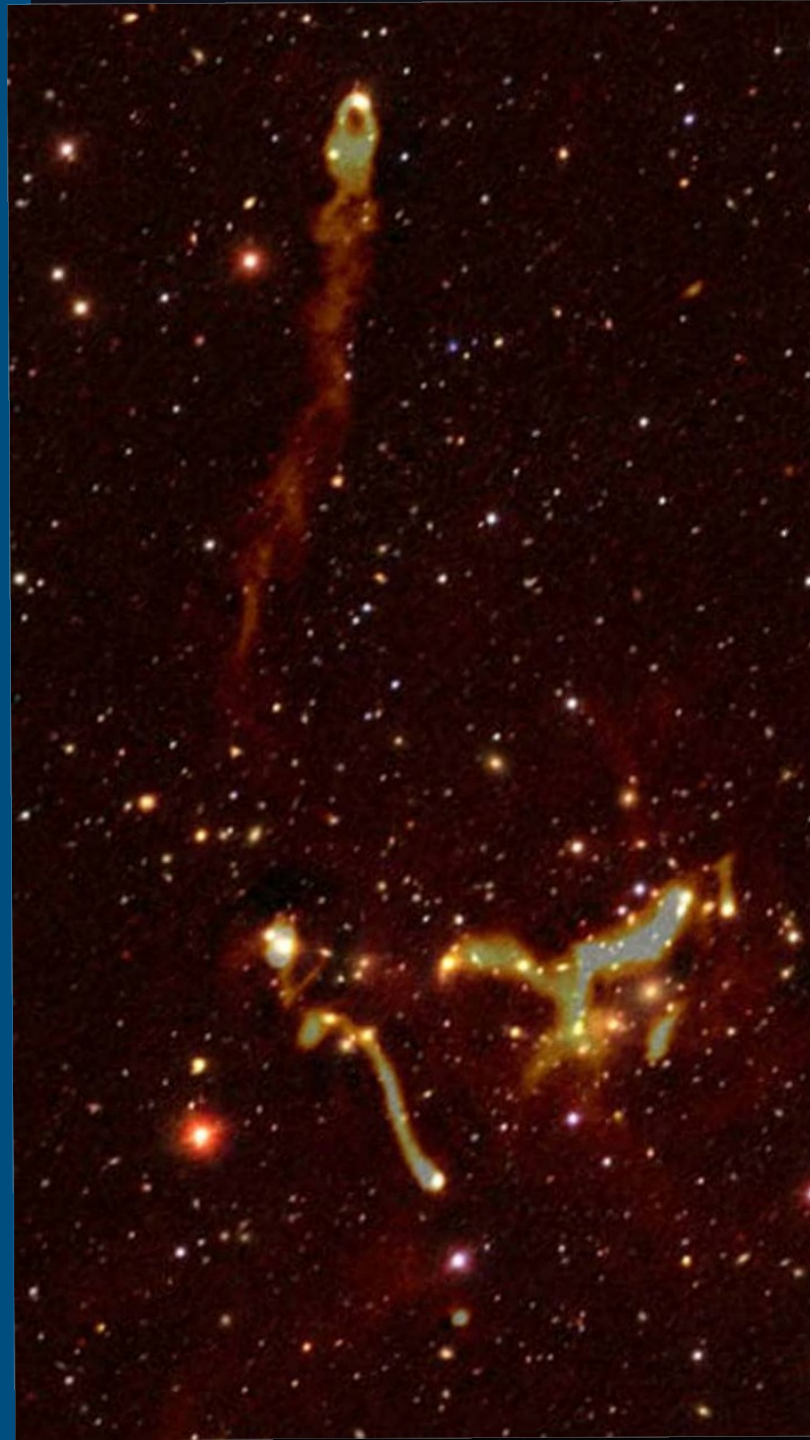

ASTRON
Netherlands Institute for Radio Astronomy

LOFAR Science Data Processing

Ágnes Mika, Hanno Holties
EGI conference, 2021 October 19

ASTRON
Netherlands Institute for Radio Astronomy



The international LOFAR telescope

- Central Processing
  - (Near) real-time processing
  - GPU & CPU clusters
  - 3+ PB temporary storage
- Long-term archive
  - 50+ PB nearline storage
  - 2+ PB online storage

Challenges for the LOFAR community

Obtaining science results

- Complex instrument – high level of expertise needed
- Handling of large data volumes

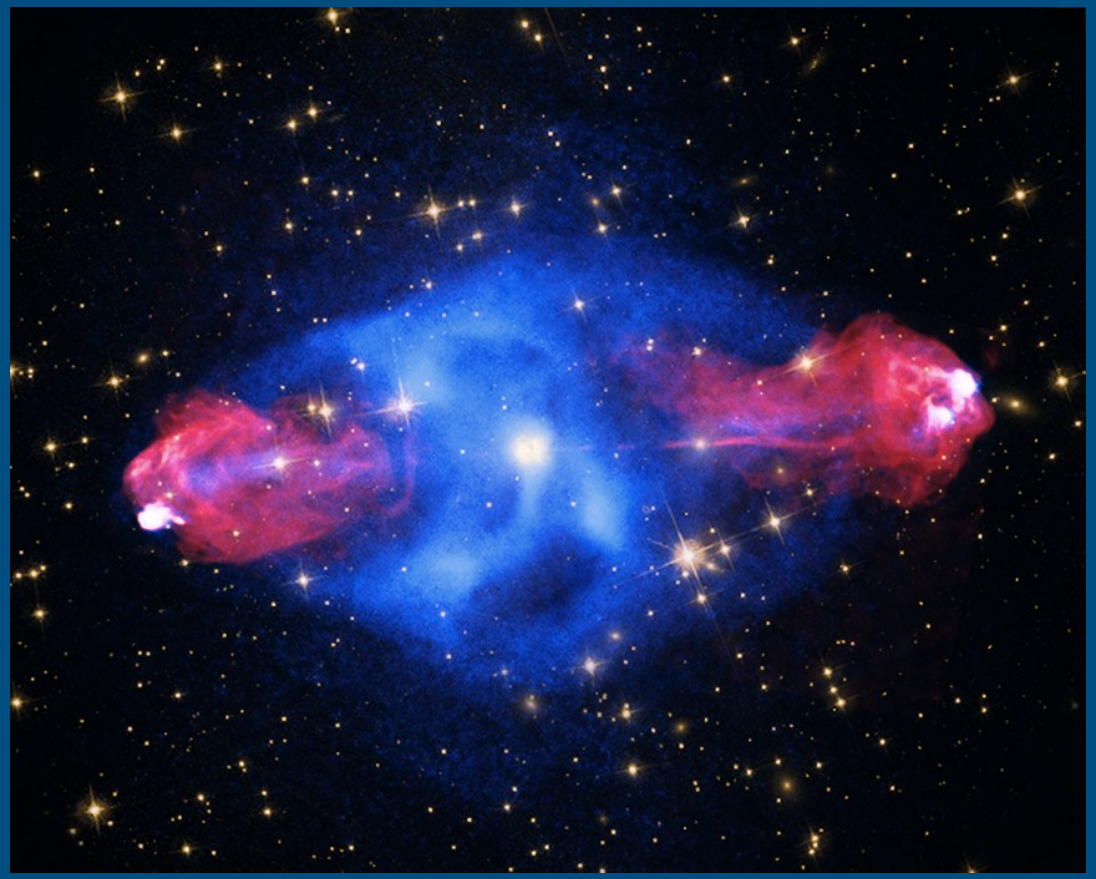
Access to large-scale resources

- Compute
- Storage

Challenges for the LOFAR community

Discoverability

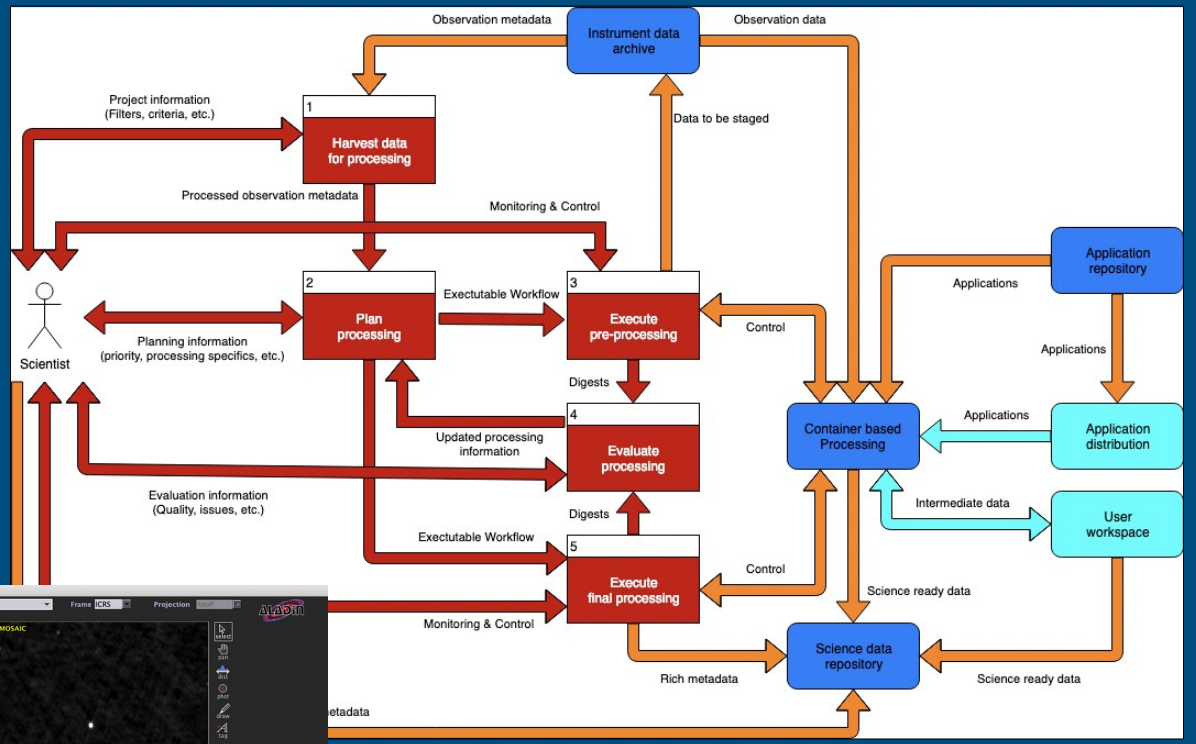
- Direct access to science-ready data
- Data exposed through astronomy standards
    - Plus: PIDs, harvesting by general-purpose EOSC data registries

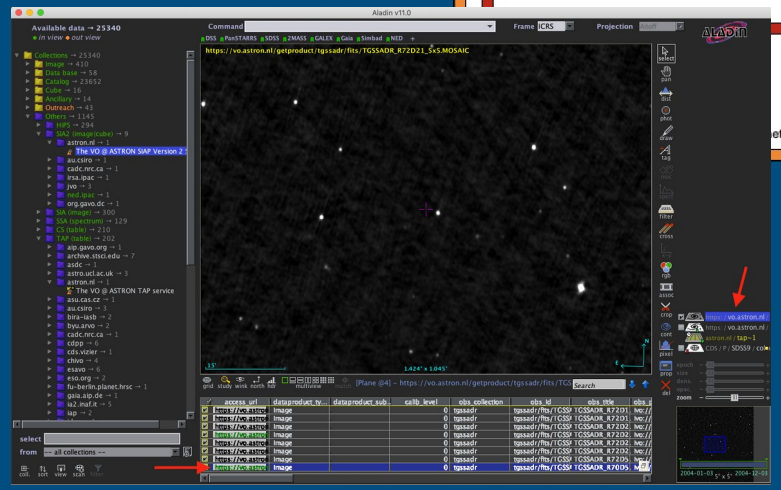


The answer: the Science Data Centre

User benefits

- Access through federated infrastructure
- Data processing to science level
- Community re-use of processing pipelines





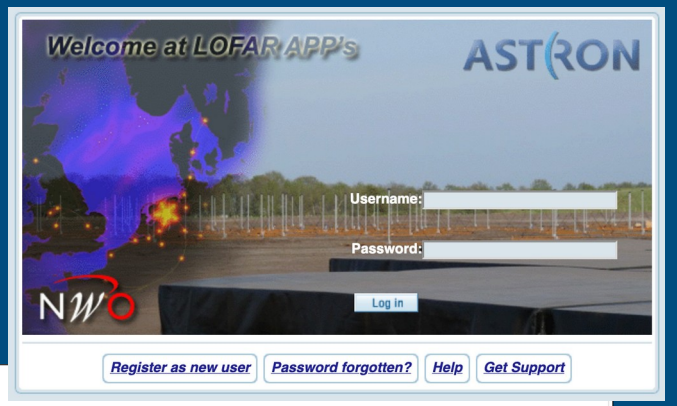
Architecture

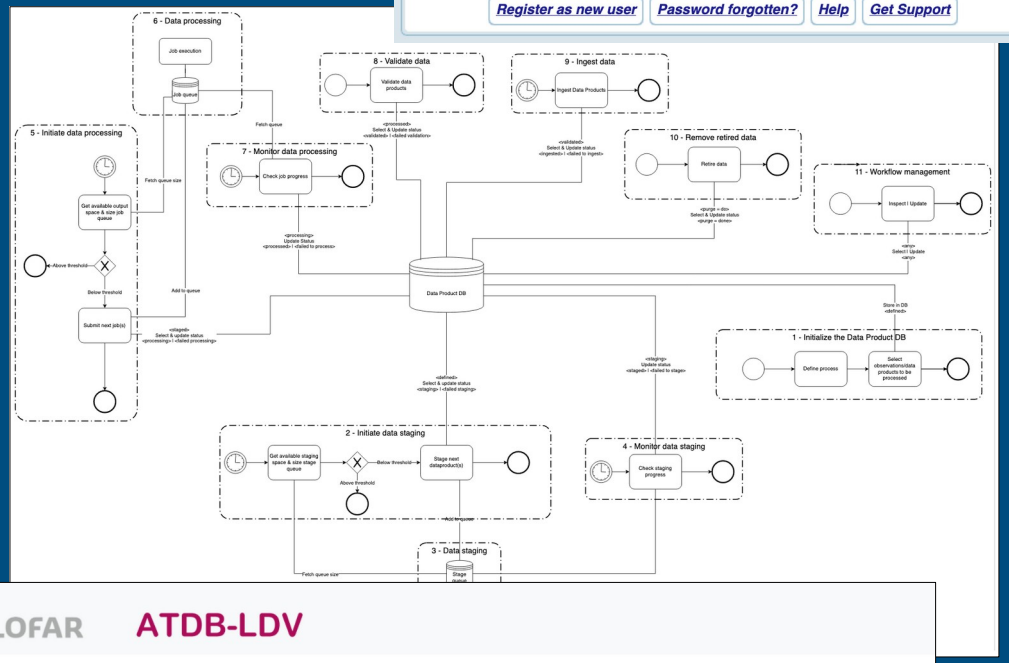
Offer managed science processing for standard pipelines

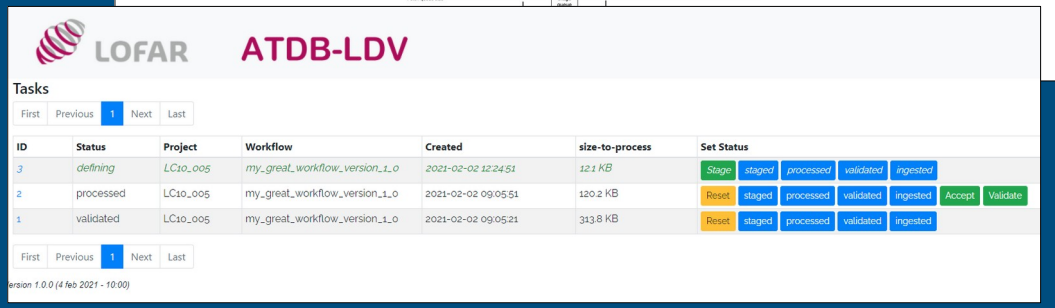
- Community request processing through LOFAR proposal calls
- Execute on existing archived data or on new observations
- Supported through the ASTRON Science Data Centre Operations

Distributed service infrastructure

Publicly available application container images







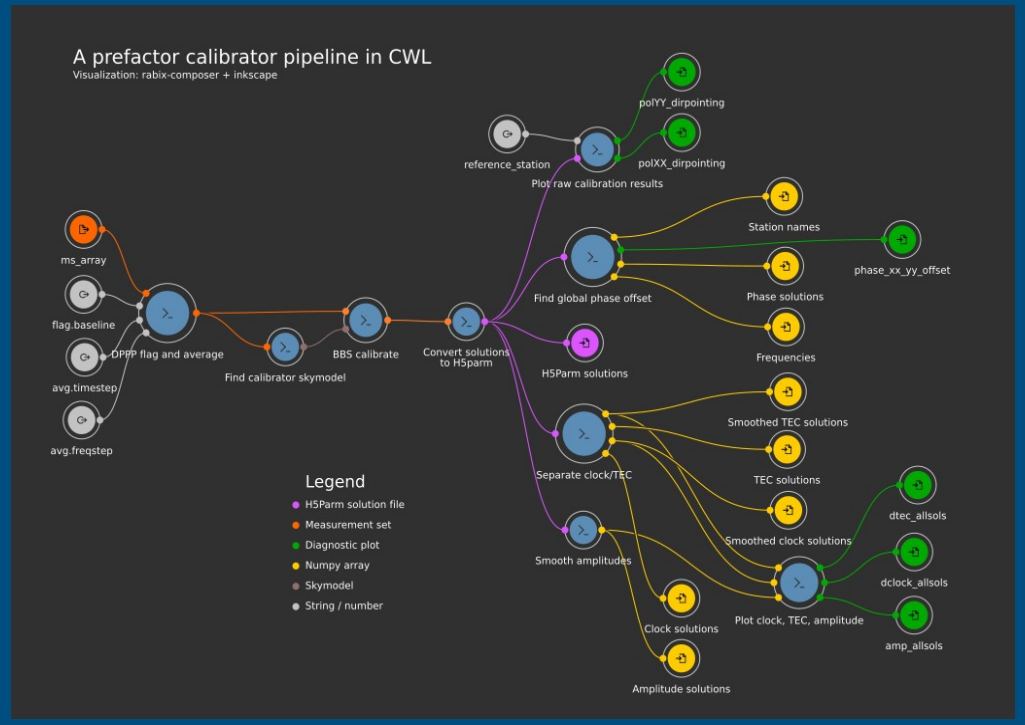
Technical requirements

Location: SURF (data locality)

- Later: all LOFAR archive sites

Compute: HTC

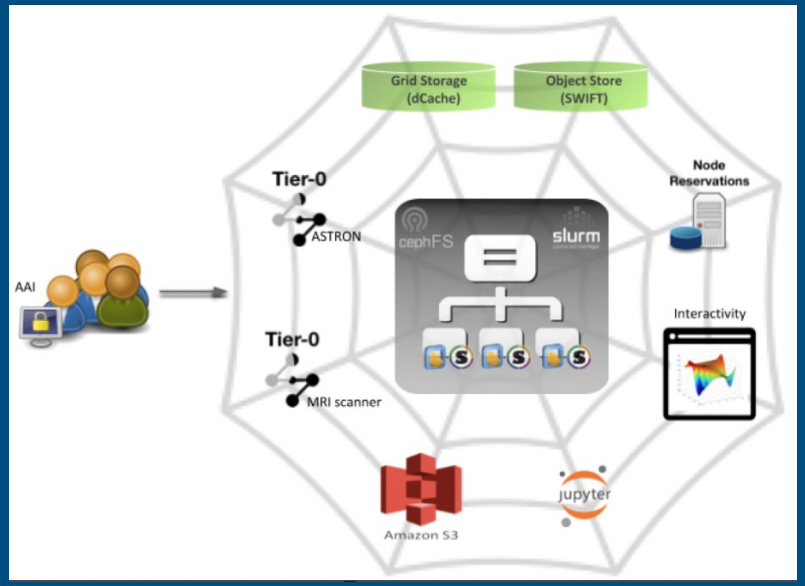
- Initial processing embarrassingly parallel with low compute over data volume ratio
- Deploy/execute: Singularity
- Slurm (later to consider DIRAC)
- Common Workflow Language support
- Next phase may benefit from HPC and GPU

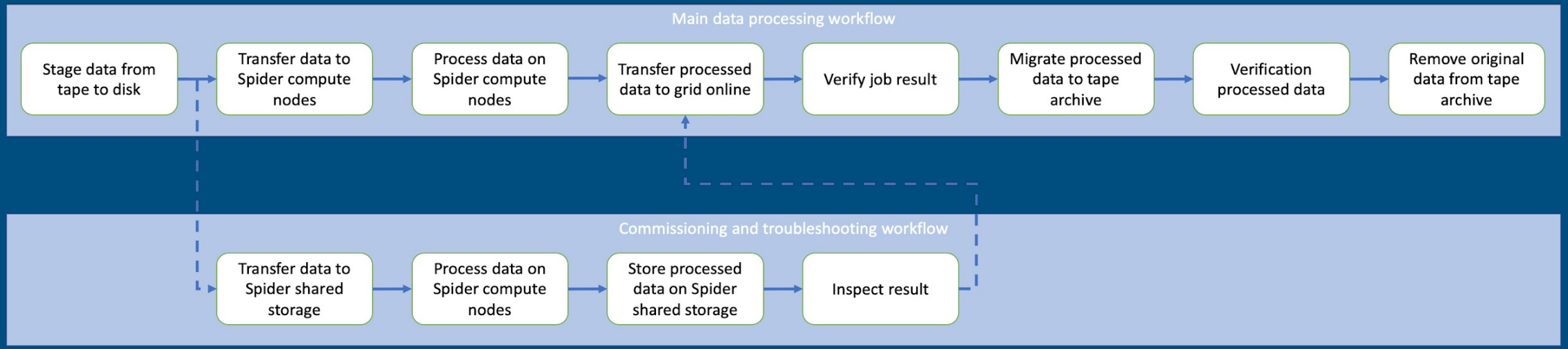


Technical requirements

Data access – dCache

- Tape – disk staging
- FTS (later to consider RUCIO)





Training for external users

LOFAR Data School (bi-yearly)

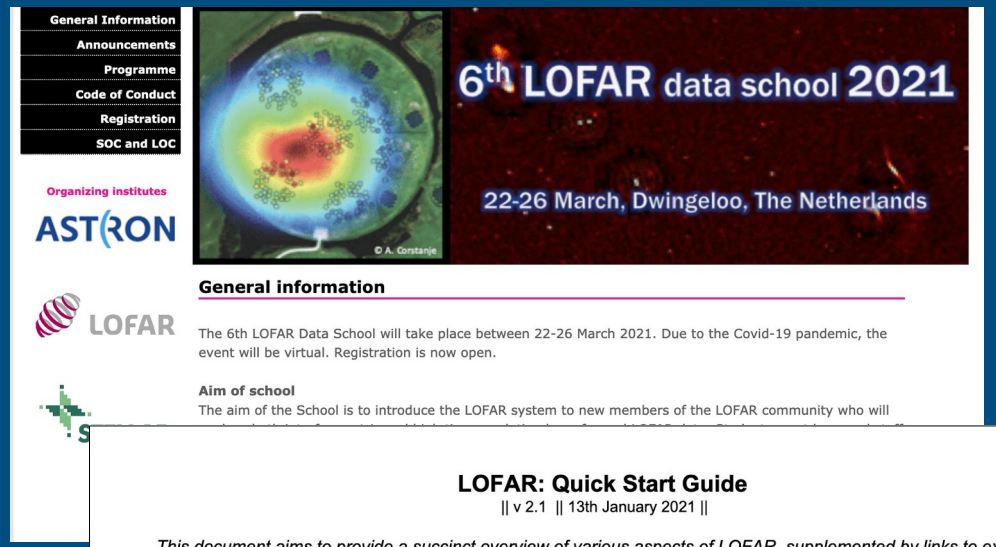
- March 2021: too early for EGI-ACE
- Next: 2023 (TBC)

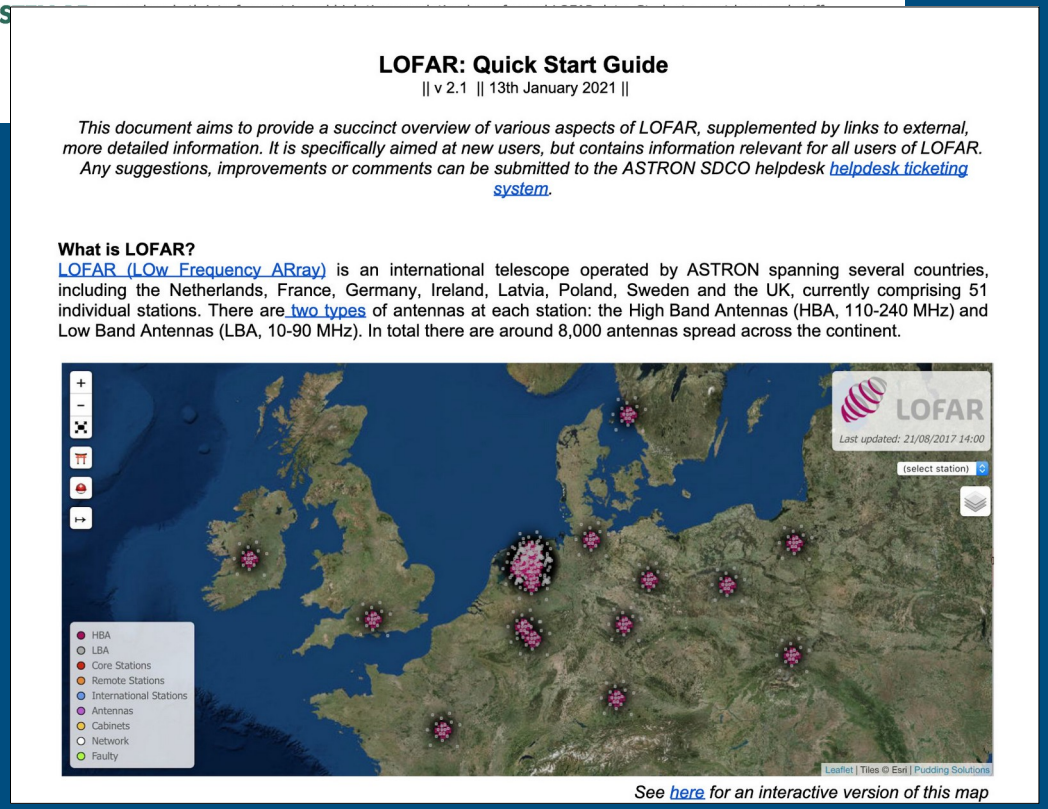
Online documentation

- System capabilities
- Science project preparation
- Data analysis cook books

Standing support organisation

International astronomy conferences





From prototyping to production

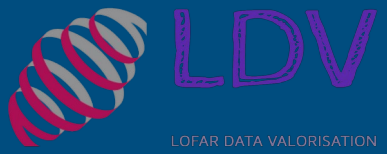
We are realising the Science Data Centre via EU-funded and internally funded projects

EOSC-Hub

LOFAR Data Valorization

EGI-ACE

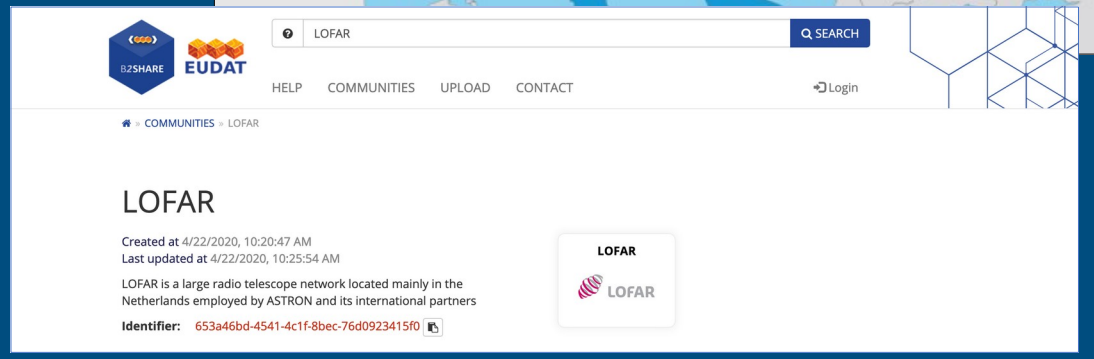
DICE



EOSC-Hub

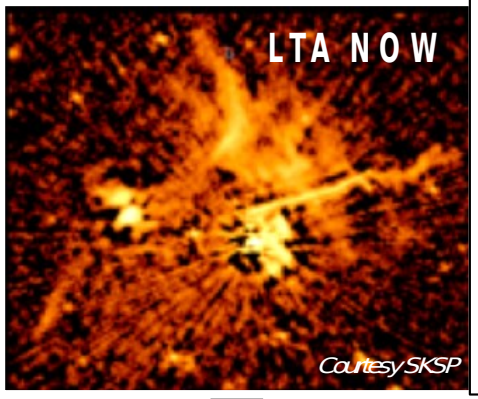
Prototyping

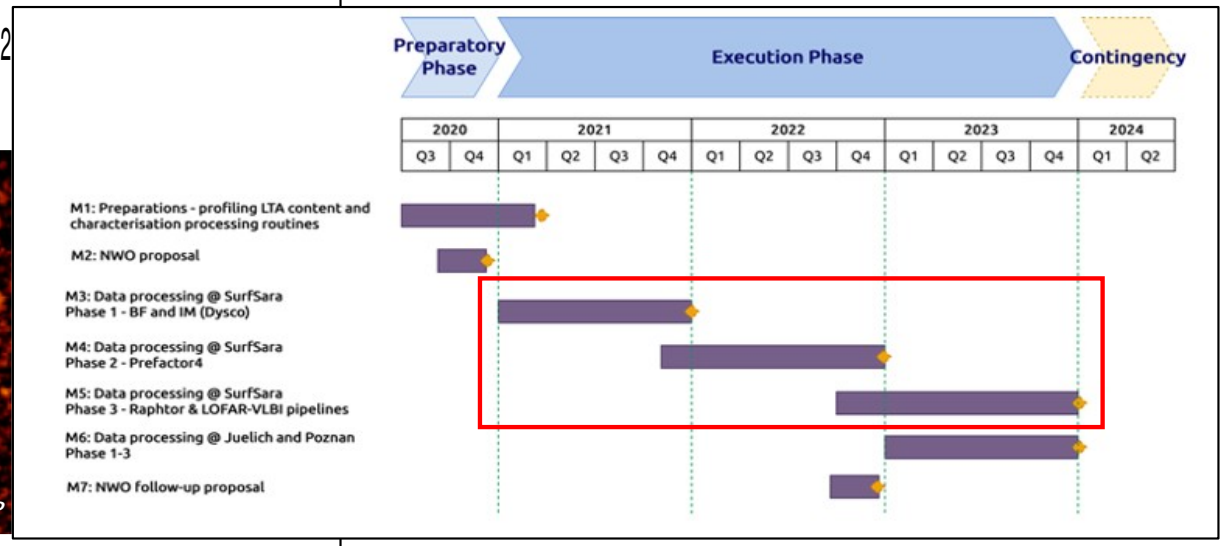
- Demonstrated integration with EOSC infrastructure
- Federated AAI
- Portable and scalable workflows
- FAIR sharing of science-level data



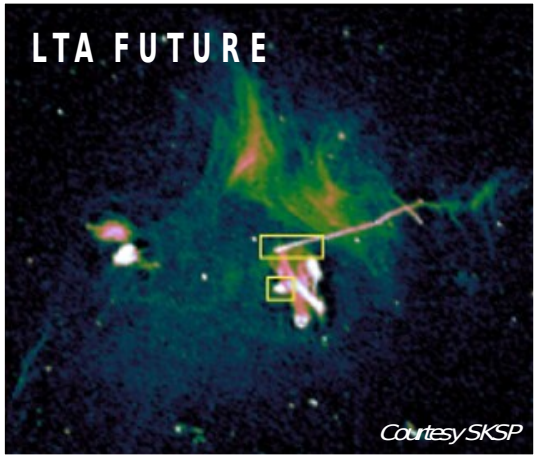
LOFAR Data Valorization

LOFAR Data Valorization (2020-202
data science-ready





LDV
LOFAR DATA VALORISATION

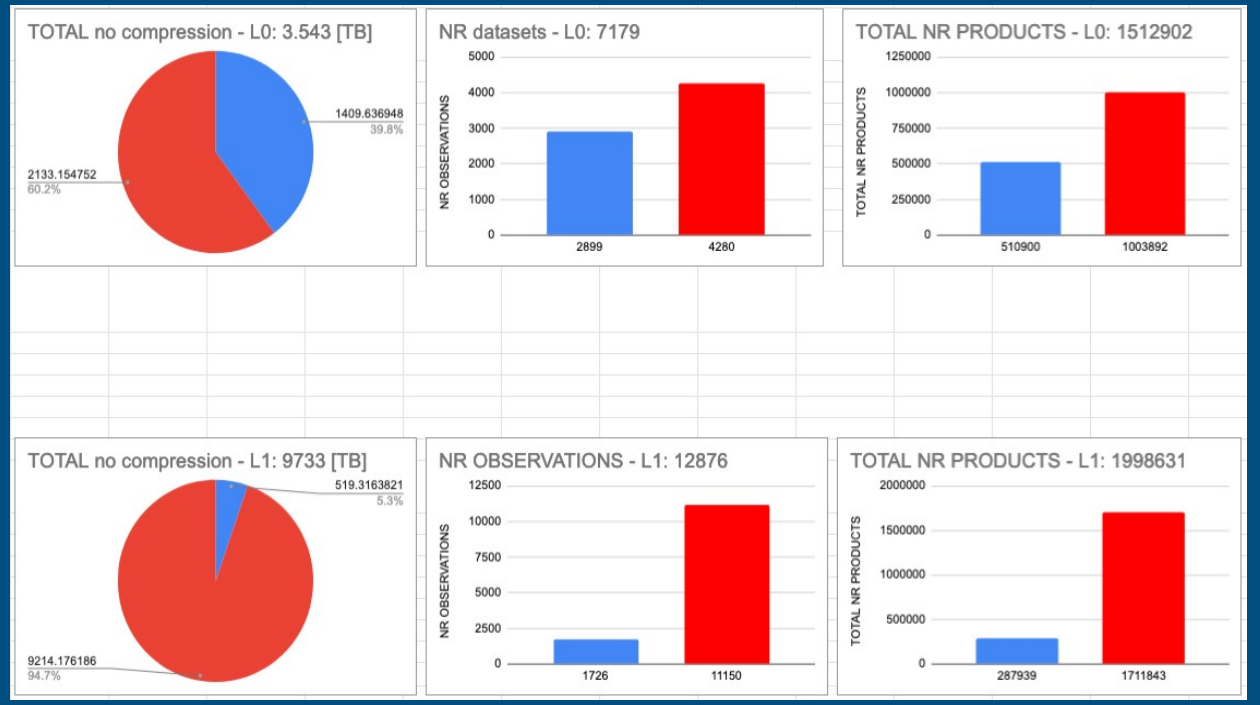


- Give added value to LOFAR data in LTA
- Reduce data volume at the LTA to reduce operational costs
  - Streamline data processing operations at the LTA
  - Prepare ASTRON for DUPLLO surveys
- 3 phases: compression, direction-independent products, science-ready products.
- Start with processing at SURF
- Data products to be made available through subsequent data releases

LOFAR Data Valorization

Phase I: Pre-processing at SURF

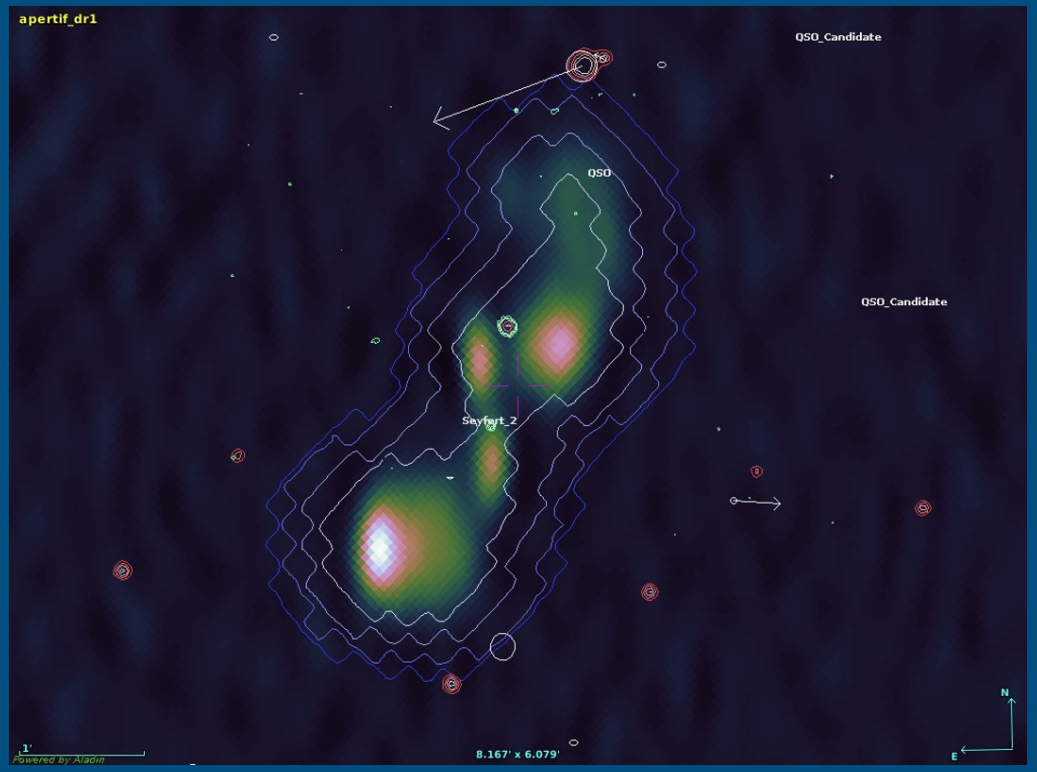
- Compression on 13 PB of data in the archive (factor 3.5 reduction)
- Homogenise data in the archive
- Apply known corrections to instrument issues
- Generate quality indicators
- Prepare for further automated processing
    - Calibration
    - Imaging
    - Source finding



EGI-ACE

Providing virtual access to radio astronomy data

- Building on EOSC architecture
- Available from EOSC portal
- Community contributions
- Data processing as a virtual access service

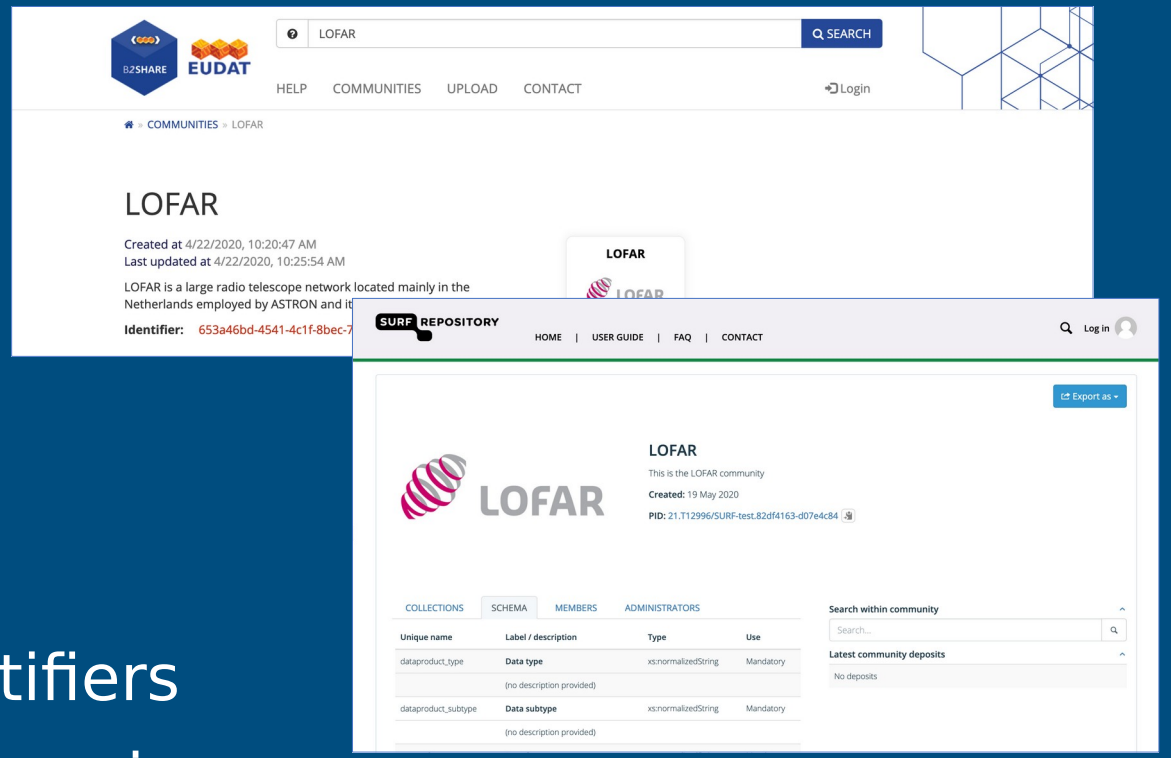


Offer LOFAR processing as a service from end 2021

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 service
- Publication compliant with Virtual Observatory standards



Offer LOFAR Science Data Repository as a service from end 2021