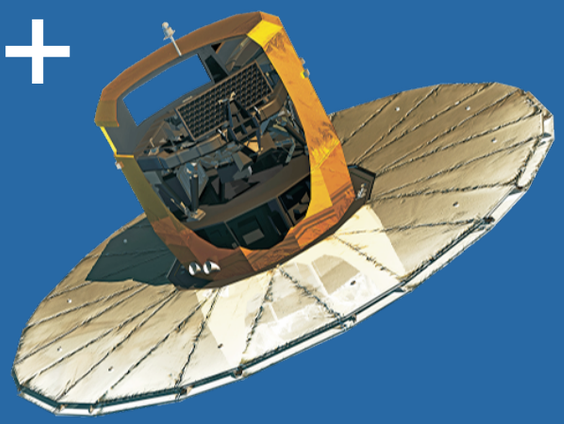
Building an automated 100 million+ variable star catalogue for Gaia

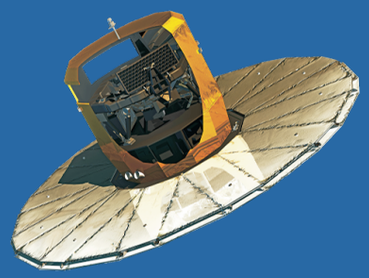


Berry Holl[1]
presented by: Laurent Eyer[1]
Nami Mowlavi[1], Dafydd W. Evans[2], Gisella Clementini[3], Jan Cuypers[4], Alessandro Lanzafame[5], Joris D. Ridder[6], Luis Sarro[7], Diego Ordoñez-Blanco[1], Krzysztof Nienartowicz[1], Jonathan Charnas[1], Leanne Guy[1], Grégory Jévardat de Fombelle[1], Isabelle Lecoeur-Taïbi[1], Lorenzo Rimoldini[1], Maria Süveges[1], François Bouchy[8,1], Sara Regibo[6], Mauro López[9], Jonas Debosscher[6], Fabio Barblan[1]

[1]Department of Astronomy, University of Geneva, Versoix, Switzerland
[2]Institute of Astronomy, University of Cambridge, Cambridge, United Kingdom
[3]INAF Osservatorio Astronomico di Bologna, Bologna, Italy
[4]Royal Observatory of Belgium, Brussels, Belgium
[5]Dipartimento di Fisica e Astronomia, Universita di Catania, Catania, Italy
[6]Instituut voor Sterrenkunde, KU Leuven, Leuven, Belgium
[7]Departamento de Inteligencia Artificial, UNED, Madrid, Spain
[8]Institut d'Astrophysique de Paris, CNRS, Paris, France

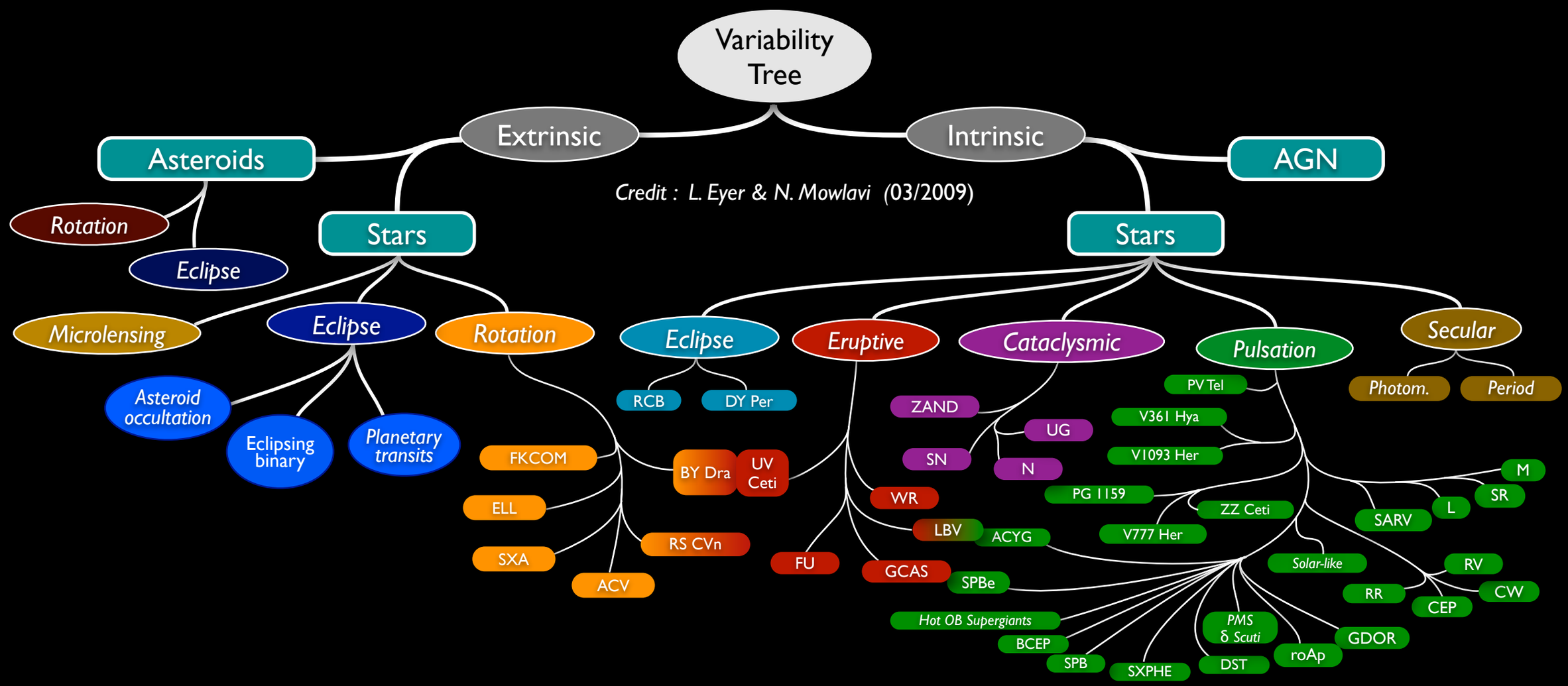
IAU XXIX General Assembly, Hawaii, Division G: Large-Scale Surveys and Variability (DG.6.02)

Gaia challenge



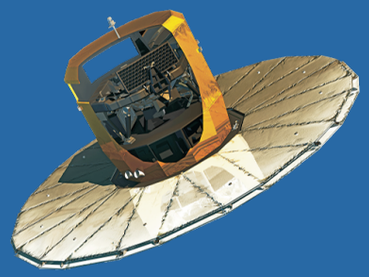
~72 epochs in 5 years

>1,000,000,000 objects, ~10% variable



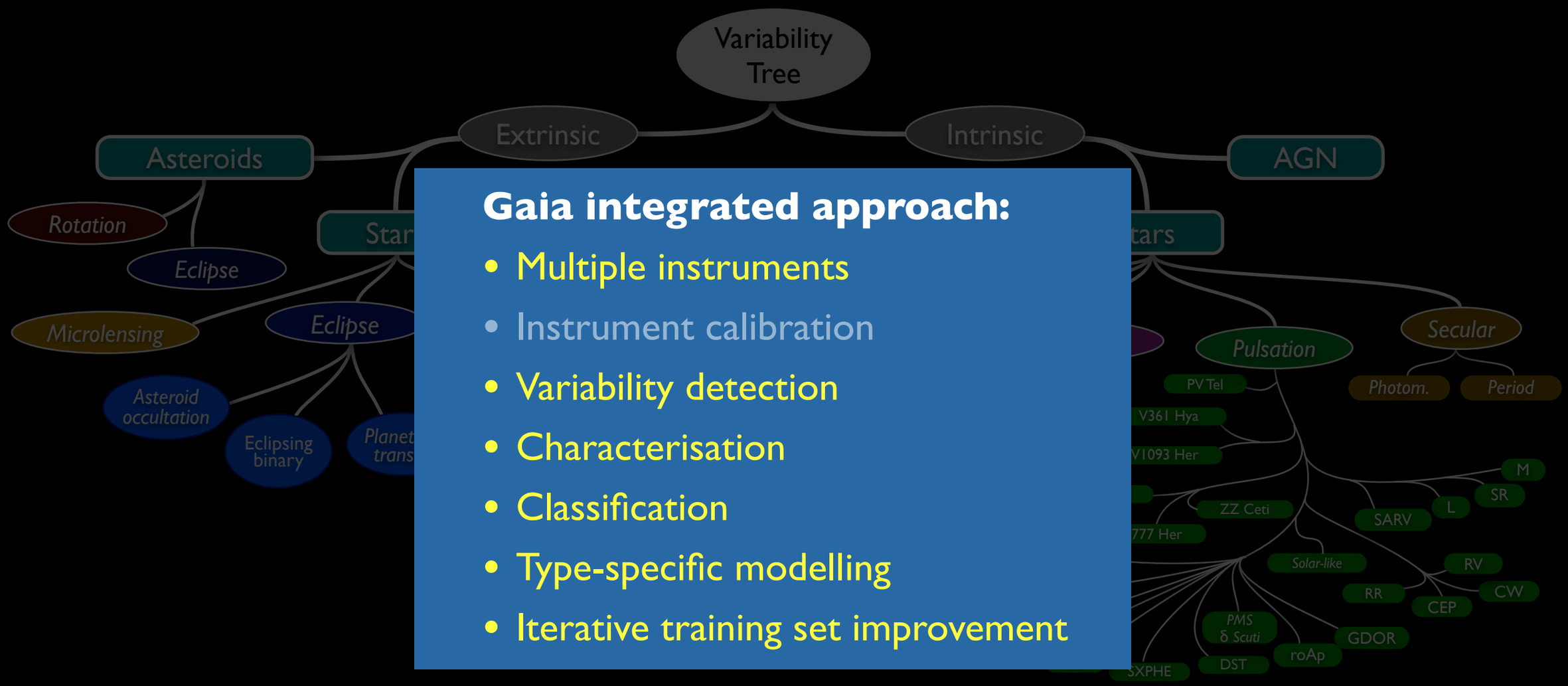
How to automatically detect, classify and characterise?

Gaia challenge



~72 epochs in 5 years

>1,000,000,000 objects, ~10% variable

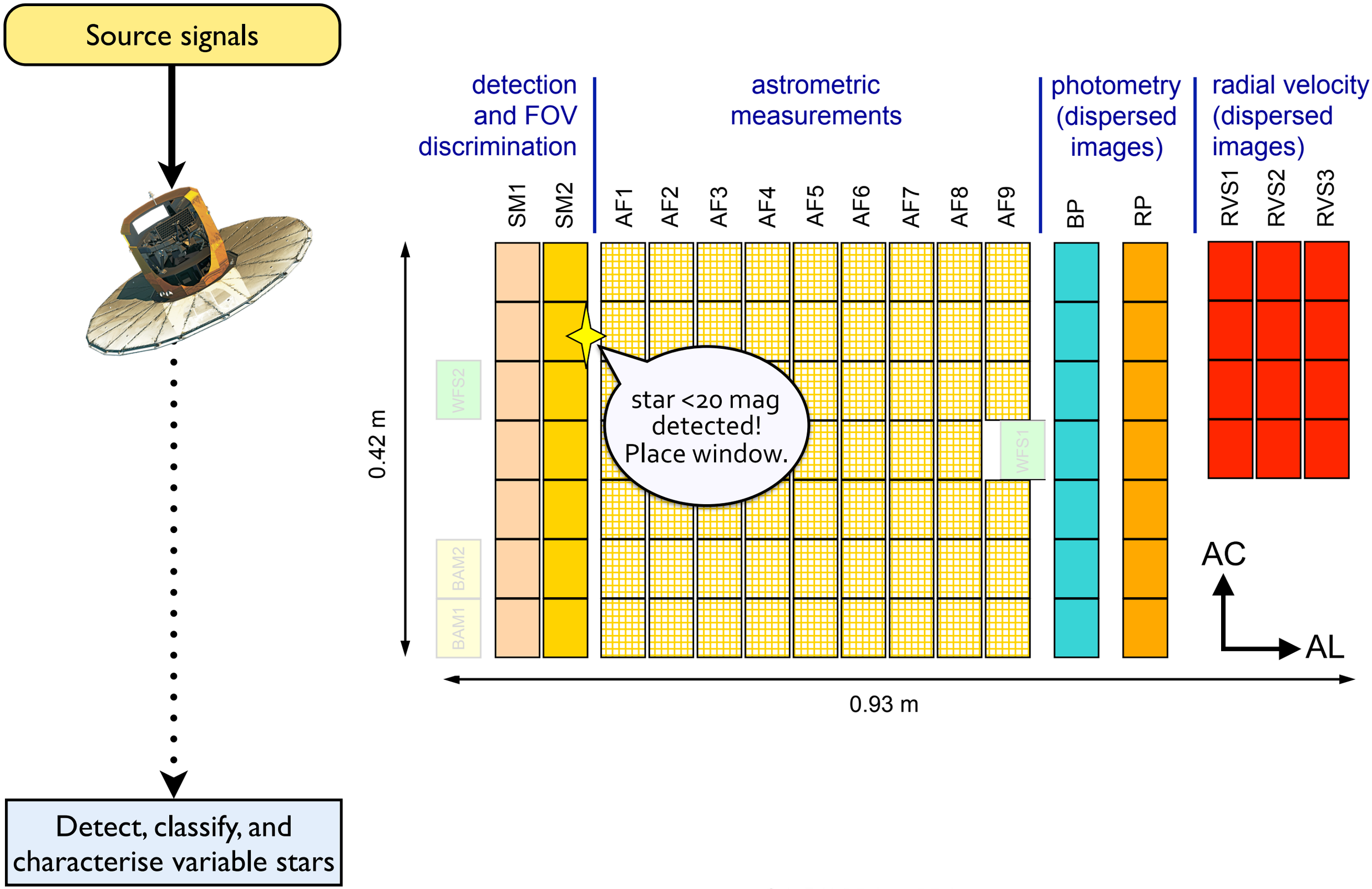


Gaia integrated approach:

- Multiple instruments
- Instrument calibration
- Variability detection
- Characterisation
- Classification
- Type-specific modelling
- Iterative training set improvement

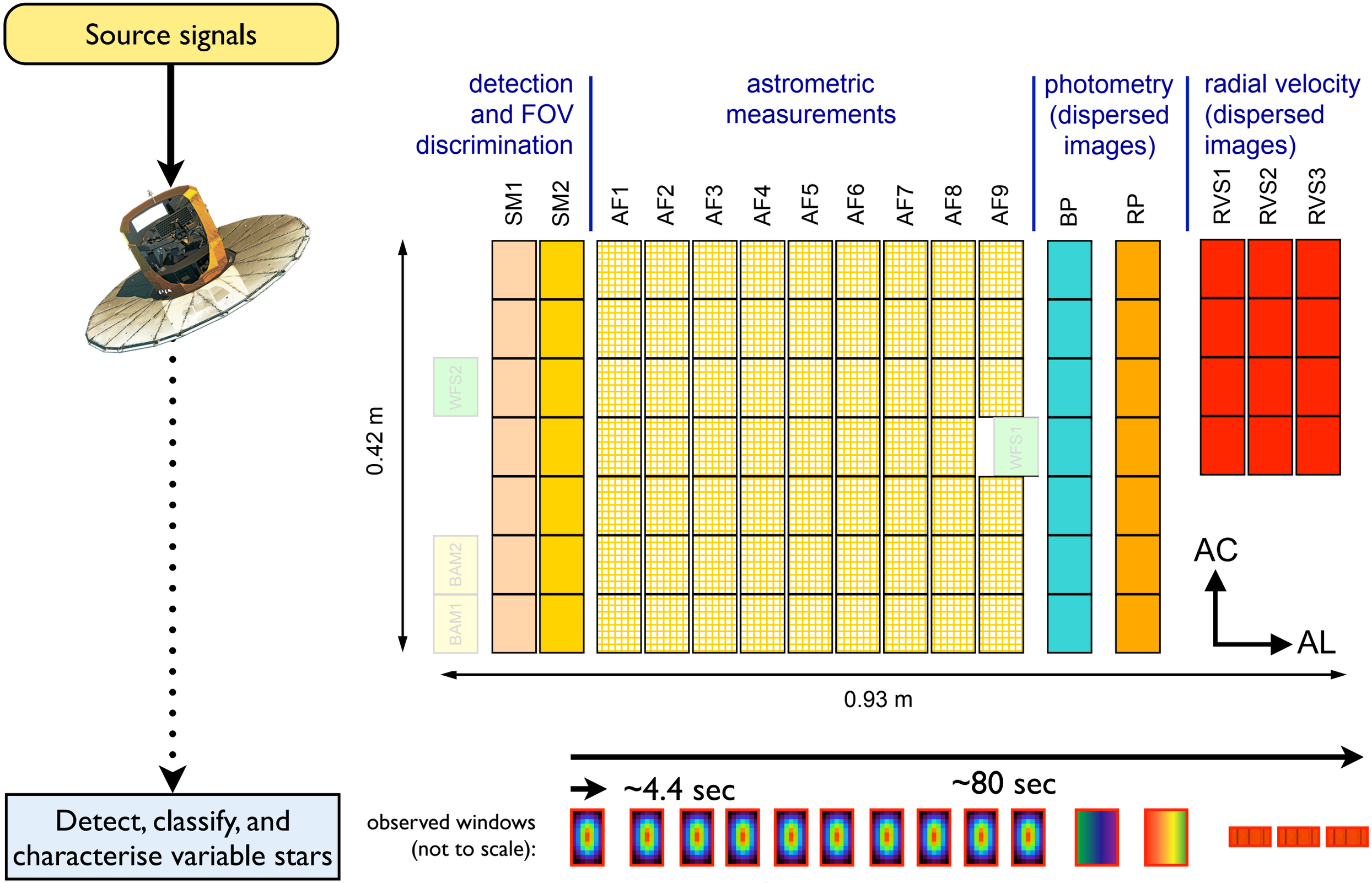
How to automatically detect, classify and characterise?

Multiple instruments



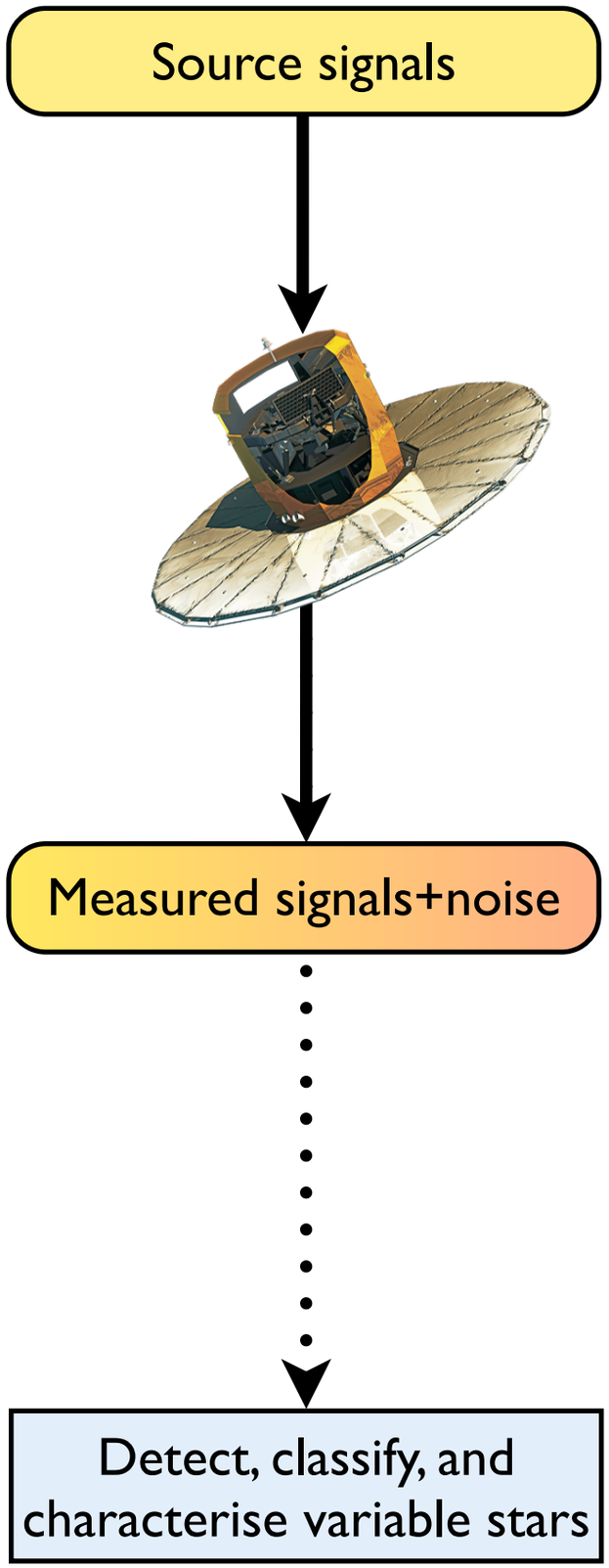
Gaia FoV: 0.7 deg x 0.7 deg
pixel: 0.059”(AL) x 0.177”(AC)

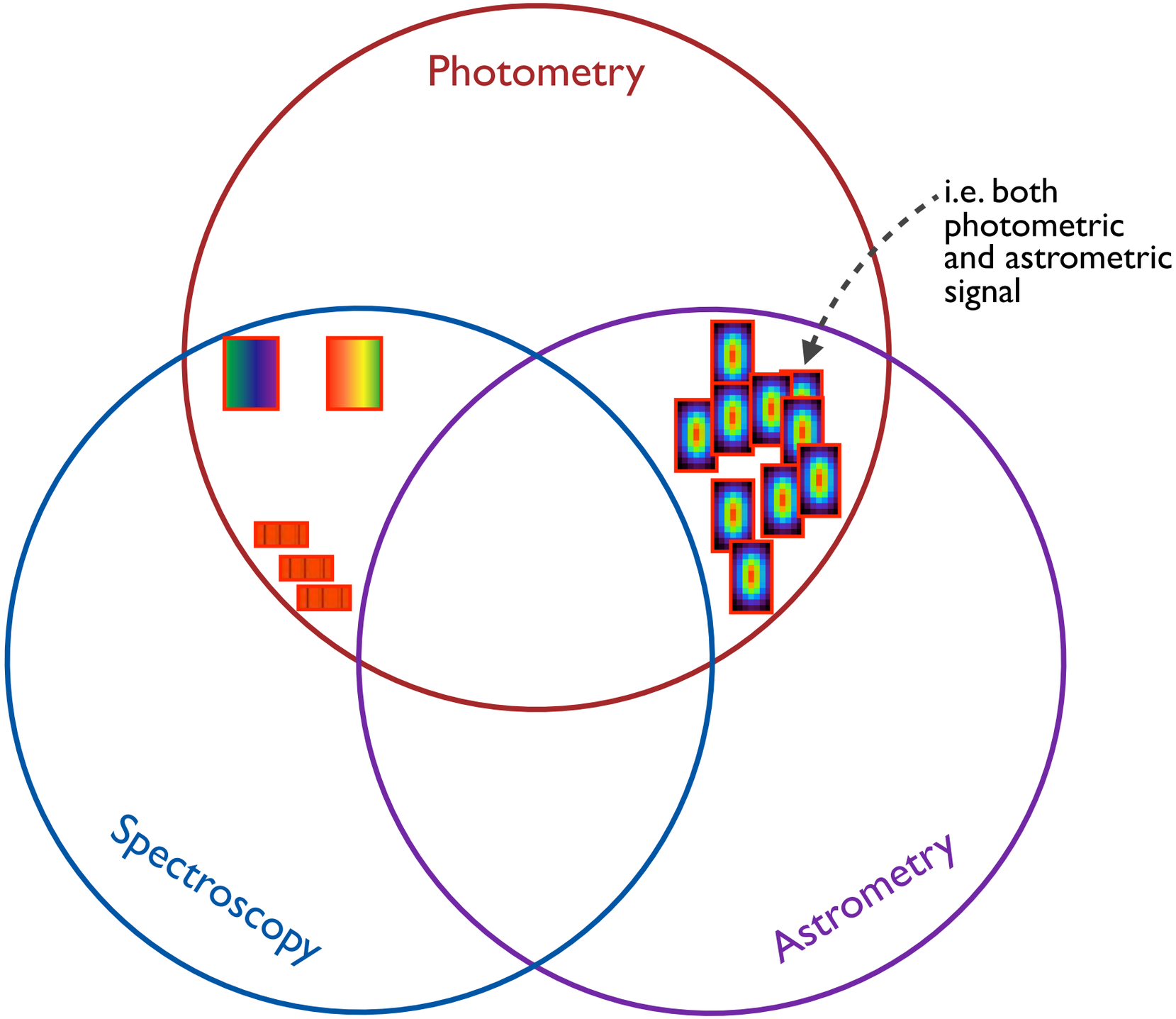
Multiple instruments



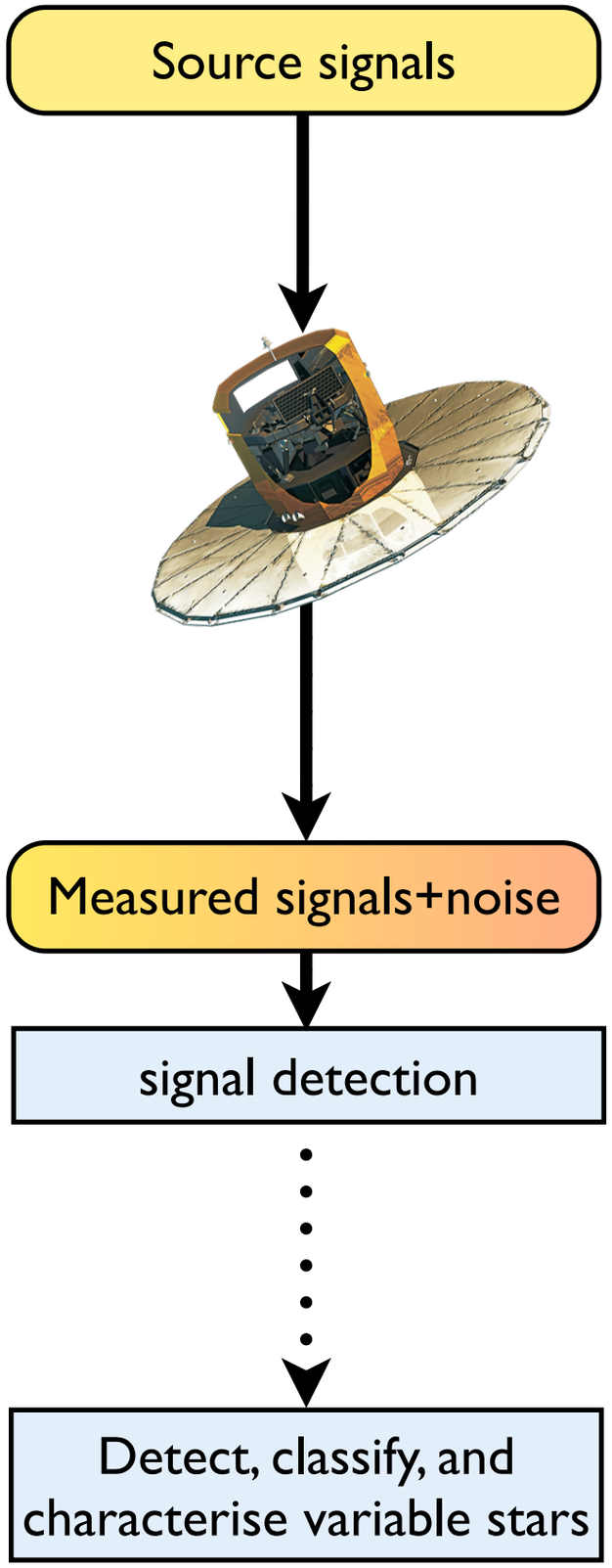
Gaia FoV: 0.7 deg x 0.7 deg
pixel: 0.059”(AL) x 0.177”(AC)

Multiple instruments





Multiple instruments



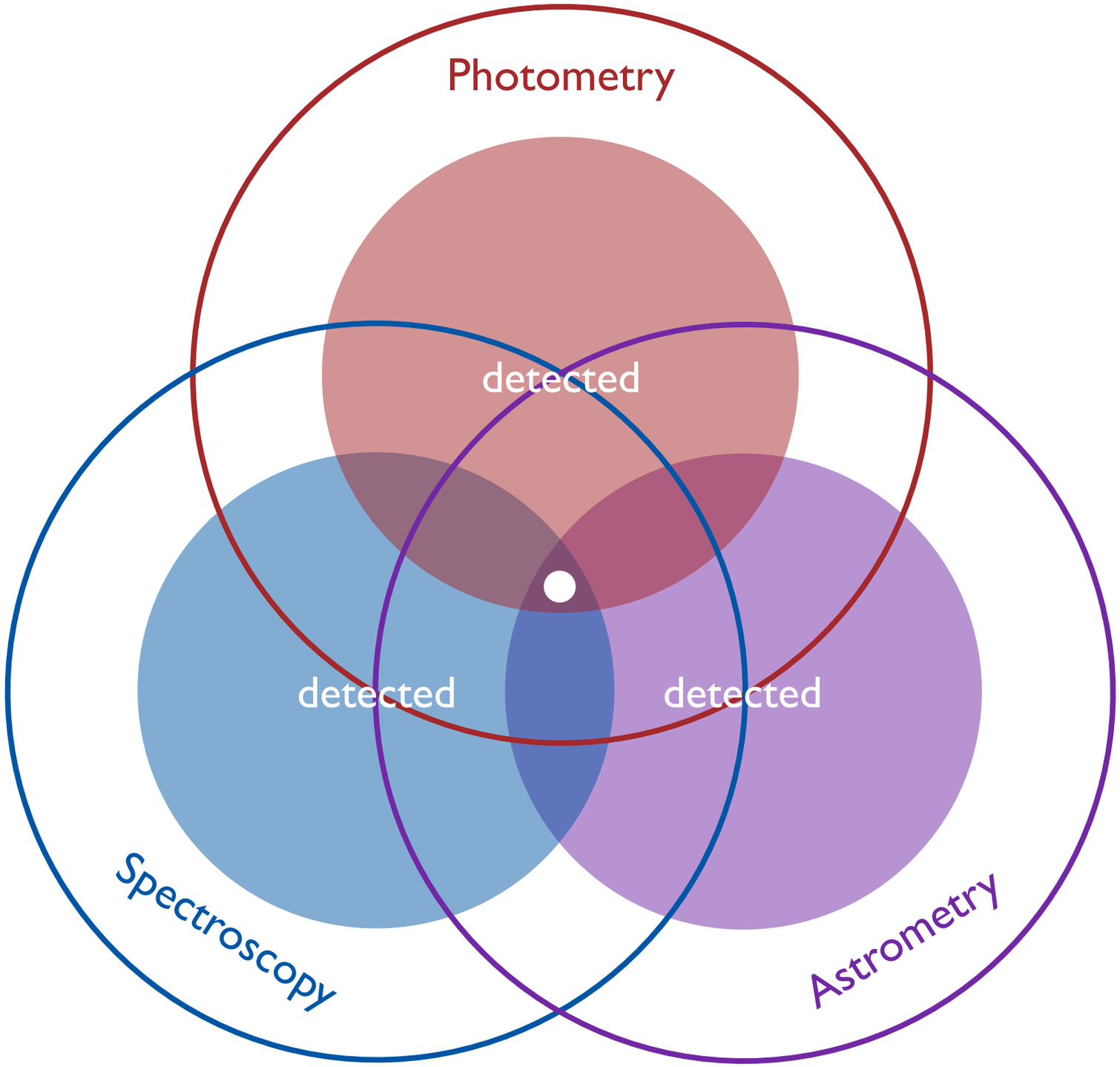
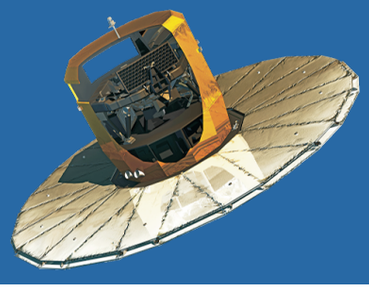


Figure adapted from Eyer & Holl, et al. (2013)

Gaia challenge

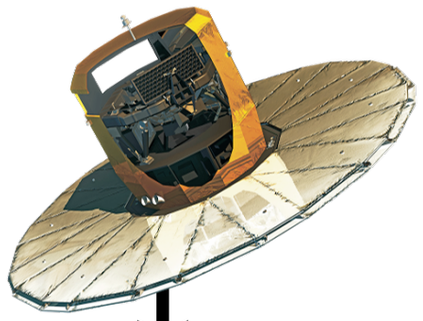


How to automatically detect, classify and characterise 100 million+ variable stars?

Gaia integrated approach:

- Multiple instruments
- Instrument calibration
- **Variability detection**
- **Characterisation**
- **Classification**
- **Type-specific modelling**
- Iterative training set improvement

Variability detection



Measured signals+noise

Variability detection

Classify and characterise variable stars

>1,000,000,000 time series
(spectro)photometric bands, radial velocity, ...

General variability detection

Constancy rejection
based on p-values of e.g.:
Chi^2, Abbe, Kurtosis, Skewness, ...

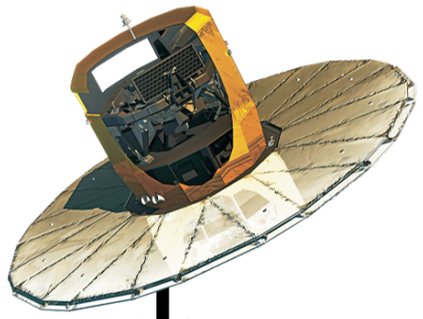
Special variability detection

Signal matching filters

- short timescale [10s, 2h]
  e.g. eclipsing binaries, pulsating stars
- solar-like activity (spots and flares)
- small periodic amplitude
- planet transits

Select for variable processing (~10%)

Signal characterisation



Measured signals+noise

Variability detection

Characterisation

Classify variable stars

>1,000,000,000 time series
(spectro)photometric bands, radial velocity, ...

General variability detection

Special variability detection

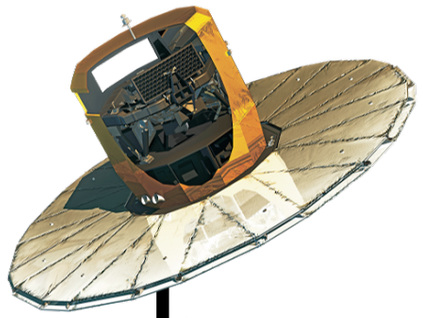
Periodicity / stochastic / ....

- **Periodic model:**
  multi-periodic harmonic Fourier with polynomial (e.g. trend)

Attribute computation

- **Many parameters available:**
  i.e. astrometric, astrophysical, radial velocity (for bright stars), etc.
- **Various attributes, e.g.:**
  absolute magnitude (using parallax),
  BP-RP (colour),
  peak-to-peak amplitude,
  period,
  skewness,
  A2/A1 (ratio of Fourier harm. ampl.),
  etc.

Classification



Measured signals+noise

Variability detection

Characterisation

Classification

>1,000,000,000 time series
(spectro)photometric bands, radial velocity, ...

General variability detection

Special variability detection

Periodicity / stochastic / ....

Attribute computation

Supervised

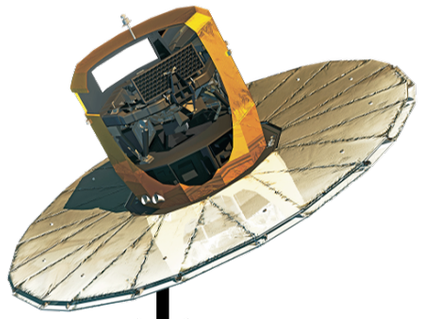
Unsupervised

- **Single and multi-stage:**
  Gaussian Mixture,
  Random Forest,
  Bayesian network, ...
- **Extractors:**
  transients,
  microlensing
- **Meta-classifier:** combination of classifiers for best performance.

Identify/confirm new classes

Hierarchical Modal Association Clustering (HMAC)

Type-specific modelling



Measured signals+noise

Variability detection

Characterisation

Classification

Detailed modelling

>1,000,000,000 time series
(spectro)photometric bands, radial velocity, ...

General variability detection

Special variability detection

Periodicity / stochastic / ....

Attribute computation

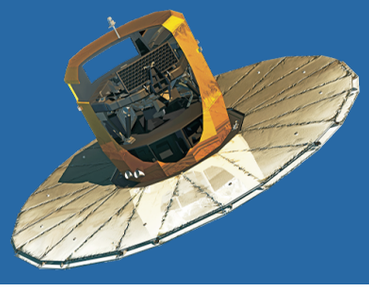
Supervised

Unsupervised

Specific Object Modelling

- **RR-Lyrae**
- **Cepheids**
- **Eclipsing binaries**
- **Long period Var.**
- **Cataclysmic**
- **BE stars**
- **Pre main sequence**
- **Rapid phase**
- **Microlensing**
- **AGN**
- **Planetary transit**
- **Solar like oscillators**
- **Flare stars**

Gaia challenge



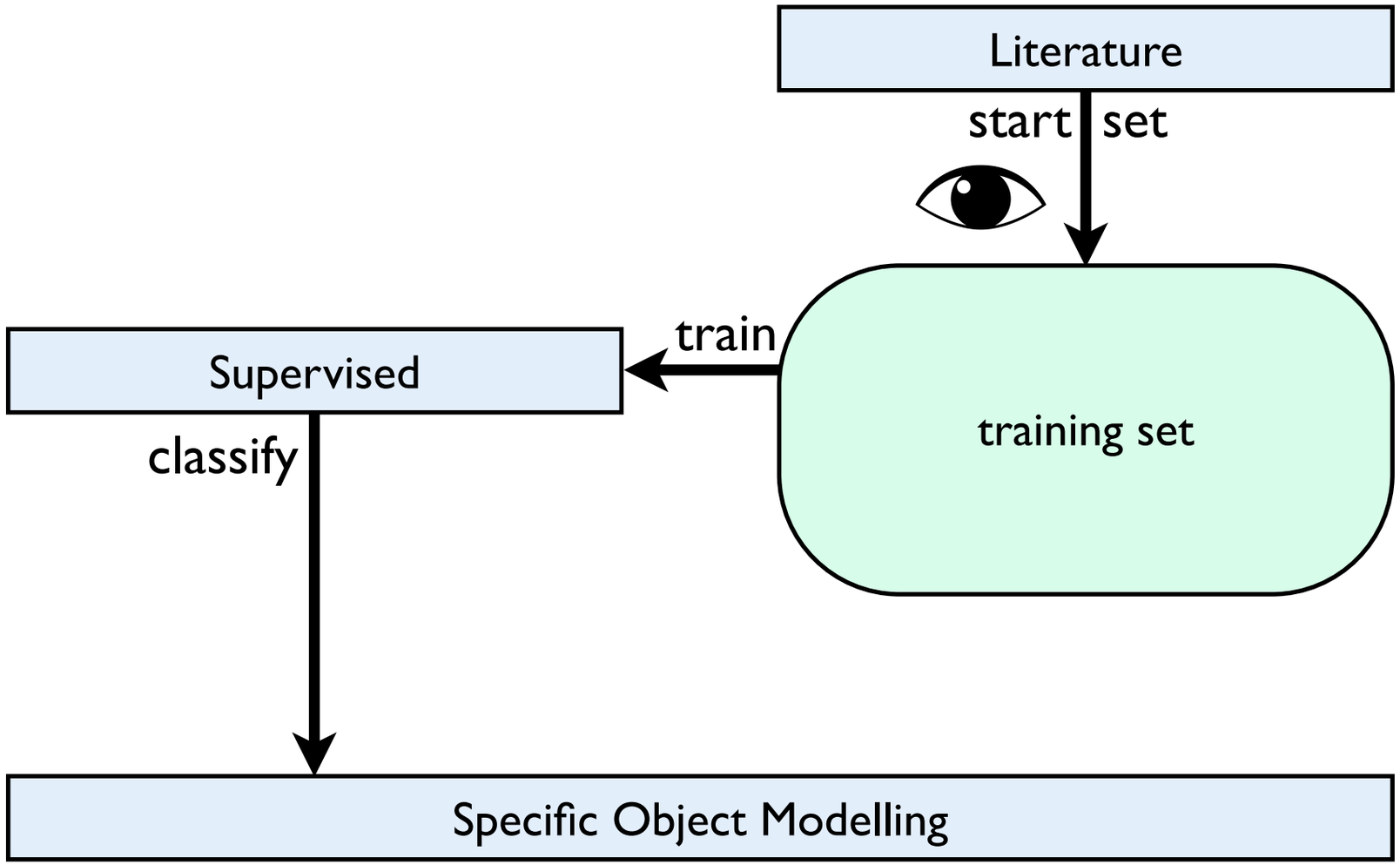
How to automatically detect, classify and characterise 100 million+ variable stars?

Gaia integrated approach:

- Multiple instruments
- Instrument calibration
- Variability detection
- Characterisation
- Classification
- Type-specific modelling
- **Iterative training set improvement**

Iterative training set improvement

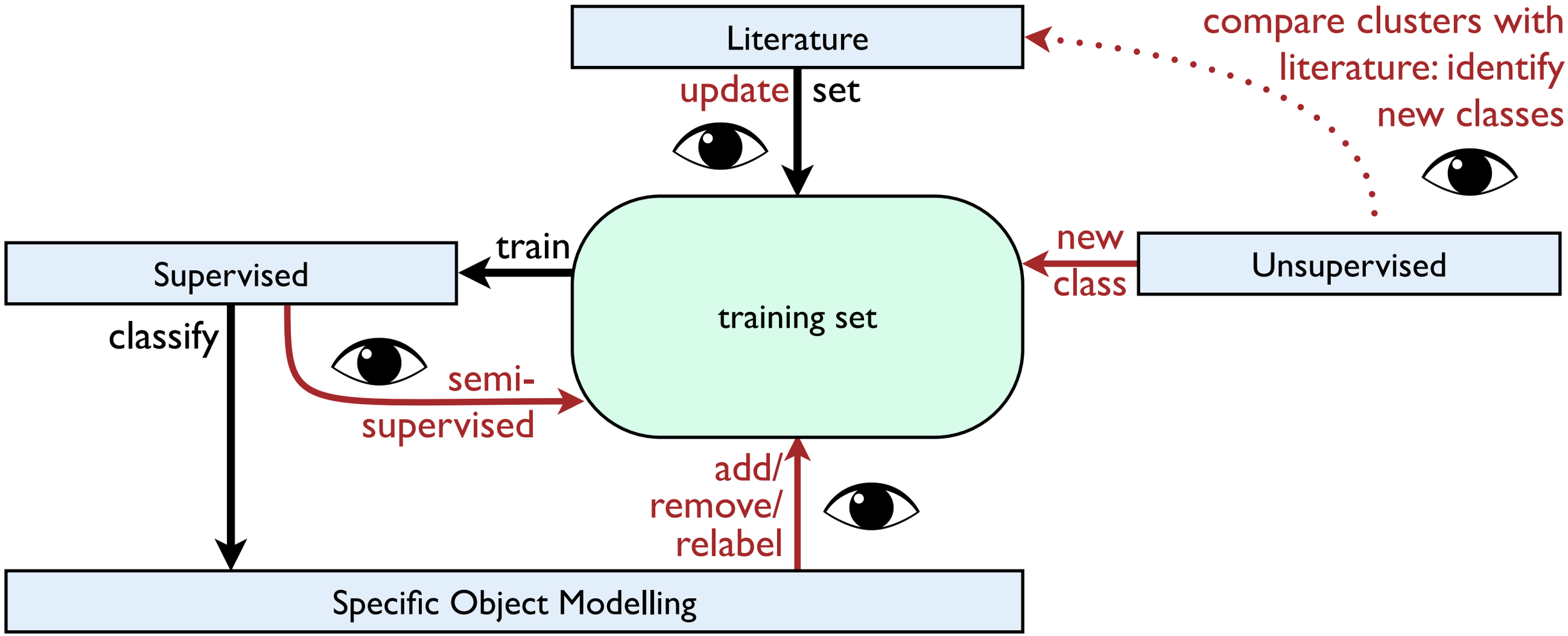
First iteration:



= requires visual inspection/confirmation

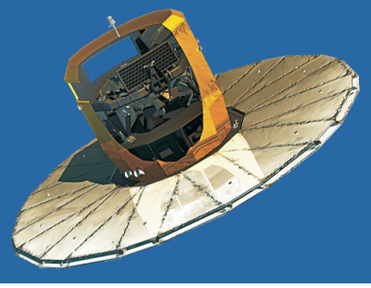
Iterative training set improvement

Further iterations:

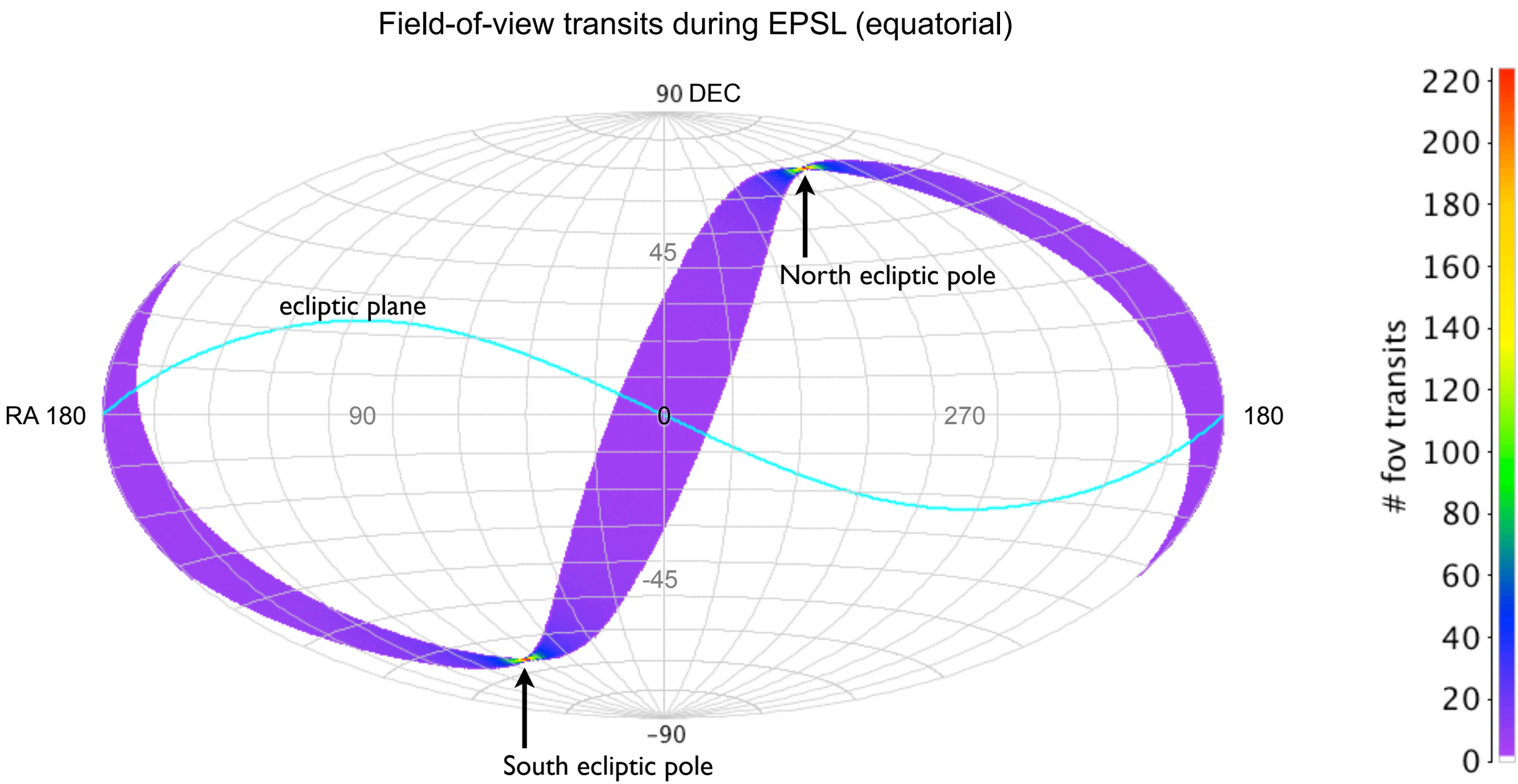


= requires visual inspection/confirmation

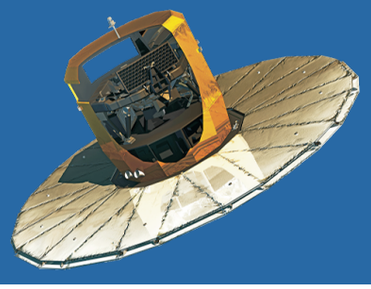
First Gaia data!



28 day Ecliptic Pole Scanning data (summer 2014)

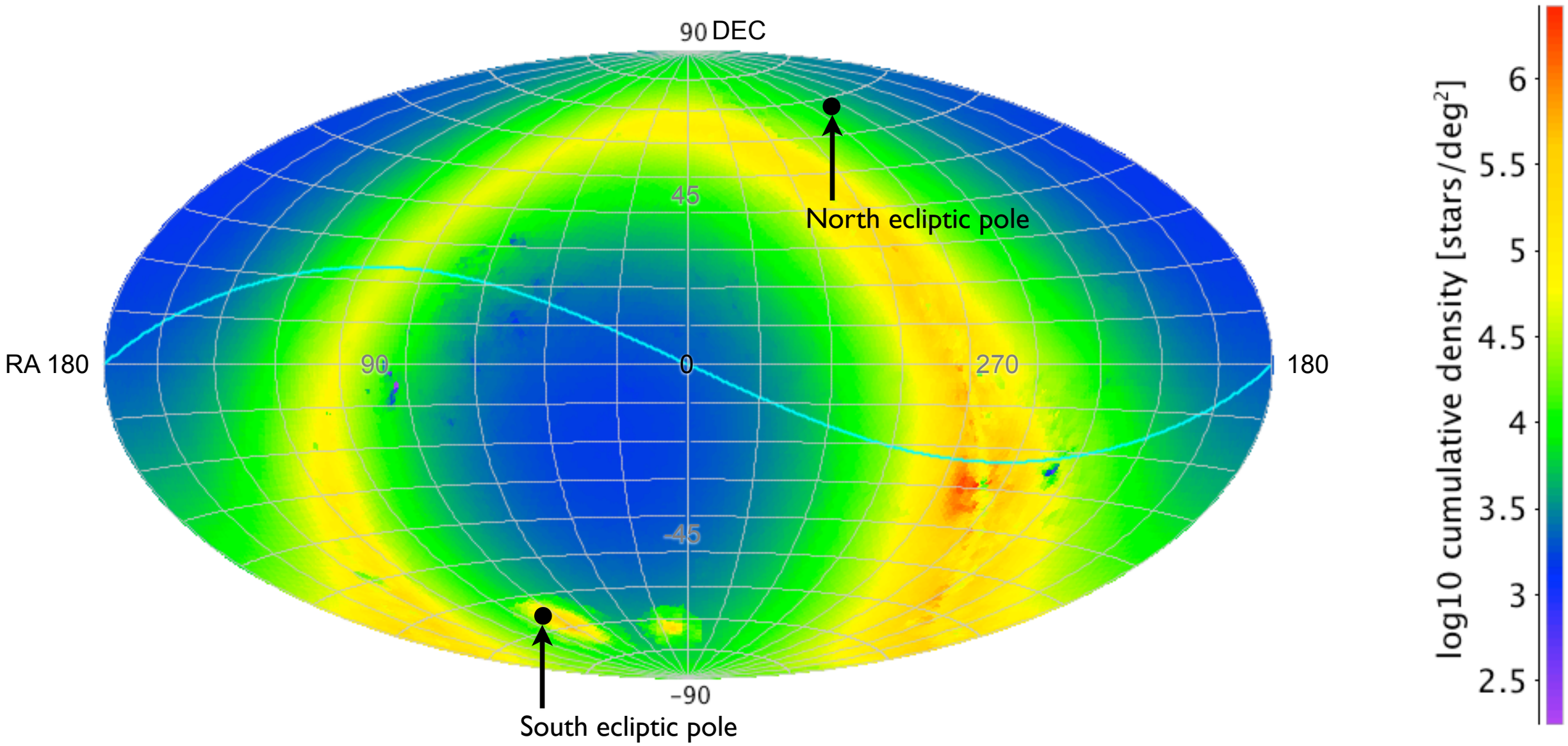


First Gaia data!

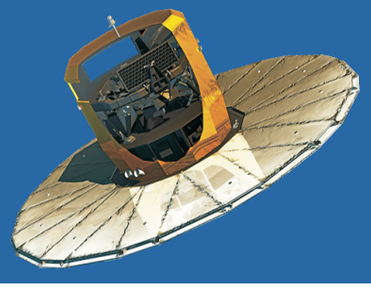


28 day Ecliptic Pole Scanning data (summer 2014)

Number density of stars up to magnitude 20

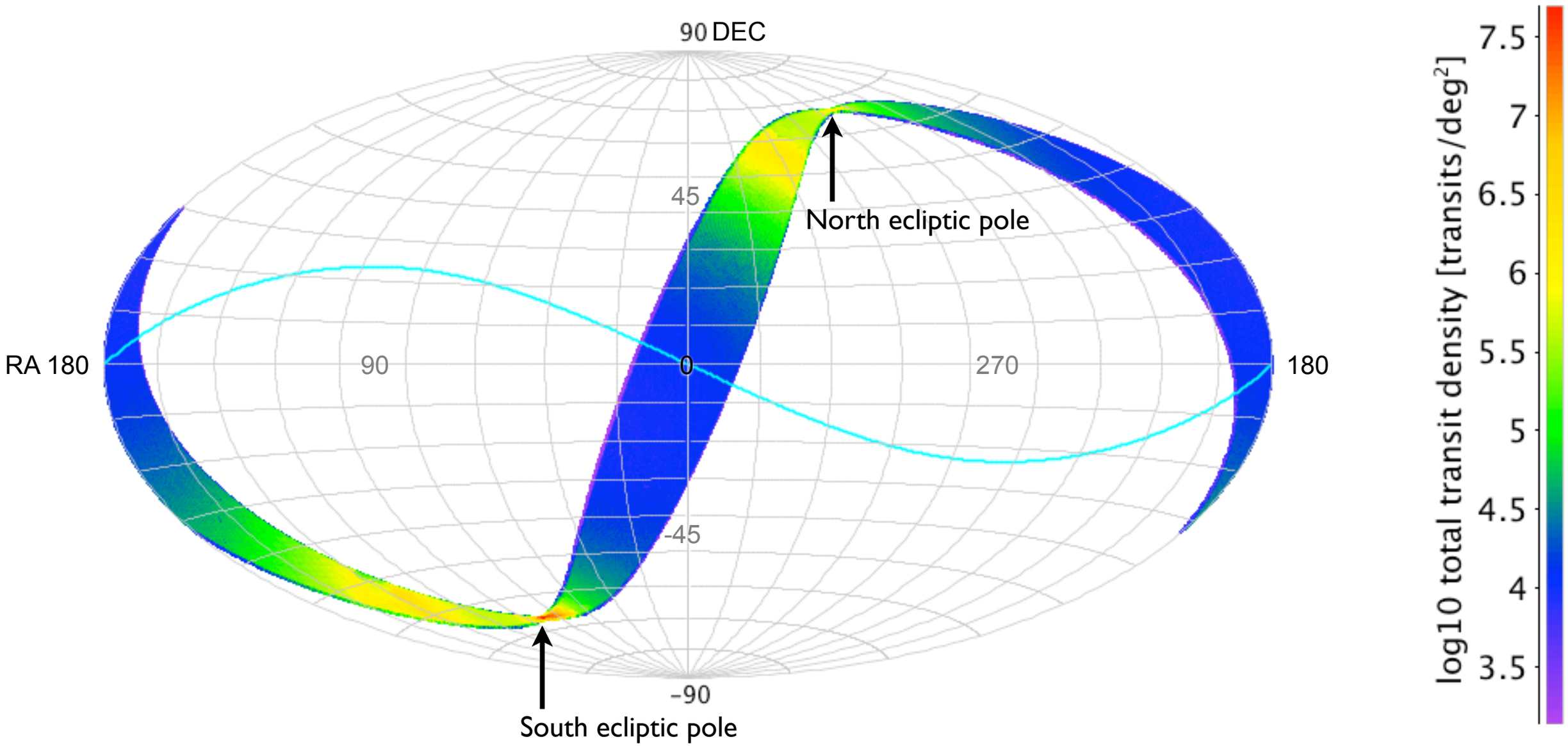


First Gaia data!



28 day Ecliptic Pole Scanning data (summer 2014)

Number density of EPSL measurements up to magnitude 20



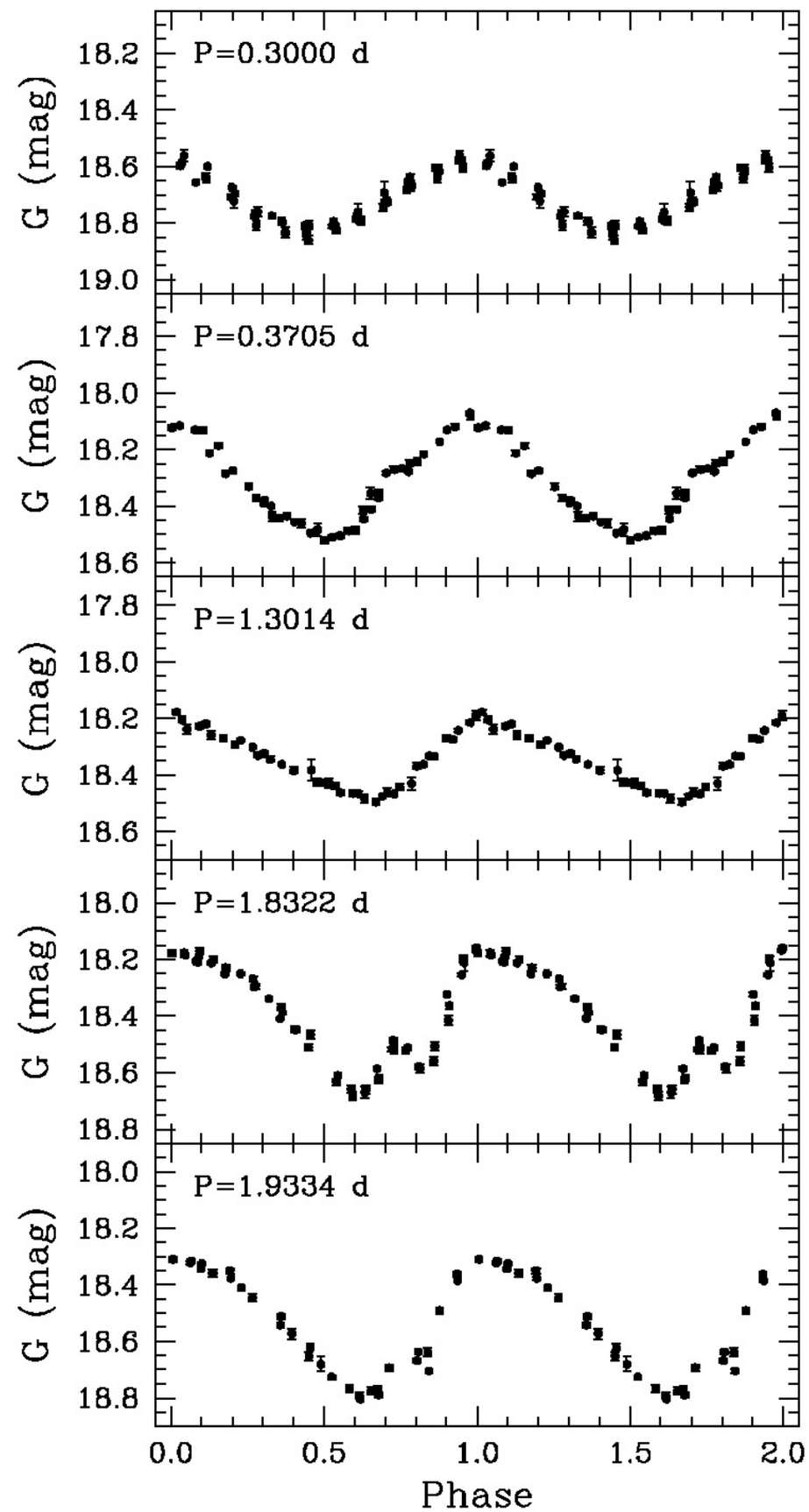
EPSL data examples

Gaia: G-band

Short period/faint magnitude Cepheids in the Large Magellanic Cloud

Gaia Image of the week:
http://www.cosmos.esa.int/web/gaia/iow_20150528

Credits: ESA/Gaia/DPAC/CU5/DPCI/CU7/INAF-OABo/INAF-OACn Gisella Clementini, Vincenzo Ripepi, Silvio Leccia, Laurent Eyer, Lorenzo Rimoldini, Isabelle Lecoeur-Taibi, Nami Mowlavi, Dafydd Evans, Geneva CU7/DPCG and the whole CU7 team. The photometric data reduction was done with the PhotPipe pipeline at DPCI; processing data were received from the IDT pipeline at DPCE.

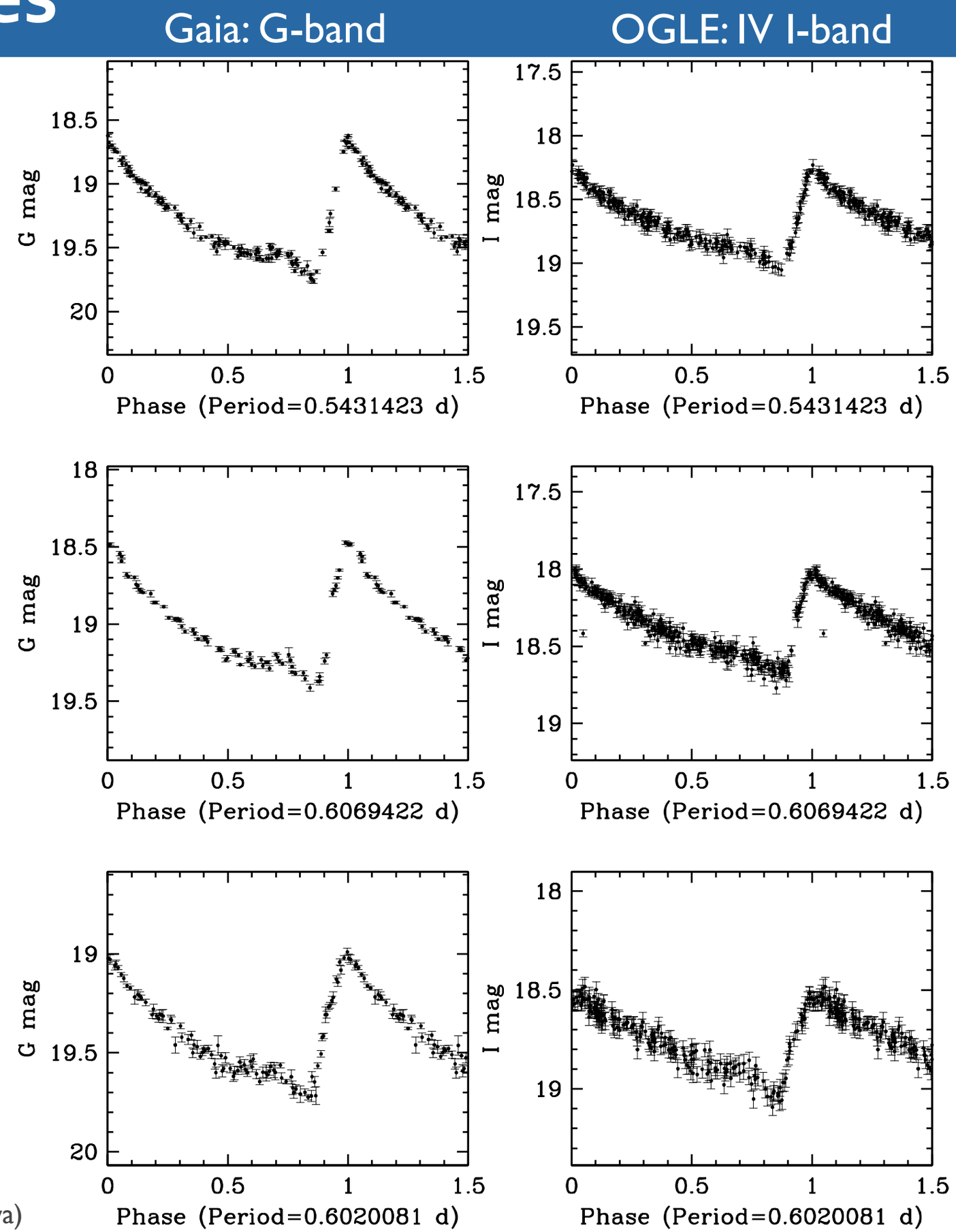


EPSL data examples

RR-Lyrae stars
in the Large Magellanic Cloud

Gaia Image of the week:
http://www.cosmos.esa.int/web/gaia/iow_20150305

Credits: ESA/Gaia/DPAC/CU5/CU7/INAF-OABo, Gisella Clementini, Dafydd Evans, Laurent Eyer, Krzysztof Nienartowicz, Lorenzo Rimoldini and the Geneva CU7/DPCG and CU7/INAF-OACN teams.



Gaia data releases