



# ASTRON

Netherlands Institute for Radio Astronomy



---

# ASTRON data provenance

Yan Grange, Mattia Mancini, Nico Vermaas,  
Vishambhar Nath Pandey, Hanno Holties

**ASTRON**

Netherlands Institute for Radio Astronomy

ESCAPE provenance workshop 2020-09-08



# ASTRON data

- Data from radio interferometers
  - WSRT, LOFAR, APERTIF
  - Therefore in essence comparable to Jive and SKA use cases
- Data archives for current instruments
  - Lofar Long-Term Archive (LTA)
  - Apertif Long-Term Archive (ALTA)
- Working towards end products in VO
  - Surveys from WSRT, LOFAR and APERTIF

**ASTRON**

**The VO @ ASTRON**

Welcome to the ASTRON VO data center.

ESCAPE provenance workshop 2020-09-08

## Search

Q Basic search

Q Advanced search ▾

☰ Browse projects ▾

## Basic search

The Basic Search module allows you to search for data within a specified pointing (coordinates) and to specify whether you want to perform your search on observations and/or pipelines. If you decide to select a project, the search will be confined to only that project. For more advanced search options per data type use the "Advanced search".

### Data product types ⓘ

☒ Observation

☐ Averaging Pipeline

☐ Calibration Pipeline

☐ Imaging Pipeline

☐ Long Baseline Pipeline

☐ Pulsar Pipeline

### Pointing ⓘ

Object  [resolve](#)

## Welcome to ALTA

Apertif Long Term Archive



**ALTA**

Version 1.1.10b (4 sept 2020)

[Please start here](#)

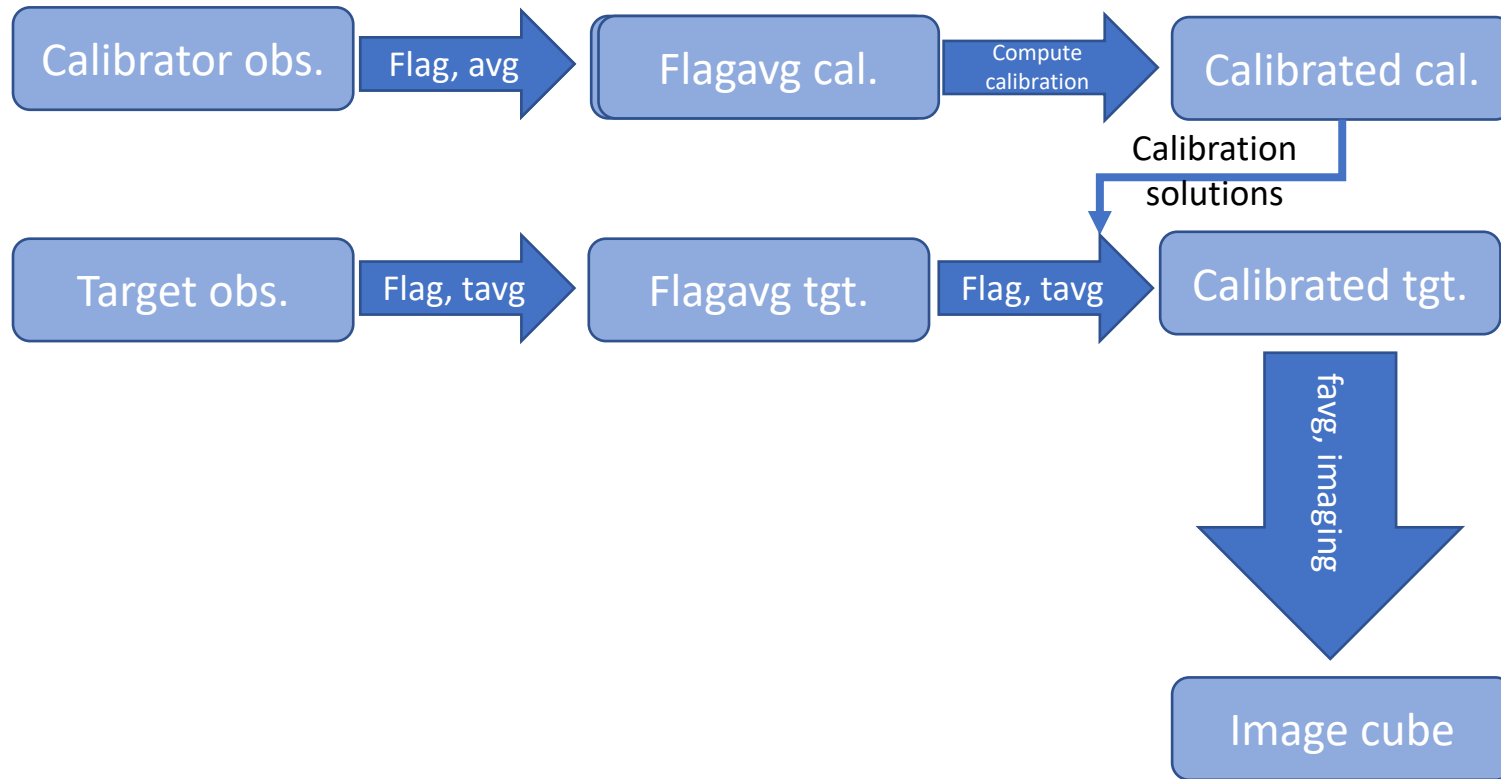
**ASTRON**

Netherlands Institute for Radio Astronomy

# Data properties

- LOFAR: 50 PB archive, with mostly raw visibility and beam formed (time series) data. Data publically accessible, but limited public usage.
  - In part because it is low-level data
  - LOFAR surveys will have data releases with images (level 3 in VO terms)
- APERTIF: Few PB Survey instrument, output are images and timing data (including triggers). Multiple VO-levels of output:
  - 1: raw visibilities and timing data, 2 (Calibrated visibilities), 3 (Images & cubes).
  - Level 3 is guaranteed to be directly accessible, and will be released through VO.
  - Other levels may be kept or deleted in the future.
- WSRT: Several levels of data present. Focus on data accessibility of VO Level 3 (images/cubes) and 4 (catalogues) through VO.

# Example processing flow



# Provenance information (current)

- LOFAR:
  - The files itself (history in the MS)
  - Lofar Long-Term Archive (LTA):
    - Core schema describes data products and processes which are linked as <observation process> - <observation data products> - [<pipeline process> - <pipeline dataproducts>]\*
    - Provenance can be queried through webform, python (AstroWise) code and database clients.
- APERTIF:
  - Files (both the MS and FITS format)
  - Apertif LTA (ALTA):
    - Provenance of archive based on the VO provenance data model (yay!)  
data model on bonus slide if you are interested
    - Observation couple to data products by ID, data products coupled to activity



# Provenance information (current)

- WSRT survey data:
  - Mostly FITS images (again: FITS history)
  - provenance via the observation management system, which does not directly link to the end products



ESCAPE provenance workshop 2020-09-08

# User access to provenance

- In general, it is non-trivial to find targets and calibrators that ‘belong together’
  - In the end, the relation is made when running the calibration pipeline (could be reasons to not use the calibrator originally observed near the target).
  - Surveys teams publish the data themselves; though most will be part of the VO
- Relevant provenance information for our users is:
  - What pipeline executed,
    - what version of each step
    - what settings
  - Observation (linking to telescope settings)
  - Relations between calibrator and target
  - The “who” seems not too relevant for us (also: GDPR)



# User access to provenance

The screenshot displays the ALTA (Apertif Long Term Archive) web interface. At the top, the header reads "ALTA - Apertif Long Term Archive" with a "Logout (vermaas)" link. Below the header is a navigation bar with links for "Home", "Science", and "Help". A secondary navigation bar includes "All Observations", "All Pipelines", "Search Results", and "Ingest Monitor". A search bar contains the text "runid", and a dropdown menu shows "Release: Commissioning2018".

The main content area features a grid of observation cards. On the left, a card for "uncalibratedVisibility" (3003\_B000.MS) is highlighted with a red box around its "Show Provenance chain" button and the "180223003" identifier. To the right, a larger grid shows four cards for the "Commissioning2018" release, each corresponding to a different data product: "180223003\_INSP", "180223003\_IMG", "180223003\_CUBE", and "180223003\_CAL". Each of these cards has a "Data" button and a "Provenance" button, with the "Provenance" buttons highlighted by red boxes. The cards display various astronomical data visualizations, including heatmaps and a 3D cube visualization labeled "CUBE pipeline".

