



# OSCARS

Open Science Clusters' Action  
for Research & Society

## Funded Project

# Creating a FAIR open Science Pipeline for high-resolution LOFAR imaging (COPLI)

Presenter: Jurjen de Jong, Leiden Observatory

Implemented by



Universiteit  
Leiden  
The Netherlands



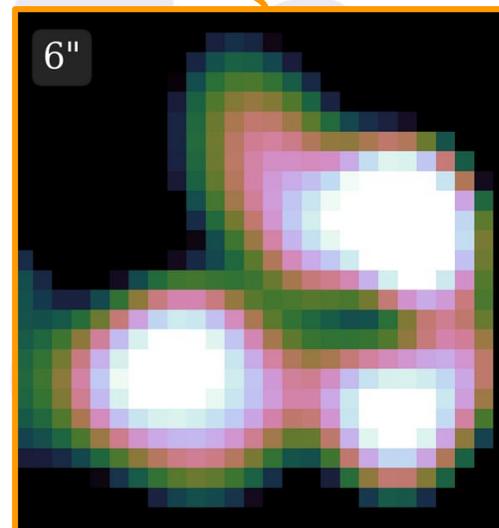
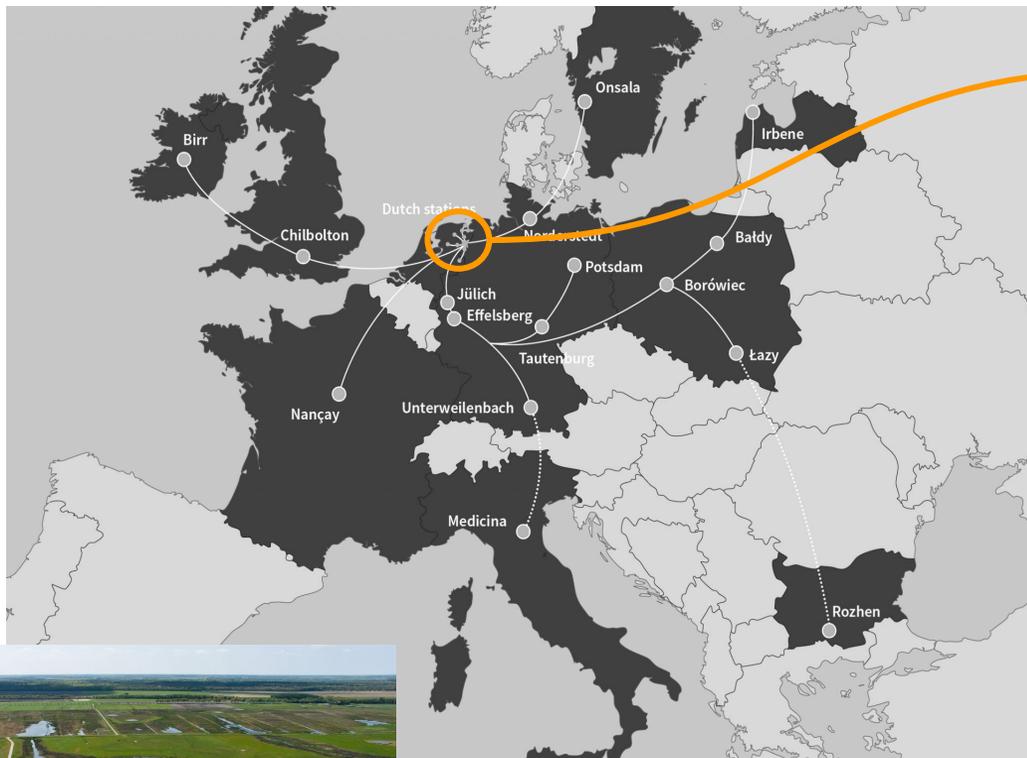
**ASTRON**

Netherlands Institute for Radio Astronomy



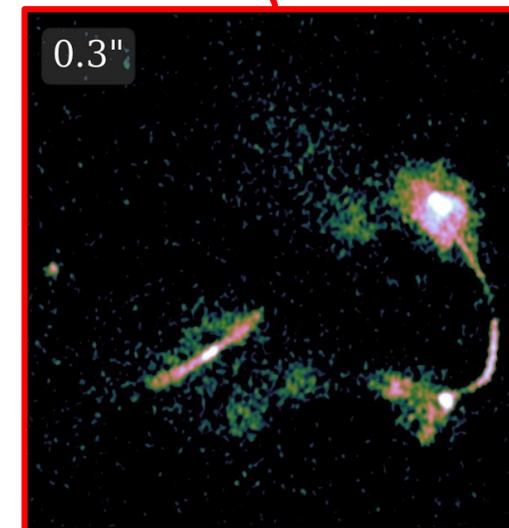
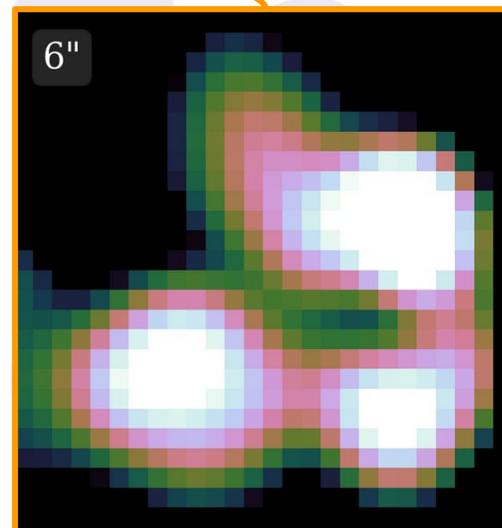
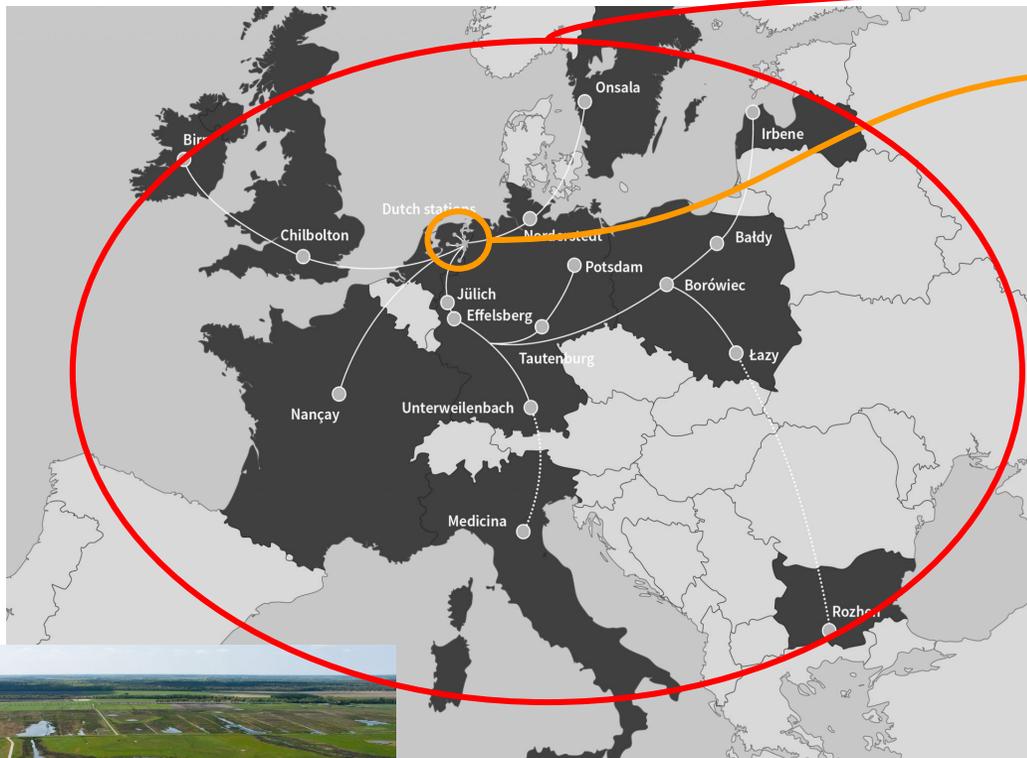
Funded by  
the European Union

# CHALLENGE ADDRESSED



LOFAR

# CHALLENGE ADDRESSED



01

**Improve data processing strategies**

- High-quality images
- Reduce computational cost
- Reduce data volume cost

02

**Develop automated open FAIR data processing pipeline**

- Findable → Output data traceable
- Accessible → Open-source software
- Interoperable → Running on various HPC systems
- Reproducible → Proper logging of processing steps

03

**Generate science-ready products**

- Findable with DOIs, PIDs
- Available in the VO
- HIPs format

01

Improve data processing strategies

- ✓ High-quality images
- ✓ Reduce computational cost
- ✓ Reduce data volume cost

02

Develop automated open FAIR data processing pipeline

- Findable → Output data traceable
- ✓ Accessible → Open-source software
- ✓ Interoperable → Running on various HPC systems
- Reproducible → Proper logging of processing steps

03

Generate science-ready products

- Findable with DOIs, PIDs
- Available in the VO
- ✓ HIPs format

- ❖ Publicly available pipeline and software
  - de Jong et al. (2025)
  - Van der Wild et al. (submitted)
- ❖ Science products in an open accessible repository with associated metadata and data provenance (EOSC)
  - de Jong et al. (in prep.)

## COPLI data processing life cycle

### 06 – Access & reuse

- All results are open and free
- Data registered in VO ( Virtual Observatory )
- Software registered in ESCAPE open-source scientific software repository ( OSSR )

### 05 – Register results

- Extended schema's
- Results hosted and registered in trusted SURF Data Repository
- DOI for Data collection
- PID for all digital objects
- Curation policy

### 04 – Data lineage

- Data+Software provenance
- Processing history
- Enhanced metadata



### 01 – Find & access data

- LOFAR long term archive
- LOFAR ERIC tooling

### 02 – Prepare software

- Github ( long baseline group )
- Optimized containerization
- Deployment via CVMFS

### 03 – Process data

- SURF HPC ( NWO 2023.040 )
- CWL + TOIL/StreamFlow
- Processing history
- Diagnostics plots
- HIPS format

## Who has been doing it?

- Dr. Reinout van Weeren (PI)  
<https://orcid.org/0000-0002-0587-1660>
- Dr. Jurjen de Jong  
<https://orcid.org/0000-0001-6876-8719>

