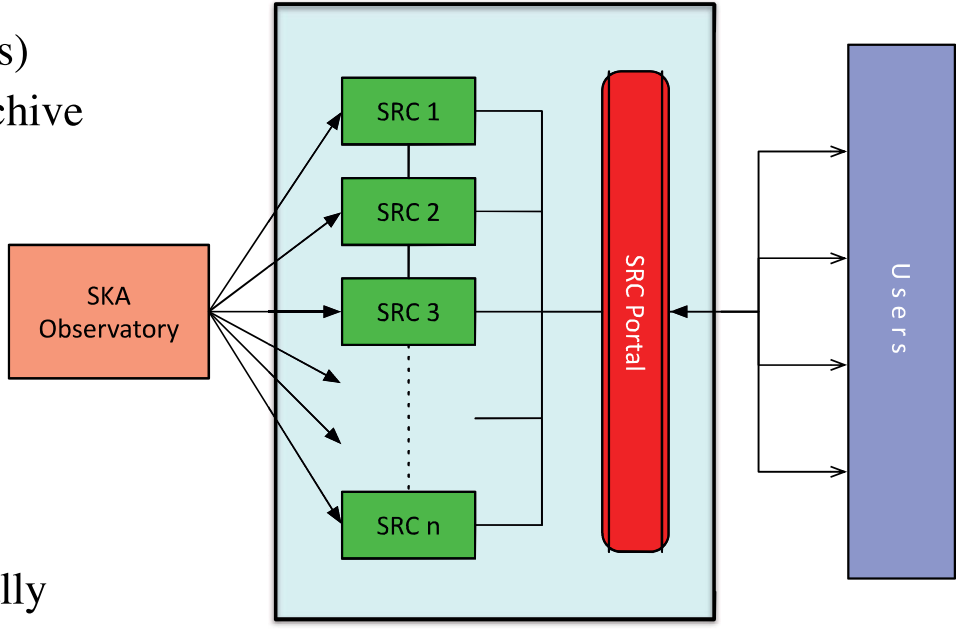


SKA cube ~ 0.9 PB

ASKAP HI Cube (Jurek et al. 2010)

SKA Regional Centres

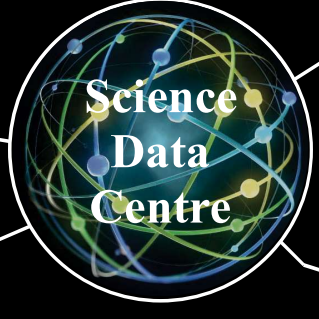
- SKA Regional Centres (SRCs) will host the SKA science archive
- Provide access and distribute data products to users
- Provide access to compute and storage resources
- Provide analysis capabilities
- Provide user support
- Multiple regional SRCs, locally resourced and staffed



Primary interface for SKA data analysis

Global Network of Centres

What is a Science Data Centre?



Astronomers

Place to find their data
Place to analyze their data
Place to find support

Funding Agencies

Way to fund science
Way to maximize investment
Way to spur innovation

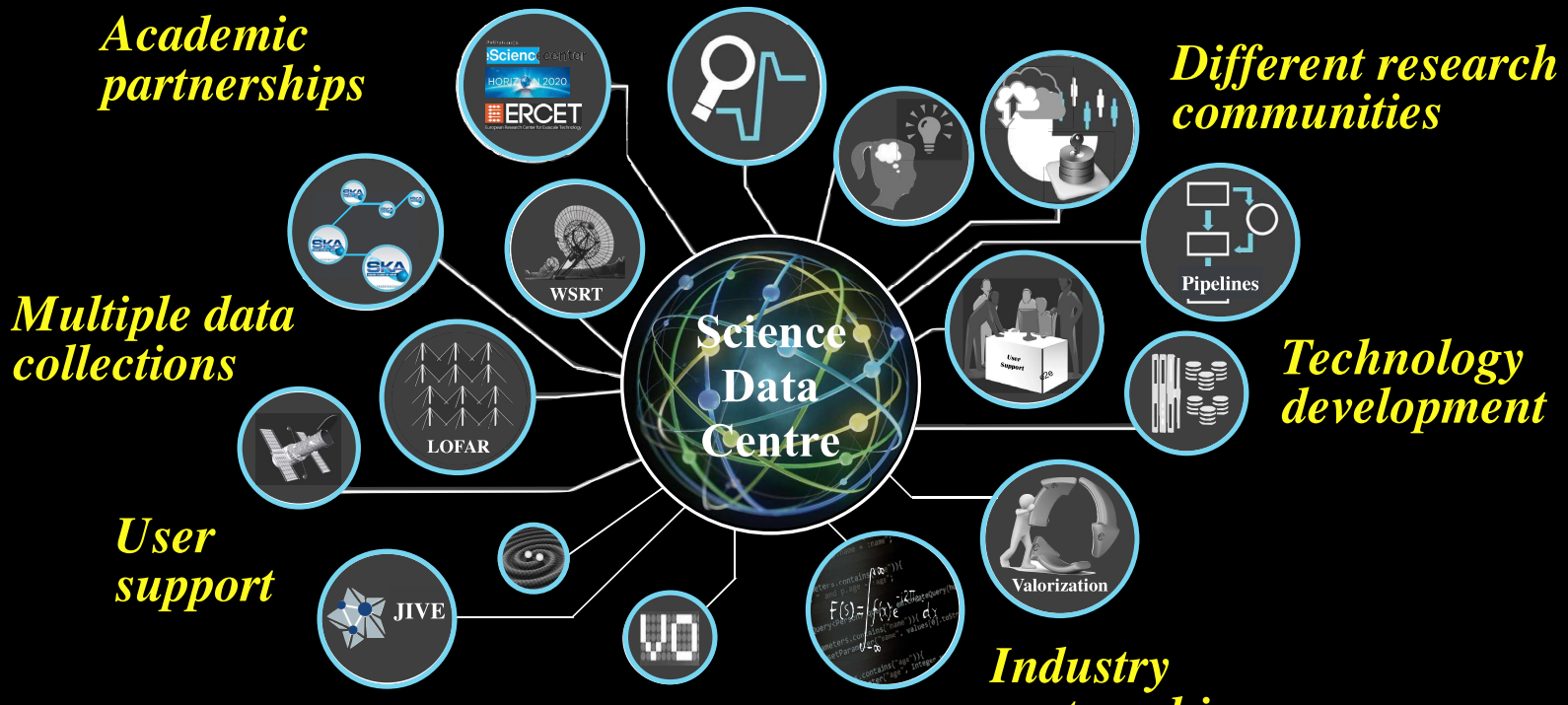
Observatories

Place to curate their data
Place to develop new capabilities
Place to support their users

Industry

Place to find new challenges
Place to collaborate with academia
Place to test new technologies

Connections Beyond Astronomy



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 Horizon 2020 (€3 million)

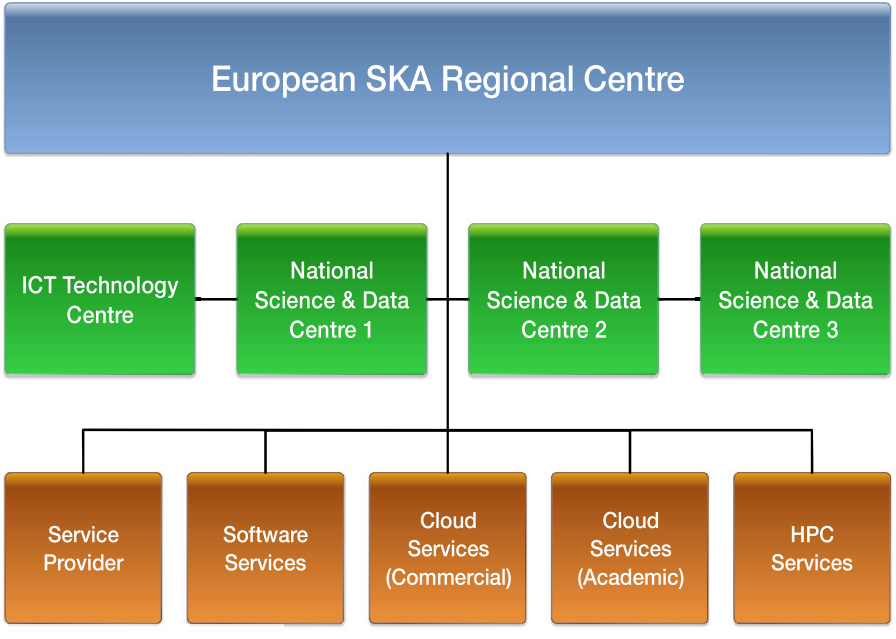
13 countries, 28 partners, SKAO, host countries, e-infrastructures (EGI, GÉANT, RDA), NREN's

Three year project (2017-2019)

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

European SKA Regional Centre

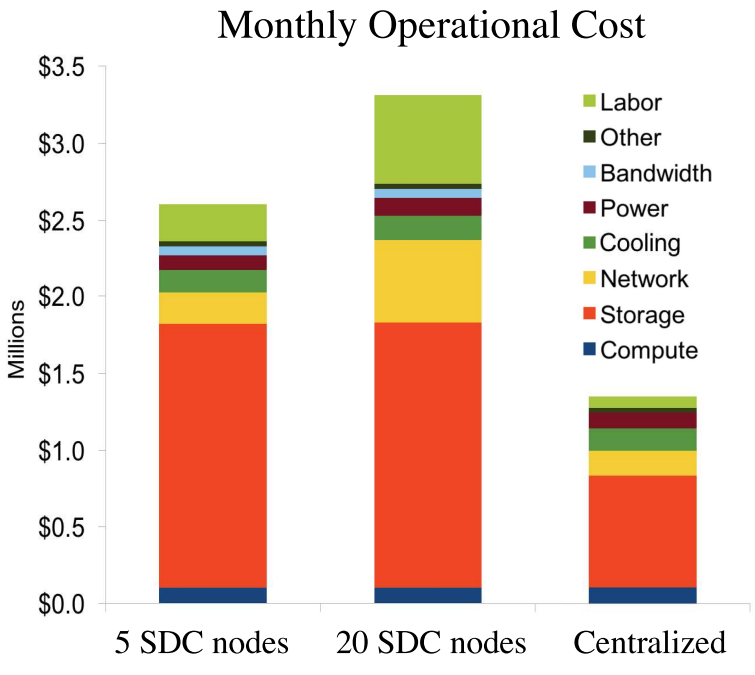
- Create a European-scale, federated Regional Centre for the SKA
- Provide resources for SKA science extraction to users
- Coordination with ICT communities, industry, and service providers
- Facilitate shared development, interoperability, and innovation
- European counterpart for engagement with other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



Organizational Challenges

- Large and distributed global user community
- Exascale computation and data management
- Integration of global research infrastructures
- Long-term development and operations
- Sustainable funding models for common infrastructure
- Clash of cultural mindsets (structured HPC versus interactive analysis)

Distributed vs. Centralized



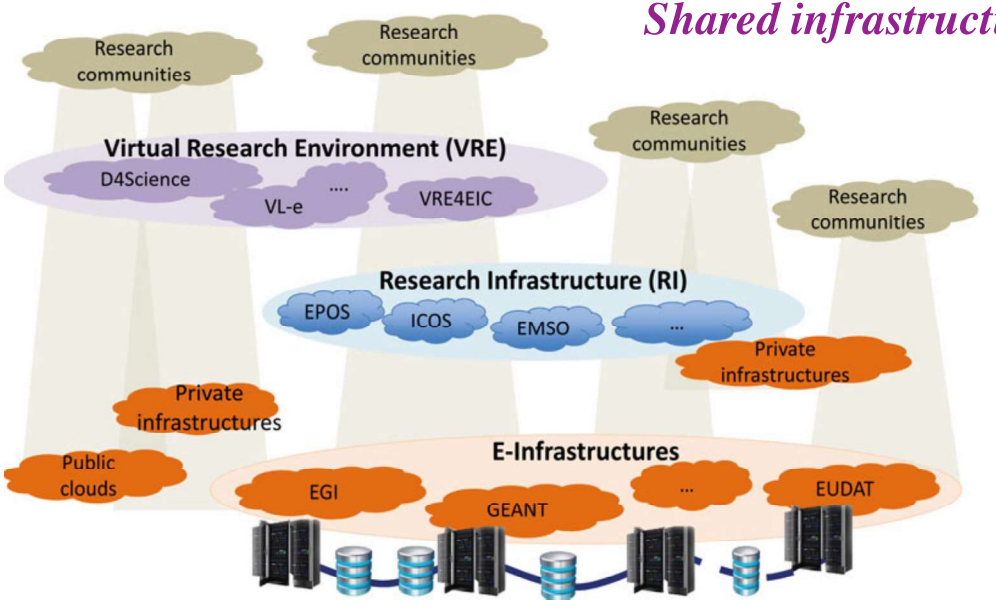
"A centralized model saves money."

"Never move the data."

- Engineers

- Research community is distributed
- Existing resources are distributed
- Need at least one archive backup
- Partners prefer investments in regional infrastructure
- Centralized model makes regional valorization difficult
- Can't afford to start from scratch!

Shared vs. Dedicated



From "VRE in the Data for Science Approach to Common Challenges", Zhao, Martin, & Jeffery

Shared infrastructure is a boundary condition, but there are issues:

- Long-term curation
- Committed resources
- Analysis latency
- HPC culture shock
- Requires outside expertise
- Mechanism to harness community investments (personal grants, etc.)
- Preserve environment for discovery and innovation

Sustainable Partnerships

- Multiple perspectives….academic, agency, project, industry
- Difficult to establish, difficult to sustain over long project lifetime

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?

GLEAM Survey (Hurley-Walker et al. 2016)