www.egi.eu
@EGI_eInfra

# Data spaces: Services for online analytics

*LOFAR Science Products*

*Agnes Mika and Hanno Holties, ASTRON*

*05th. February 2021*

The work of the EGI Foundation
*is partly funded by the European Commission*
*under H2020 Framework Programme*

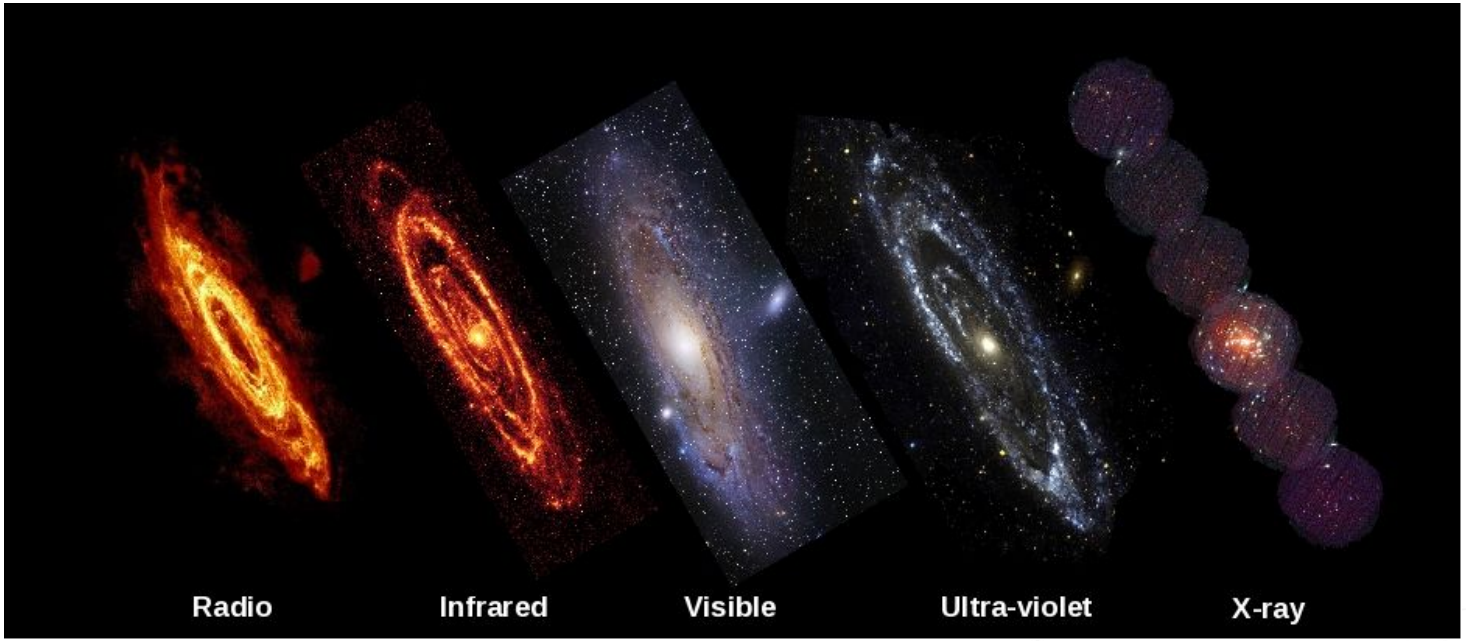
# Outline

- What is LOFAR?
- What are the computing/data challenges?
- How to address the computing/data challenges?
- How to increase the science impact of LOFAR data?
- Towards an operational service: EGI-ACE and DICE

# Radio astronomy

*Astronomy - wavelengths*

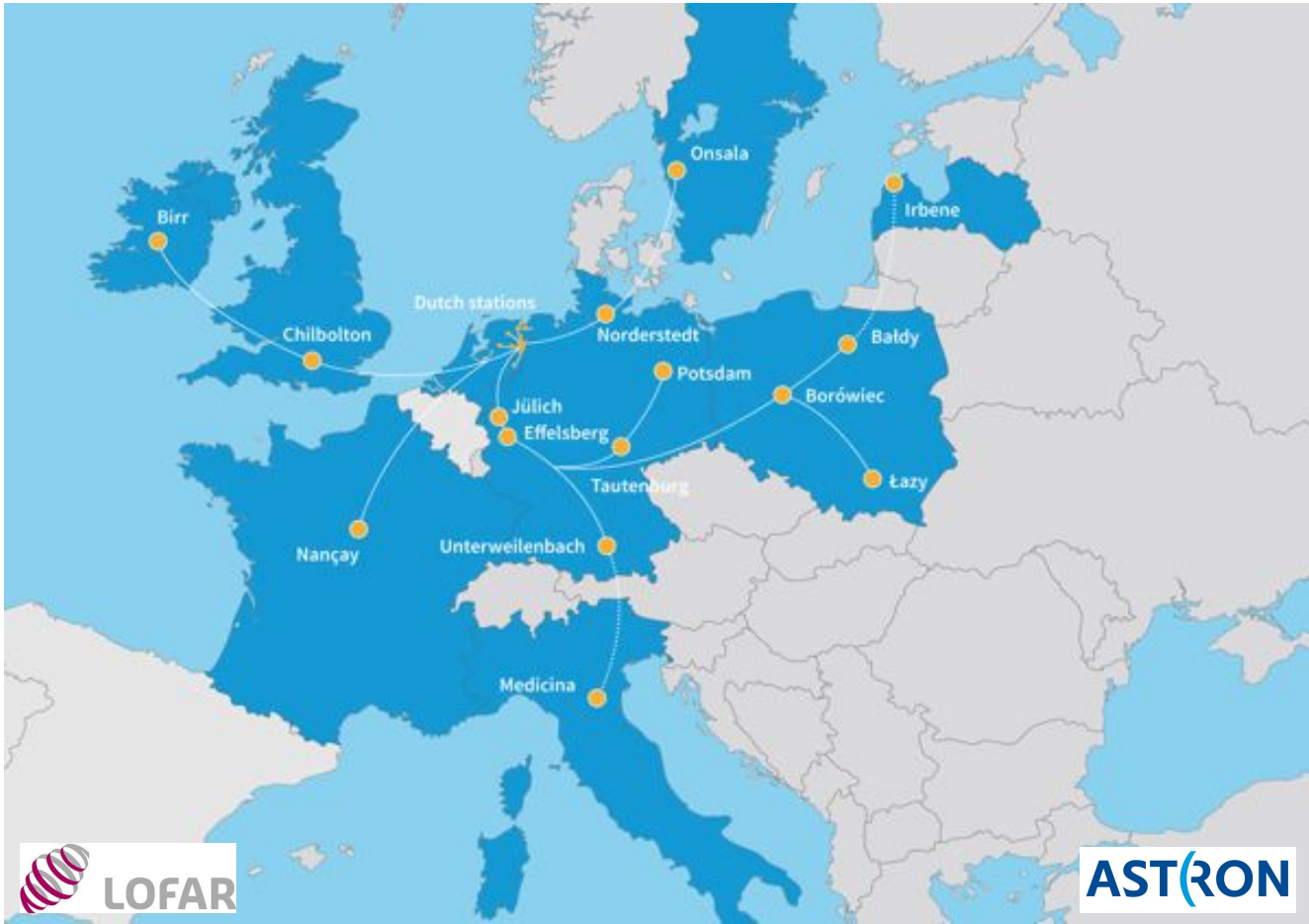
Andromeda galaxy (Multi-wavelength view)




Longer ← Electromagnetic wavelength → Shorter

# The international LOFAR telescope

- 18 stations in the Netherlands
- 13 international stations, from Birr (Ierland) to Irbene (Latvia)
- Italy to join soon
- Longest baseline: ~2000 km



# The international LOFAR telescope

Data Centers 🔴

- (Near) real-time processing
- GPU & CPU clusters
- 3+ PB temporary storage

Long-term archive 🟢

- 50+ PB nearline storage
- 2+ PB online storage

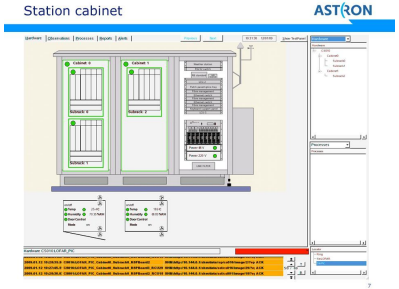
# LOFAR: data intensive science

Digitised dipole signals

20 Tb/s raw data

Central beamforming, etc.

92 Tflops real-time processing



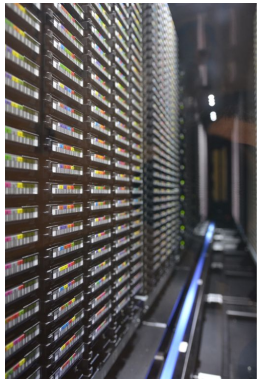

Station beamforming

225 Gb/s real-time data

Flagging, averaging, etc.

96 Tflops offline processing

2 Gb/s to archive



50+ PB stored

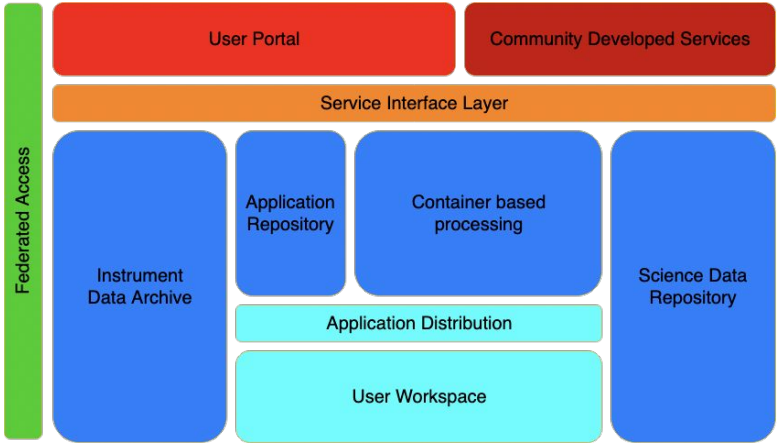
# LOFAR: from data to science

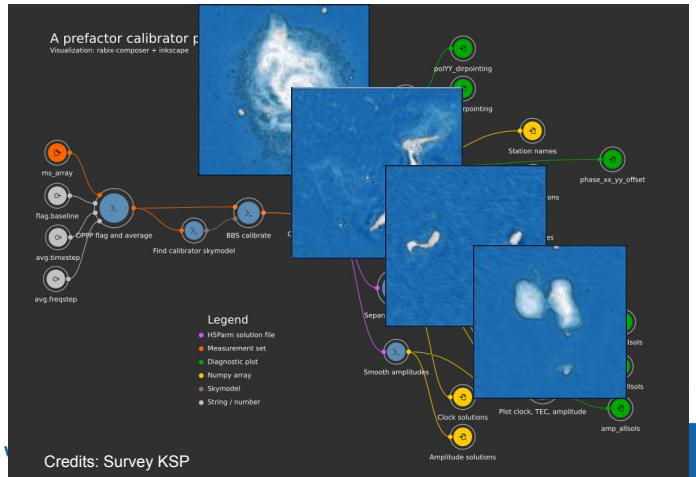
- Data archive: 50 PB, mostly in raw format
- To achieve science results, we need:
  - Trained people
  - Further data processing
  - Data access
- Access to resources
  - Compute
  - Storage
- Discoverability
  - Apply standards
  - Direct access to science-ready data

# Towards a Science Data Centre

Aim: maximise the scientific return of observing facilities

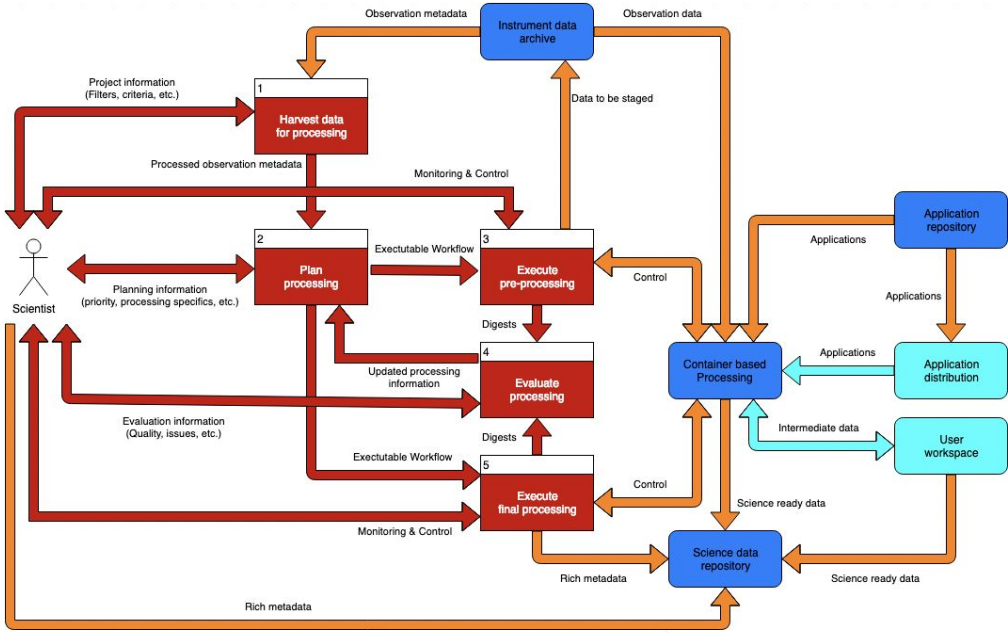
- Global federated AAI
  - LOFAR collaboration
- Workflow Portability
  - Standardisation
    - Pipeline definition (CWL)
    - Execution framework
  - Containerisation
  - Application distribution
- Infrastructure
  - Storage
  - Compute
- FAIR



# Towards a Science Data Centre

- Science Data Repository
  - Standard science data products
  - Community generated data
  - Open & FAIR
- Data Discovery
  - Annotation: Quality, Instrument, Physics, ...
- Interoperability:
  - VO, Linked data, Data Formats
- Science Analysis Services
  - Data centric
  - Distributed data management
  - Data locality
  - Application Portability



# LOFAR in EOSC

*Towards an operational system*

EUROPEAN OPEN SCIENCE CLOUD

*"...ensure that European scientists reap the full benefits of data-driven science."*

*- European Cloud Initiative publication*

- EOSCpilot - Technology exploration
  - LOFAR was a science demonstrator
- EOSC-hub - Prototyping
  - LOFAR is a competence center
- ESCAPE ESFRI - Research platform
  - LOFAR provides a test platform
- DICE - Operational service
  - LOFAR provides a community platform
- EGI-ACE - Operational service
  - LOFAR provides a federated dataspace

EOSC-hub

EOSCpilot www.eoscpilot.eu

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

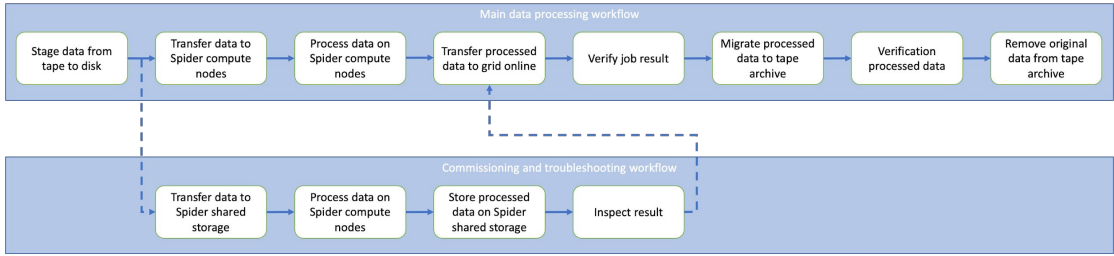
EGI-ACE

DICE Data Infrastructure Capacity for EOSC

# LOFAR in EGI-ACE

LOFAR Data Valorisation: Adding value to the LOFAR data archive

- Large scale processing for
  - Reduction of data volumes
  - Science-ready data



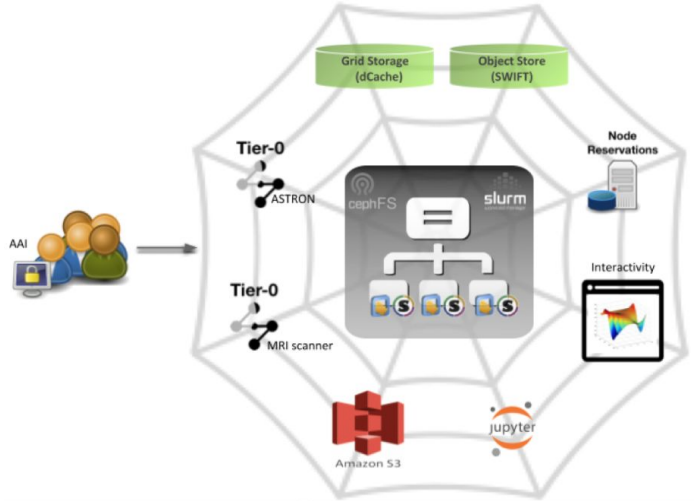

Providing Virtual Access to radio astronomical data

- Building on EOSC infrastructure
- Available from EOSC Portal
- Community contributions
- Data processing as a VA service

# Required resources in EGI-ACE

Resources

- Millions of core hours (~6 million) on the SURFsara compute cluster
- Few hundred TB storage to support the processing workflow



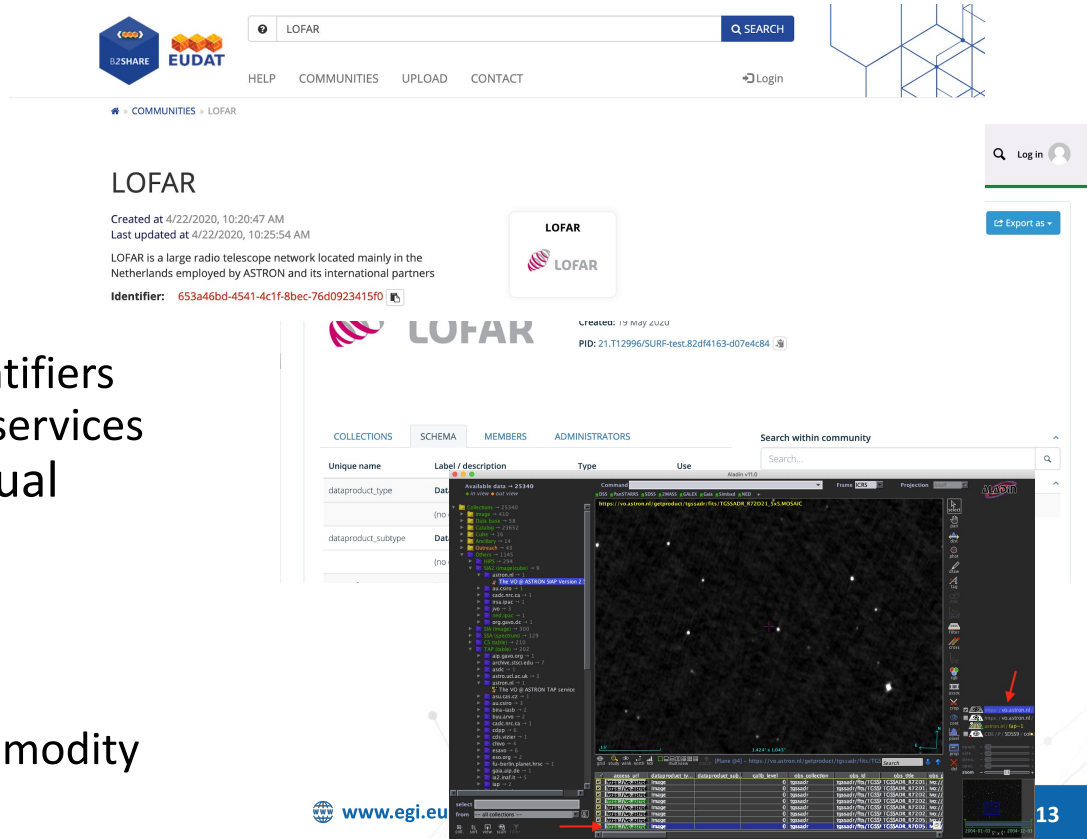

Timeline

- Integration during 2021
- Service exploitation to start at the end of 2021

# Link to LOFAR in DICE

*LOFAR Science Data Repository*

- Science-level data
  - Generated through EGI-ACE
- Applying FAIR principles
  - Rich metadata
  - Provenance
  - Registration of Persistent Identifiers
  - Harvesting by data discovery services
- Publication compliant with Virtual Observatory standards
  - Community standard
  - Supporting multi-wavelength astronomical research
  - Enabling science through commodity community tools



# Thank you!