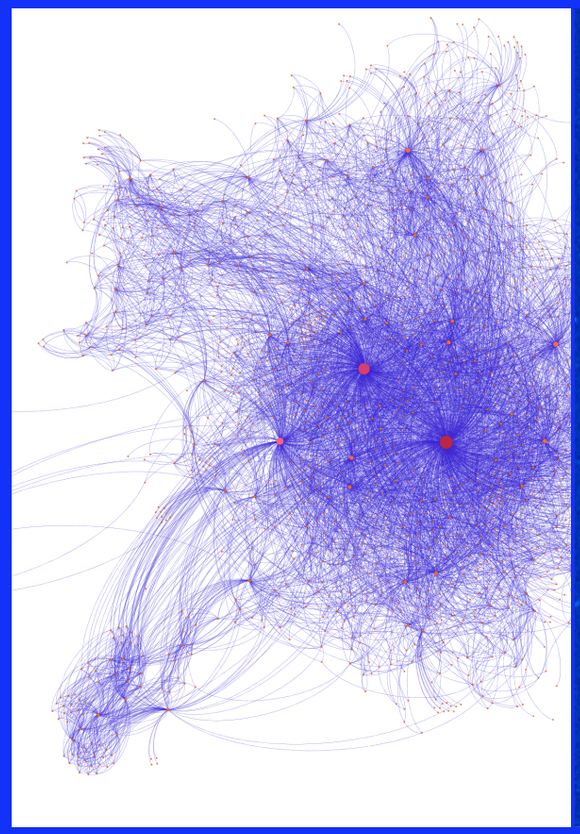
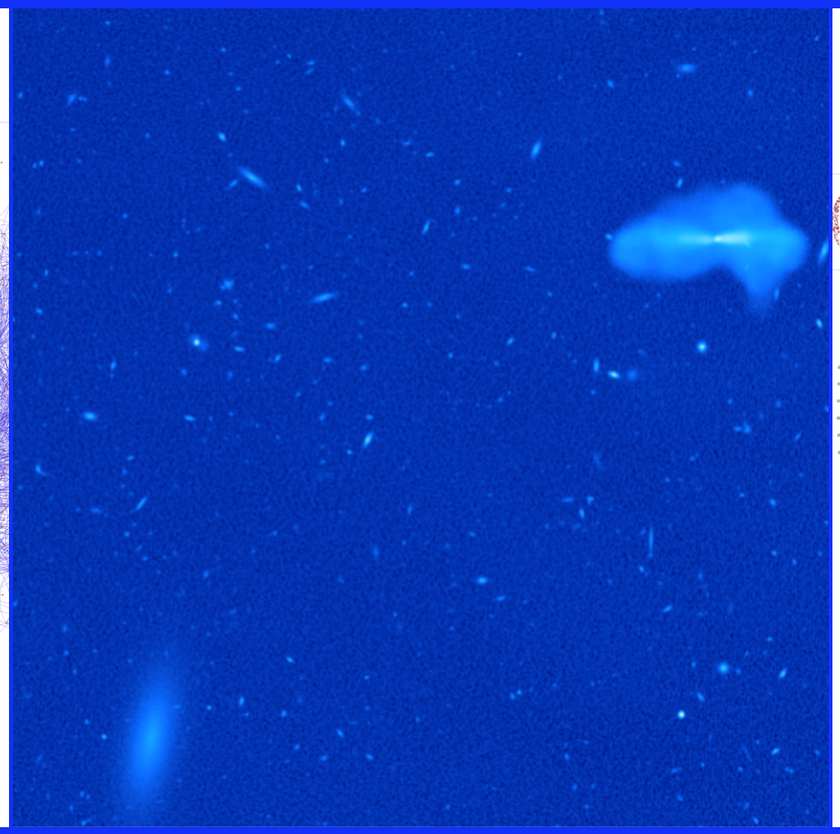
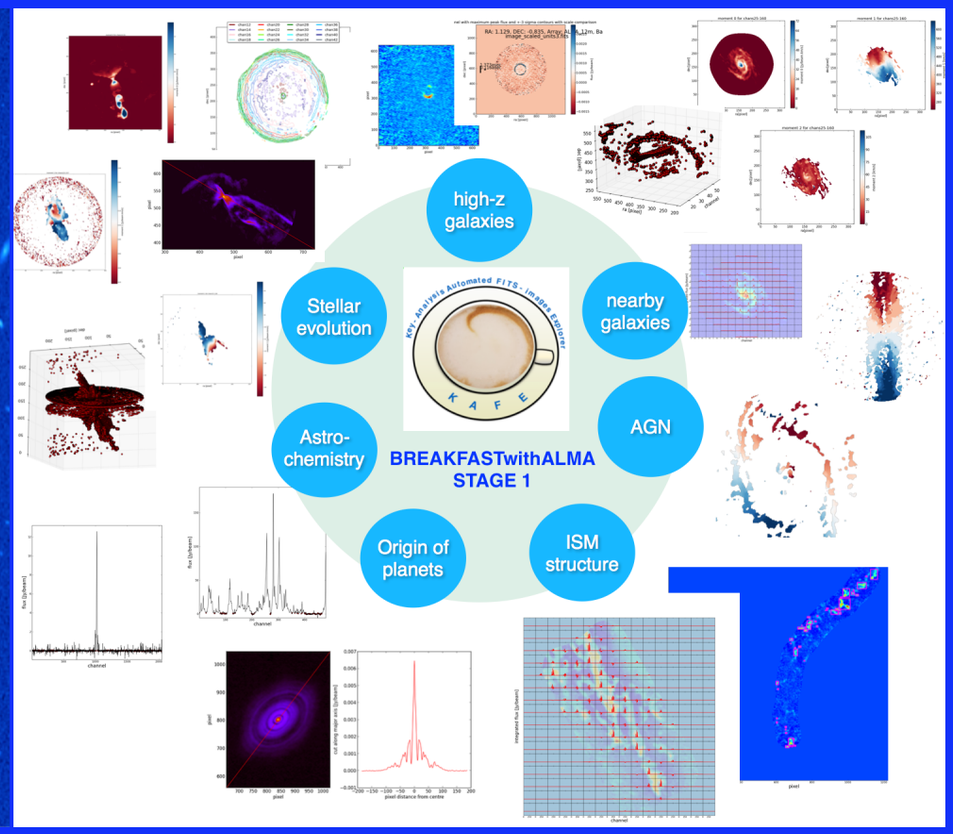




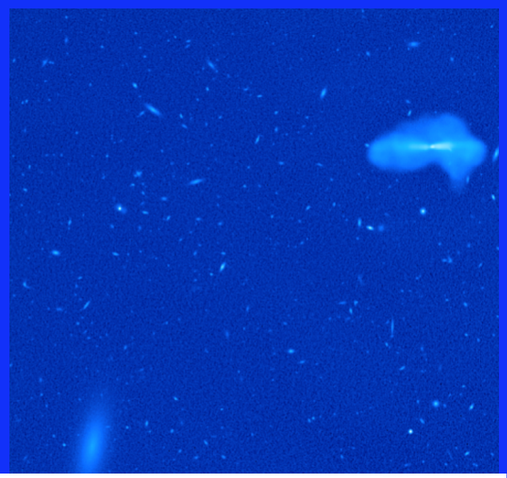


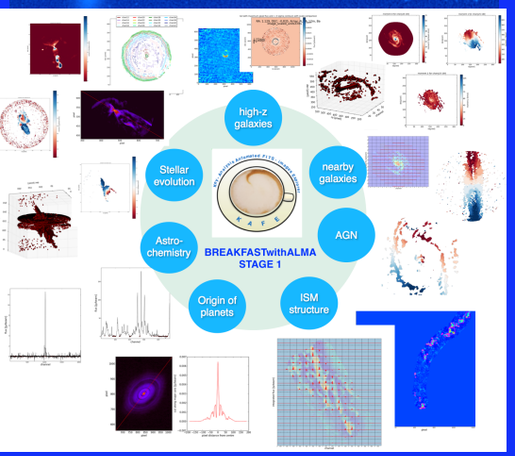
Dealing with data complexity as astronomy approaches Big Data

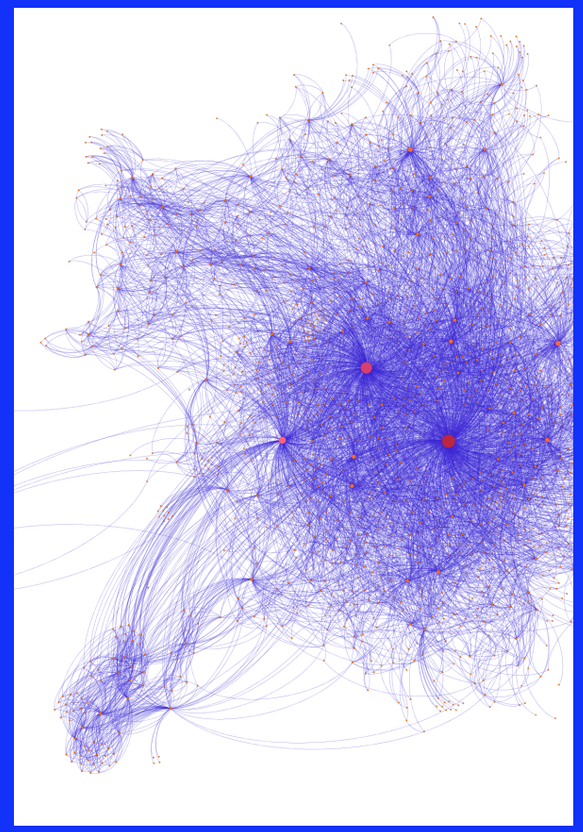
Sandra Burkutean
INAF-IRA, Italian ALMA Regional Centre, Bologna, Italy

EUROPEAN ARC
ALMA Regional Centre || Italian

How to get the most out of your data







data utility

data analysis routines

data access

link to theory/ modeling

synergy with other telescopes

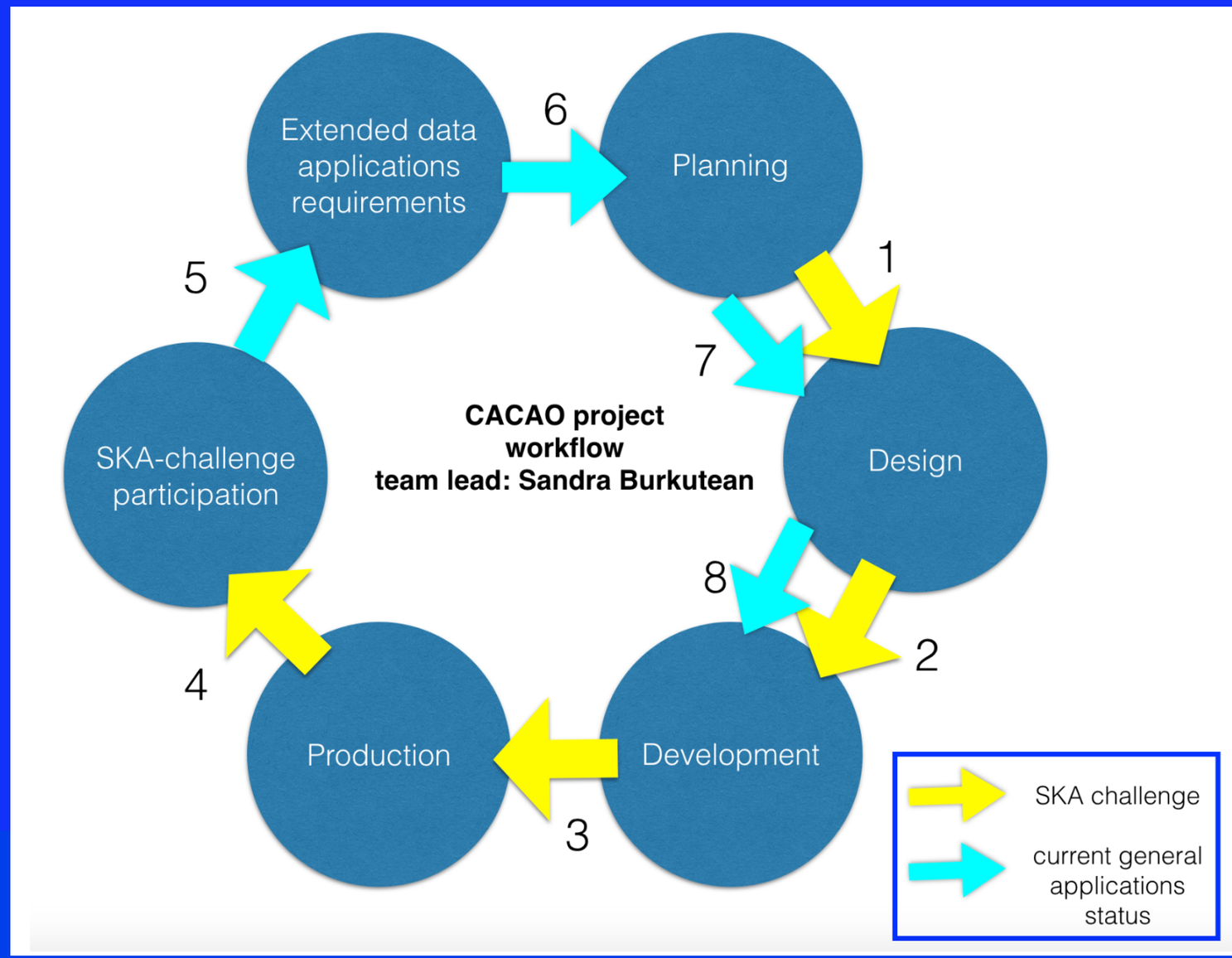
data visualization

data taxonomy

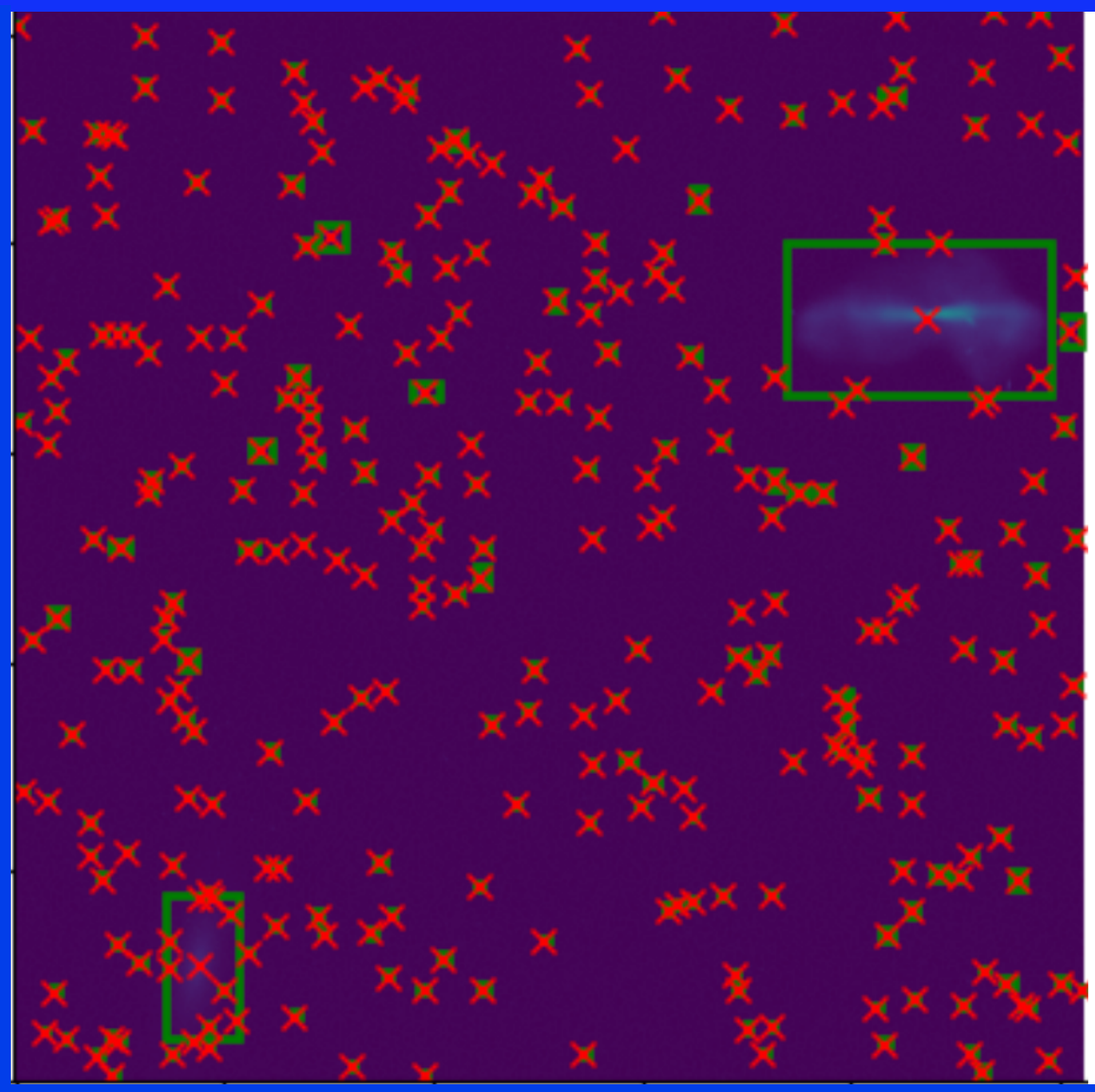
software tools developments

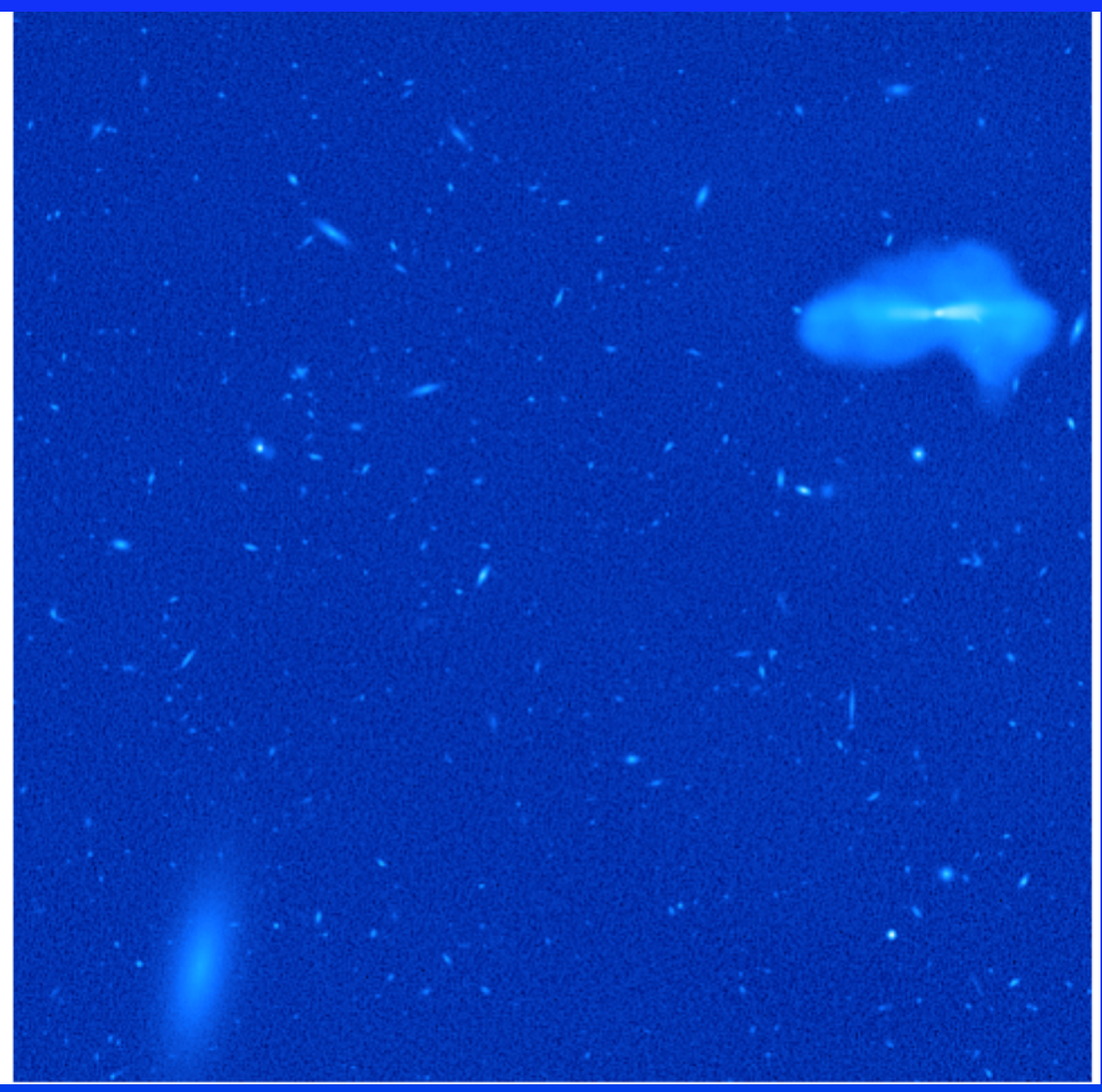
ARCIt-CACAO

Italian SKA data challenges experience



CACAO: The Complete Automated Classification of Astronomical Objects Tool

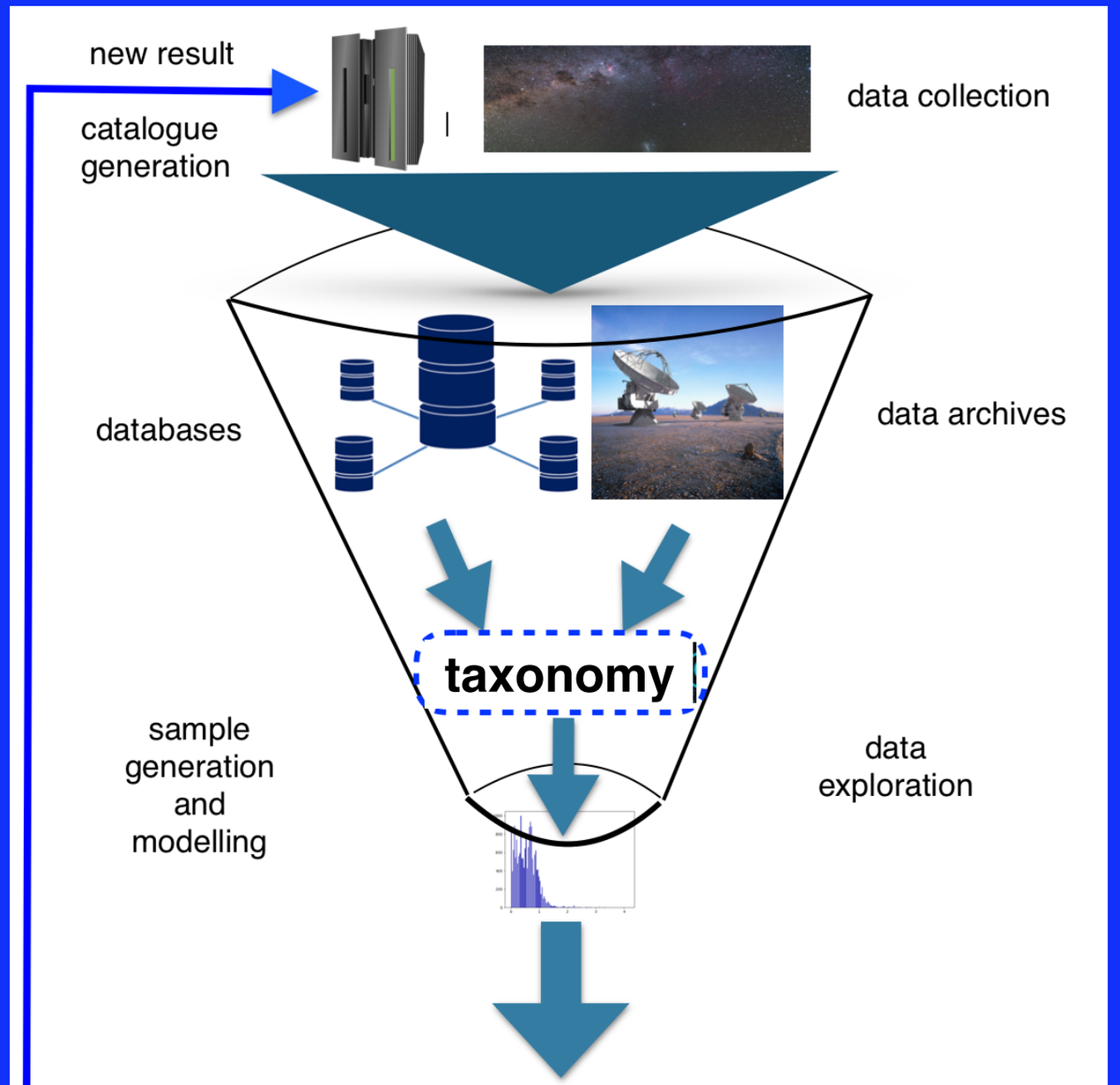




source finding, description, classification

developped for the 1st SKA data challenge, tests on real data

fully parallelized, further extension towards machine learning ongoing



often, the dataflow stops here once the catalogues have been produced

Proof of concept: BREAKFASTwithALMA

Burkutean in prep.

data taxonomy

sub-sample generation

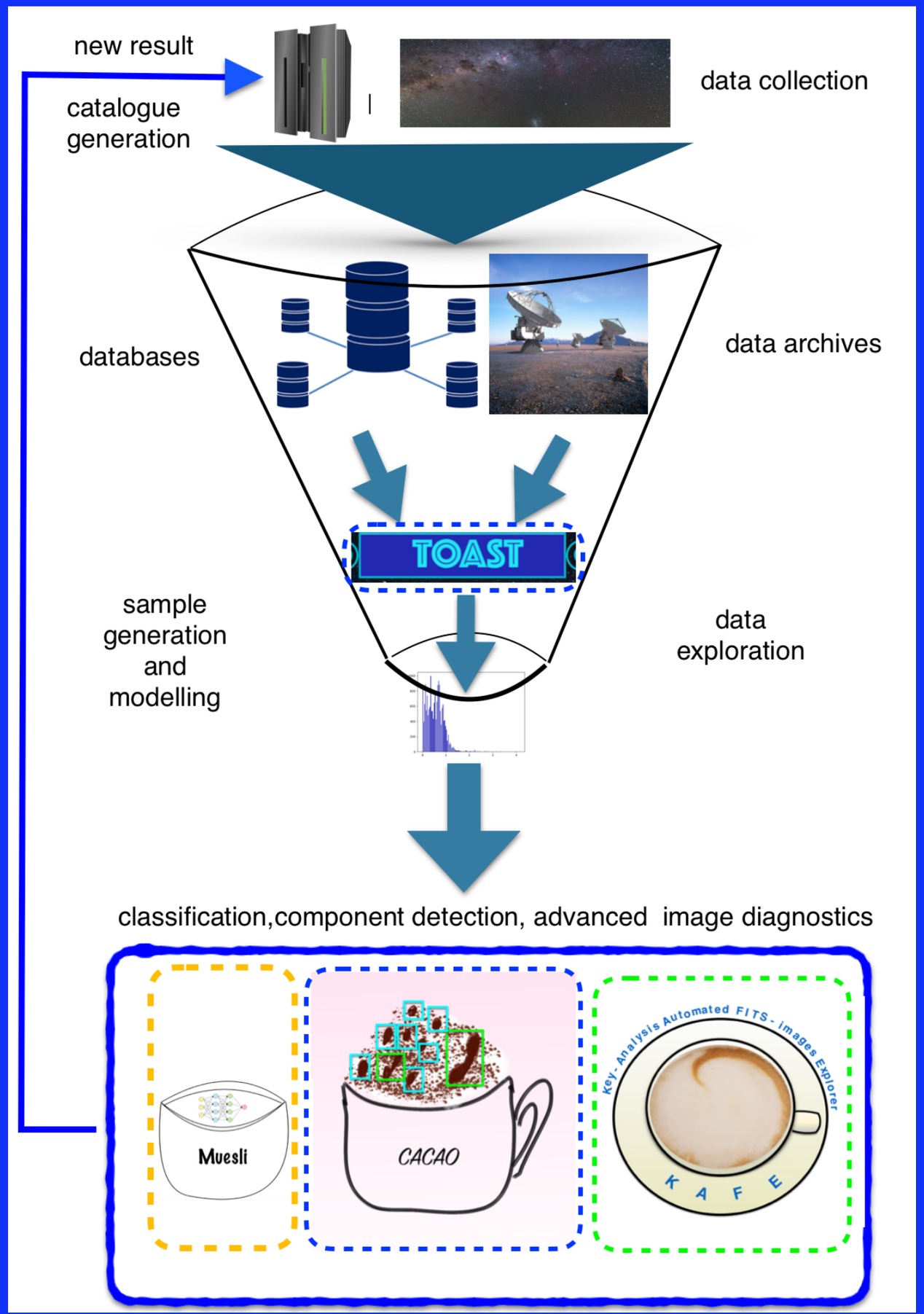
multi-project data combination

advanced data product generation

modeling of physical parameters

Automation through machine learning

Proof of concept: BREAKFASTwithALMA



TOAST: Telescope Observational Archive Sample Tool

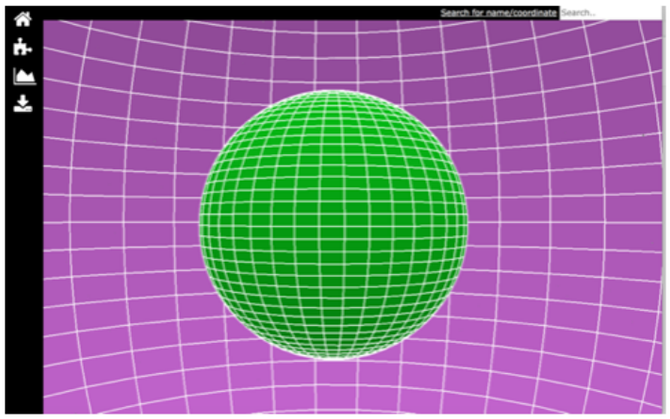
Burkutean in prep.

+

+

+

=

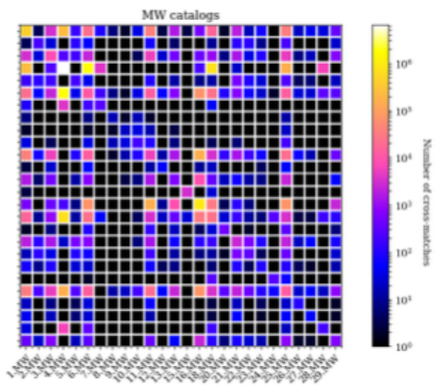


Big DATA visualisation

+

Big DATA analysis

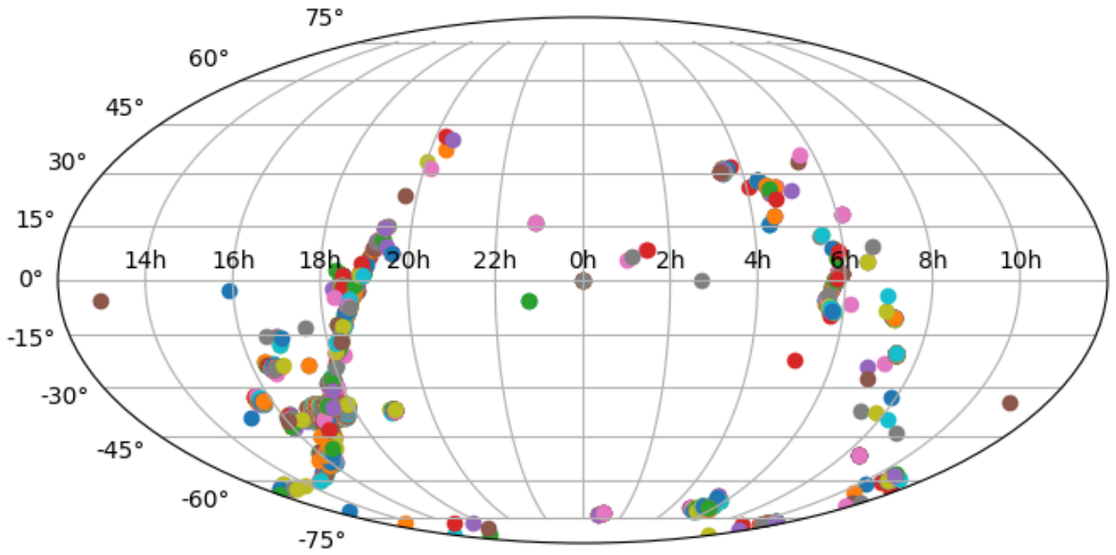
+

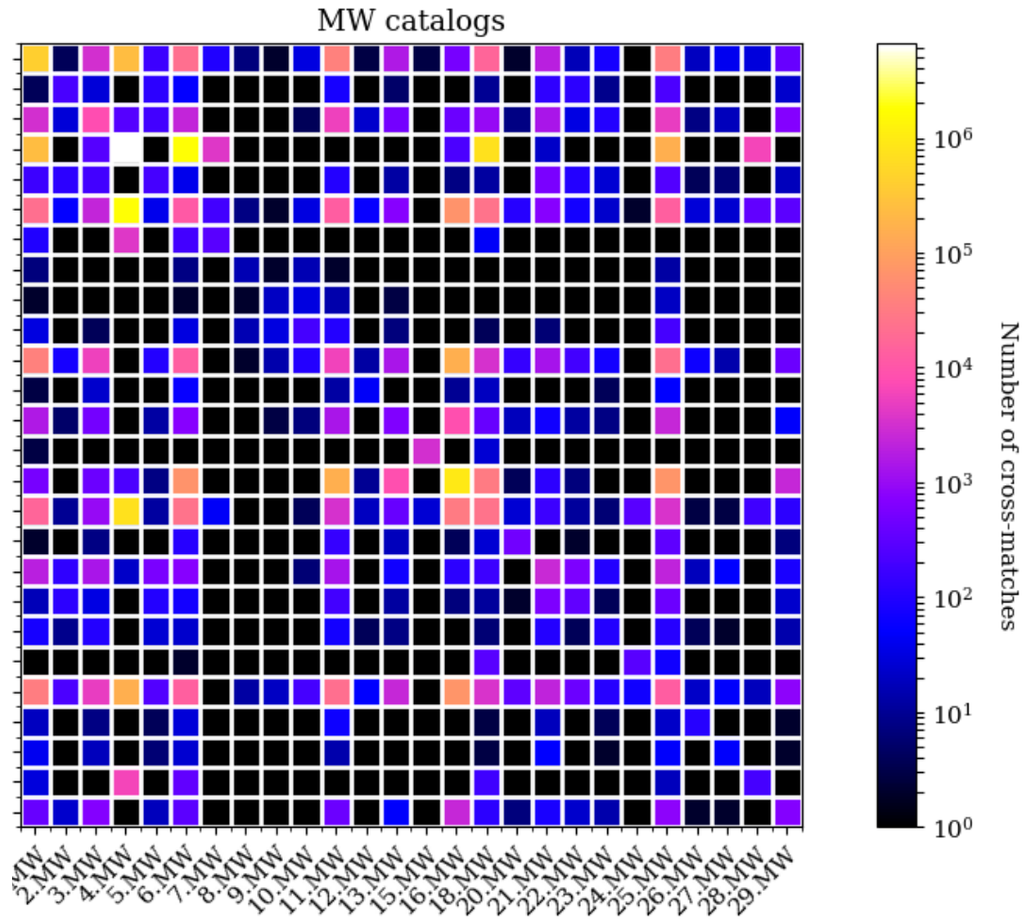


Big DATA database statistics

BREAKFASTwithALMA and TOAST

Burkutean in prep.

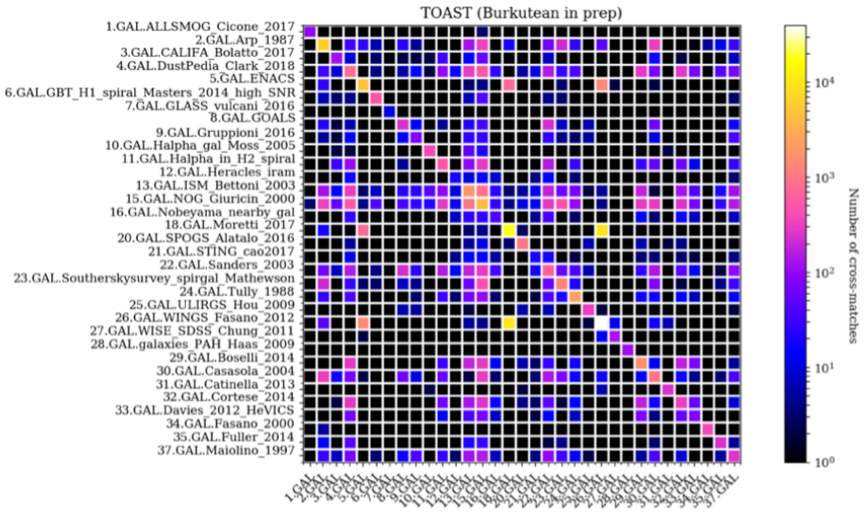


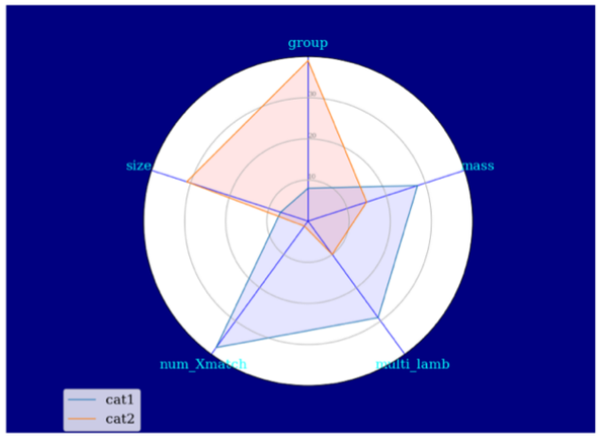


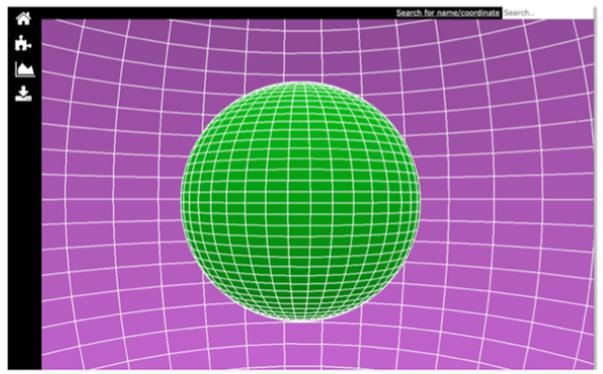
using synergy with catalogues+NED+SIMBAD to create sub-samples

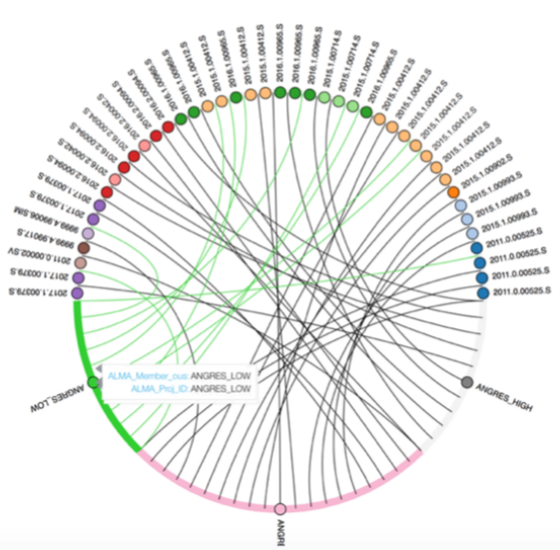
TOAST: Telescope Observational Archive Sample Tool

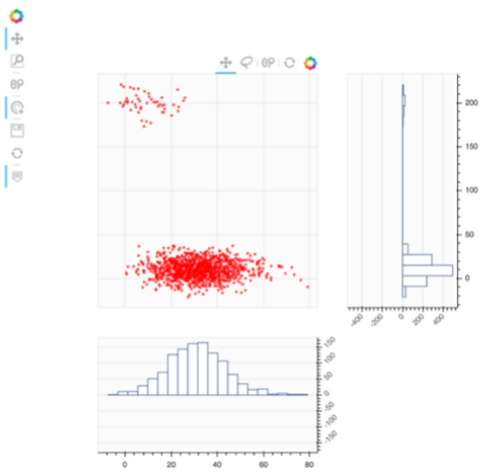
Burkutean in prep.







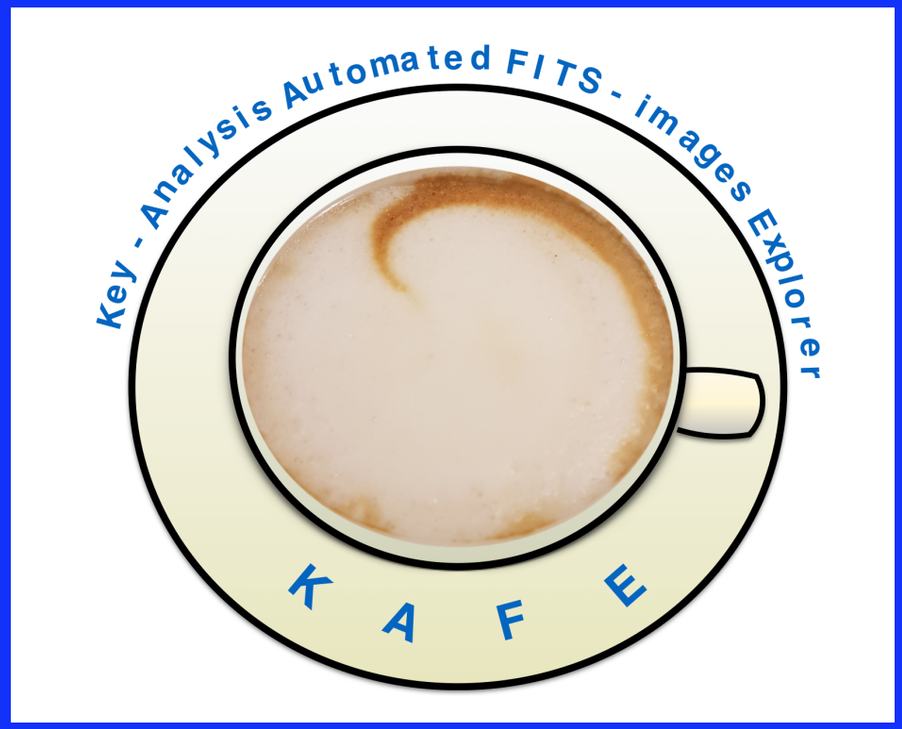




fully interactive graphs on web interface with partial 3D visualization

KAFE: automated FITS image analysis + visualisation

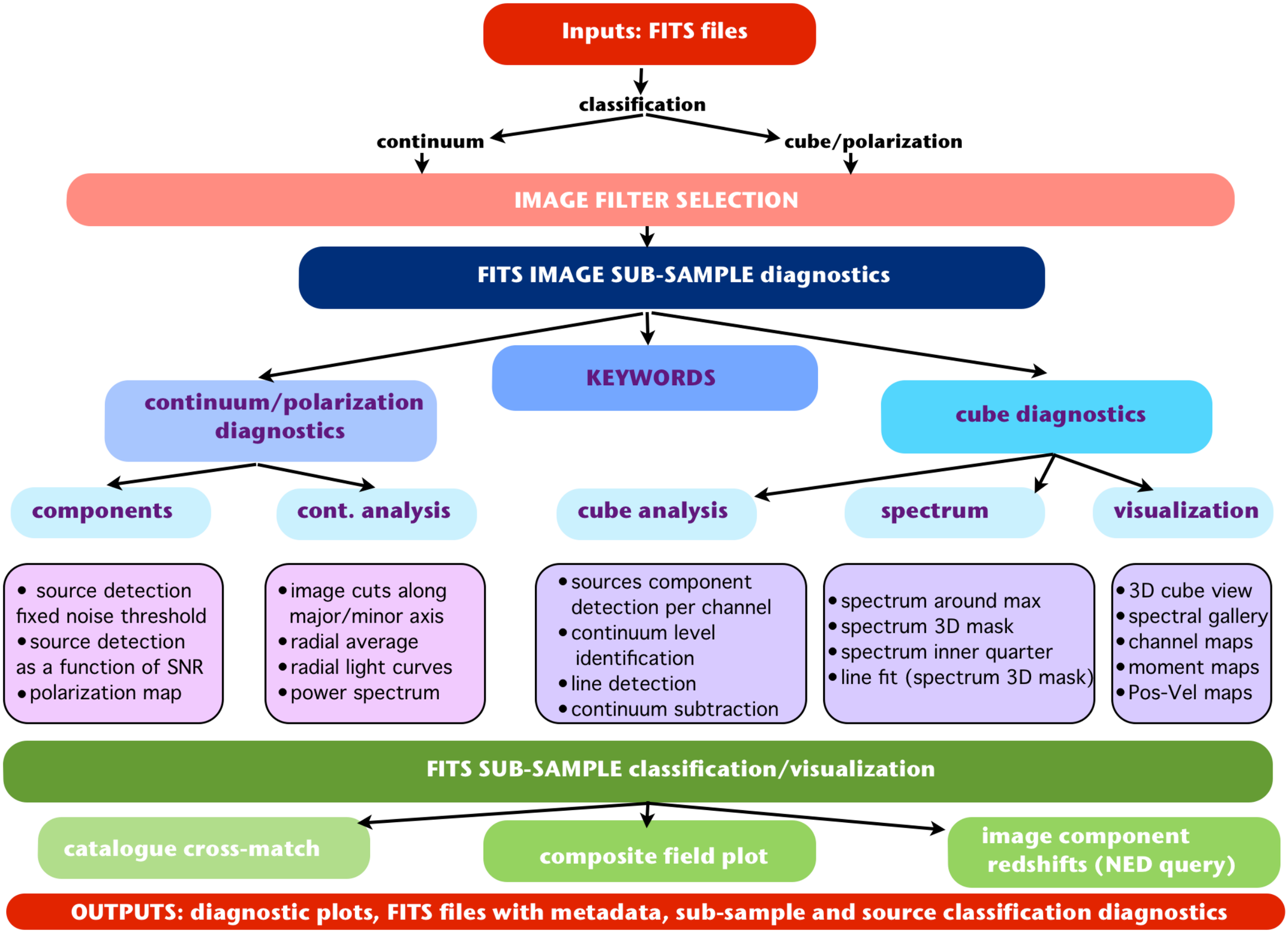
Burkutean et al., J. Astron. Telesc. Instrum. Syst. 4(2), 028001 (2018)



user-generated/archival FITS-images

image analysis for ALMA,JVLA,PdB etc.

fully automated


Inputs: FITS files
classification
continuum
cube/polarization
IMAGE FILTER SELECTION
FITS IMAGE SUB-SAMPLE diagnostics
KEYWORDS
continuum/polarization diagnostics
cube diagnostics
components
cont. analysis
cube analysis
spectrum
visualization
• source detection fixed noise threshold
• source detection as a function of SNR
• polarization map
• image cuts along major/minor axis
• radial average
• radial light curves
• power spectrum
• sources component detection per channel
• continuum level identification
• line detection
• continuum subtraction
• spectrum around max
• spectrum 3D mask
• spectrum inner quarter
• line fit (spectrum 3D mask)
• 3D cube view
• spectral gallery
• channel maps
• moment maps
• Pos-Vel maps
FITS SUB-SAMPLE classification/visualization
catalogue cross-match
composite field plot
image component redshifts (NED query)
OUTPUTS: diagnostic plots, FITS files with metadata, sub-sample and source classification diagnostics

Proof of concept: BREAKFASTwithALMA

Public archival images in the ALMA archive

KAFE

nearby galaxies

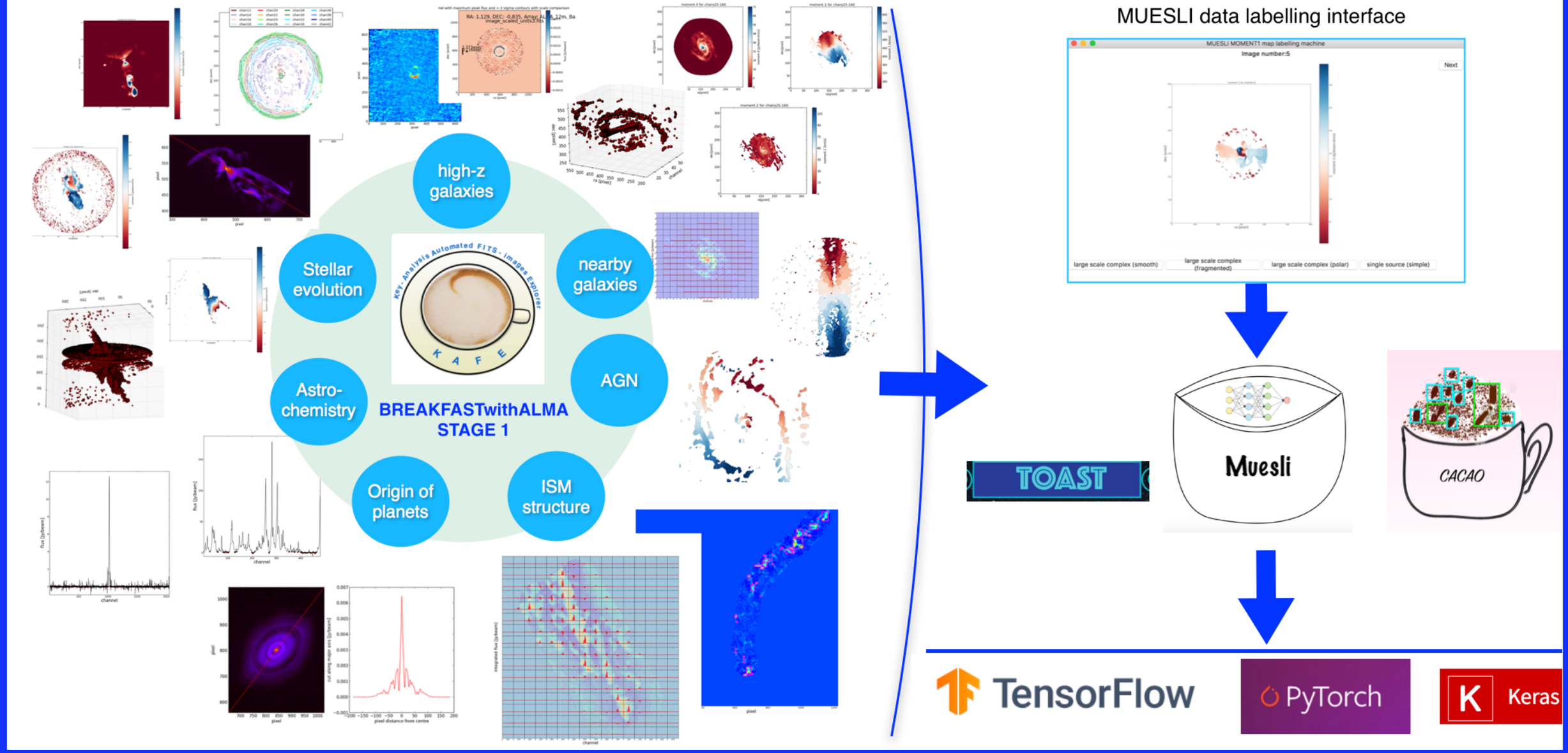
galactic

AGN

high-z

on Italian ARC cluster + HPC computing time at CHIPP (Trieste)
+ GPU time (CINECA)

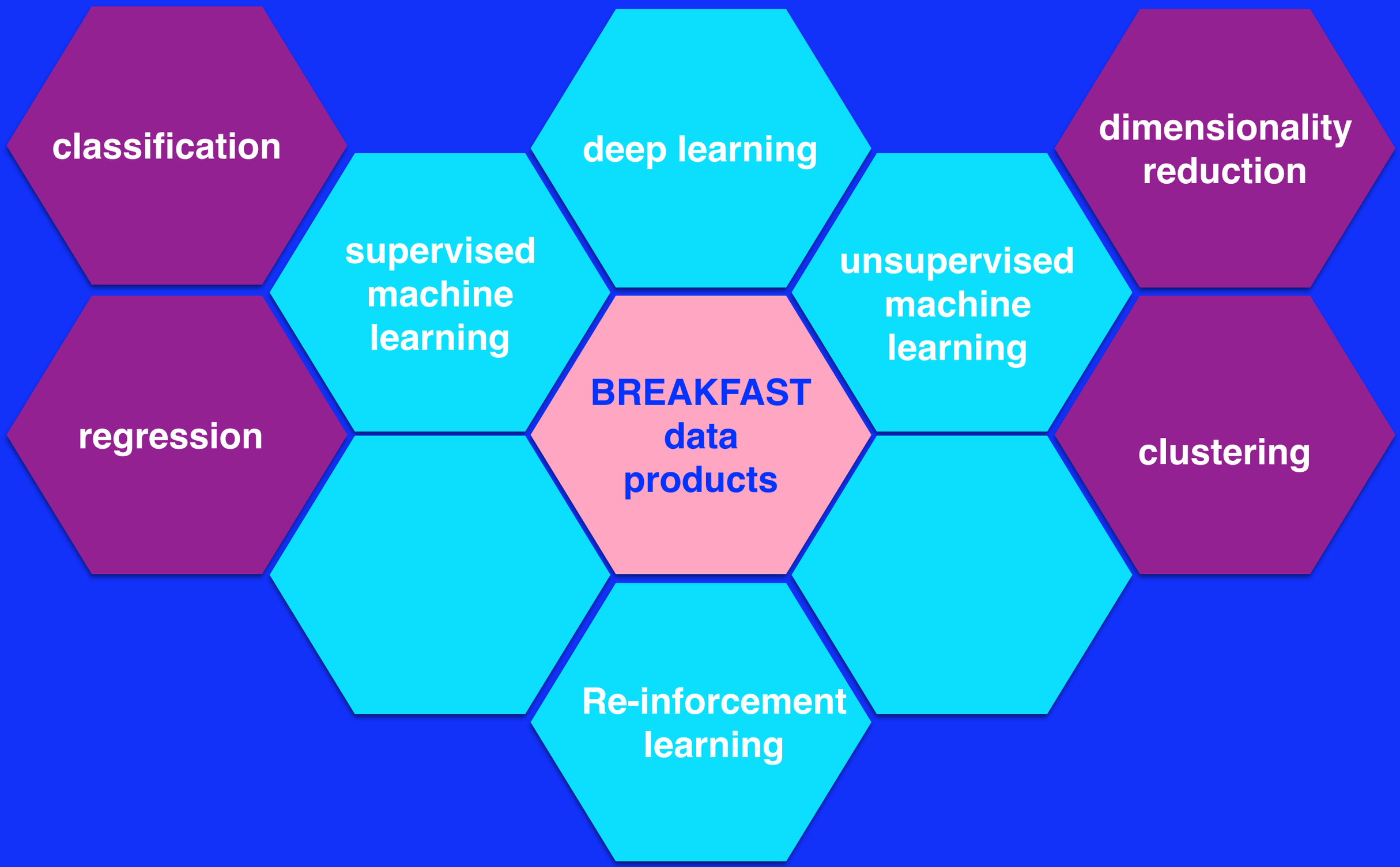
Proof of concept: BREAKFASTwithALMA



Automation is the key !

MUESLI and SCONES: Machine and Deep Learning applications

Burkutean in prep.

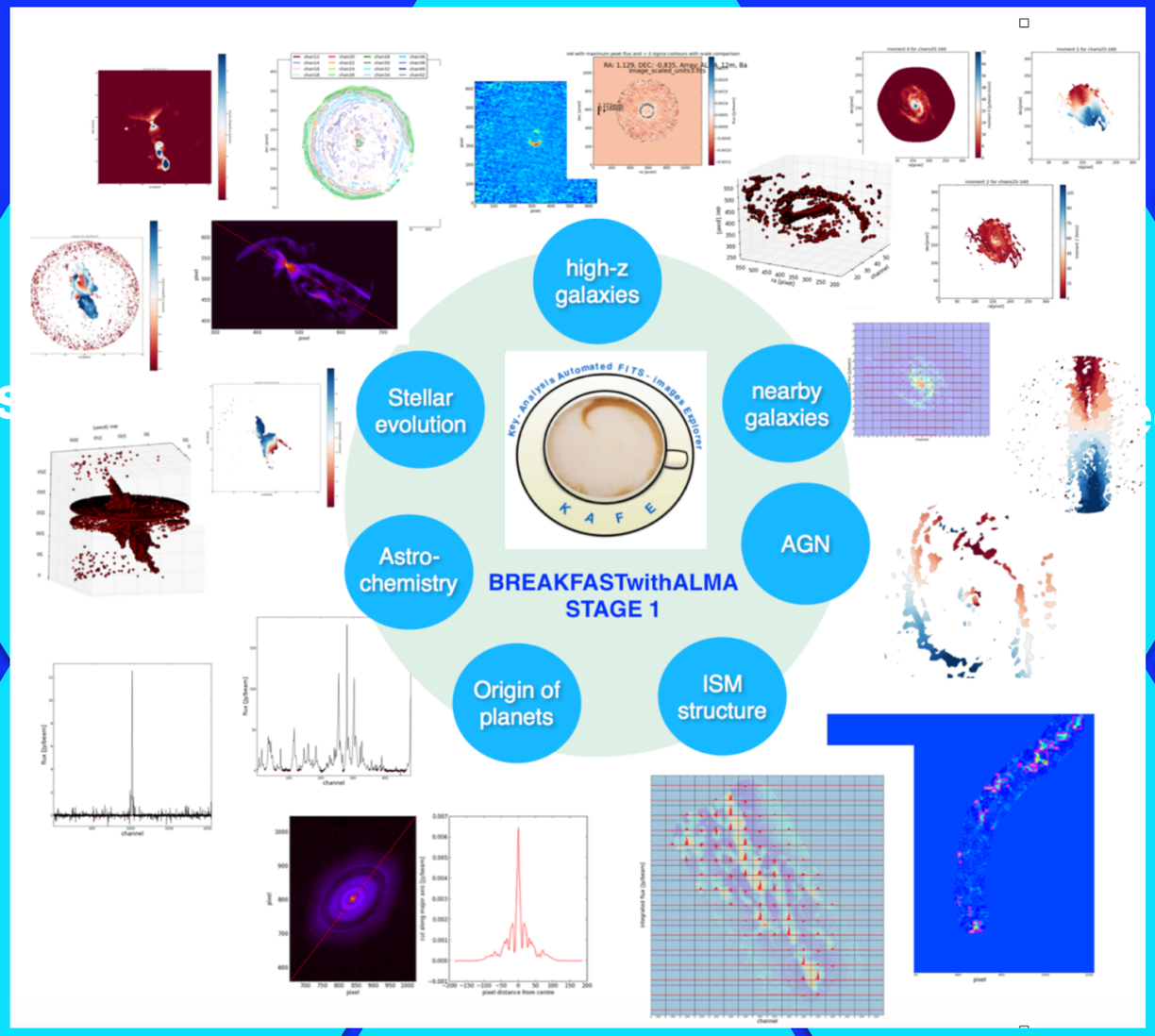


MUESLI and SCONES: Machine and Deep Learning applications

Burkutean in prep.

classification

regression



dimensionality reduction

clustering

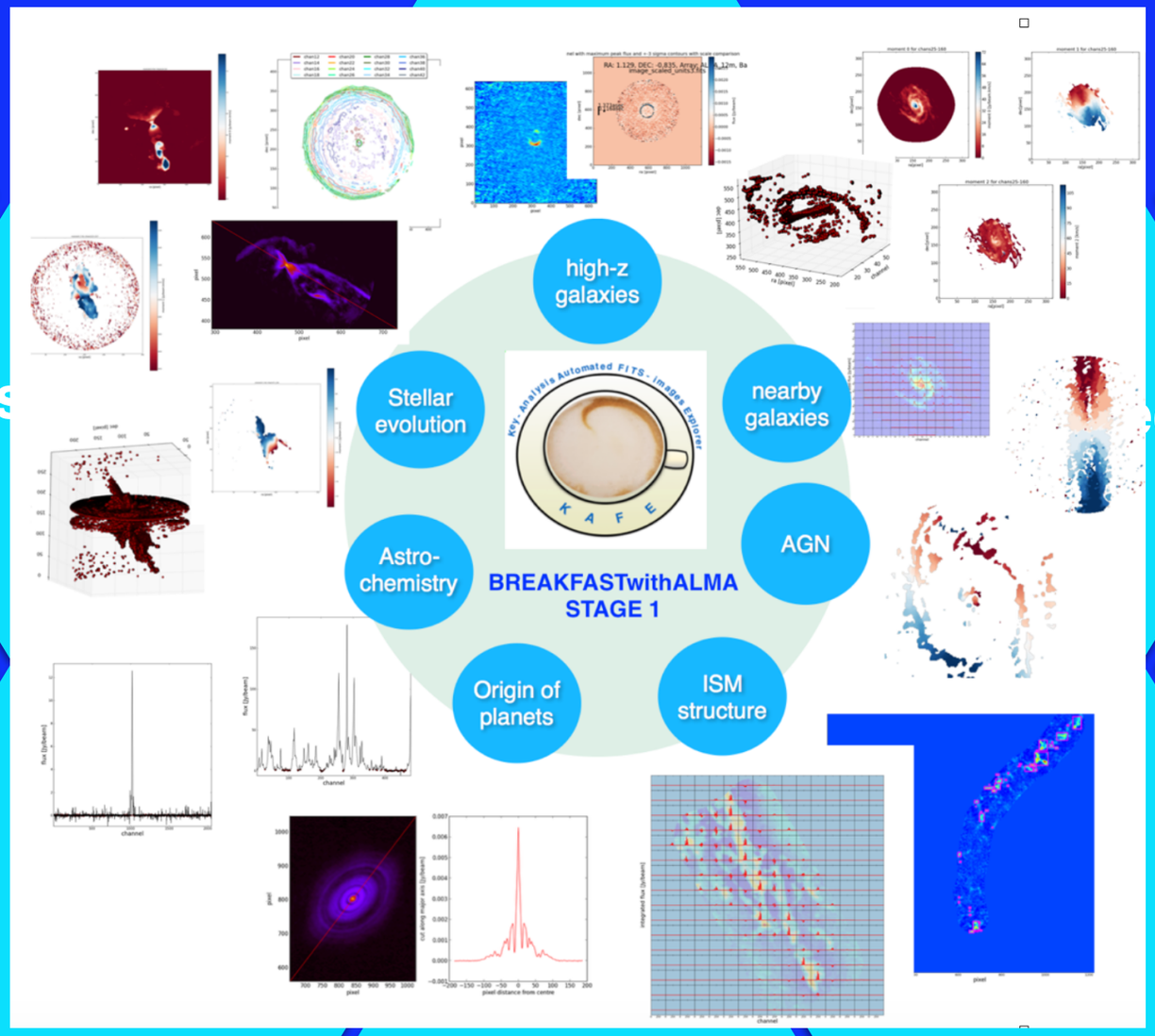
Automatically generated advanced data products and their associated metadata are vital !!!

MUESLI and SCONES: Machine and Deep Learning applications

Burkutean in prep.

classification

regression



dimensionality reduction

clustering

Ultimate aim: build a database system/archive that can stand the test of time and whose search capabillities exploit AI techniques !

Summary

Pathfinder projects such as BREAKFASTwithALMA can simulate dataflow at the analysis stage

We need to think of long-term advanced data product maintainance.

AI will be essential for pinpointing scientifically interesting regions in the data product parameter space

Stay tuned for BREAKFAST!