



# ASTRON

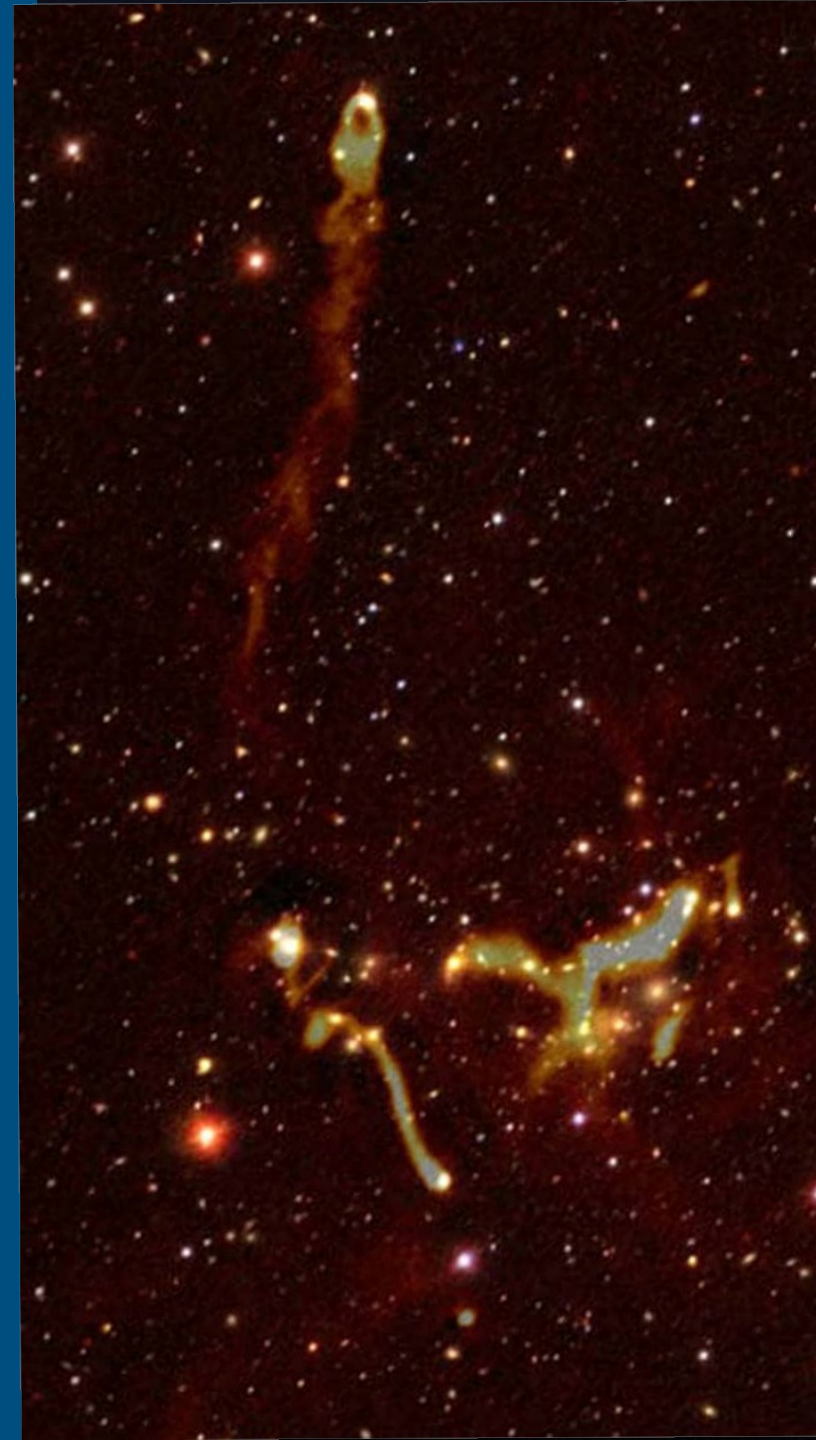
Netherlands Institute for Radio Astronomy

# 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

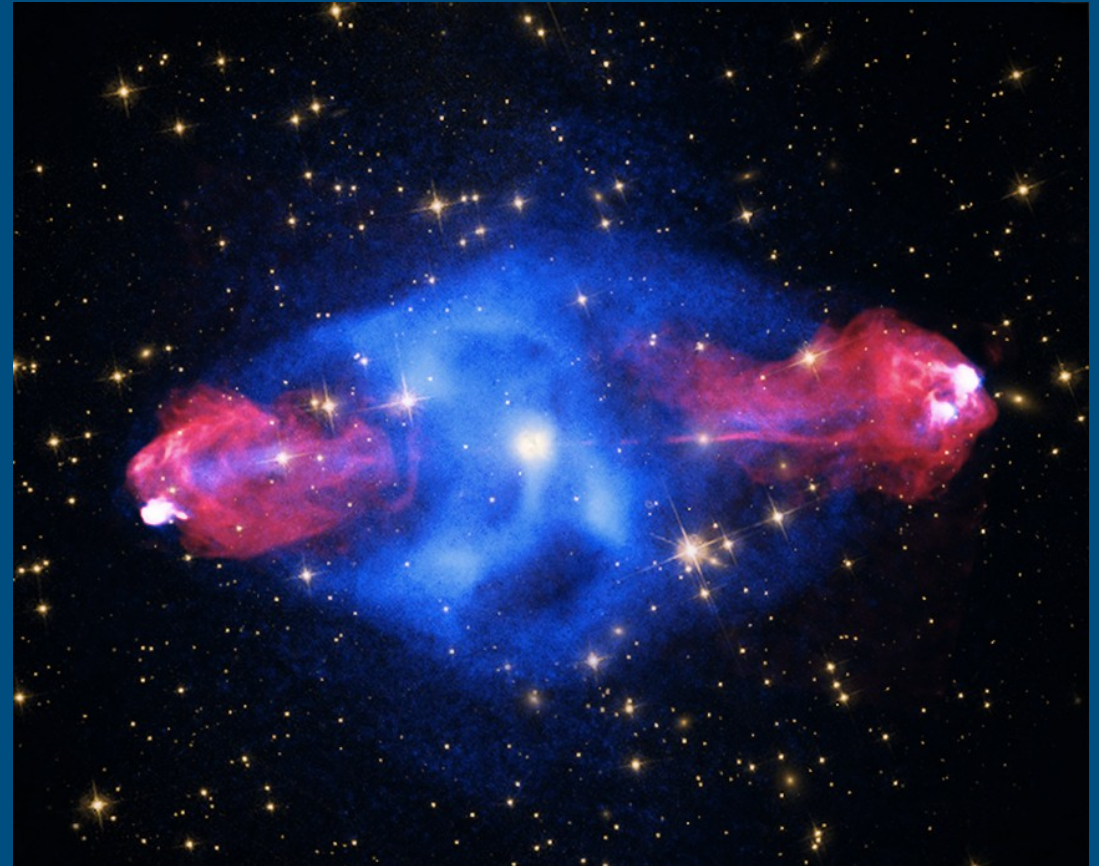
- Compute
- Storage



# Challenges for the LOFAR community

## Discoverability

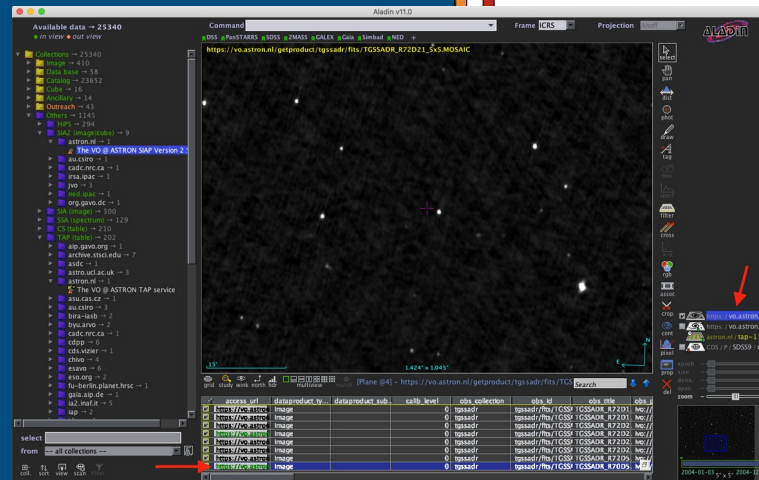
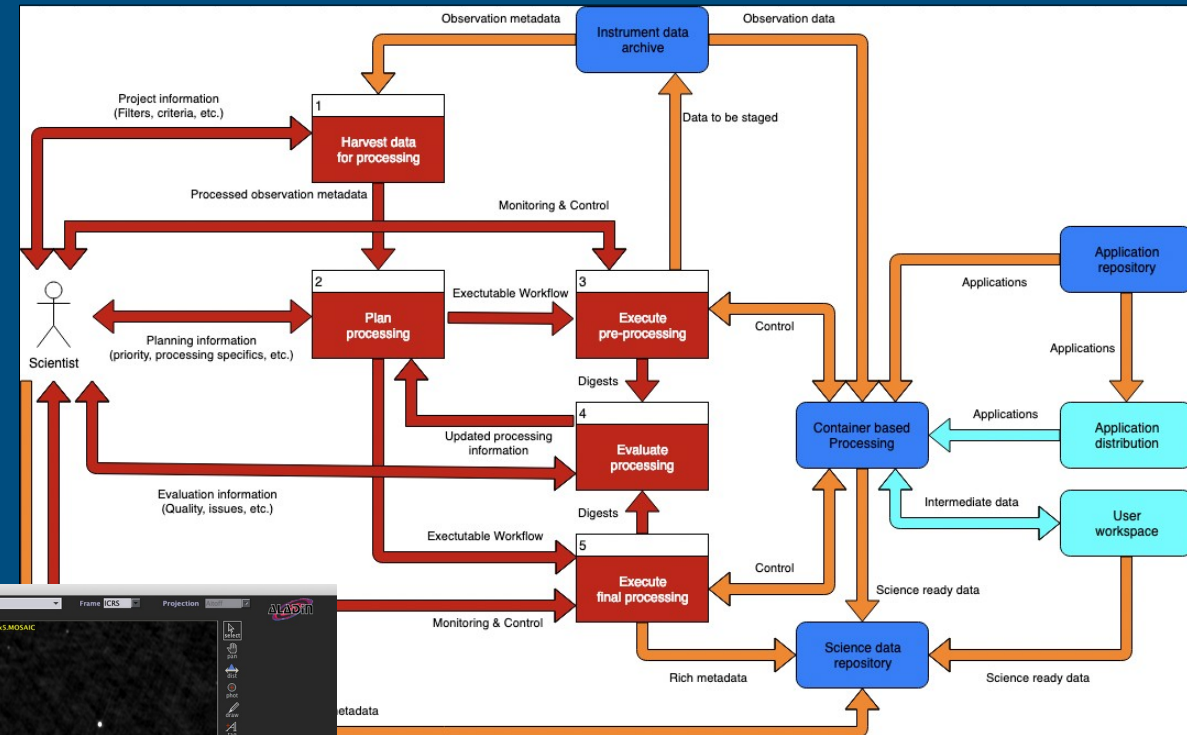
- Direct access to science-ready data
- Data exposed through astronomy standards
  - Plus: PIDs, harvesting by general-purpose EOSC data registries



# The answer: the Science Data Centre

## User benefits

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

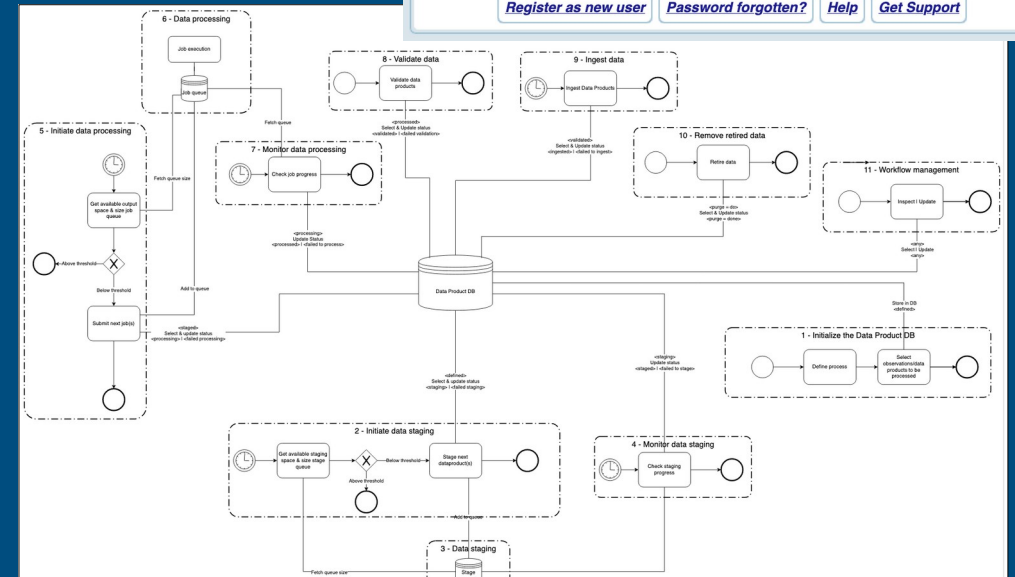
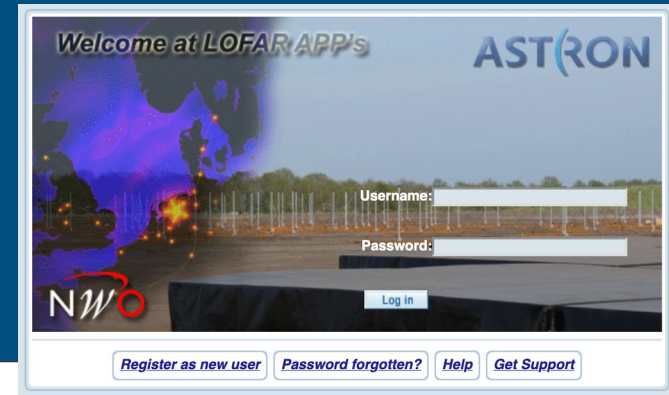



# Architecture

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



 **LOFAR** **ATDB-LDV**

Tasks

First Previous **4** Next Last

ID	Status	Project	Workflow	Created	size-to-process	Set Status
3	defining	LC10_005	my_great_workflow_version_1_0	2021-02-02 12:24:51	12.1 KB	<a href="#">Stage</a> <a href="#">staged</a> <a href="#">processed</a> <a href="#">validated</a> <a href="#">ingested</a>
2	processed	LC10_005	my_great_workflow_version_1_0	2021-02-02 09:05:51	120.2 KB	<a href="#">Reset</a> <a href="#">staged</a> <a href="#">processed</a> <a href="#">validated</a> <a href="#">ingested</a> <a href="#">Accept</a> <a href="#">Validate</a>
1	validated	LC10_005	my_great_workflow_version_1_0	2021-02-02 09:05:21	313.8 KB	<a href="#">Reset</a> <a href="#">staged</a> <a href="#">processed</a> <a href="#">validated</a> <a href="#">ingested</a>

First Previous **4** Next Last

Version 1.0.0 (4 Feb 2021 - 10:00)

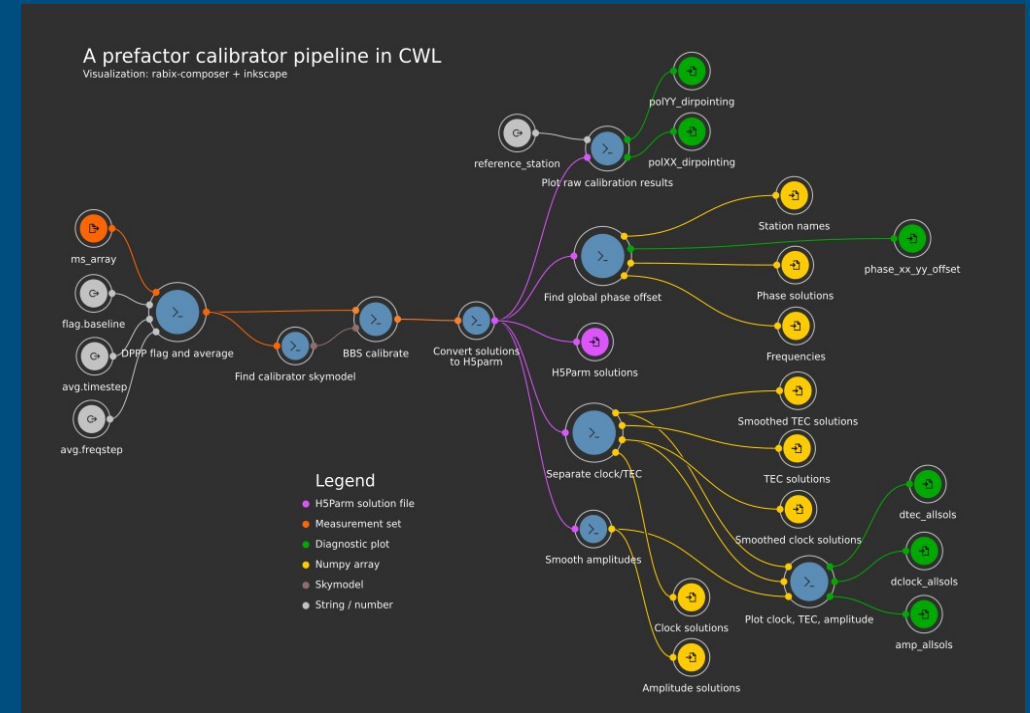
# Technical requirements

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

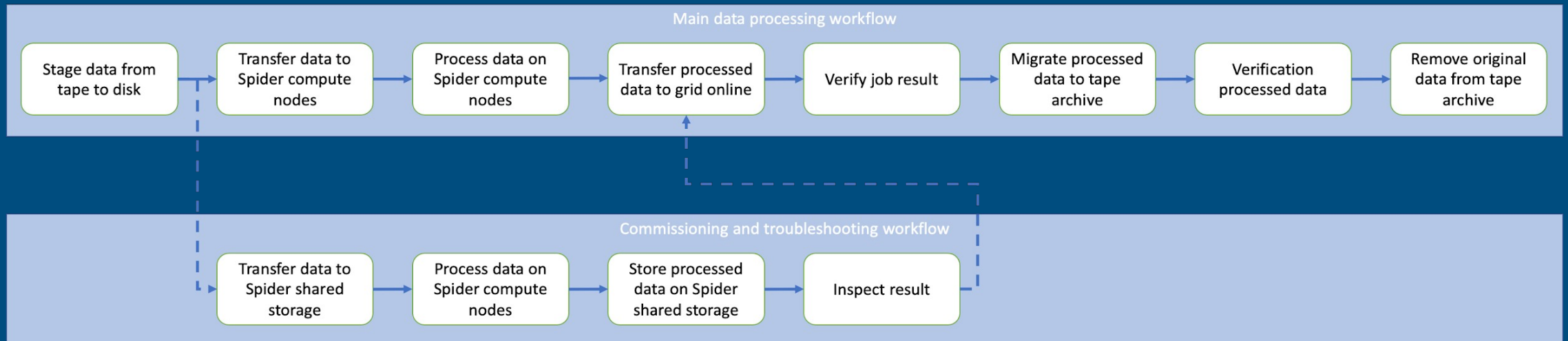
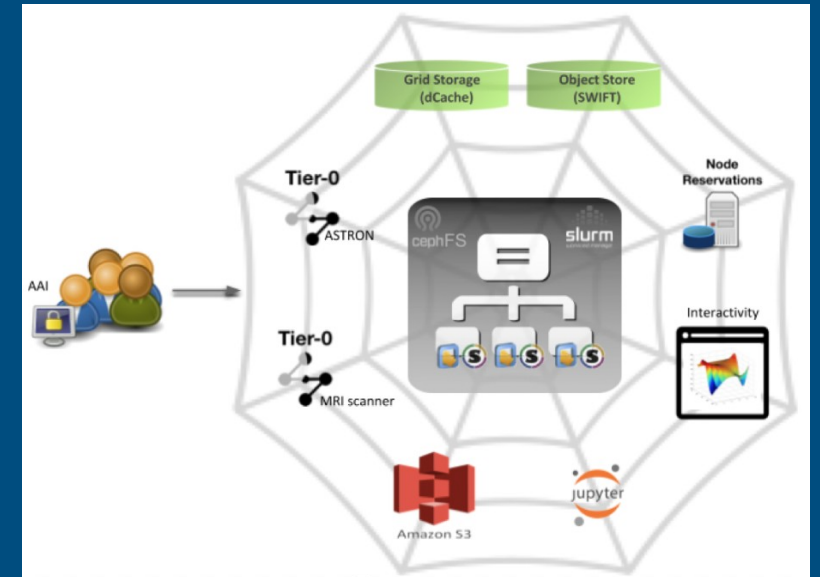




# Technical requirements

## Data access – dCache

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



# Training for external users

## LOFAR Data School (bi-yearly)

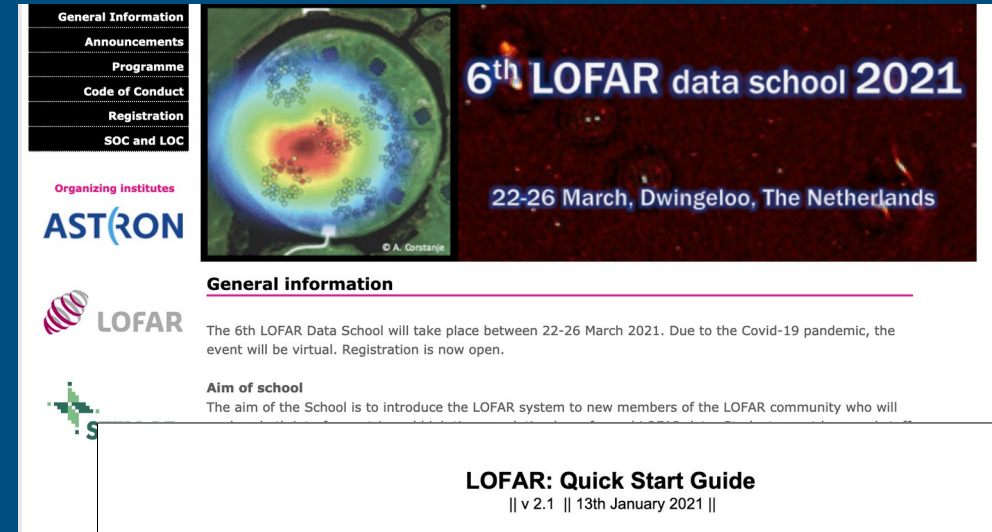
- March 2021: too early for EGI-ACE
- Next: 2023 (TBC)

## Online documentation

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

## Standing support organisation

## International astronomy conferences



The screenshot shows the LOFAR website's navigation menu on the left, including links for General Information, Announcements, Programme, Code of Conduct, Registration, and SOC and LOC. Below the menu is the ASTRON logo and the LOFAR logo. A large banner on the right features a colorful image of the LOFAR antenna array and text for the "6th LOFAR data school 2021" held from 22-26 March in Dwingeloo, The Netherlands. Below the banner, there is a "General information" section with a sub-header "General information" and a paragraph stating that the 6th LOFAR Data School will be virtual due to the COVID-19 pandemic. An "Aim of school" section follows, stating the goal is to introduce the LOFAR system to new members.

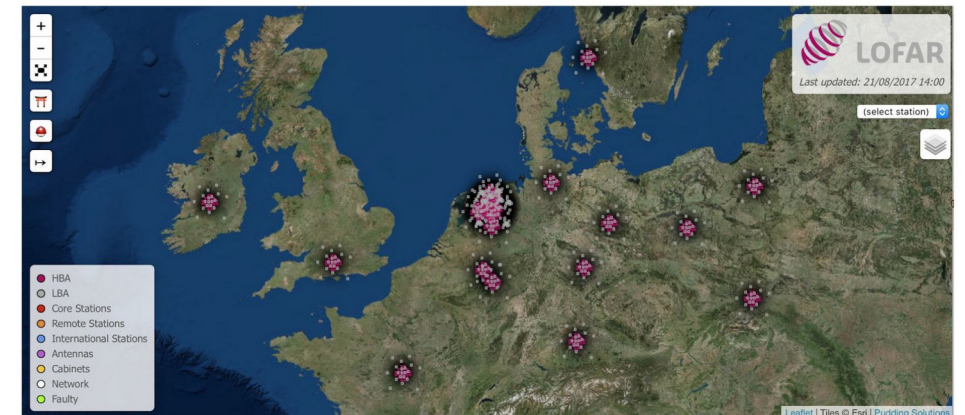
### LOFAR: Quick Start Guide

|| v 2.1 || 13th January 2021 ||

*This document aims to provide a succinct overview of various aspects of LOFAR, supplemented by links to external, more detailed information. It is specifically aimed at new users, but contains information relevant for all users of LOFAR. Any suggestions, improvements or comments can be submitted to the ASTRON SDCO helpdesk [helpdesk ticketing system](#).*

#### What is LOFAR?

[LOFAR \(Low Frequency ARray\)](#) is an international telescope operated by ASTRON spanning several countries, including the Netherlands, France, Germany, Ireland, Latvia, Poland, Sweden and the UK, currently comprising 51 individual stations. There are [two types](#) of antennas at each station: the High Band Antennas (HBA, 110-240 MHz) and Low Band Antennas (LBA, 10-90 MHz). In total there are around 8,000 antennas spread across the continent.



See [here](#) for an interactive version of this map

# 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

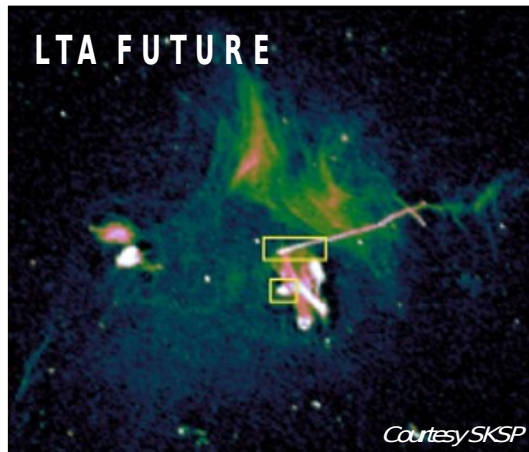
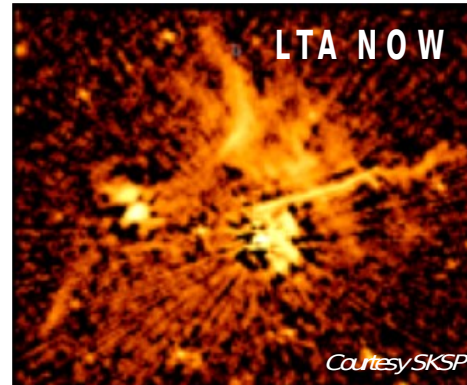
- Demonstrated integration with EOSC infrastructure
- Federated AAI
- Portable and scalable workflows
- FAIR sharing of science-level data



A screenshot of the EOSC Hub interface. The top navigation bar includes the B2SHARE and EUDAT logos, a search bar with "LOFAR" entered, and a "SEARCH" button. Below the navigation bar are links for "HELP", "COMMUNITIES", "UPLOAD", and "CONTACT", along with a "Login" button. The main content area displays the "LOFAR" community page, which includes the text: "Created at 4/22/2020, 10:20:47 AM", "Last updated at 4/22/2020, 10:25:54 AM", and "LOFAR is a large radio telescope network located mainly in the Netherlands employed by ASTRON and its international partners". An identifier is provided: "Identifier: 653a46bd-4541-4c1f-8bec-76d0923415f0". A small LOFAR logo is also visible on the right side of the page.

# LOFAR Data Valorization

LOFAR Data Valorization (2020-2022)  
data science-ready



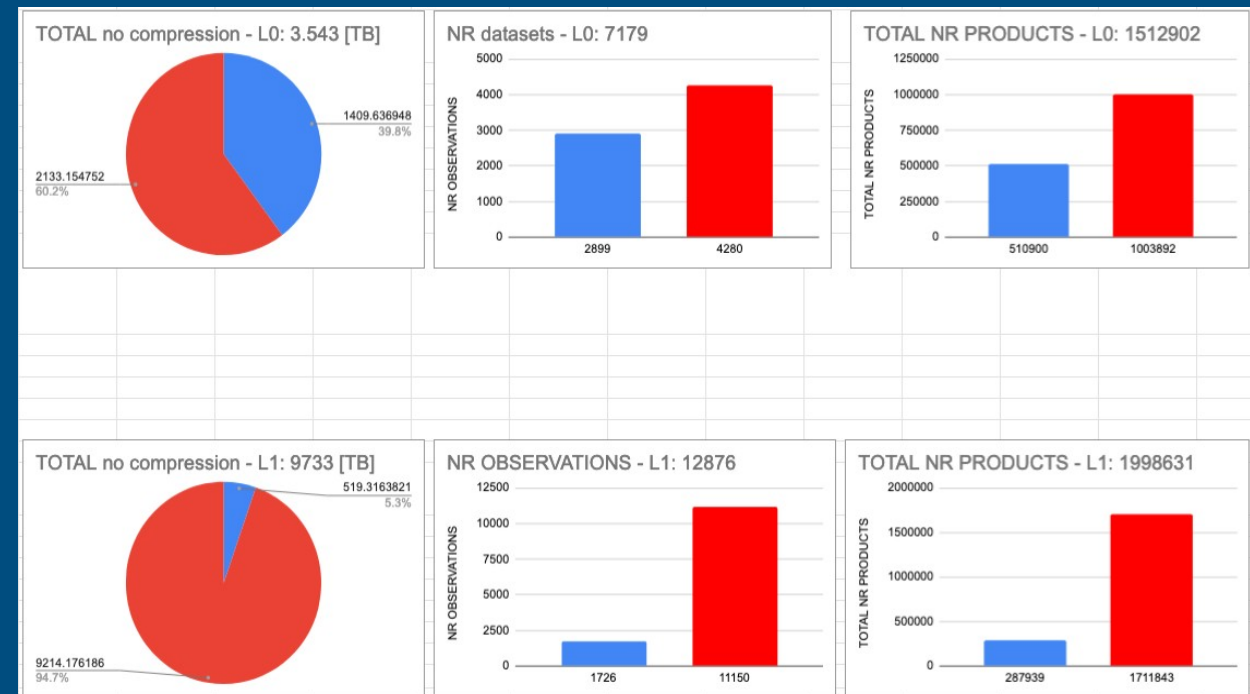
- ∅ Give added value to LOFAR data in LTA
- ∅ Reduce data volume at the LTA to reduce operational costs
  - § Stream line data processing operations at the LTA
  - § Prepare ASTRON for DUPLLO surveys
- ∅ 3 phases: compression, direction-independent products, science-ready products.
- ∅ Start with processing 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



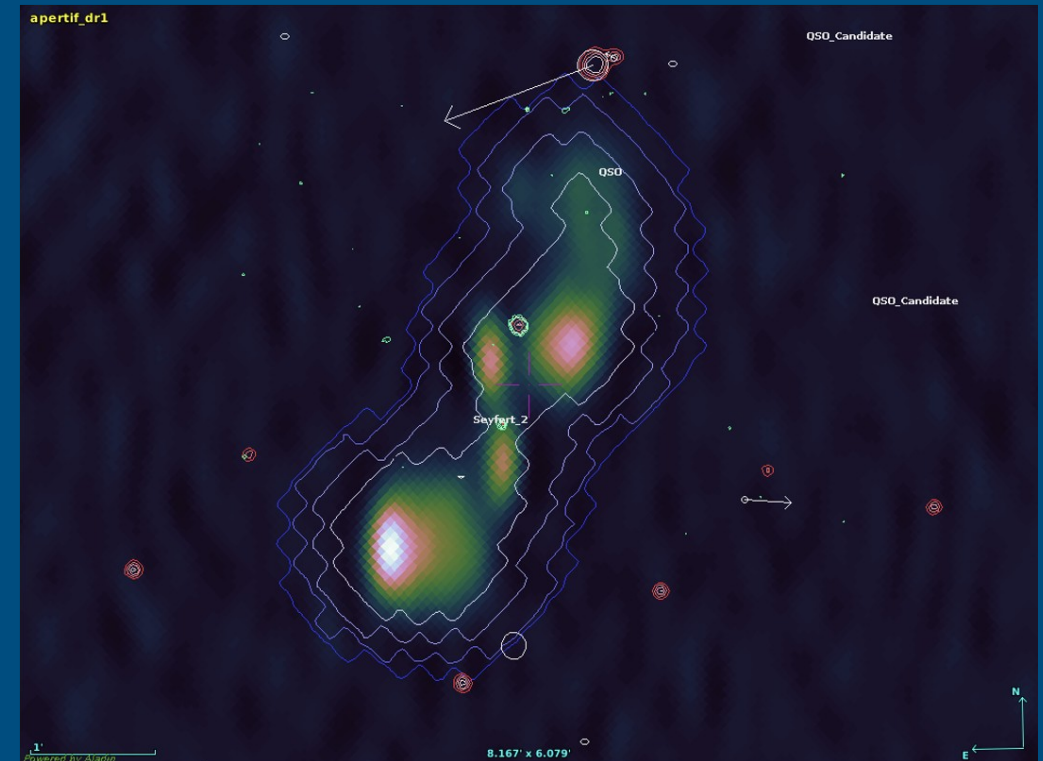
# EGI-ACE



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**

The top screenshot shows the LOFAR community page on the SURF REPOSITORY. It includes the B2SHARE and EUDAT logos, a search bar, and navigation links. The page title is 'LOFAR' and it displays metadata: 'Created at 4/22/2020, 10:20:47 AM' and 'Last updated at 4/22/2020, 10:25:54 AM'. A description states: 'LOFAR is a large radio telescope network located mainly in the Netherlands employed by ASTRON and its partners'. An identifier is provided: 'Identifier: 653a46bd-4541-4c1f-8bec-7...'. The bottom screenshot shows the SURF REPOSITORY interface for the LOFAR community. It features the SURF REPOSITORY logo, navigation links (HOME, USER GUIDE, FAQ, CONTACT), and a search bar. The page title is 'LOFAR' and it displays the community description: 'This is the LOFAR community', 'Created: 19 May 2020', and 'PID: 21.T12996/SURF-test.82df4163-d07e4c84'. Below this is a table with columns 'Unique name', 'Label / description', 'Type', and 'Use'. The table contains two rows: 'dataprodct\_type' with 'Data type', 'xs:normalizedString', and 'Mandatory'; and 'dataprodct\_subtype' with 'Data subtype', 'xs:normalizedString', and 'Mandatory'. A search bar and 'Latest community deposits' section are also visible.

Unique name	Label / description	Type	Use
dataprodct_type	Data type (no description provided)	xs:normalizedString	Mandatory
dataprodct_subtype	Data subtype (no description provided)	xs:normalizedString	Mandatory