

An end-to-end data reduction pipeline for MeerKAT and other radio telescopes

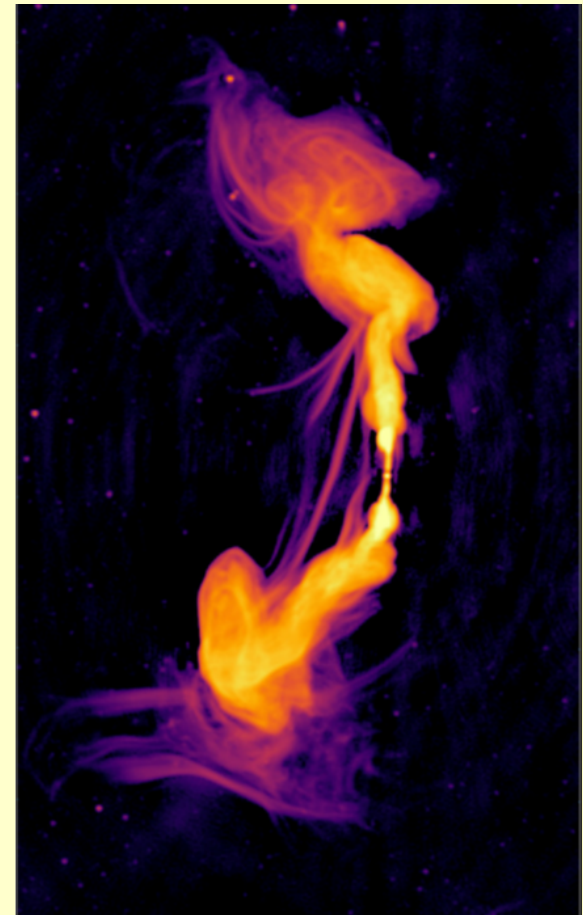
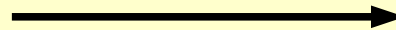


G. I. G. Józsa, S. V. White, K. Thorat, O. M. Smirnov., P. Serra, M. Ramatsoku, A. J. T. Ramaila, S. Perkins, D. Molnár, S. Makhathini, F. M. Maccagni, D. Kleiner, P. Kamphuis, B. V. Hugo, W. J. G. de Blok, L. A. L. Andati

<https://caracal.readthedocs.io>



MeerKAT
(+ Deep2 field, Mauch et al. 2019)

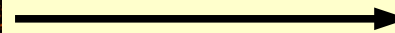


ESO 137-006
(Ramatsoku et al. 2020)

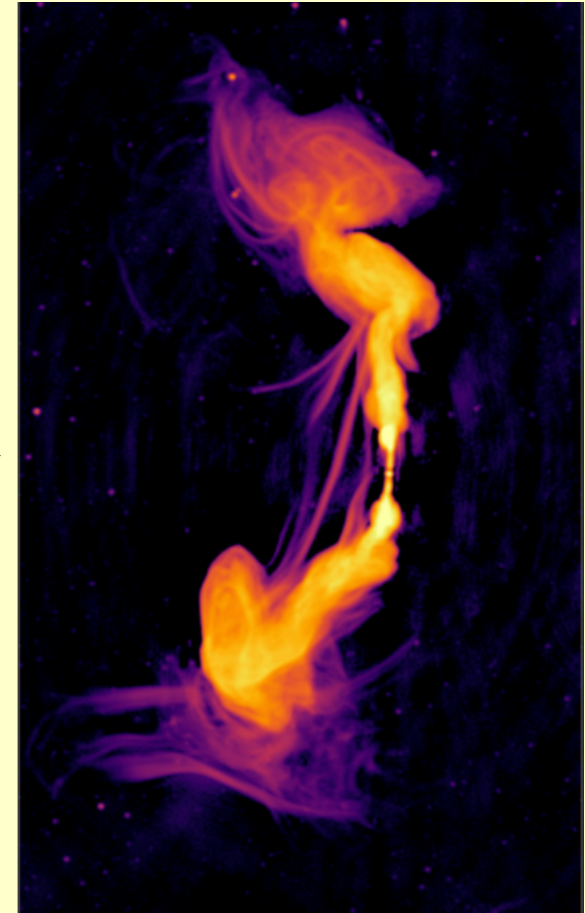
ESO 137-006
(Ramatsoku et al. 2020)



MeerKAT
(+ Deep2 field, Mauch et al. 2019)



Observation
Storage
Processing
Archiving

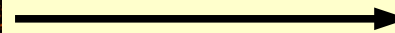


ESO 137-006
(Ramatsoku et al. 2020)

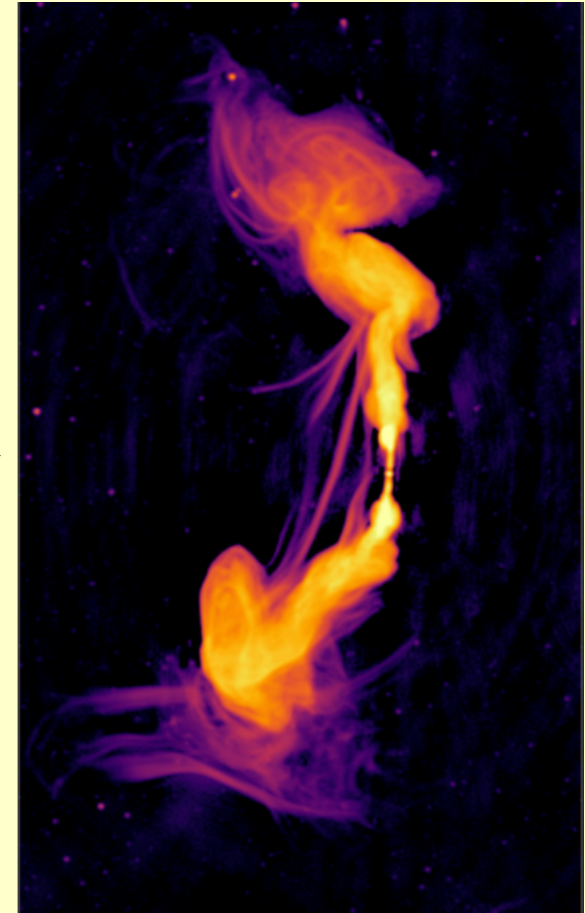
ESO 137-006
(Ramatsoku et al. 2020)



MeerKAT
(+ Deep2 field, Mauch et al. 2019)



Observation
Storage
Processing
Archiving



ESO 137-006
(Ramatsoku et al. 2020)

- Data volume makes interactive processing impractical
 - New telescopes and receivers require improved data reduction
- > Requirement of modern processing pipelines (and computing infrastructure)

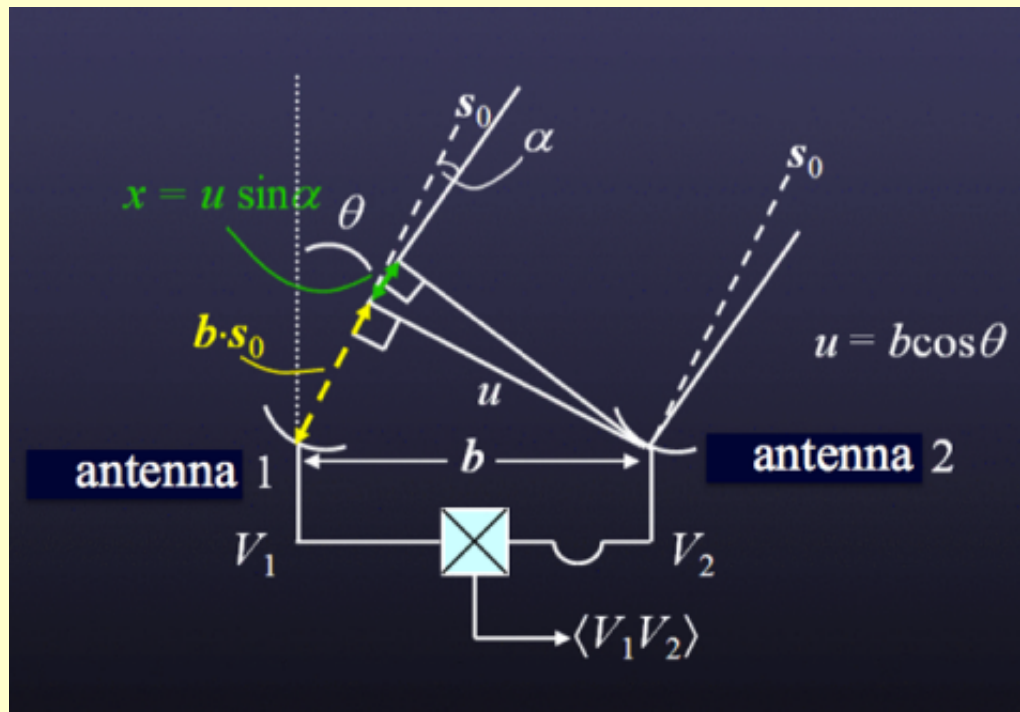


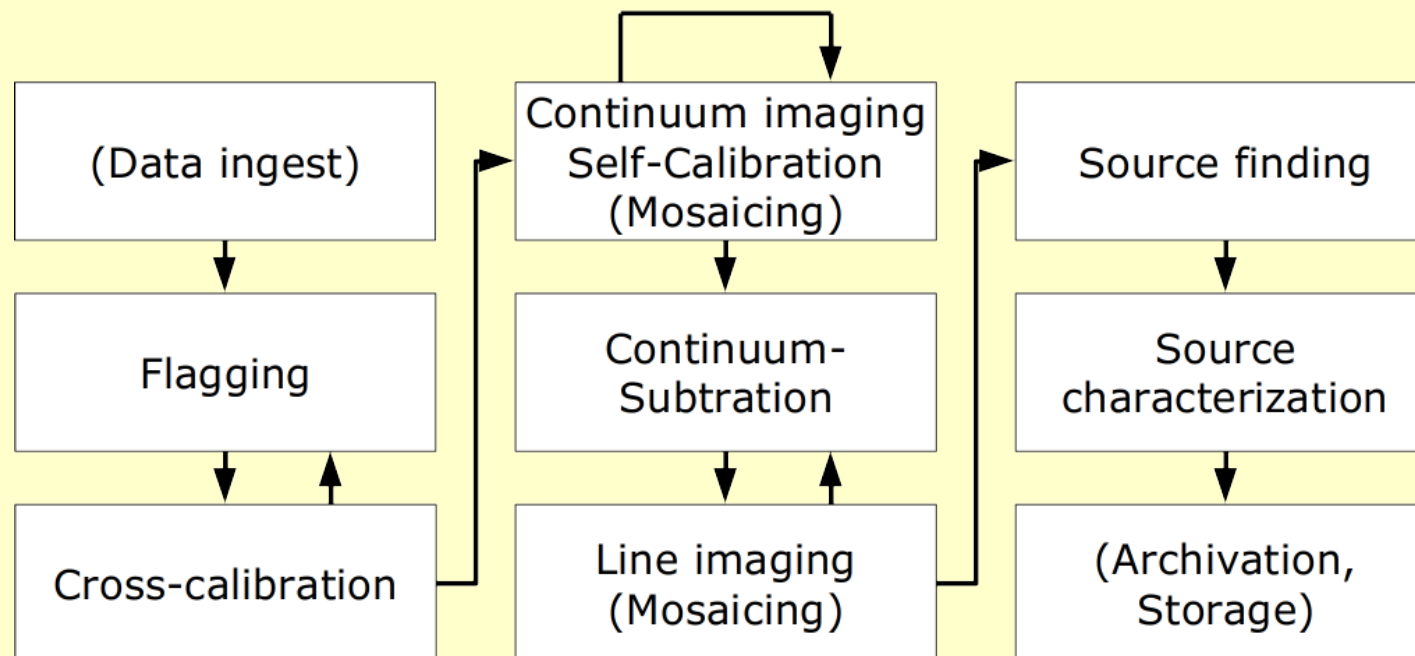
Image Credit: ALMA



Image Credit: SARA

Containerized Automated Radio Astronomy Calibration

- End-to-End data reduction pipeline
- Source accessible and easy to change
- Make use of any best available software (for the specific telescope)
- Portable
- Easy to install
- Tunable and adjustable to science case and telescope (with emphasis on MeerKAT)
- Scalable (working for large data sets in distributed environment)

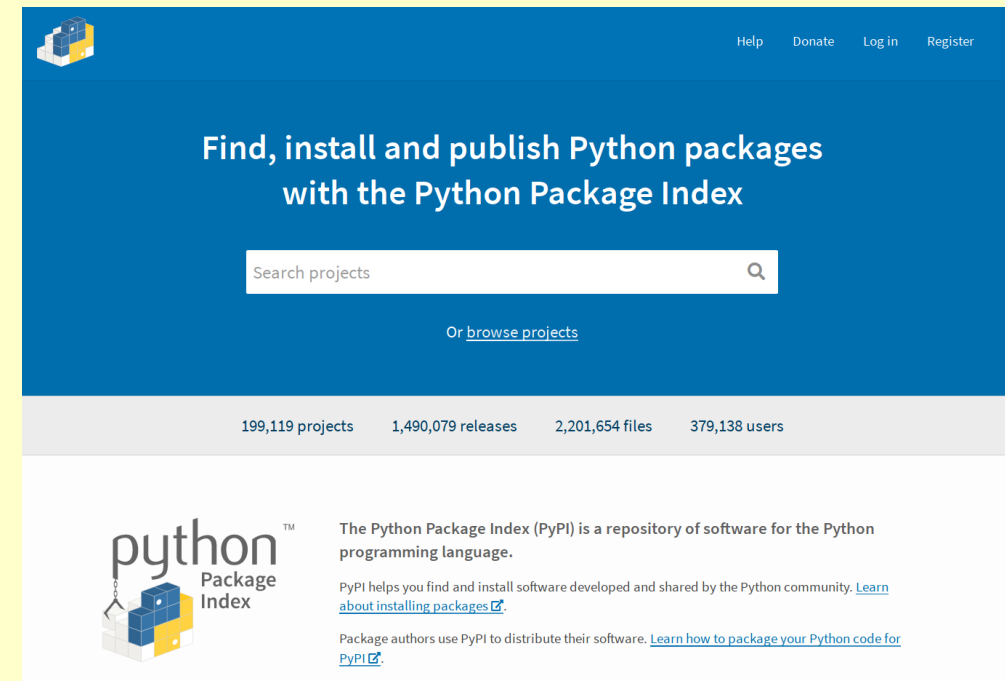
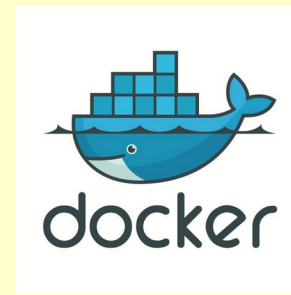
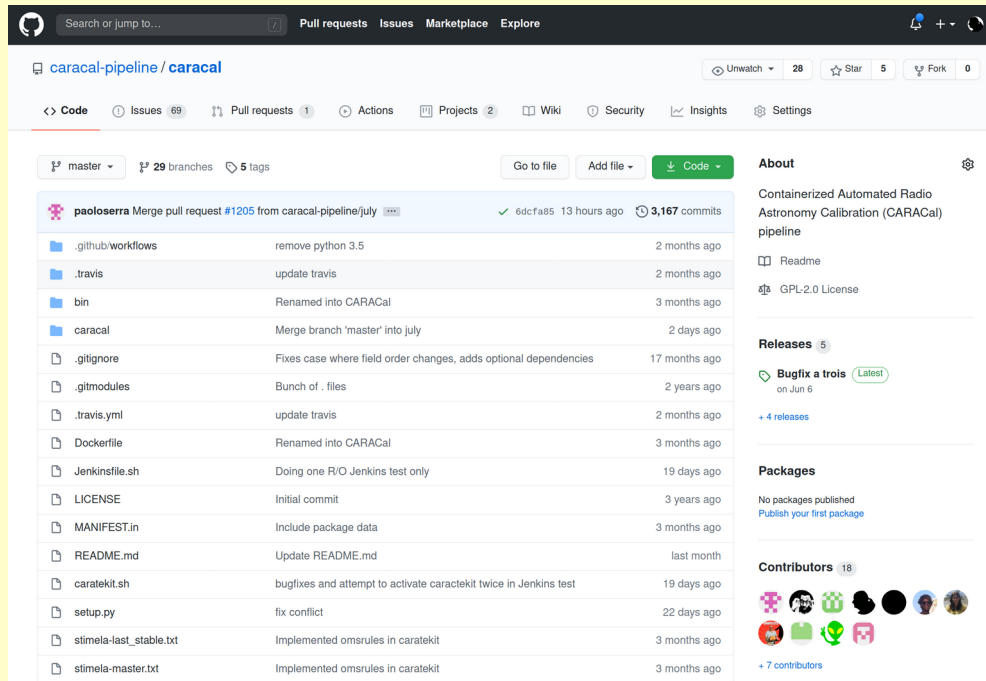


The team



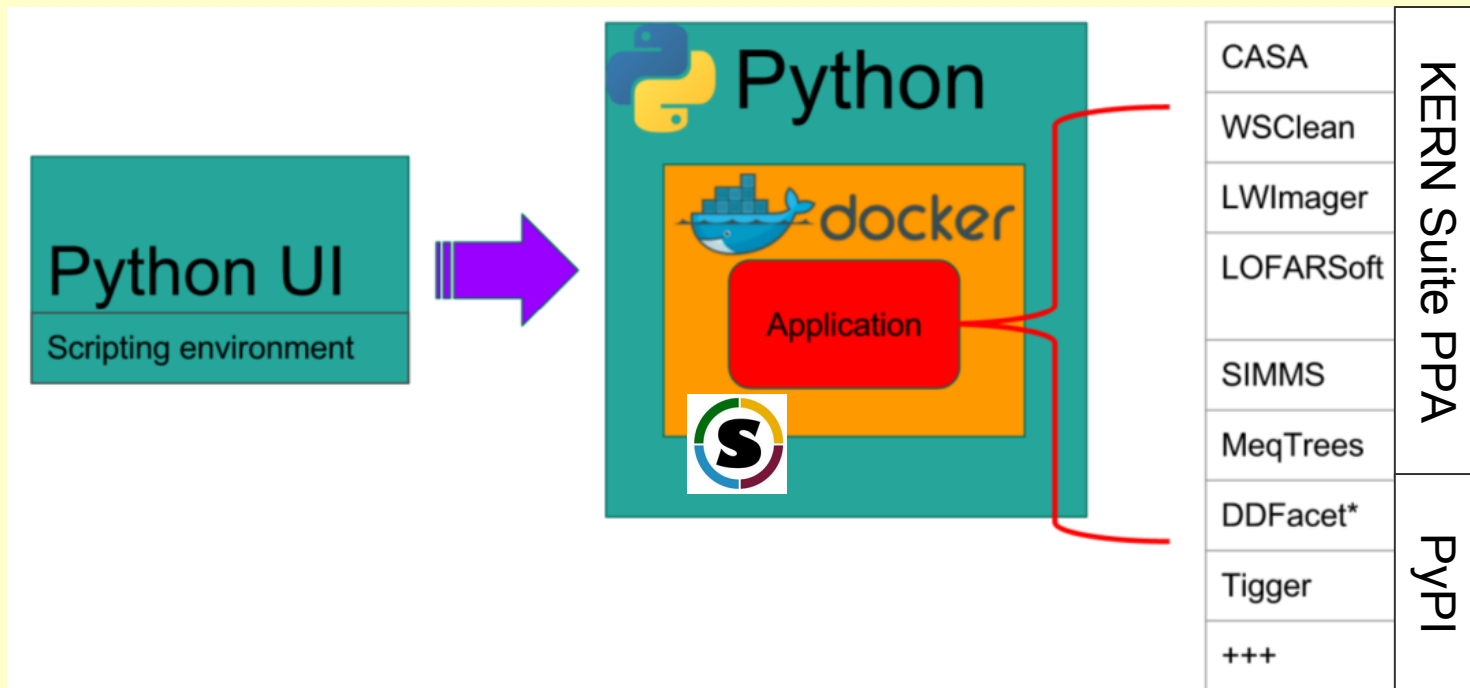
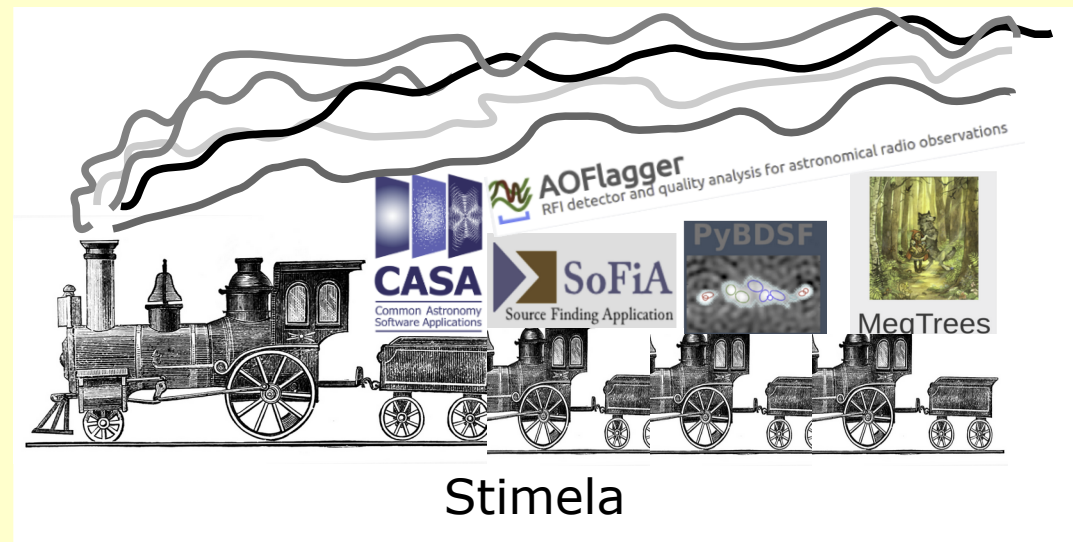
- 16 developers
- Astronomers -> simple implementation/language required
- Computer scientists -> good communication of physical problems
- Some of them students -> good time management / fast turnaround

- Python (simple)
- PyPI (accessible)
- Stimela framework (portability)



Architecture and building blocks

- Python (simple)
- PyPI (accessible)
- Stimela framework
 - Several containerization methods
 - Common interface
 - Variety of most recent software



- Python (simple)
- PyPI (accessible)
- Stimela framework
 - Several containerization methods
 - Common interface
 - Variety of most recent software
- Quality Assessment and remote inspection



```

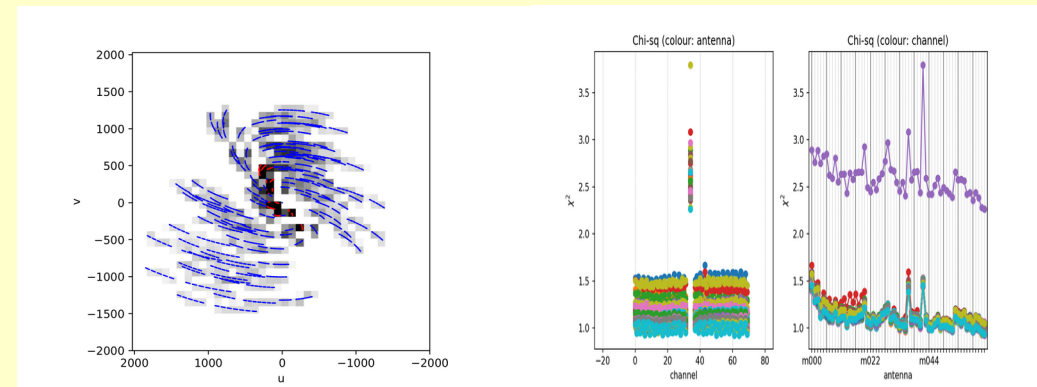
emacs@com06.science.kat.ac.za (on com06.science.kat.ac.za)
File Edit Options Buffers Tools Help
schema_version: 0.2.0

general:
  prefix: ic1459_30MHz_comb
  rawdatadir: '/scratch2/jozsa/ic1459/rawdata'

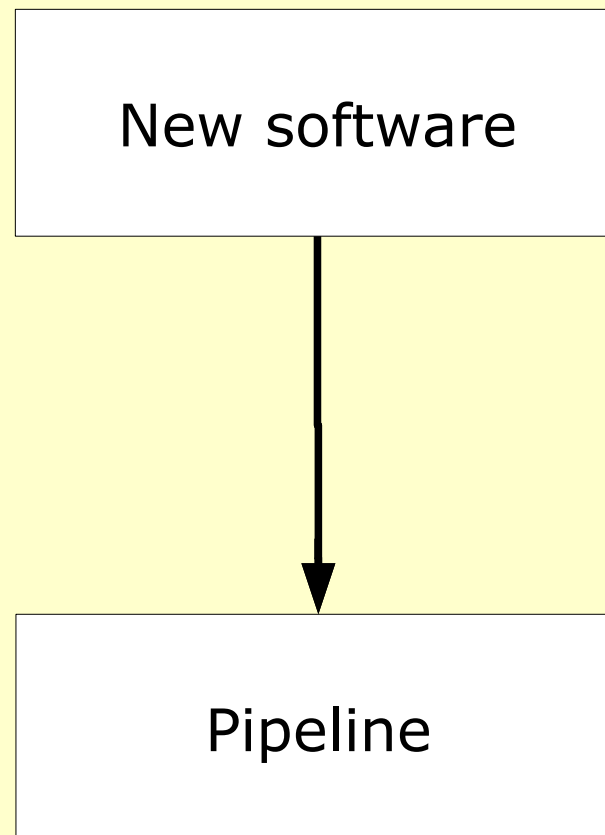
getdata:
  dataid: ['ic1459_30MHz_old', 'ic1459_30MHz_new']

obsconf:
  refant: 'm010'

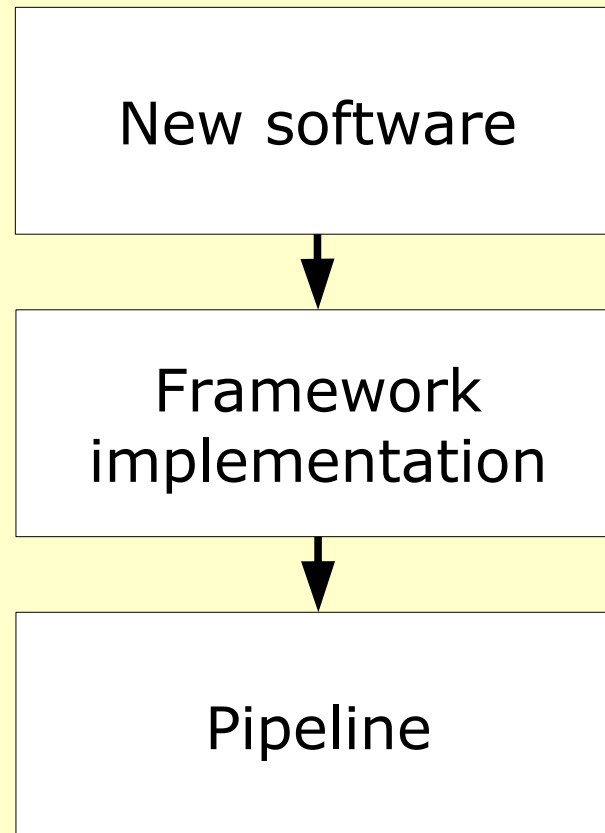
line:
  enable: true
  label_in: 'l'
  restfreq: '1.420405752GHz'
  subtractmodelcol:
    enable: false
  make_cube:
    enable: false
    npix: [1800]
    cell: 6
    taper: 20
    robust: 0
  mstransform:
    enable: true
    doppler:
      enable: true
      telescope: meerkat
    uvlin:
      enable: true
      fitorder: 4
      fitspw: '*:1405.5~1406.5MHz, *:1407.4~1409.9MHz, *:1414.7~1419.7MHz, *:142.5~1.4~1425.0MHz'
    flag_mst_errors:
      enable: true
    pb_cube:
      enable: false
  -:--- ic1459_comb_a.yml Top (3,8) (Fundamental)
  
```



- Portability
 - Minimize dependencies
 - Usage of containerization
 - Implementation of new techniques/software

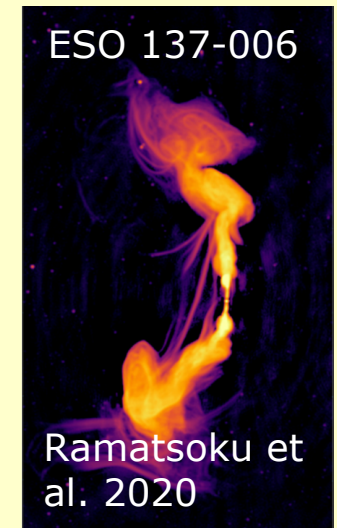
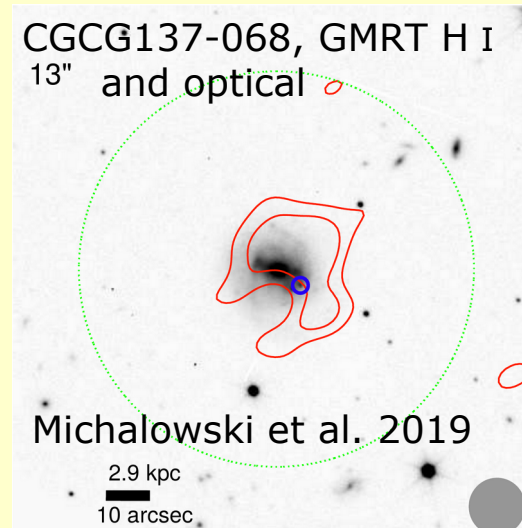
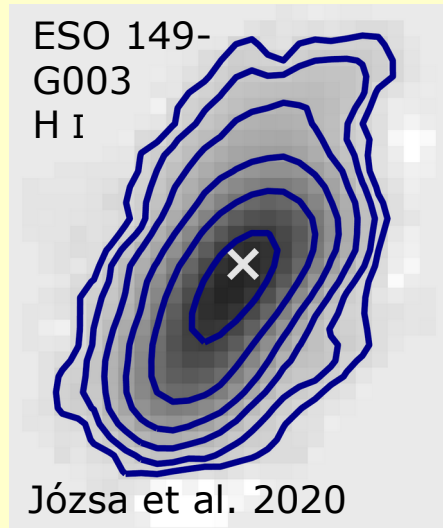
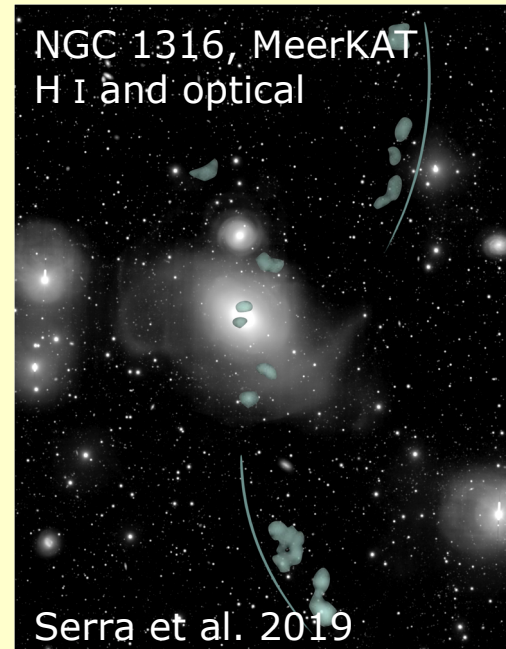
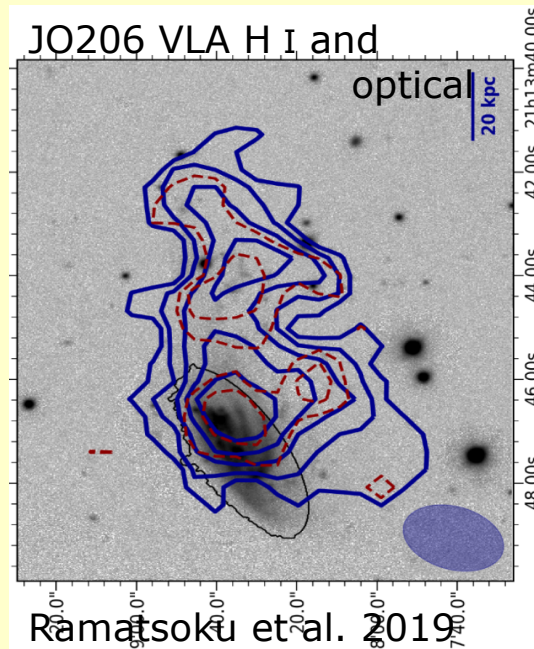


- Portability
 - Minimize dependencies
 - Usage of containerization
 - Implementation of new techniques/software depends on integration in framework
- **Has to be maintained with short turnaround times**



Time lines

- Start in May 2017
- 5 Busy weeks
- Public release in May 2020
- Several publications based on CARACal



- CARACal is an end to end data reduction pipeline for radiointerferometric data
- Based on the scripting and containerization framework Stimela it was possible to release a first version on a relatively short time scale
- The framework and particular software choices made an efficient communication between developers with very different background possible
- One requirement for this to work is constant and good communication and a fast turnaround for the framework implementation
- Stay tuned for future developments

<https://caracal.readthedocs.io>

