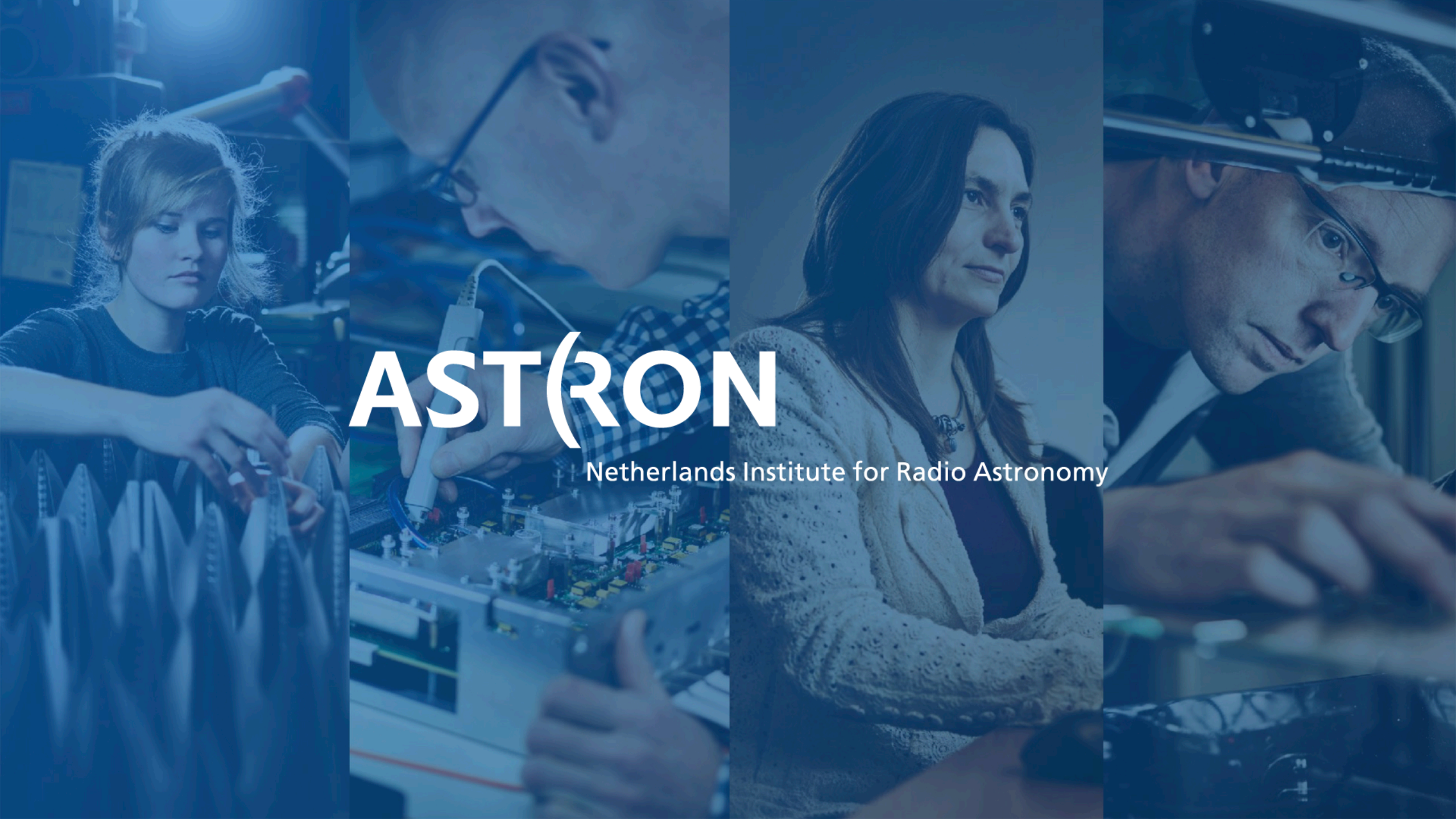

ASTRON
Netherlands Institute for Radio Astronomy

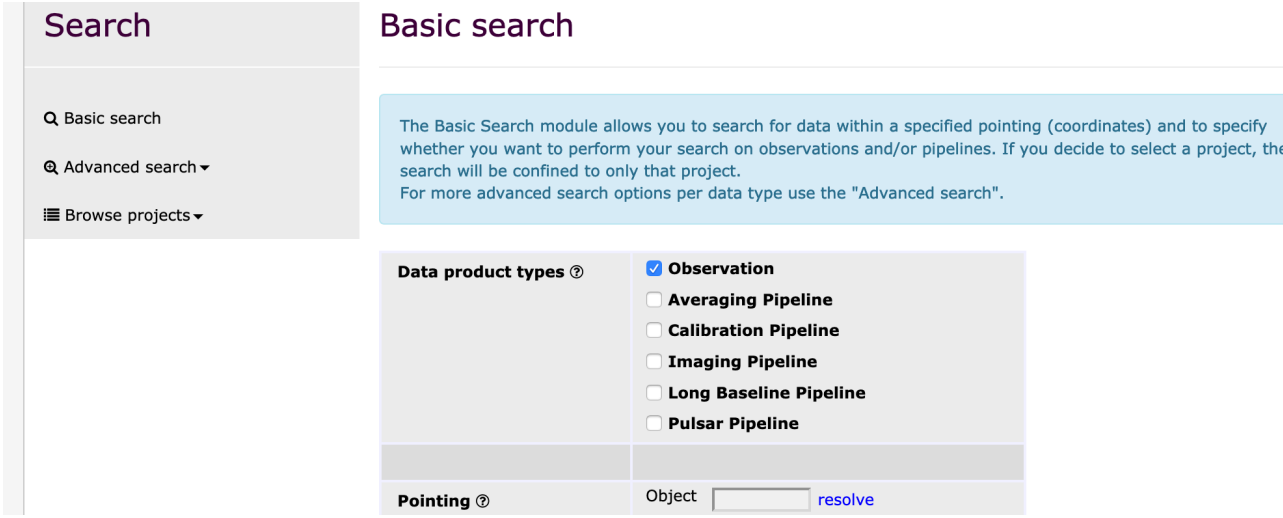
ASTRON data provenance

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

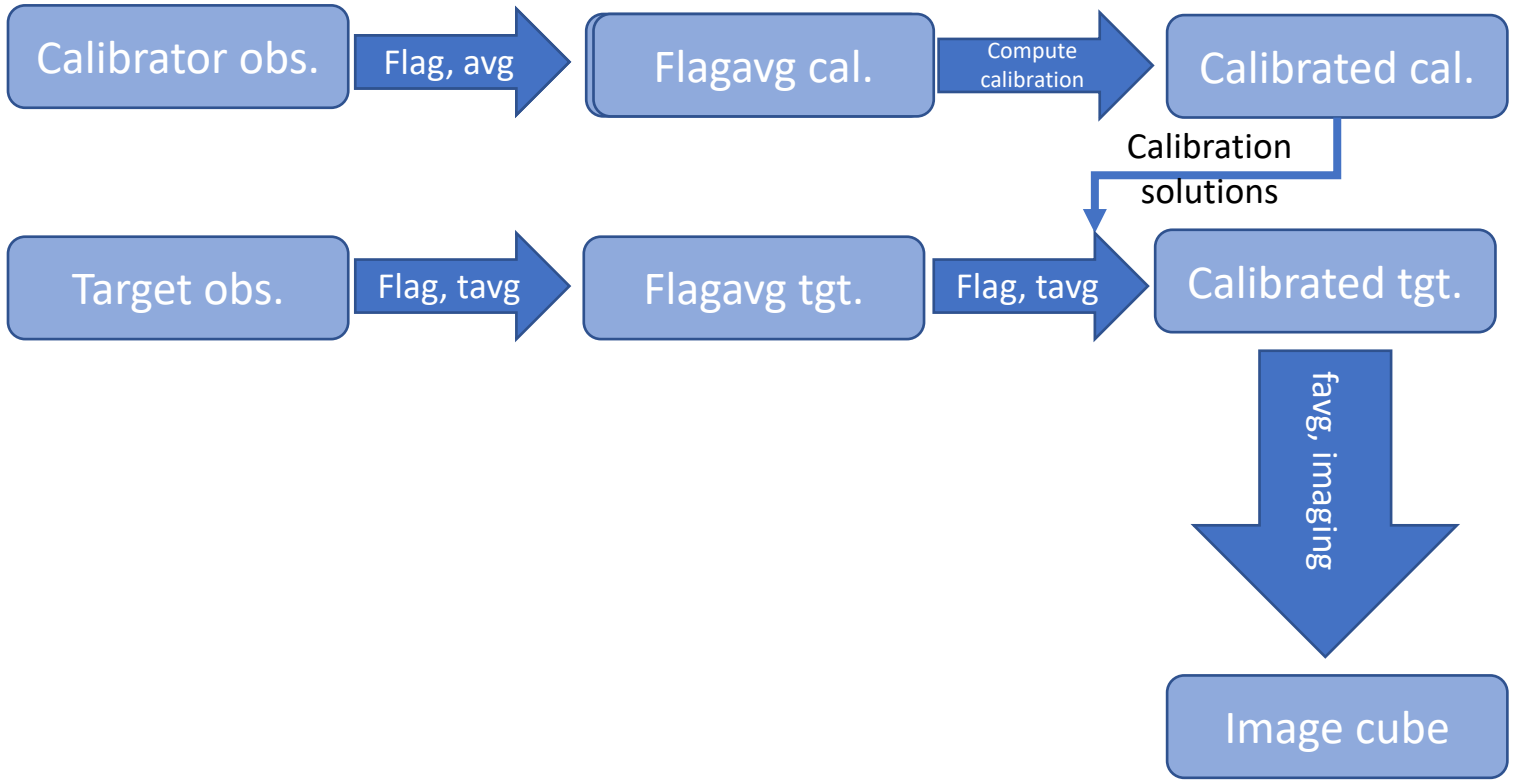
Welcome to the ASTRON VO data center.

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

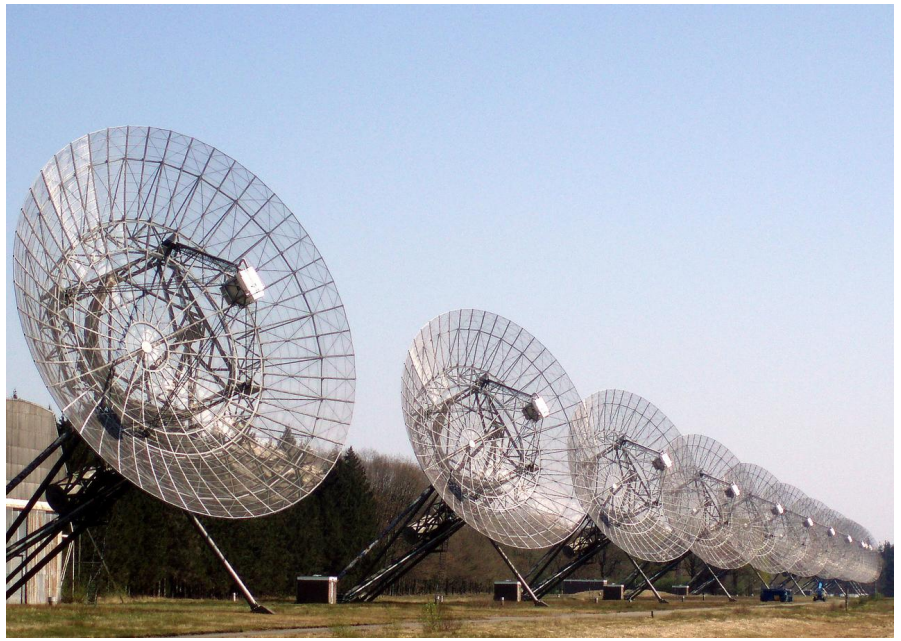


Provenance information (current)

- 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)

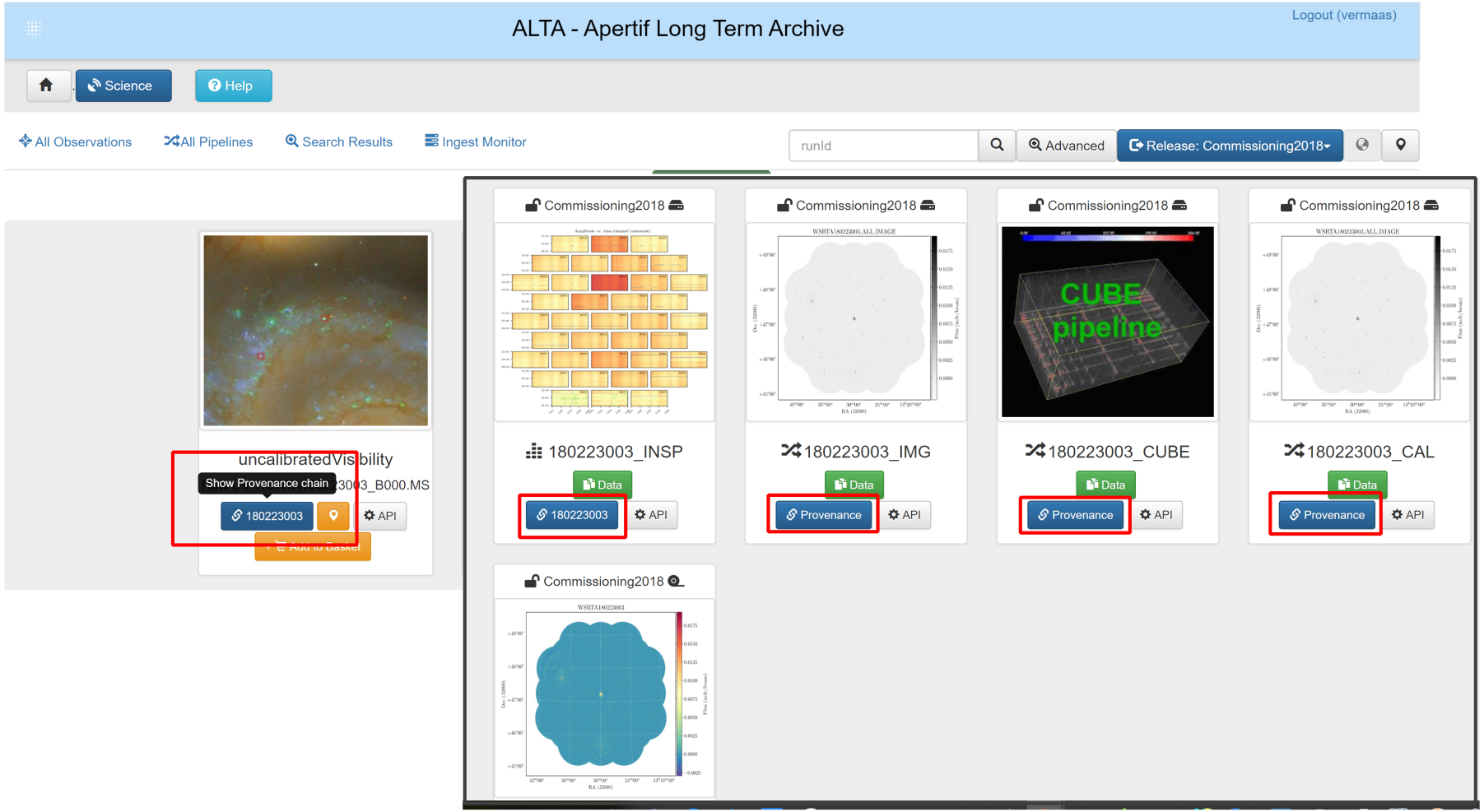
- WSRT survey data:
  - Mostly FITS images (again: FITS history)
  - provenance via the observation management system, which does not direclty link to the end products



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



Models

ALTA-Django-Model-Class-Diagram

latest update: nv 19 nov 2019