
Data Processing Pipelines

LTA Development Meeting

2026-06-30

ASTRON

Netherlands Institute for Radio Astronomy



The Vision

- The LTA is *the* repository for scientific data produced by LOFAR.
- This includes *both* instrumental data (when retained) and “advanced” / “science ready” data products (when available).
- Modulo proprietary time, advanced data products are accessible to all users
 - They have the technical ability to download them.
 - They have appropriate documentation & support to understand them.

The Consequence

Either

data is processed within the LTA (i.e. on compute systems at the LTA sites — “*centralized processing*”) and from there stored directly to the archive;

or

data is processed elsewhere (“*federated processing*”) and from there uploaded to the archive (“*user data ingest*”).

In this session we focus on the former.

Operational Oversight

- The Observatory Operations team is responsible for managing execution across the LTA network.
- That is, they are responsible for determining:
 - Which data should be processed,
 - On which systems,
 - With which software,
 - When.
- Observing proposals will be allocated a quota of compute/storage time; Observatory Ops will track aggregated usage across LTA sites.



Click to Filter ()

✖ Clear Filter

☰ Active

☰ Failed

☰ On Hold

defining

defined

staging

staged

fetches

processing

processed

stored

validated

scrubbed

(pre) archiving

(pre) archived

finished

suspended

Search for...

First

Previous

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

...

Next

Last

▲ ID ▼	▲ type ▼	▲ Spec ▼	Details	▲ WF ▼	▲ Status ▼	▲ Project ▼	▲ SAS_ID ▼	▲ Processing ▼	▲ Ingest Policy ▼	▲ Filter ▼	▲ Size ▼	Actions
470360	aggregation	735		 36	idle	lc4_034	346946 => 202600000432	lofar-srm.fz-juelich.de	lofar-srm.fz-juelich.de	lc4_034_fzj-aa-comp	0 bytes	
470359	regular	735		 36	staging	lc4_034	346946 => 202600000432	lofar-srm.fz-juelich.de	lofar-srm.fz-juelich.de	lc4_034_fzj-aa-comp	2.2 GB	
470358	regular	735		 36	staged_failed	lc4_034	346946 => 202600000432	lofar-srm.fz-juelich.de	lofar-srm.fz-juelich.de	lc4_034_fzj-aa-comp	4.3 GB	 Retry
470357	regular	735		 36	staging	lc4_034	346946 => 202600000432	lofar-srm.fz-juelich.de	lofar-srm.fz-juelich.de	lc4_034_fzj-aa-comp	4.3 GB	
470356	regular	735		 36	staging	lc4_034	346946 => 202600000432	lofar-srm.fz-juelich.de	lofar-srm.fz-juelich.de	lc4_034_fzj-aa-comp	4.3 GB	
470355	regular	735		 36	staging	lc4_034	346946 => 202600000432	lofar-srm.fz-juelich.de	lofar-srm.fz-juelich.de	lc4_034_fzj-aa-comp	4.3 GB	
470354	regular	735		 36	staging	lc4_034	346946 => 202600000432	lofar-srm.fz-juelich.de	lofar-srm.fz-juelich.de	lc4_034_fzj-aa-comp	4.3 GB	
470353	regular	735		 36	staging	lc4_034	346946 => 202600000432	lofar-srm.fz-juelich.de	lofar-srm.fz-juelich.de	lc4_034_fzj-aa-comp	4.3 GB	



NARCIS

Nimble Application for Resource Control of Interactive System

John Swinbank

JS

NARCIS > PROJECTS

Filter

Project

Search

Name or project title



Code



Author(s)



Select Cycle(s)

☐ In open cycle

Project source

☐ TULP☐ TMSS

Proposal

Type (category)

☒ ACCEPTED☒ REJECTED☒ No status

Projects (160) ^

Cycle(s): LOFAR2_Array_B

[L2CC_007](#) Marco Iacobelli <iacobelli@astron.nl> ...

LOFAR2-793 - LOFAR shall be capable of derivin...

ACCEPTED

REGULAR

Cycle(s): LOFAR2_Array_B

[L2CB_016](#) (No authors listed)LOFAR2-779 Raw visibilities resolution ≤ 1 s

ACCEPTED

REGULAR

Cycle(s): LOFAR2_Array_B

[L2CB_008](#) (No authors listed)

LOFAR2-814-LOFAR shall be capable of observa...

ACCEPTED

REGULAR

Cycle(s): LOFAR2_Array_B

[L2CB_003](#) Aleksandar Shulevski <shulevski@astr...

LOFAR2-832-LOFAR stations shall continuously ...

ACCEPTED

REGULAR

Cycle(s): TULP Commissioning Cycle

[MPS_002](#) Noa Peters <peters@astron.nl> (PI)

Best Proposal Ever - TEST

ACCEPTED

REGULAR

Cycle(s): LOFAR2_Array_B

Cycle: LOFAR2_Array_C

[Export Cycle CSV](#)

Total used/allocated resources:

Cycle: TULP Commissioning Cycle

[Export Cycle CSV](#)

Total used/allocated resources:

Cycle: LOFAR2_Array_B

[Export Cycle CSV](#)

Total used/allocated resources:

Cycle: Cycle 20B

[Export Cycle CSV](#)

Total used/allocated resources:

Data ingested per site

Data ingested per category

Cycle: Cycle 20

[Export Cycle CSV](#)

Total used/allocated resources:

Data ingested per site

Data ingested per category

What are they running?

- As of now, there is not a tightly-defined format for a “pipeline”.
- However, all major pipelines are currently (being) defined as a series of Python and/or C++ components wrapped in CWL.
 - We anticipate evolution of this framework in future
 - (And indeed I personally think we need to make jumping frameworks as lightweight as possible in future)
- ASTRON is not writing all of the pipelines itself; we expect them to be sourced from across the LOFAR community.

What are they running on?



LOFAR LTA Site Requirements

Auteur(s) / Author(s):	Organisatie / Organization	Datum / Date
John D. Swinbank Hanno Holties Roberto Pizzo	ASTRON ASTRON ASTRON	2025-07-29
Controle / Checked:		
Michiel P. van Haarlem	LOFAR/ERC	2025-05-08
Goedkeuring / Approval:	...	...
Autorisatie / Authorization:	...	...
Handtekening / Signature:	...	...

This technical note expresses the opinions of the authors on the date of writing.
It is not binding on the SDC Program, and may not accurately reflect future developments.

© ASTRON 2025.
All rights are reserved. Reproduction in whole or in part is prohibited without written consent of the copyright owner.

1 / 5

SDC

Doc. Nr.: SDC-038

Rev.: 2.0-1-gla68514

Date: 2025-07-29

- Provide a Linux-based compute cluster.
- Support Singularity/Apptainer based containers on both head and compute nodes.
- Support Python-based daemonic services on head nodes.
- The SLURM workload manager is preferred; use of others is subject to evaluation & test.
- Provide a shared (POSIX) filesystem, accessible from all nodes.
- Provide at least 30 x86-64 compute cores per node.
- Provide at least 8 GB RAM per core.
- Provide at least 200 GB fast local “scratch” storage per core.

What are they made of?

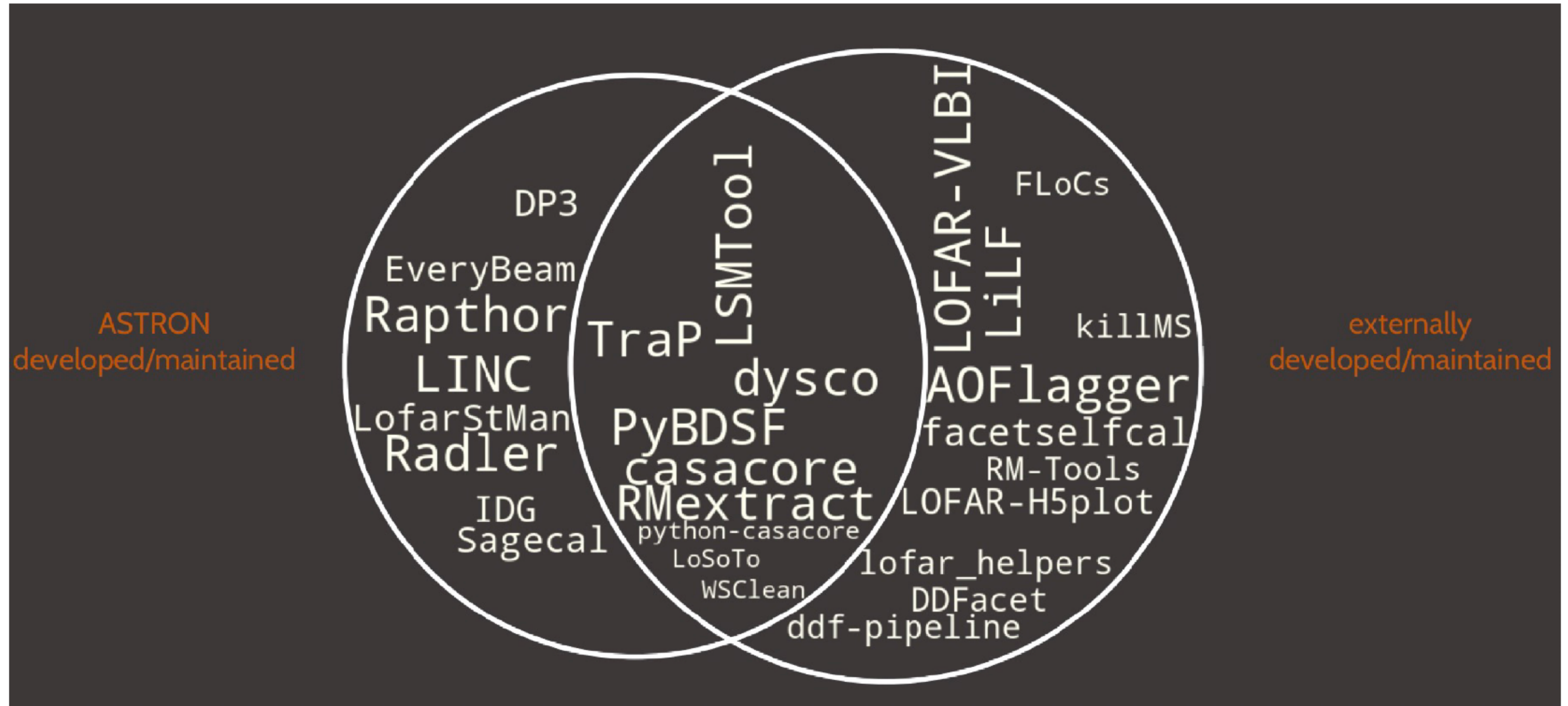


Figure: Sweijen (via Shimwell)

Observatory-Supported Pipelines



Observatory-Supported Pipelines for LOFAR

	Organisatie / Organization	Datum / Date
Auteur(s) / Author(s):		
John D. Swinbank	ASTRON	2025-07-25
Controle / Checked:		
...	...	...
Goedgekeuring / Approval:		
...	...	...
Autorisatie / Authorization:		
...	...	...
Handtekening / Signature:		

© ASTRON 2025.
All rights are reserved. Reproduction in whole or in part is
prohibited without written consent of the copyright owner.

- Technical Requirements
 - Including frameworks, engineering standards, licensing, etc
- Acceptance Process
 - Selecting which pipelines to make available
- Pipeline Lifecycle
 - From conception to retirement
- Applies to *all* pipeline authors — ASTRON does not get “shortcut”.

ASTRON's Pipeline Portfolio



- ASTRON R&D and our partners are investing in our “core pipeline portfolio”. This includes:
 - Preprocessing (flagging, compression, demixing)
 - LINC-HBA (initial calibration)
 - Rapthor-HBA-NL (direction-dependent calibration & imaging)
 - PULPv2 (pulsar timing)
 - TraP (image-plane transient detection & monitoring).
- These pipelines are intended to enable “core” LOFAR science
- In addition, we will directly contribute resources to work with LiLF (LBA) and PILOT (VLBI) pipelines, which are primarily developed outside ASTRON, with the goal of ensuring they are available for general use.
- ASTRON will commit resources to making sure these pipelines meet the technical/lifecycle requirements for Observatory-Supported Pipelines.
 - We believe this will make them the *de facto* choice for most science cases where they are applicable.
 - But we note that the acceptance of pipelines into operations is a LOFAR ERIC process.

The Wider Community

- There are a wide range of science cases that ASTRON developers are not resourced to support directly (lightning... solar physics... many more).
- We will encourage and facilitate community members to step up where possible.
- Note that the start of observations for a large programme will only take place when a credible data processing plan is in place.

Performance & Optimization

- In the LOFAR2.0 era, pipeline performance becomes a real limiting factor on our scientific output.
- Broadly think of that in three areas:
 1. Use of less than optimal algorithms and approaches
 2. Use of less than optimal hardware (i.e. CPUs)
 3. Inefficient I/O & filesystem interactions

Algorithmic Approaches

- This is a matter of active research in the scientific community.
- Probably the most computationally-demanding LOFAR processing task (wide-field calibration and imaging on European baselines) has seen ~an order of magnitude improvement in performance in ~4 years!
- Even after 15+ years of operations, figuring out how to get the best results from an instrument like LOFAR is a research problem.
- *Question for infrastructure providers:*
 - Radio astronomy algorithms are often very memory intensive. How scared should we be about the direction of hardware evolution? How can we best make sure our needs are met?

Accelerator Hardware

- The *current* generation of pipelines are almost all CPU based.
- This will change over the coming years — many algorithms have been or will be ported to run on GPUs.
 - Longer term, folks at ASTRON and elsewhere are exploring quantum, neuromorphic, etc.
- *Questions for infrastructure providers:*
 - How much advance notice do you need about changing hardware targets?
 - Are there other types of accelerator we should be tracking?

Inefficient I/O...

- ...or other less-than-optimal use of hardware.
- Some things already in progress:
 - Latest version of DP3 supports “streaming” imaging.
 - New data formats under development (MSv4) build on current best-practice standards like Xarray rather than legacy CASA formats.
- But this is probably where there's the greatest opportunity for **you** to get involved — how can we tune our pipelines to make best use of your hardware?
 - Where “you” is plural — we need pipelines that are applicable across different physical locations.

Future Architecture

- In the medium/long term, we hope that LOFAR pipelines are as similar-as-possible to SKA pipelines (and ngVLA, ASKAP, WSRT, etc).
 - Easier for us, easier for you, easier for the community.
- Discussion with SKAO/NRAO/SARAO/CSIRO underway to think about architectures.
 - Hanno will speak to this.
- ASTRON-SURF MoU specifically commits us to converge on a “common framework”.
 - Initial thinking is around Dask and Prefect.
 - Again, this is likely something where **you** could help.

Conclusions

- In (or next-to) archive processing is an important part of the LOFAR2.0 strategy (and beyond).
- That raises a bunch of interesting questions...
 - ...some of which we have good answers for
 - ...some of which we'd love your help with.
- Comments / questions / etc welcome!