LOFAR IF data processing flow

An overview

Marco Iacobelli (ASTRON)

7th LOFAR data processing school
April 16 2024

Outline

Lecture

Imaging mode data editing → info & challenges
Pipeline vs workflow → def & pros/cons
Data editing flow → an overview
Focused view

Demo

QA inspection

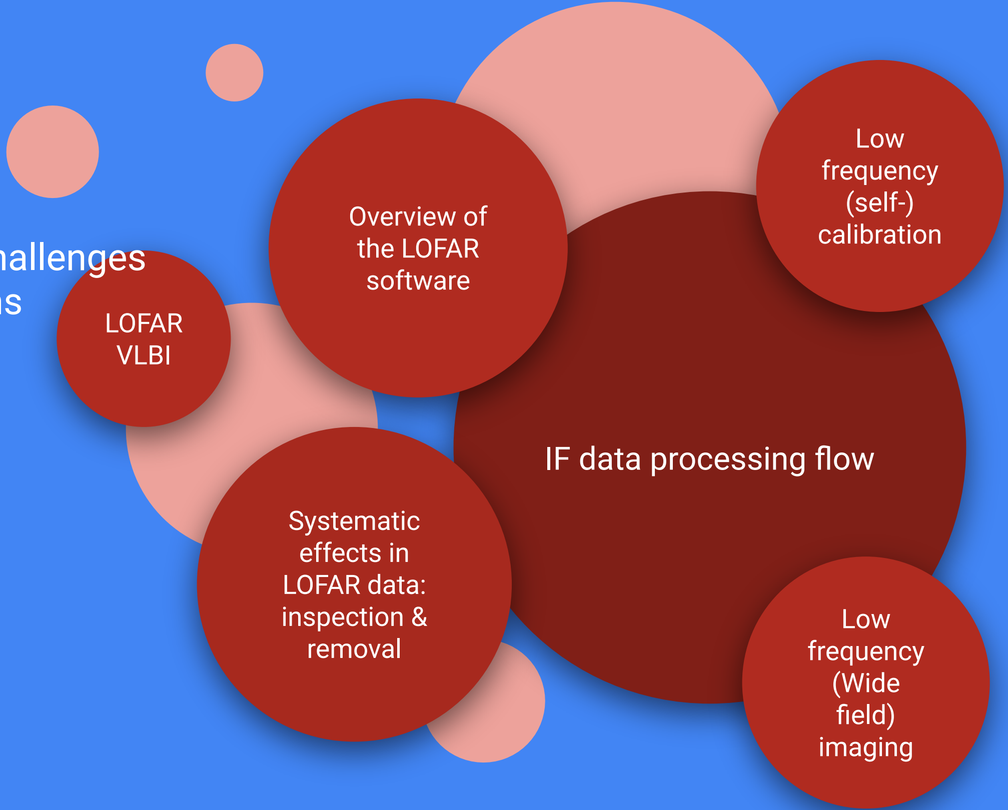
Outline

Lecture

Imaging mode data editing → info & challenges
Pipeline vs workflow → def & pros/cons
Data editing flow → an overview
Focused view

Demo

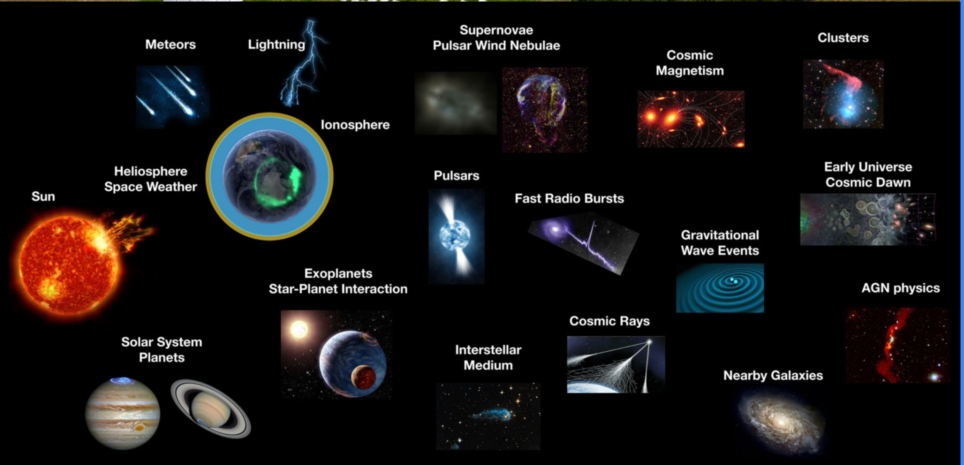
QA inspection



A next-gen facility

LOFAR is the most flexible, complex and data-intensive radio telescope currently in existence

Generates huge amounts of data which requires a lot of supercomputing power and innovative data solutions



A next-gen facility

LOFAR is the most flexible, complex and data-intensive radio telescope currently in existence

Generates huge amounts of data which requires a lot of supercomputing power and innovative data solutions

Enables astronomers to investigate a large spectrum of science use cases (from cosmology to solar-physics)

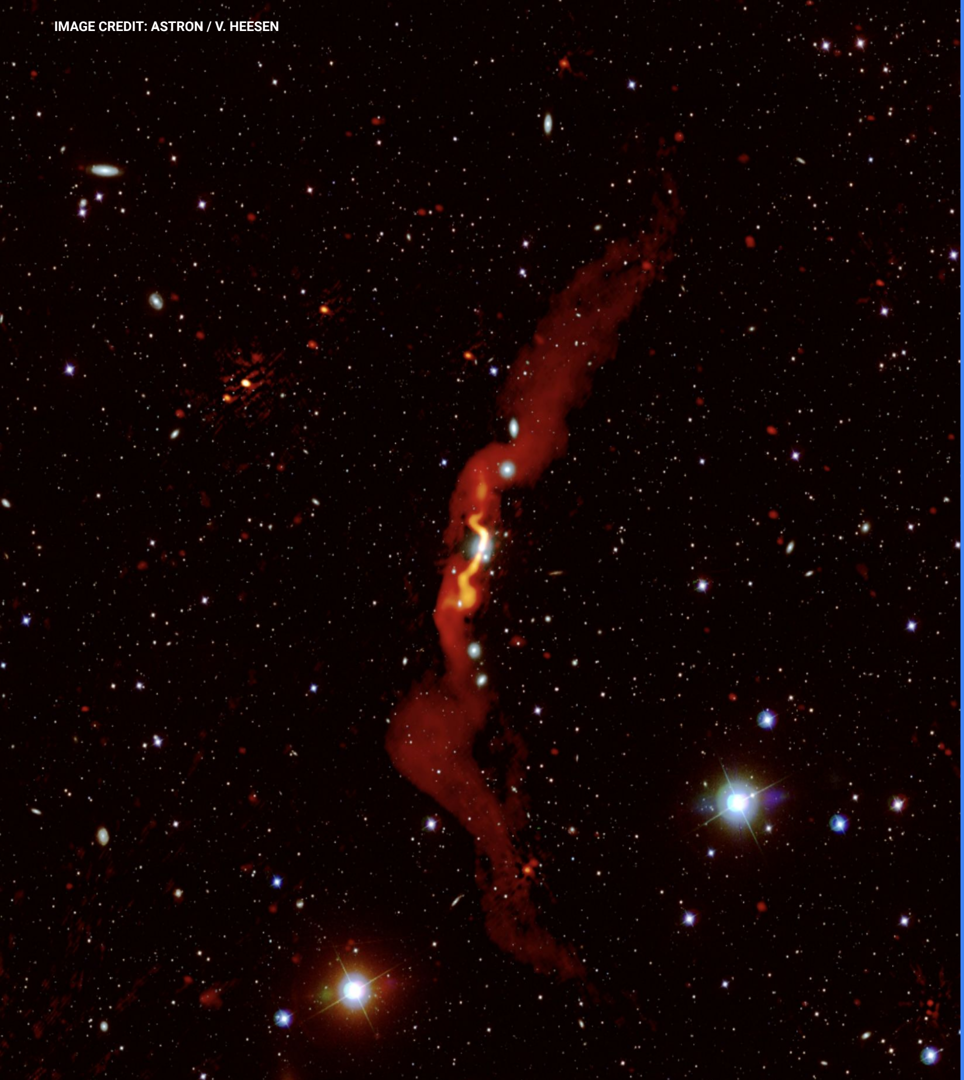
Imaging mode data editing

(Some of the) challenges in LOFAR calibration & imaging

- Large data volumes [CAL | IMG]
- LOFAR beam(s) time dependent & difficult to model [CAL | IMG]
- Low S/N regime → calibration errors [CAL]
- Large fractional bandwidth
    - Requires multi-frequency approaches [CAL | IMG]
- Large FOV
    - Direction-dependent calibration approaches needed [CAL]
    - Large w-values [IMG]
    - Deconvolution complex [IMG]

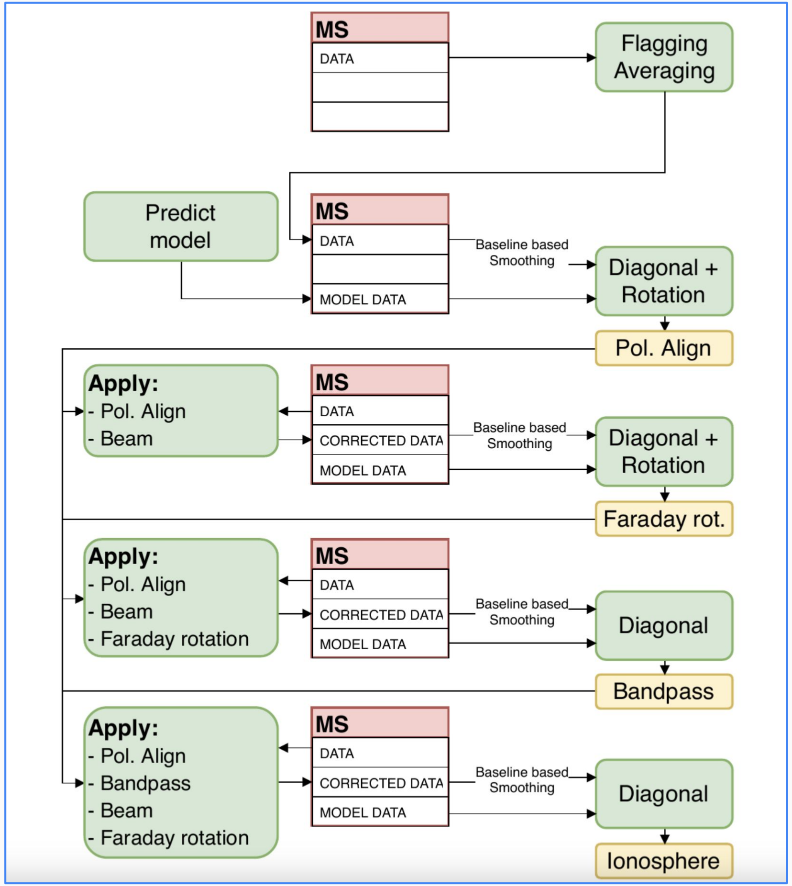
Operation	Lecturer(s)
CAL	M. Mevius & C. Groeneveld R. van Weeren R. Timmerman
IMG	A. Offringa

IMAGE CREDIT: ASTRON / V. HEESEN



Pipeline vs workflow

Data processing → process to generate from raw data science ready data products

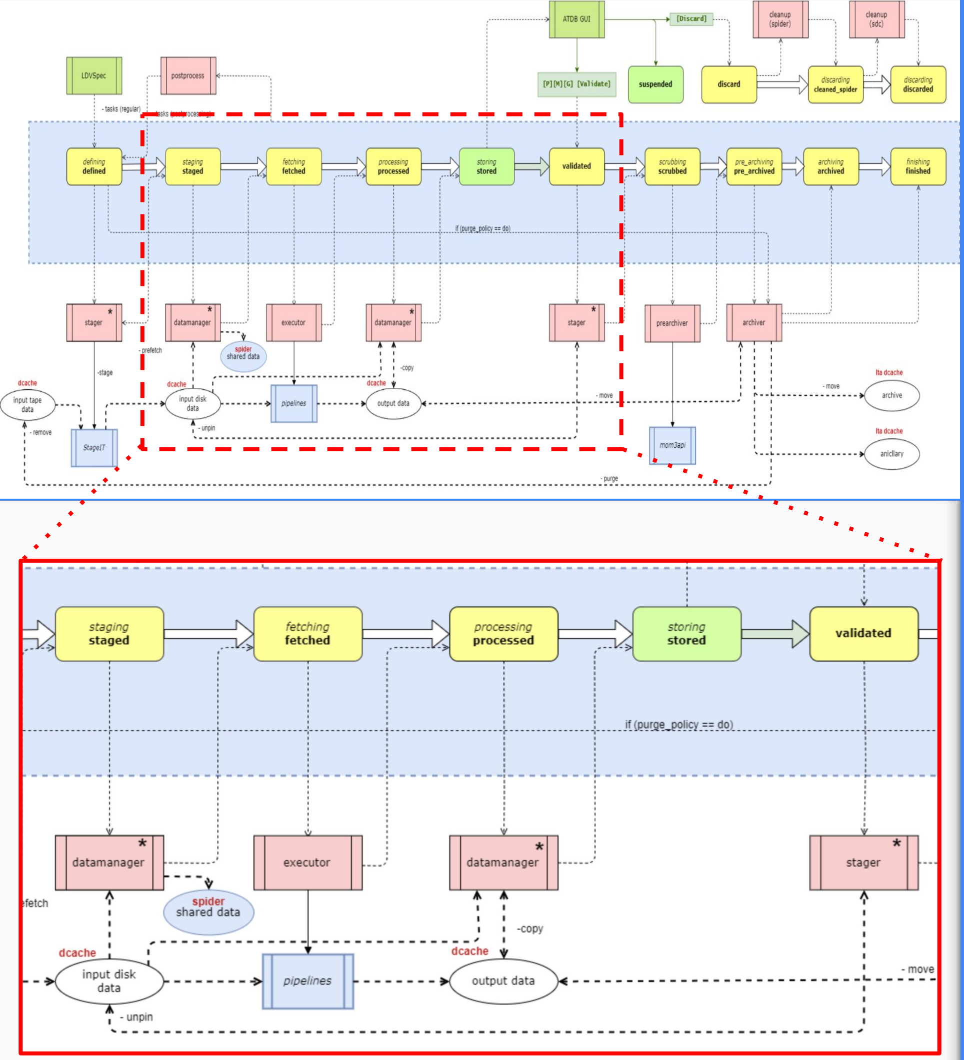


Pipeline vs workflow

Data processing → process to generate from raw data science ready data products

Pipeline → a series of processes/steps to filter or transform data

- stand alone tool



Pipeline vs workflow

Data processing → process to generate from raw data science ready data products

Pipeline → a series of processes/steps to filter or transform data

- user stand alone tool

Workflow → pipeline(s) integrated in a data processing framework including move data from a source, to a destination, based on quality validation

- supported user service

Pipeline vs workflow

LOFAR pipelines designed to perform an incremental data editing

- complexity (self-) calibration strategies → demanding hardware requirements
- data complexity and sizes O(TB) → demanding storage requirements O(10TB)

Pipeline vs workflow

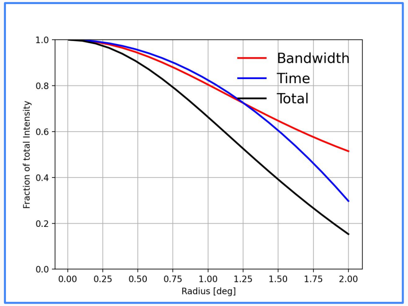
LOFAR pipelines designed to perform an incremental data editing

- complexity (self-) calibration strategies → demanding hardware requirements
- data complexity and sizes O(TB) → demanding storage requirements O(10TB)

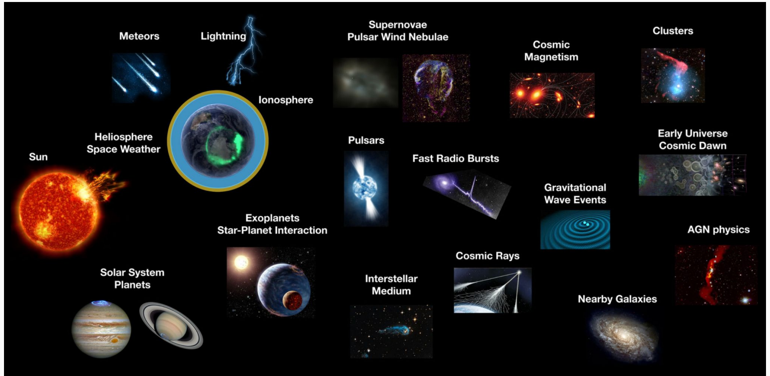
Need for multiple QA checkpoints

Q

Processing strategies vs science use case



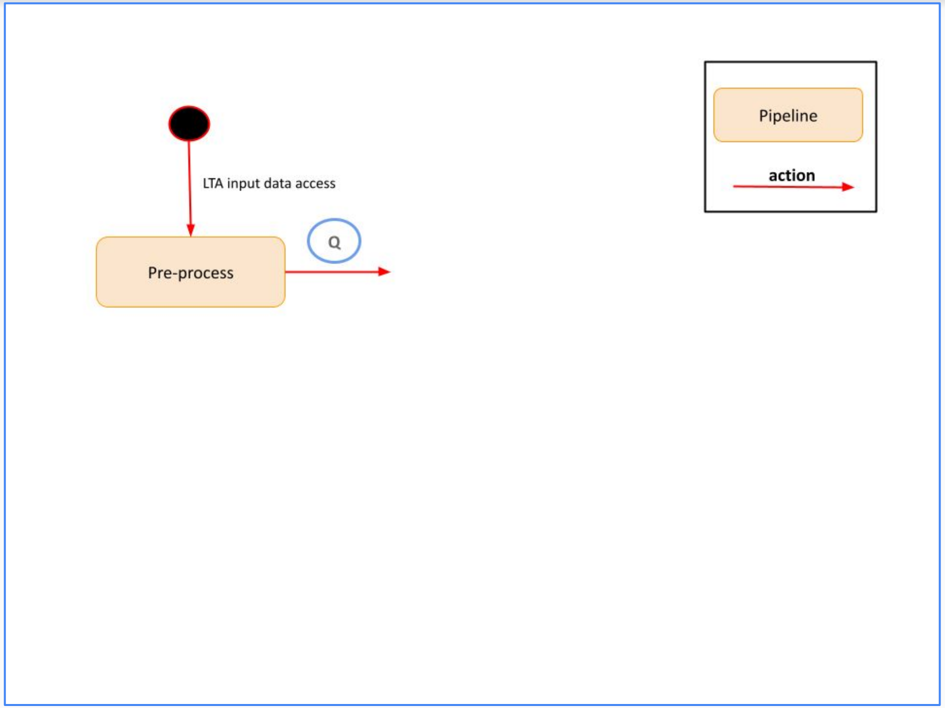
Choices at the start of the data editing constrain generation of final output → e.g. smearing



Pipeline vs workflow

Name	Run as	CPU [hrs] | MEM budget	Life cycle phase	Notes
Pre-process	workflow	$O(10^2)$ | >32 GB	Mature and supported	New version under devs
LINC HBA LINC LBA	pipeline & workflow pipeline & workflow	$O(10^3)$ | >32 GB	Mature, released & supported Devs ongoing close to 1st release	https://git.astron.nl/RD/LINC
DDF-pipeline	pipeline	$O(10^4)$ | >256 GB	Mature, released	https://github.com/mhardcastle/ddf-pipeline
LiLF	pipeline	$O(10^4)$ | >192 GB	Mature, released	https://github.com/revoltek/LiLF
RAPTHOR HBA	pipeline & workflow	$O(10^4)$ | >192 GB	Devs ongoing, released & supported	Support use of facets / screens https://git.astron.nl/RD/rapthor/-/tree/master
LOFAR-VLBI HBA	pipeline (→ workflow)	$O(10^4\text{-}10^5)$ | >192 GB	Devs ongoing, released	Targeted / wide field imaging https://github.com/LOFAR-VLBI/lofar-vlbi-pipeline

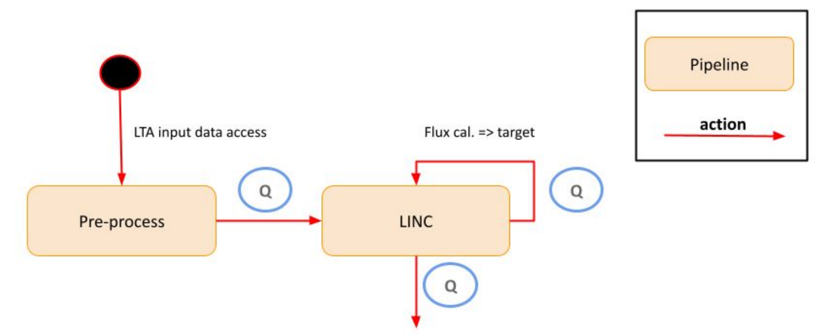
Data editing flow: pre-process



Pre-process in LOFAR data is needed to:

- decrease data size via
    - freq./time averaging
    - visibilities compression
- removal of interfering signals
    - bright off-axis sources
    - radio-frequency interference (RFI)
- Main software packages used:
    - Default PreProcessing Pipeline (DPPP) to average, demix, (RFI) flag, compress

Data editing flow: DIE cal (/img)

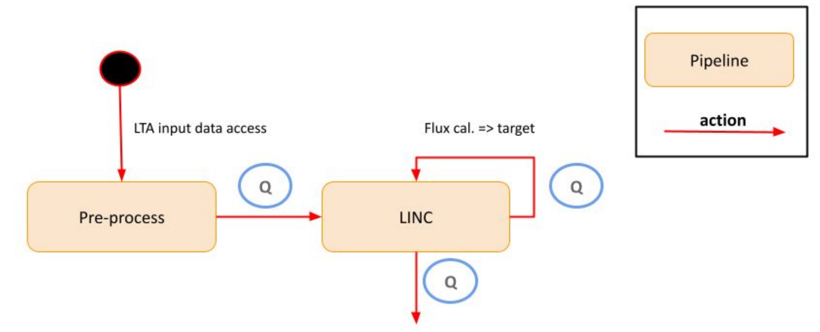


Direction-Independent Effects in LOFAR data are primarily caused by:

- The ionosphere → mostly phase effects (vary quickly in time)
  - Faraday rotation
- The instrumental effects → amplitude effects (vary slowly in time)
  - Polarisation alignment
  - Element beam
  - Bandpass
  - Clock drift

Direction-Independent calibration (mainly) attempts to correct for these effects

Focused view: LINC

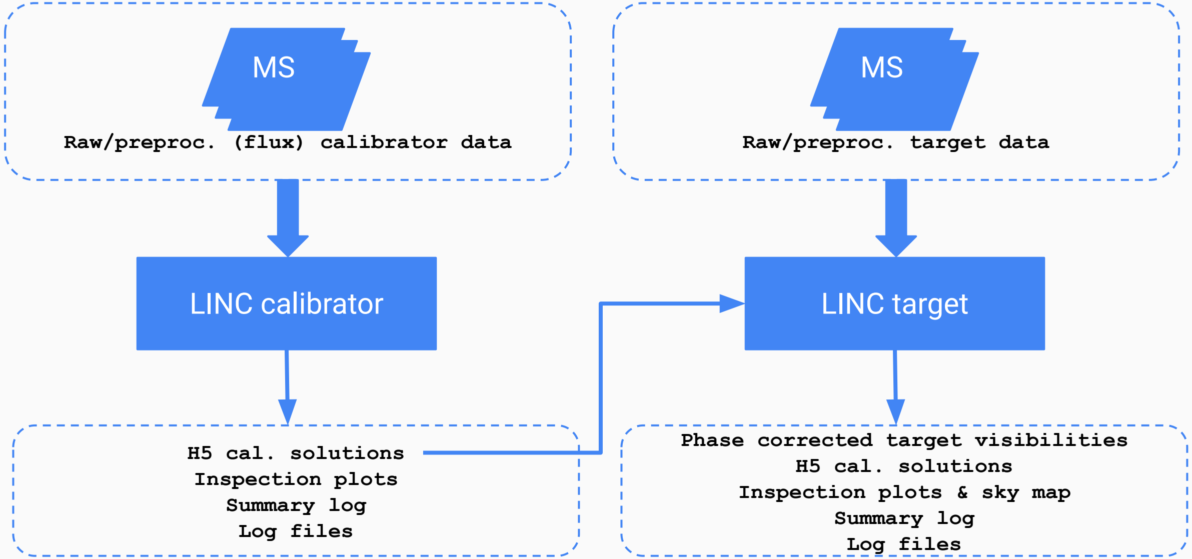


[LINC](#) perform Direction-Independent cal. (& img.)

- Based on the unified-calibration scheme of de Gasperin et al. (2019)
- Supports multi-epoch datasets (interleaved or multiple nights)
- Prepare data to use any DDE calibration software (e.g RAPTHOR, DDF-pipeline, LOFAR-VLBI)
- Main software packages used:
  - Default PreProcessing Pipeline (DPPP) to average, flag, calibrate, apply calib. solutions
  - LOFAR Solution tools (LoSoTo) to analyse/extract parameters from calib. solutions

Focused view: LINC

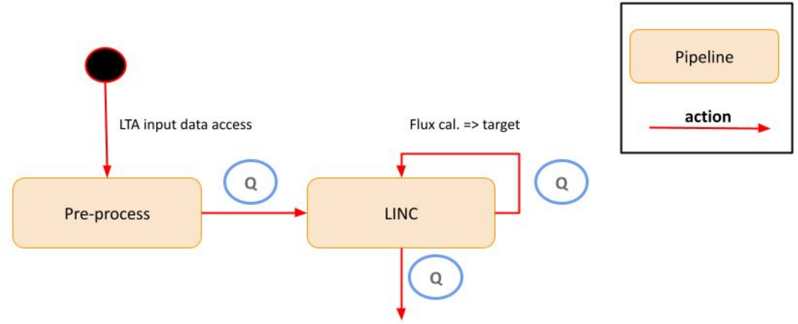
LINC perform Direction-Independent cal. (& img.)



Pipeline consists of 2 workflows:

- LINC calibrator: processes the (flux) calibrator to derive DIE corrections.
- LINC target: Transfers the DIE corrections to the target; does phase self-calibration of the target.
- DIE calibrated target visibilities to be used for further DDE processing.

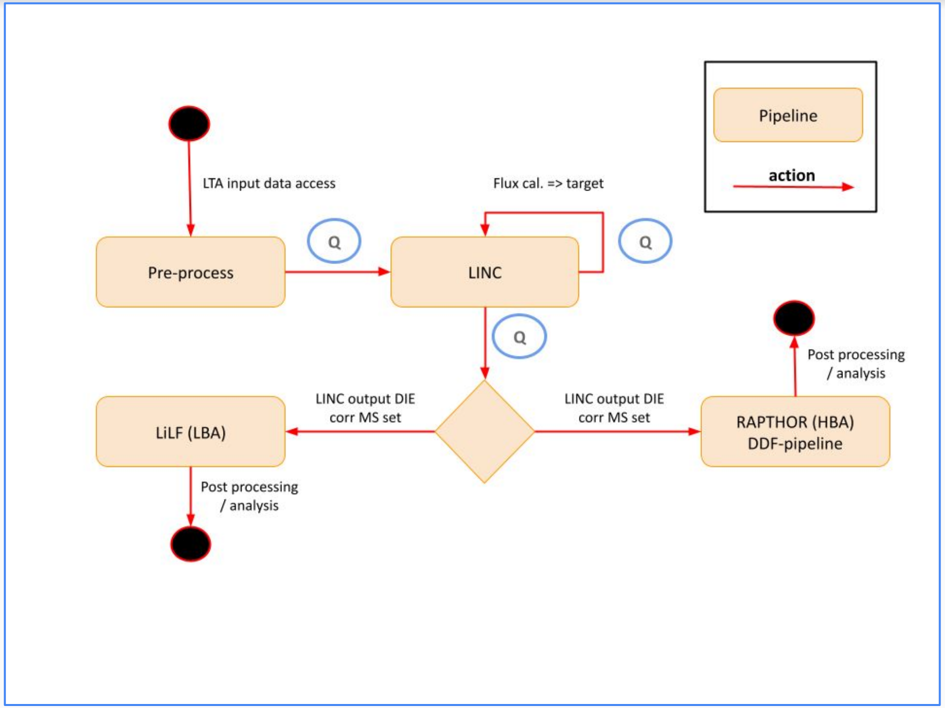
Focused view: LINC



LINC perform Direction-Independent cal. (& img.)

- LINC uses CWL pipeline framework as backend:
  - Allows distribution over cluster nodes
  - Allows resuming of interrupted jobs
  - Integrated Docker support

Data editing flow: DDE cal/img

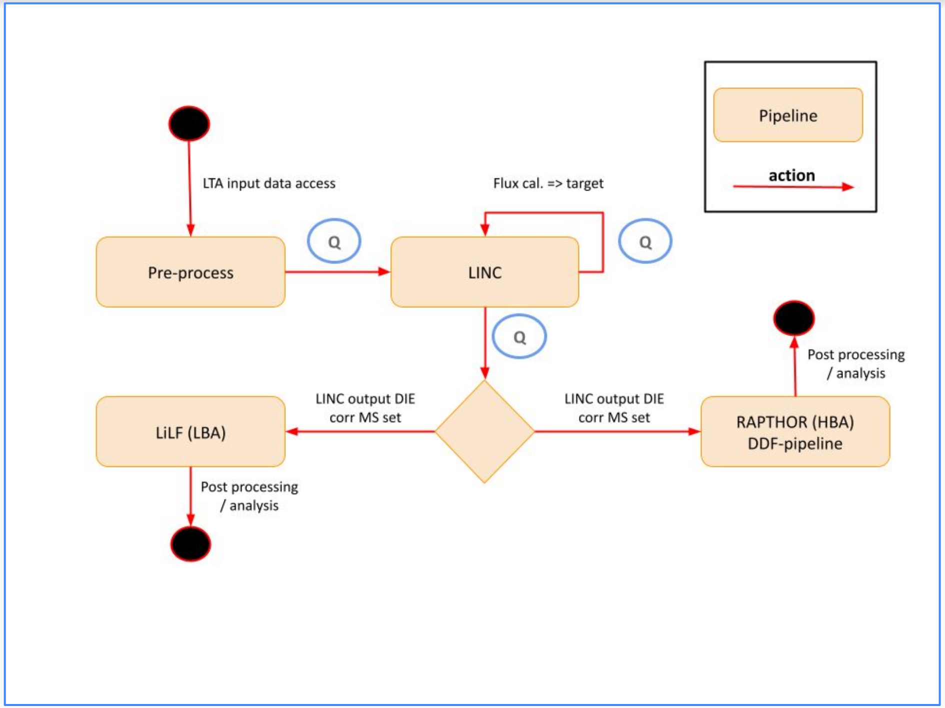


Direction-Dependent Effects in LOFAR data are primarily caused by:

- The ionosphere → mostly phase effects (vary quickly in time)
  - Dispersive delays
- The LOFAR beam → mostly amplitude effects (vary slowly in time)
  - Dipole beam & array factor

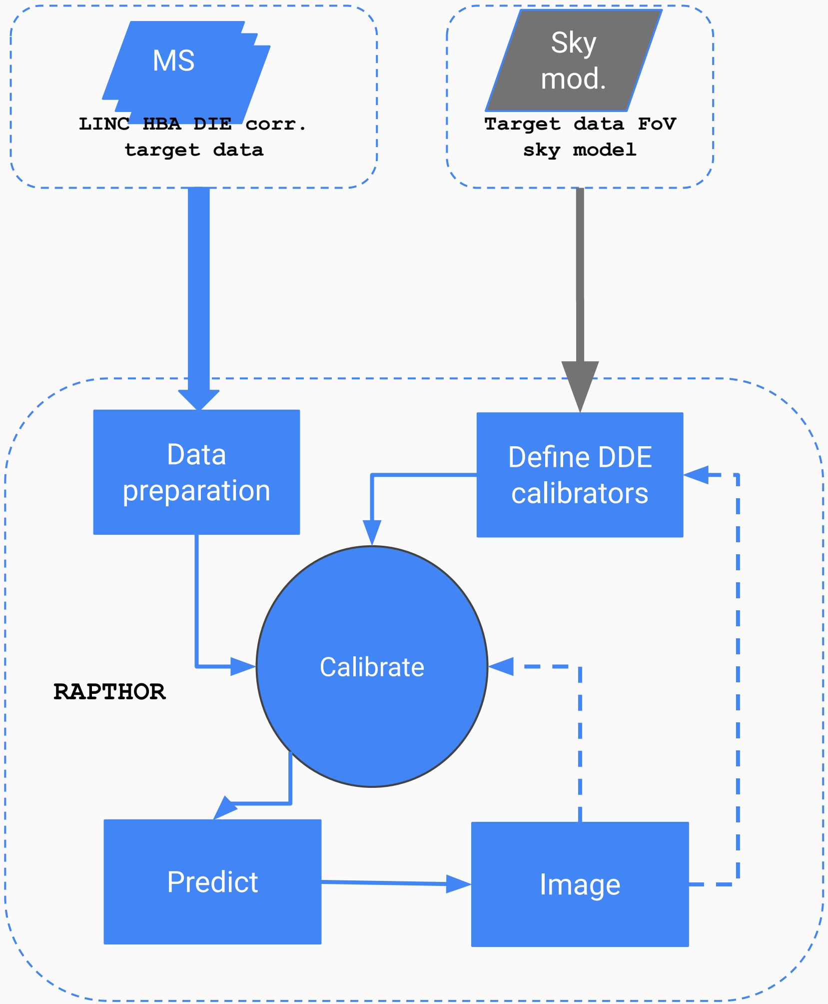
Direction-Dependent calibration (mainly) attempts to correct for these effects

Focused view: RAPTHOR



RAPTHOR performs CS&RS array Direction-(In)Dependent calib. & img.

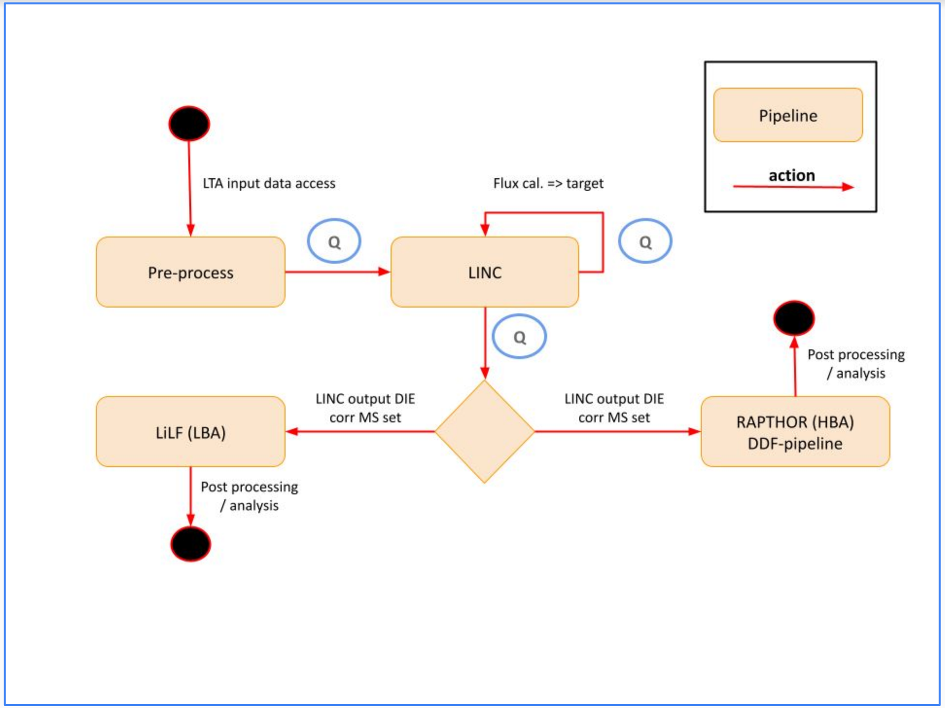
- Based on the calibration scheme of de Gasperin et al. (2020)
- Supports multi-epoch datasets (interleaved or multiple nights)
- Designed to operate on HBA and LBA data
- Enable full-field of view / targeted processing
- Sub-optimal for:
  - very extended (e.g. >=1 deg) target sources (i.e. >1 facet needed)



Focused view: RAPTHOR

- Get list of DDE calibrators from sky model
    - Divide field into facets
- Iterative self-calibration:
    - Performed in multiple directions (facets) <u>simultaneously</u>
    - Each facet gets a single calibration solution
    - Designed to enable usage of 2-D screens for both CAL & IMG (in progress)
    - Get a sky model to (eventually) loop

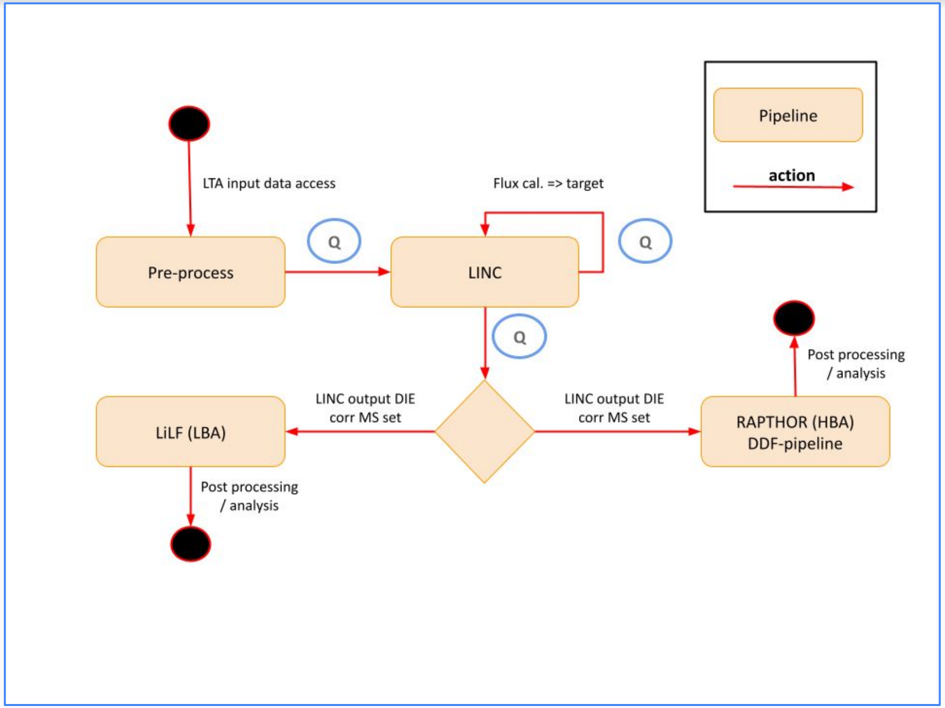
Focused view: RAPTHOR



RAPTHOR performs CS&RS array Direction-(In)Dependent calib. & img.

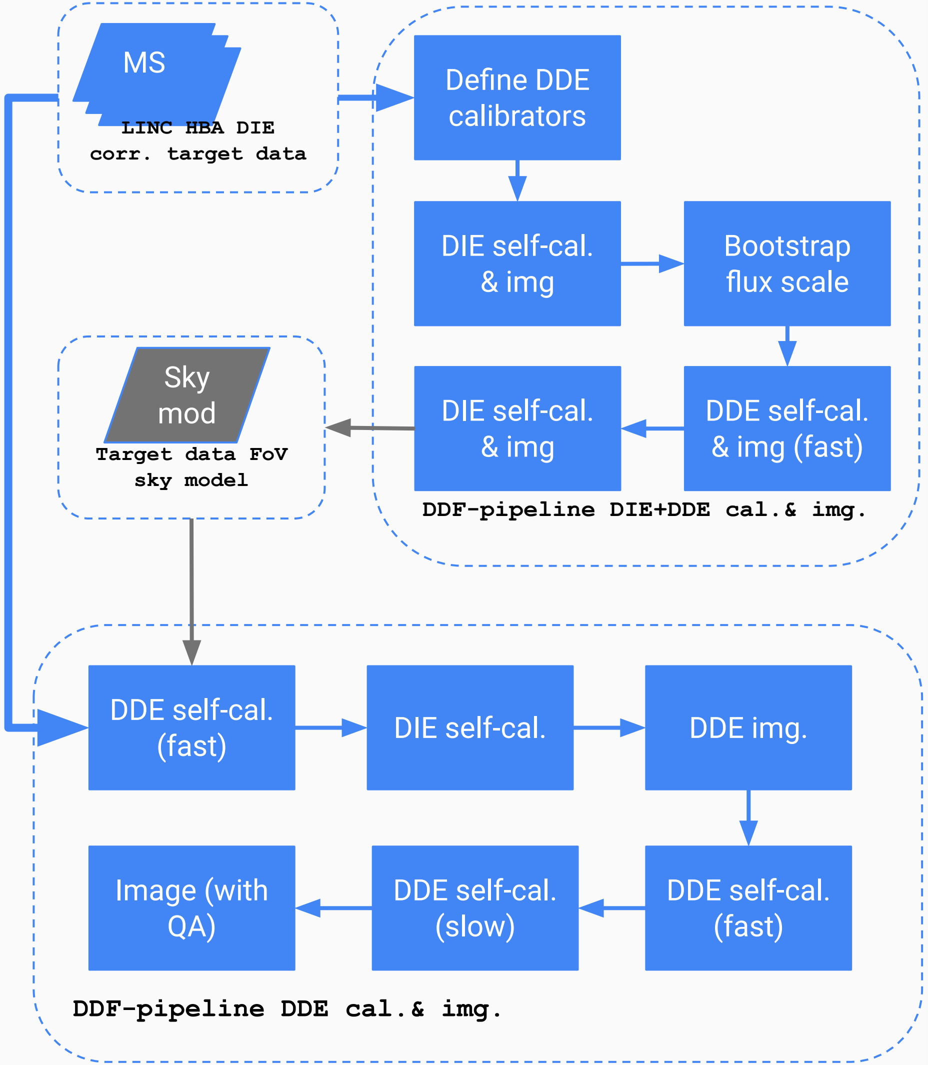
- Uses a Python wrapper around CWL pipelines as backend:
  - The wrapper sets up & executes the pipelines as "operations" to perform the actual processing
  - Allows distribution over cluster nodes
  - Allows resuming of interrupted jobs
- Main software packages used:
  - Default PreProcessing Pipeline (DPPP) to average, flag, calibrate, apply calib. sols
  - LOFAR Solution tools (LoSoTo) to analyse/extract parameters from calib. sols
  - WSClean to apply calib. sols & perform imaging

Focused view: DDF-pipeline



DDF-pipeline performs CS&RS array Direction-(In)Dependent calib. & img.

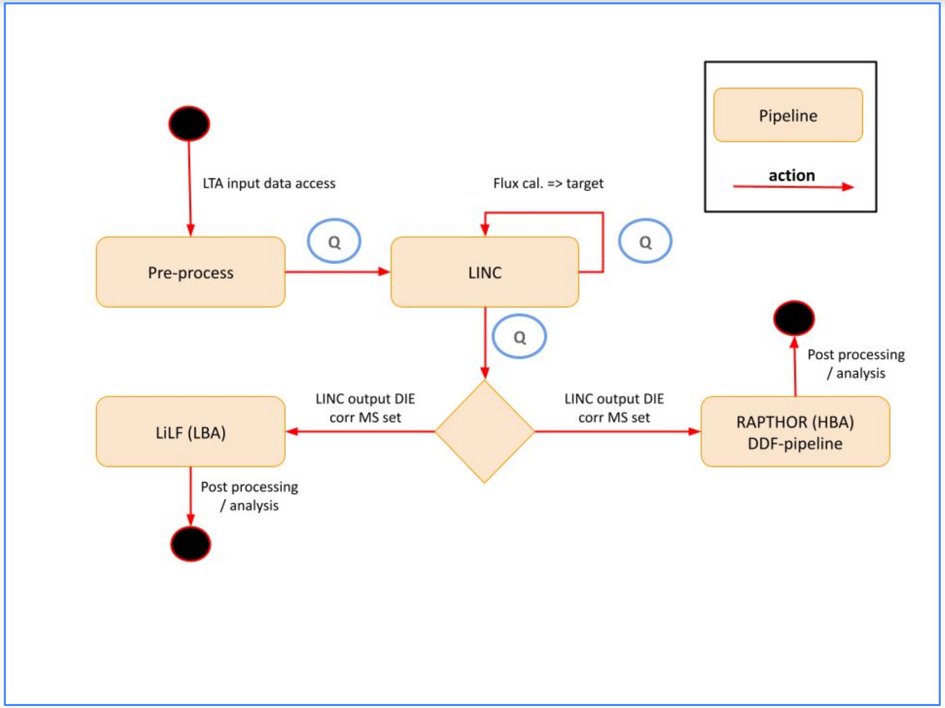
- Based on the calibration scheme of Tasse et al. (2021)
- Supports multi-epoch datasets (interleaved or multiple nights)
- Designed to operate on HBA data
- Recommended for full-field of view processing



Focused view: DDF-pipeline

- Sub-dataset self-cal. & imaging:
  - Get DDE calibrators & facets
  - DIE self-cal. & imaging
  - DDE self-cal. (phase only) & img.
  - DIE self-cal. & imaging
  - Improved DIE solutions & sky model
- Full dataset self-cal. & imaging:
  - DDE+DIE self-cal.
  - DDE imaging
  - Fast DDE self-cal.
  - Slow DDE self-cal.
  - Full field imaging (& quality checks)

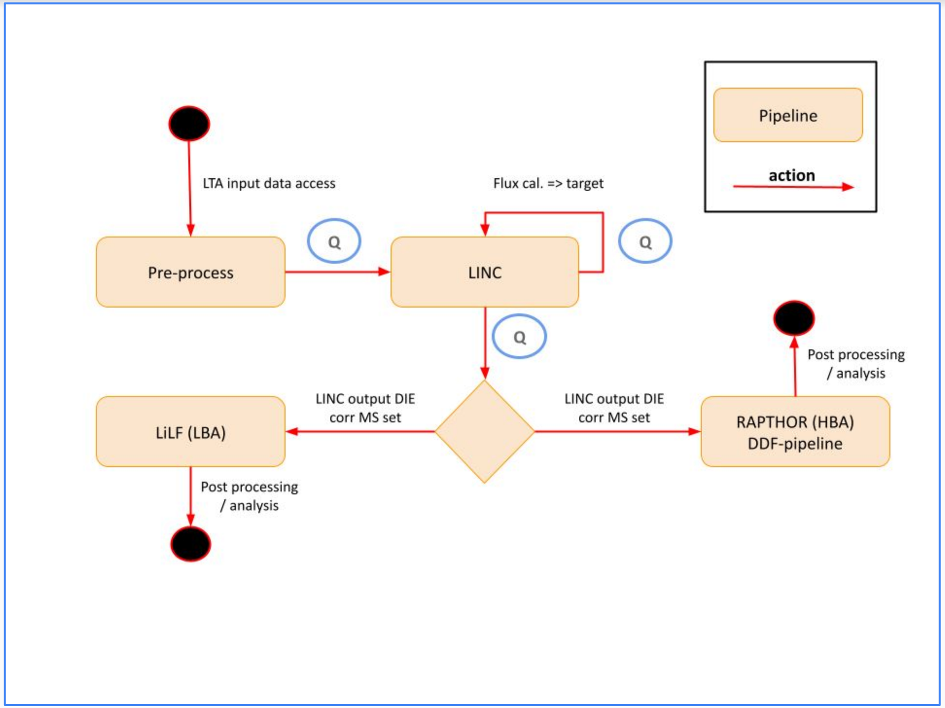
Focused view: DDF-pipeline



DDF-pipeline performs CS&RS array Direction-(In)Dependent calib. & img.

- Uses a Python for both backend and frontend:
  - Allows distribution over cluster nodes
  - Allows resuming of interrupted jobs
- Large resources required: 32 cores, 192 GB memory, 10TB of disk space
- Man software packages used:
  - DDFacet
  - KillMS

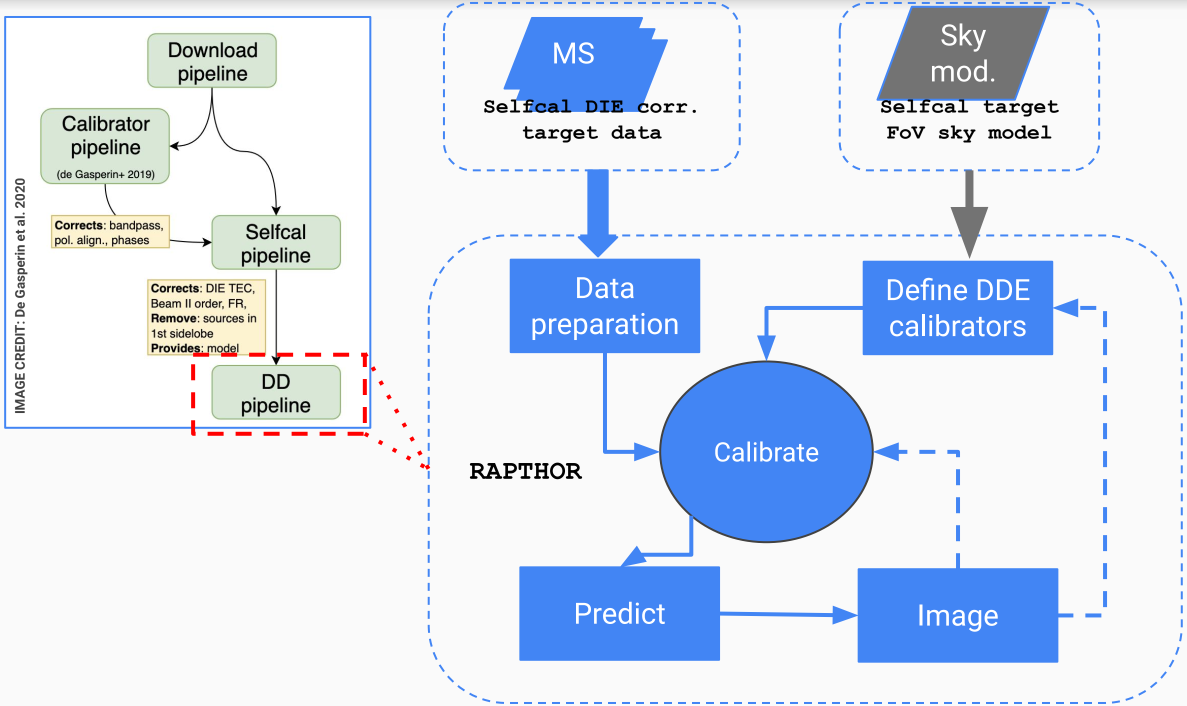
Focused view: LILF



LiLF performs CS&RS array Direction-(In)Dependent calib. & img.

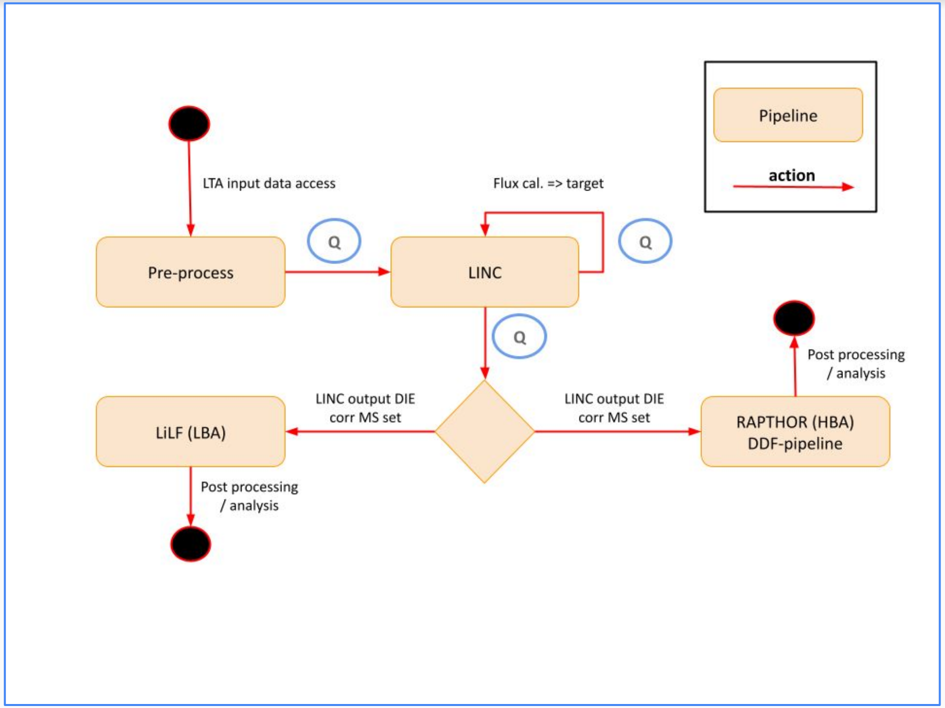
- Based on the calibration scheme of de Gasperin et al. (2020)
- Supports multi-epoch datasets (interleaved or multiple nights)
- Designed to operate on LBA data
- Enable full-field of view / targeted processing
- Sub-optimal for:
  - LBA data <30 MHz

Focused view: LILF



- Get list of DDE calibrators from sky model
  - Divide field into facets
- Iterative self-calibration:
  - Performed in multiple directions (facets) simultaneously
  - Each facet gets a single calibration solution
  - Get a sky model to (eventually) loop

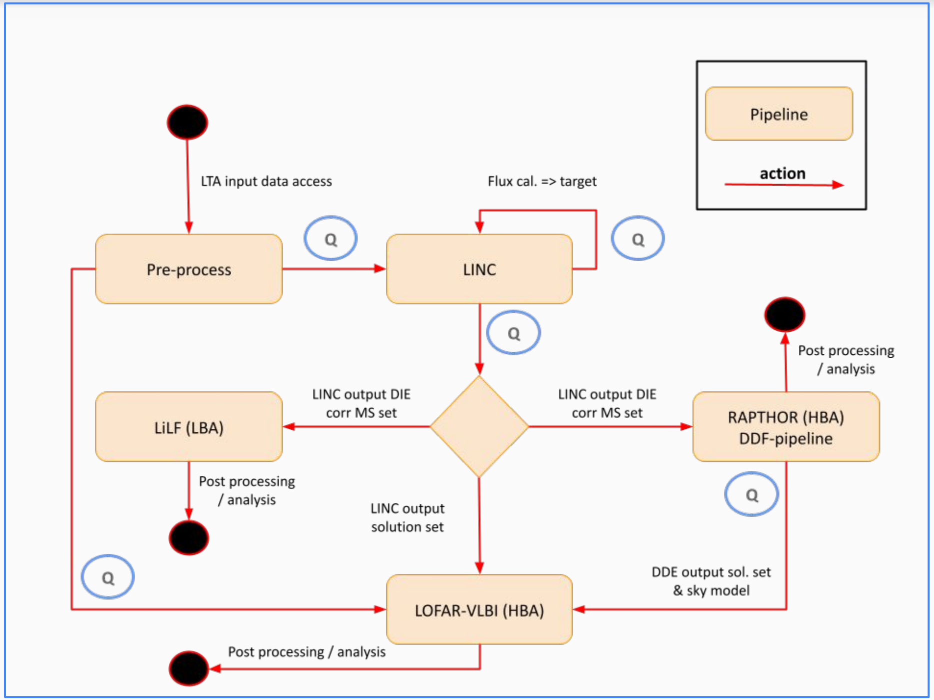
Focused view: LILF



LiLF performs CS&RS array Direction-(In)Dependent calib. & img.

- Uses a Python for both backend and frontend:
  - Allows distribution over cluster nodes
  - Allows resuming of interrupted jobs
- Main software packages used:
  - Default PreProcessing Pipeline (DPPP) to average, flag, calibrate, apply calib. sols
  - LOFAR Solution tools (LoSoTo) to analyse/extract parameters from calib. sols
  - WSClean to apply calib. sols & perform

Data editing flow: DDE cal/img (VLBI)

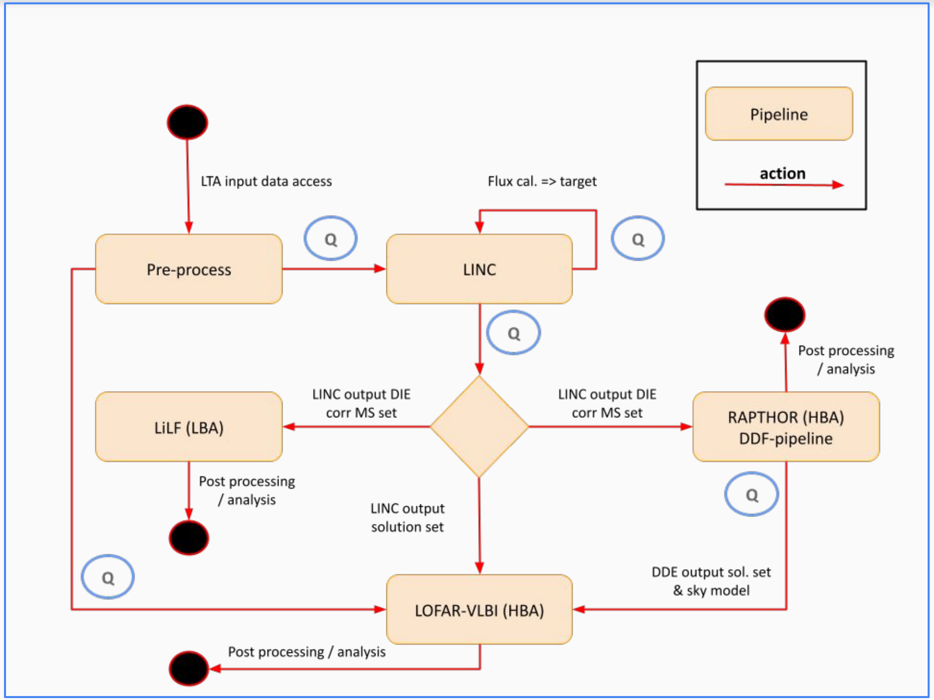


Direction-Dependent Effects in LOFAR data are primarily caused by:

- The ionosphere → mostly phase effects (vary quickly in time)
  - Dispersive delays
- The LOFAR beam → mostly amplitude effects (vary slowly in time)
  - Dipole beam & array factor

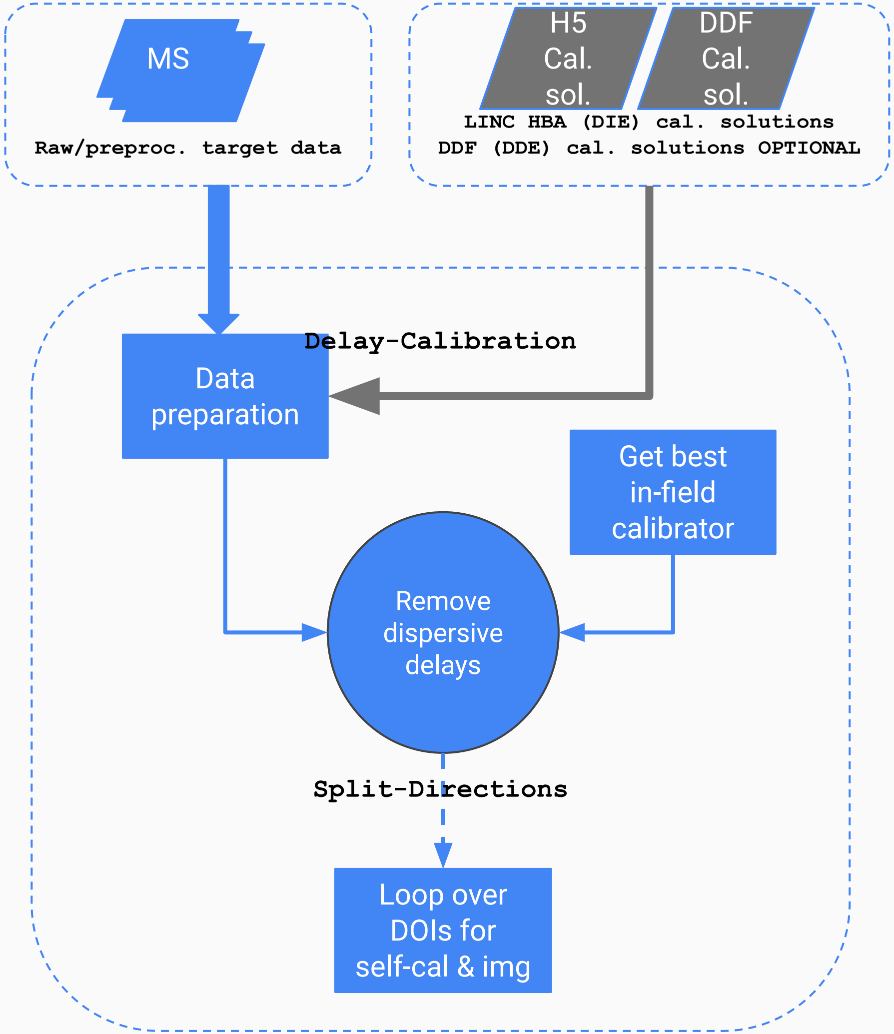
Direction-Dependent calibration (mainly) attempts to correct for these effects

Focused view: LOFAR-VLBI



LOFAR-VLBI performs full array Direction-(In)Dependent calib. & img.

- Based on the VLBI principles & LOFAR adapted scheme of Morabito et al. (2021) → see also Sweijen et al. 2022 and Ye et al. 2024
- Supports multi-epoch datasets (interleaved or multiple nights)
- Designed to operate on HBA data
- Recommended for targeted exposures
  - full-field of view intensive processing
- Sub-optimal for:
  - faint (i.e. S/N<10) target sources away (i.e. >1.5deg) from the delay-calibrator



Focused view: LOFAR-VLBI

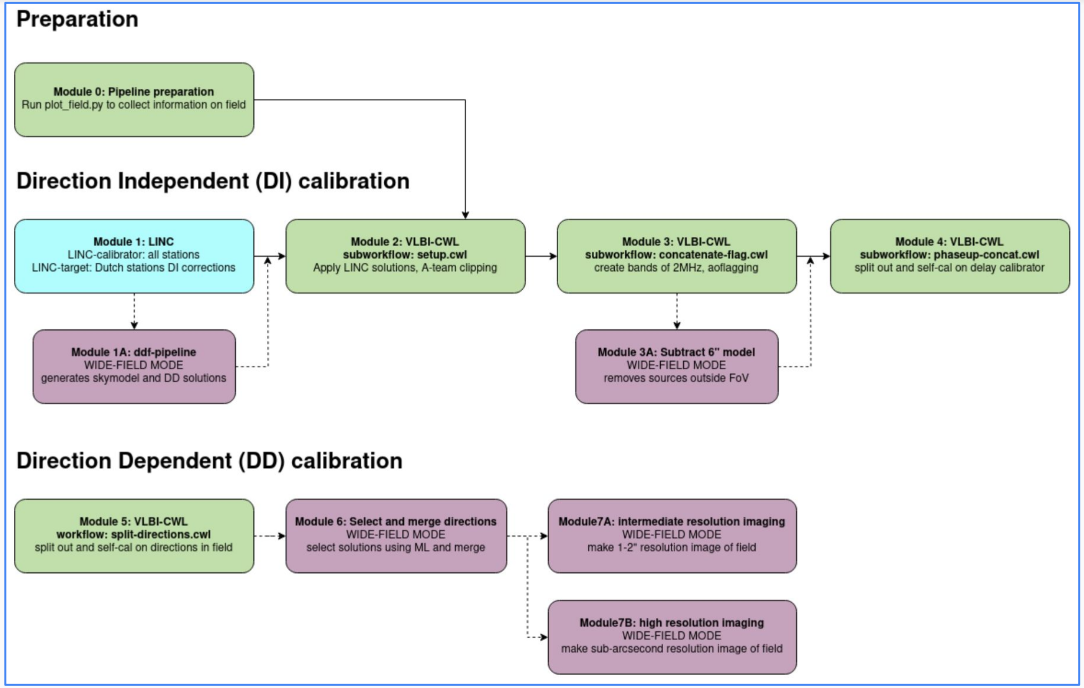
- apply solutions
- build a list of potential in-field calibrators
- for the best guessed in-field calibrator:
  - phase-shift to the direction
  - average time/freq
  - correct beam array-factor
  - combine CS into a ST & remove CS
- solve for dispersive delays (dTEC)

$$\Delta\phi_{\nu,t} = \phi_0 + \left(\frac{d\phi}{d\nu}\Delta\nu + \frac{d\phi}{dt}\Delta t\right)$$

Delay **Rate**

- apply delay calibrator solutions
- for the selected DOIs:
  - phase-shift to the direction
  - solve for residual dispersive delays
  - self-cal (amp.&phase) with imaging

Focused view: LOFAR-VLBI



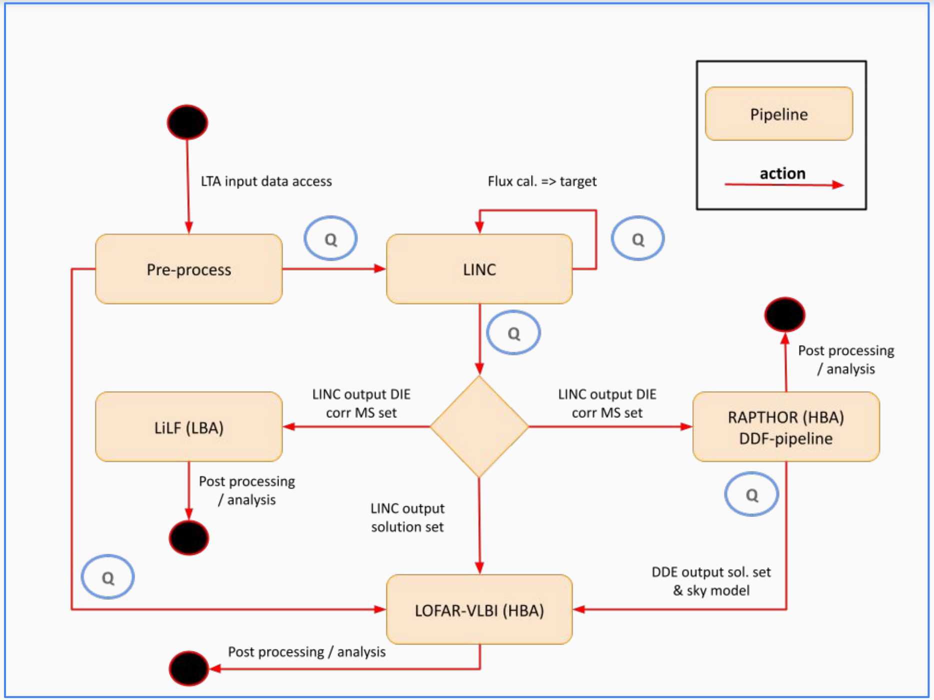
LOFAR-VLBI performs full array Direction-(In)Dependent calib. & img.

For WIDE-FIELD MODE dependency on ddf-pipeline

Mandatory step before any data access / editing is to run the `plot_field.py` script to:

- query online LotSS & LBCS databases to construct catalogues of sources in the field
- provide a short summary of the observation parameters → help the user understand if the data is suitable for LOFAR-VLBI processing !

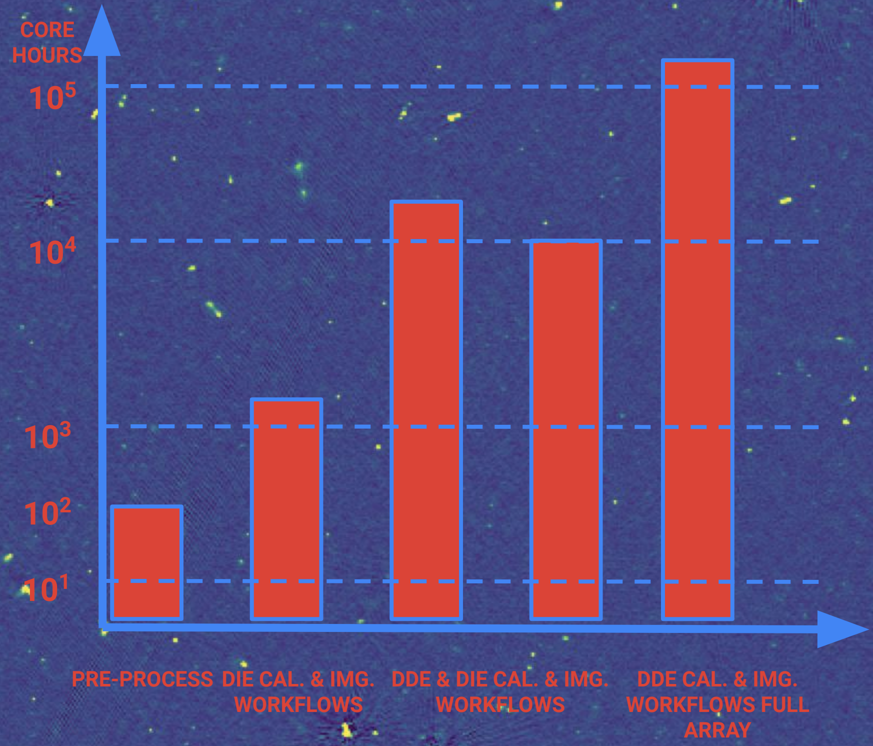
Focused view: LOFAR-VLBI

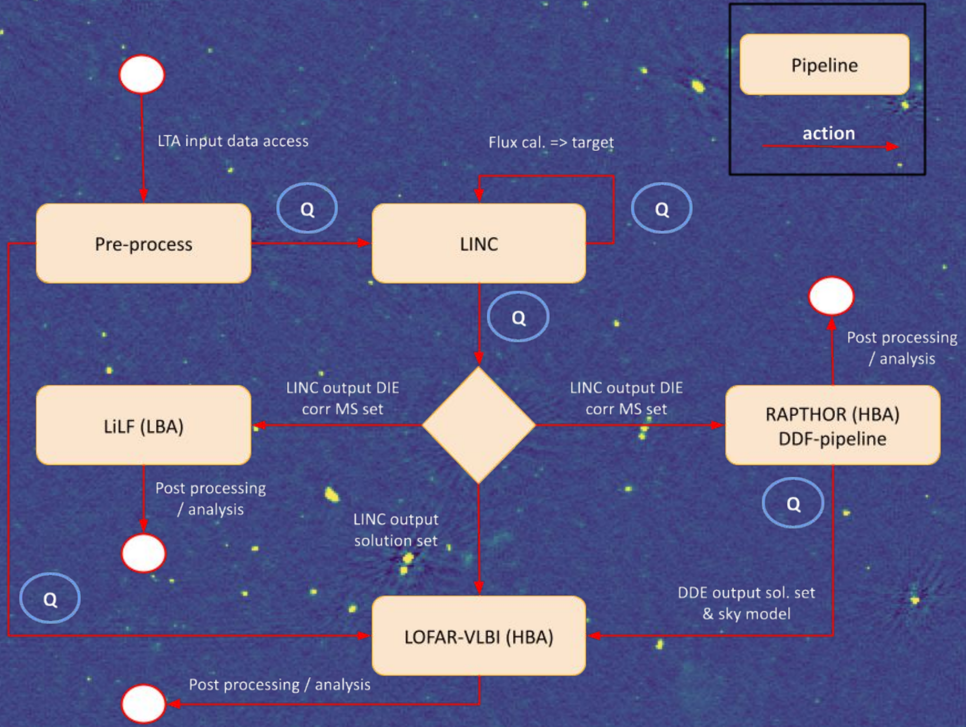


LOFAR-VLBI performs full array Direction-(In)Dependent calib. & img.

- Uses a Python wrapper around CWL pipelines as backend:
  - The wrapper sets up & executes the pipelines as "operations" to perform the actual processing
  - Allows distribution over cluster nodes
  - Allows resuming of interrupted jobs
- Main software packages used:
  - Default PreProcessing Pipeline (DPPP) to average, flag, calibrate, apply calib. sols
  - LOFAR Solution tools (LoSoTo) to analyse/extract parameters from calib. sols
  - WSClean to apply calib. sols & perform imaging

Take home messages . . .





7th LOFAR data processing school
April 16 2024