Gaia, the Galaxy in One Petabyte

Carme Jordi

University of Barcelona, Spain

gaia

UNIVERSITAT DE BARCELONA

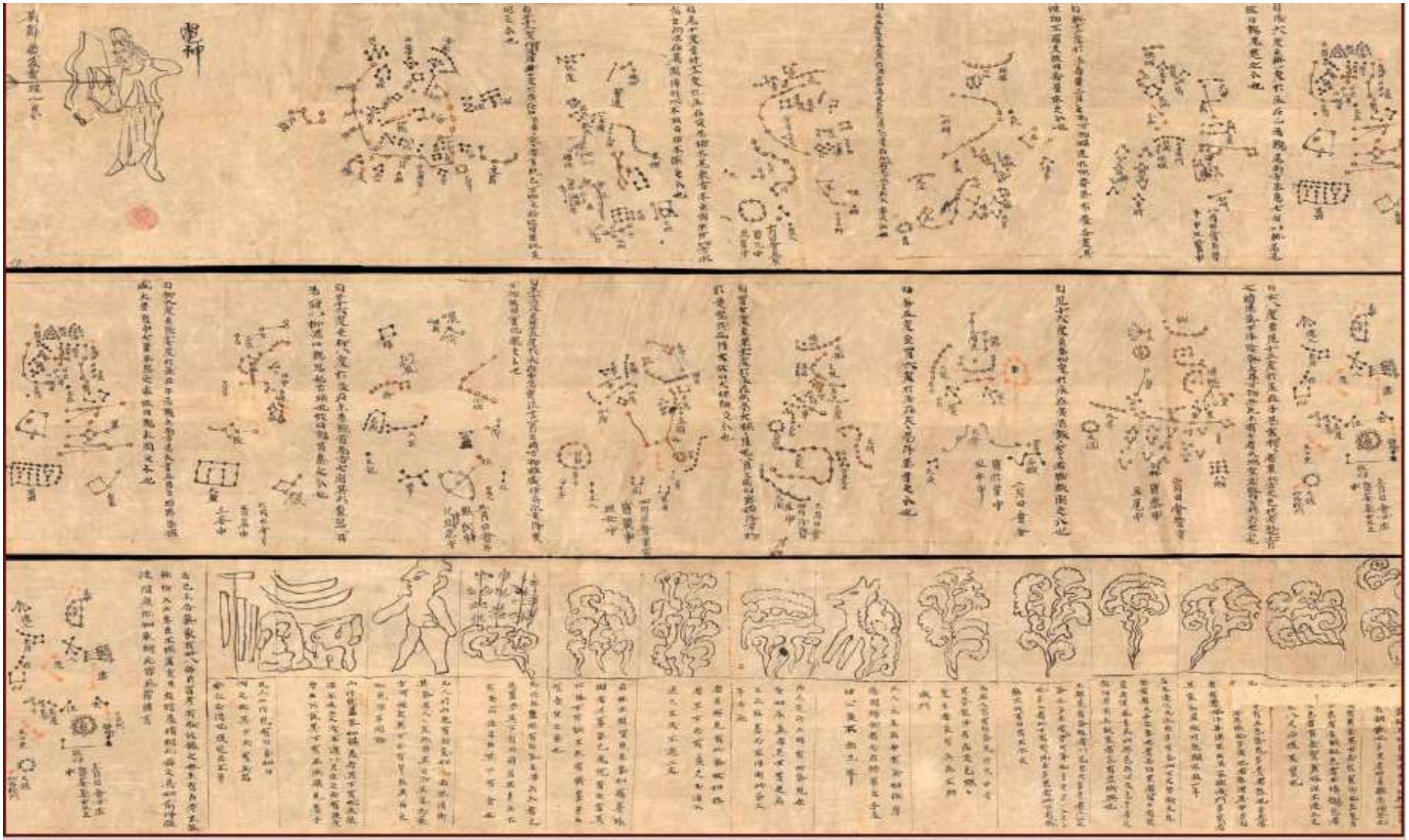
IEEC

esa

Gaia DPAC

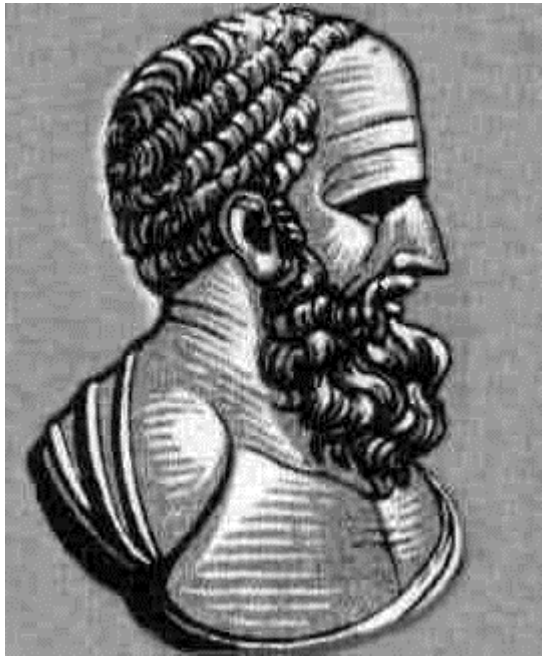

How many stars are there ?

Star catalogues



Oldest known sky map of the northern hemisphere (Tang dynasty, China 649-684). Dunhuang manuscripts (http://idp.bl.uk)

Star catalogues



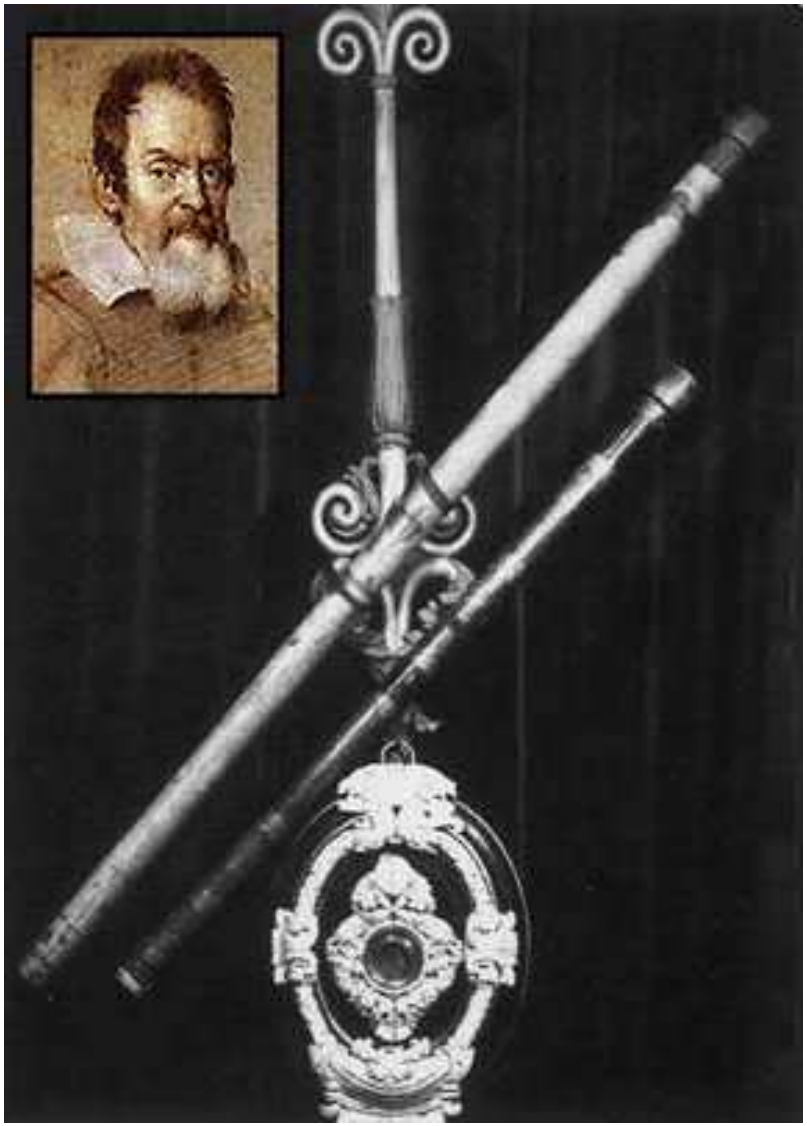
Hipparchus at about 150 BCE

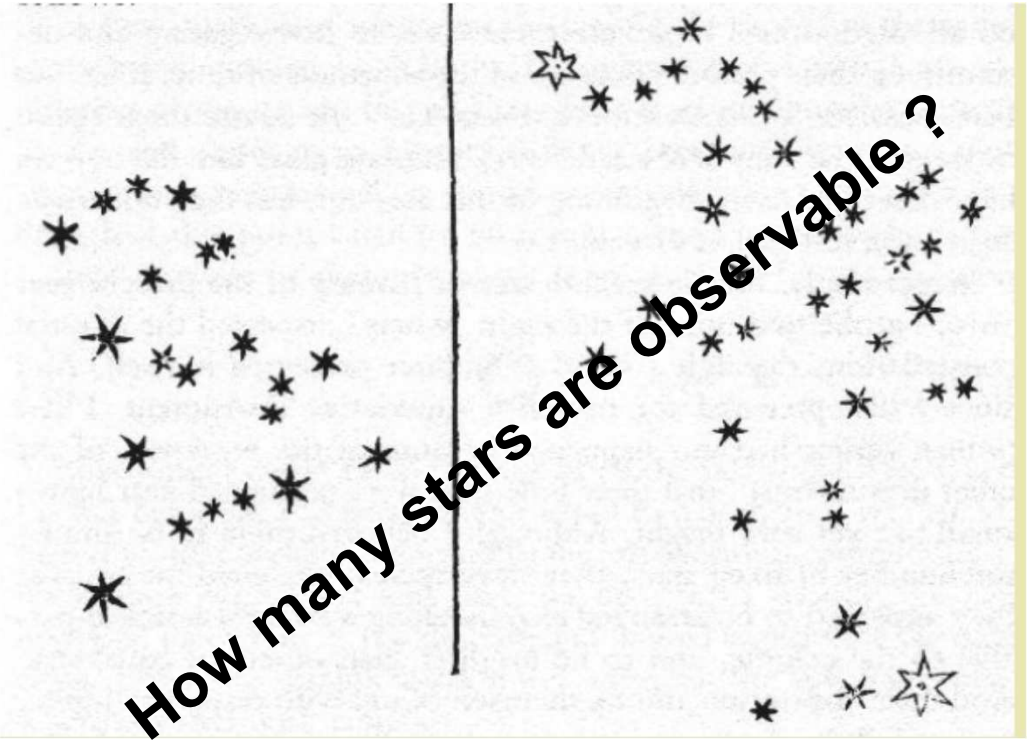
the first comprehensive catalogue of the western world

1 000 stars
positions on the sky
magnitude

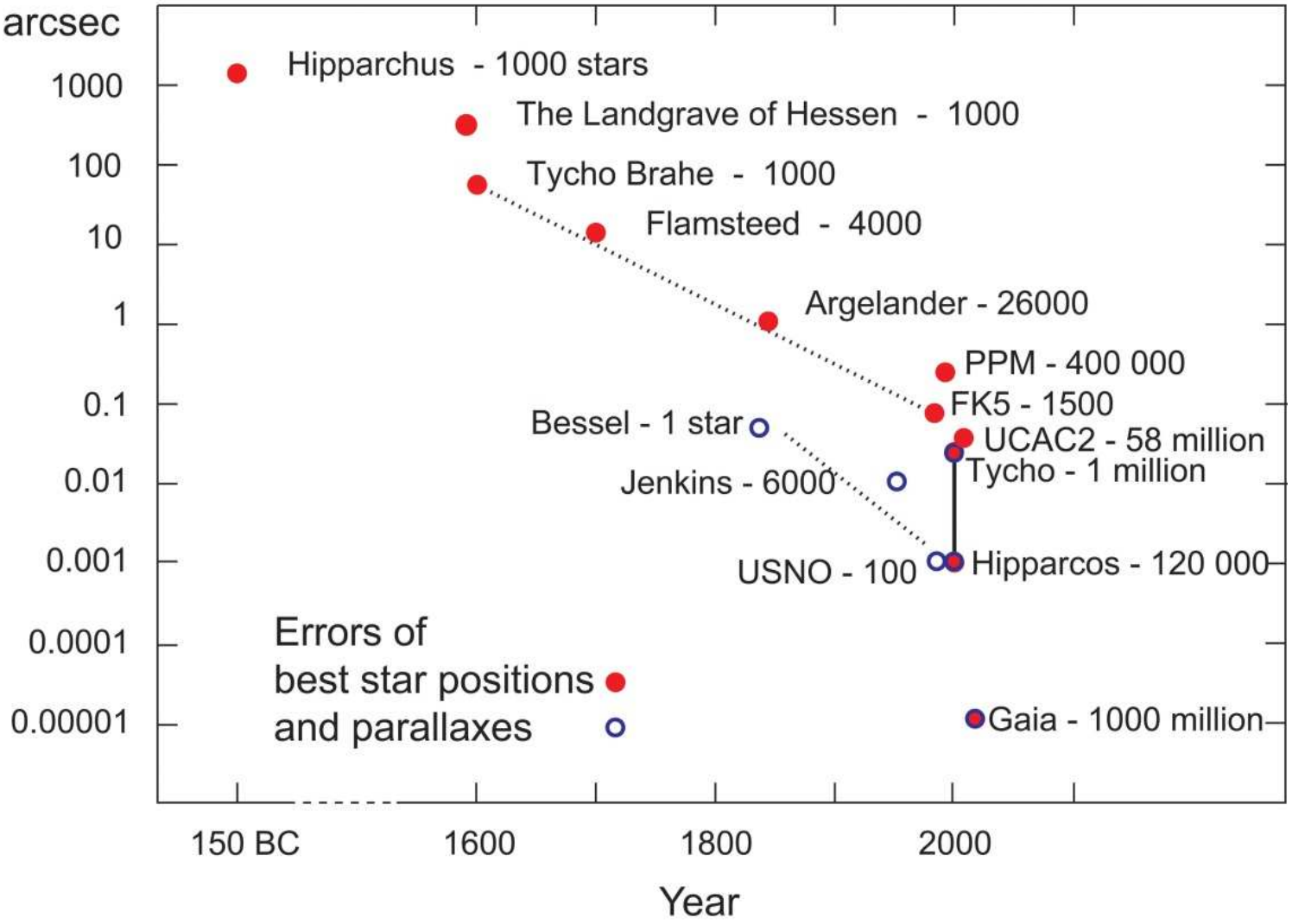
Tycho Brahe, XVI century

Galileo Galilei and the telescope



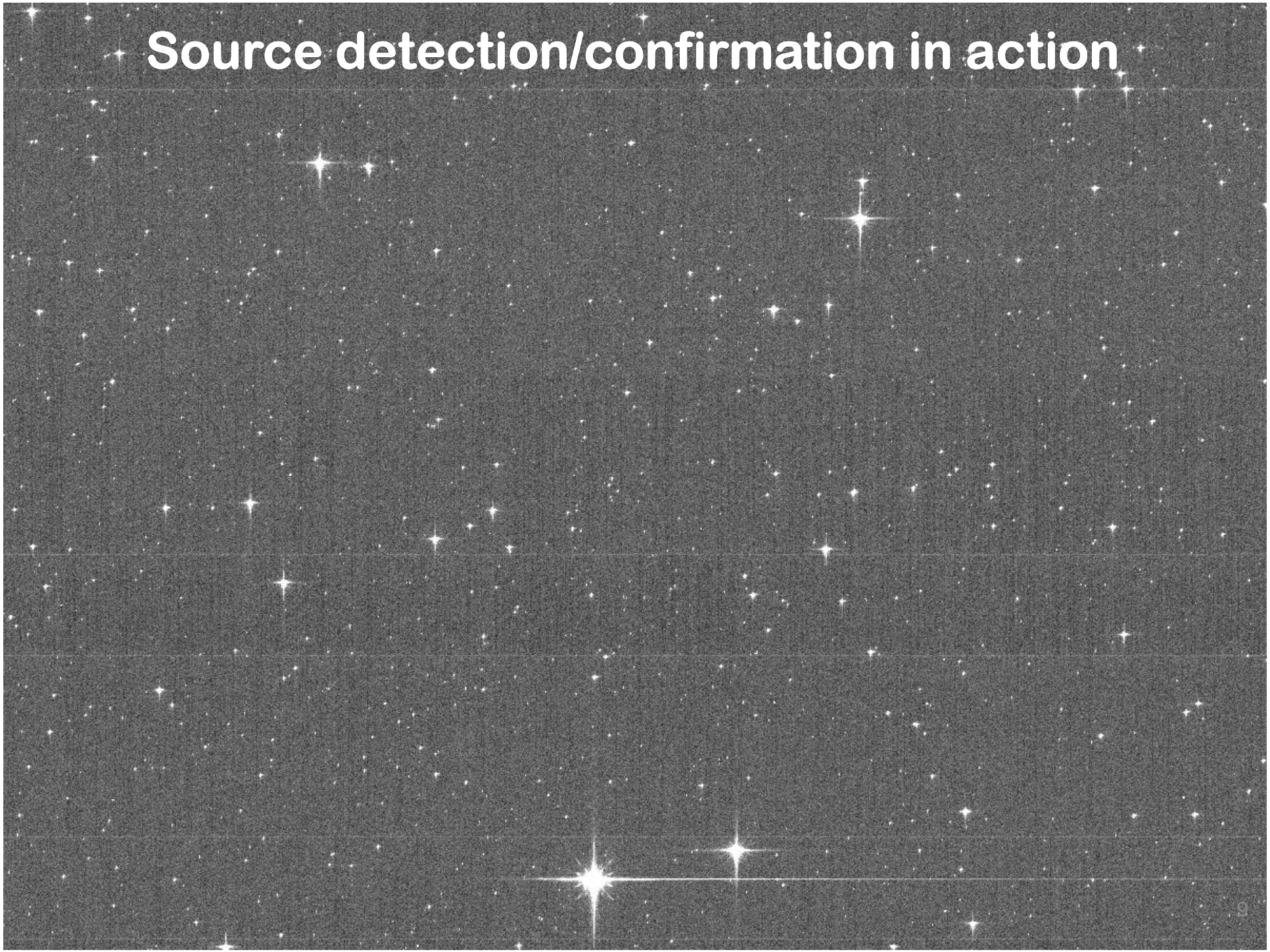


Star catalogues

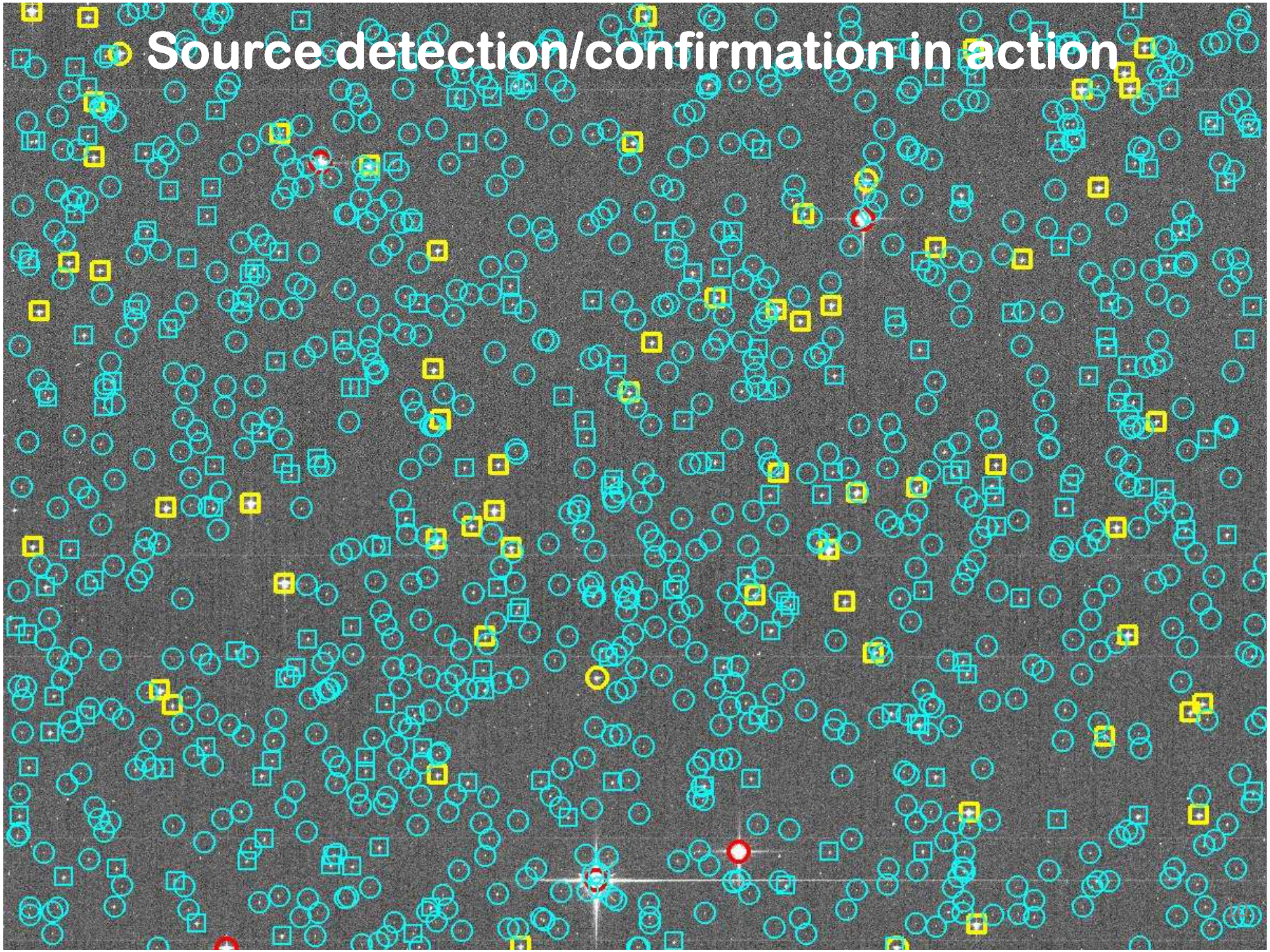


How does Gaia measures the sky ?

Source detection/confirmation in action

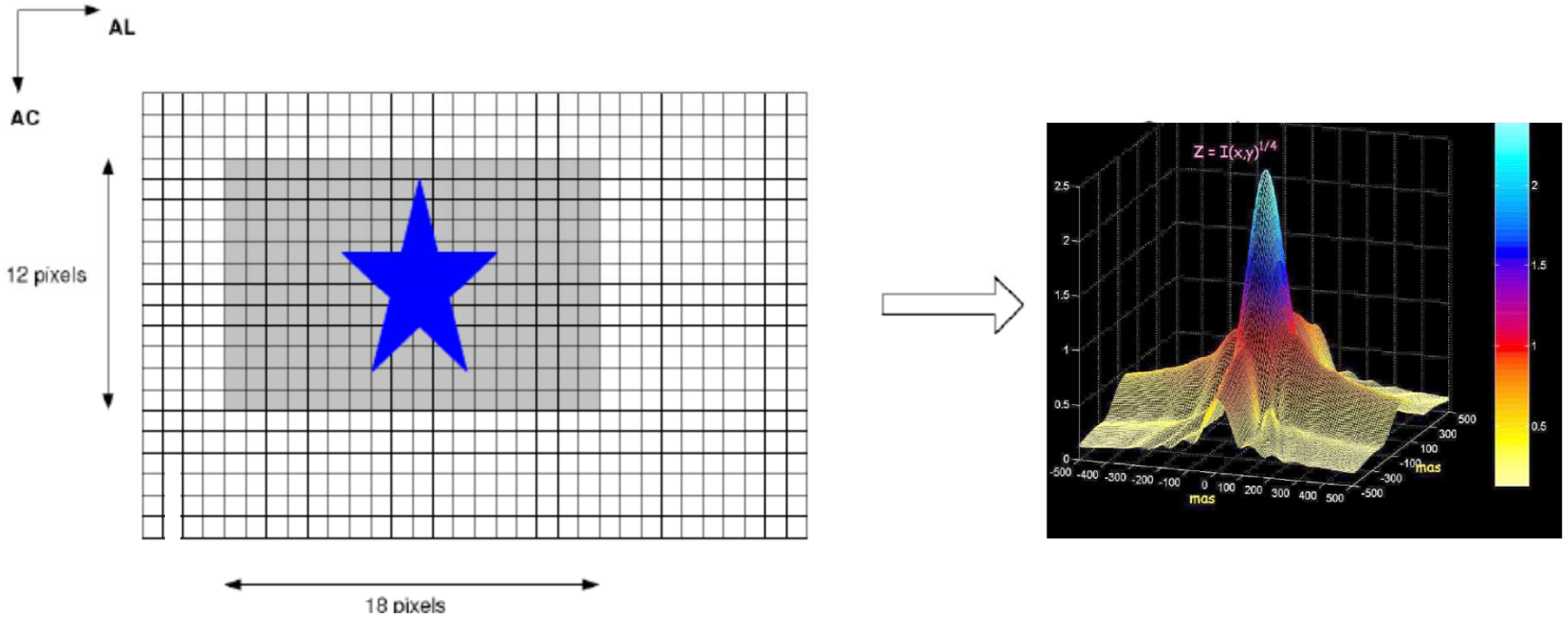


Source detection/confirmation in action



Data reception

Complexity of the data



Photometry

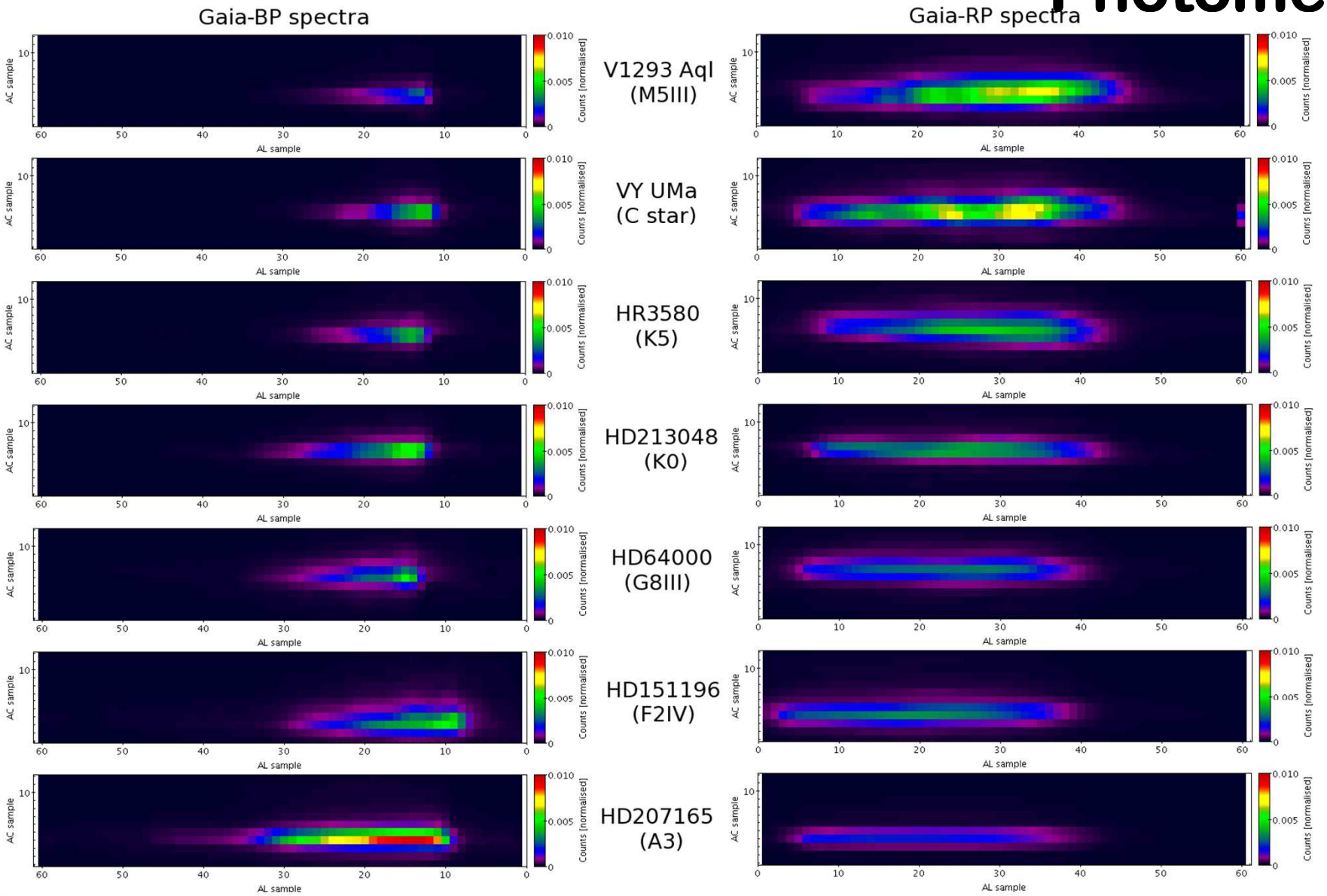
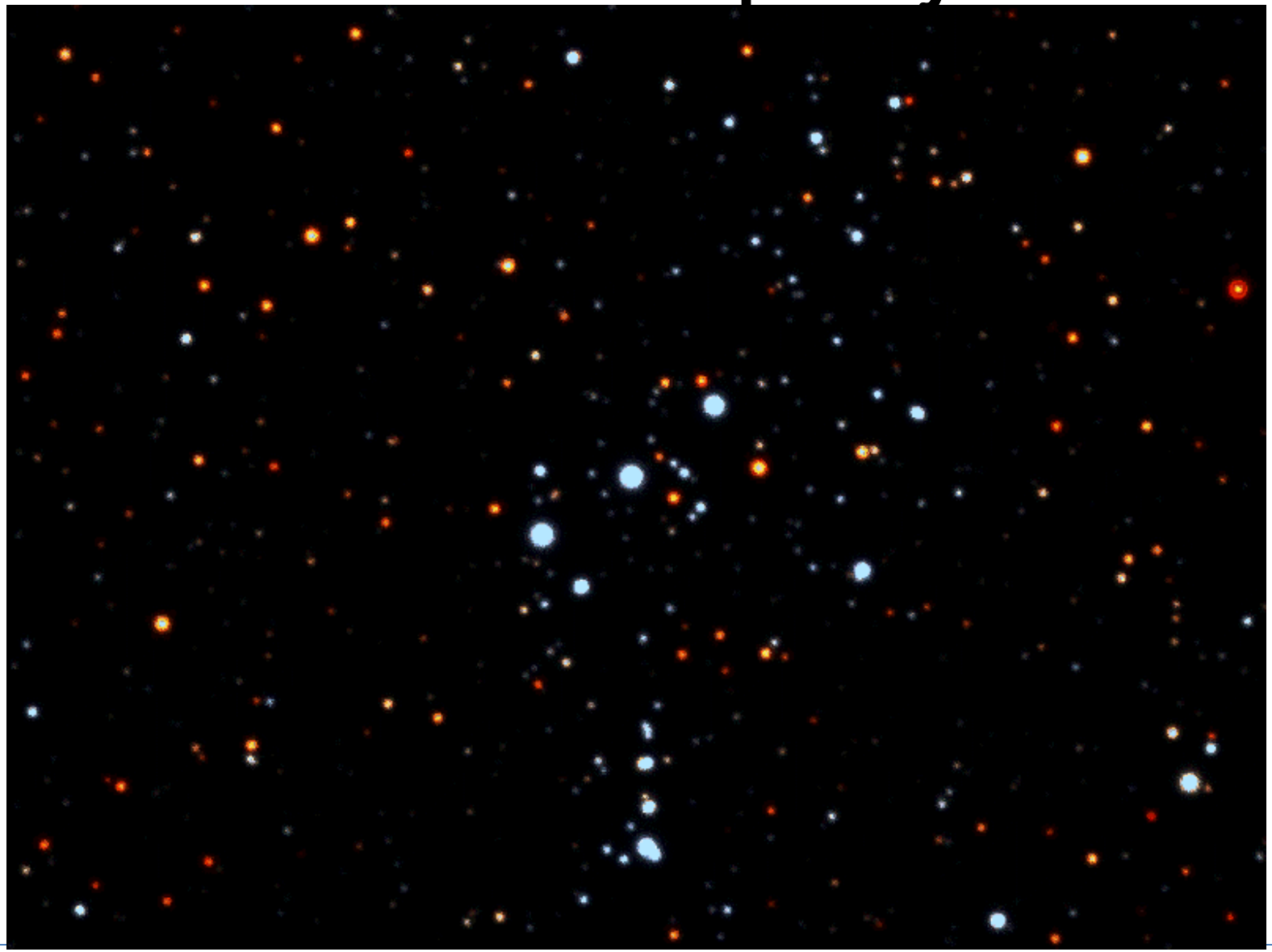


Figure by JM Carrasco

Complexity of the data



Spectroscopy

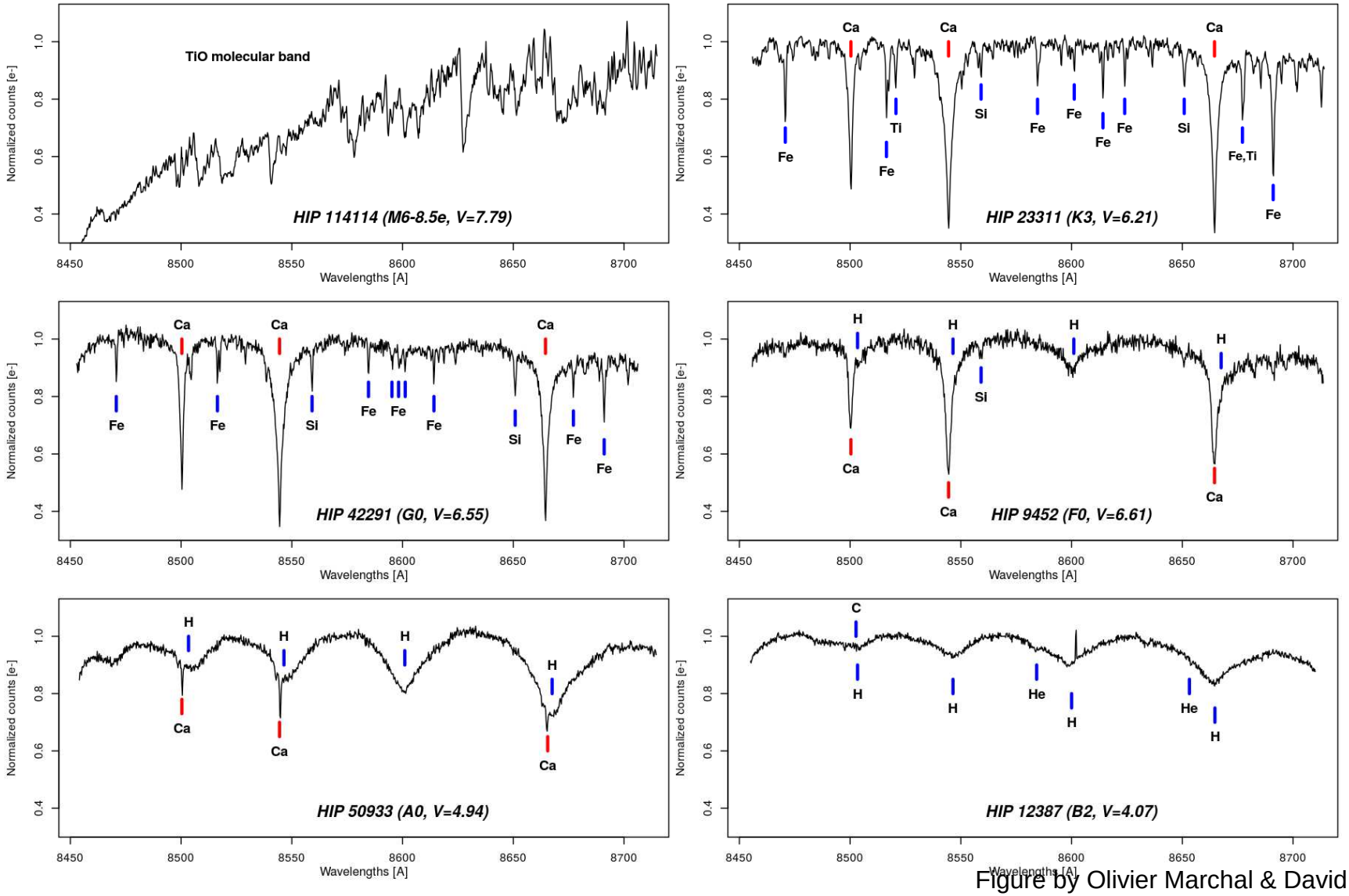
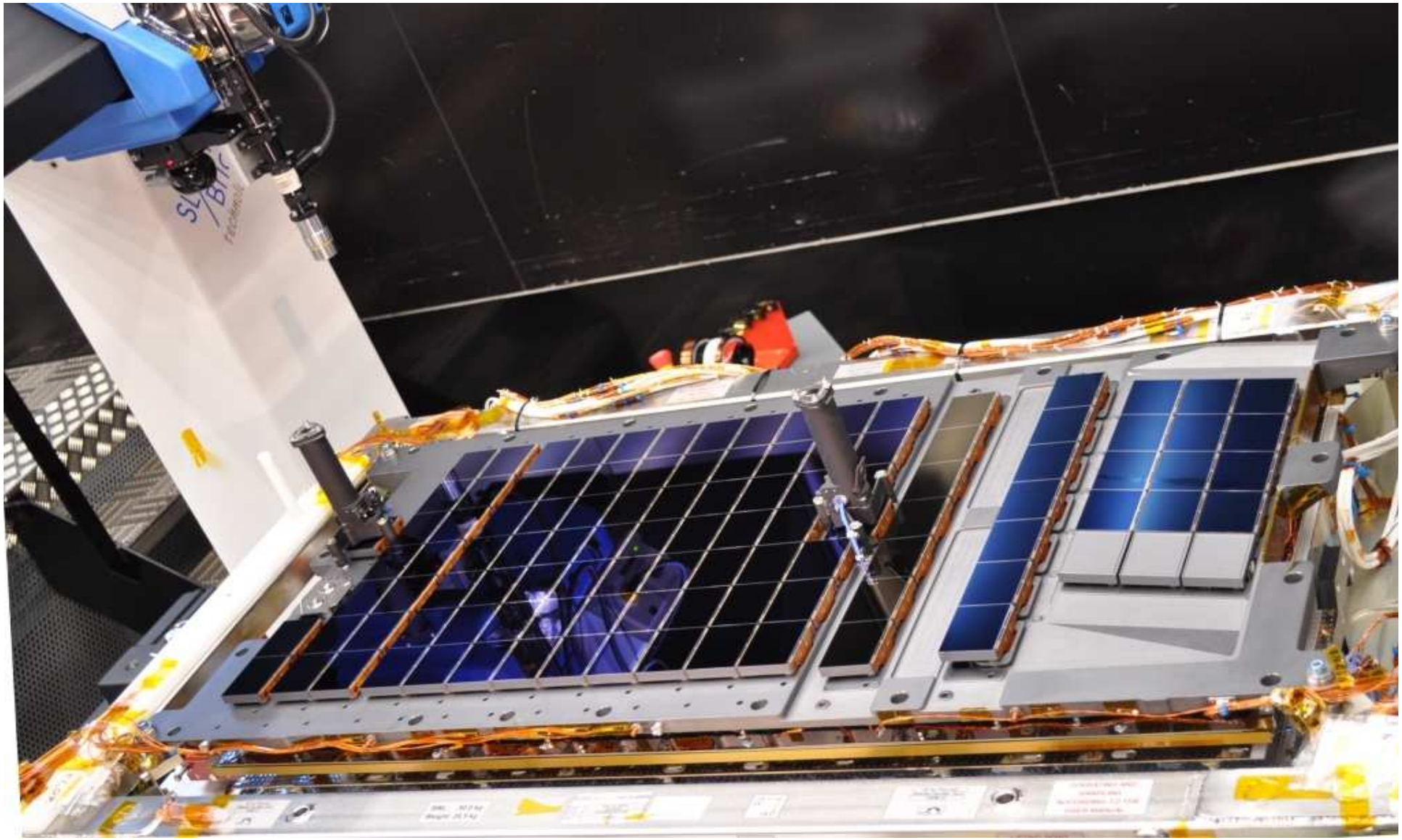


Figure by Olivier Marchal & David Katz

Complexity of the data



Data Processing and Analysis Consortium

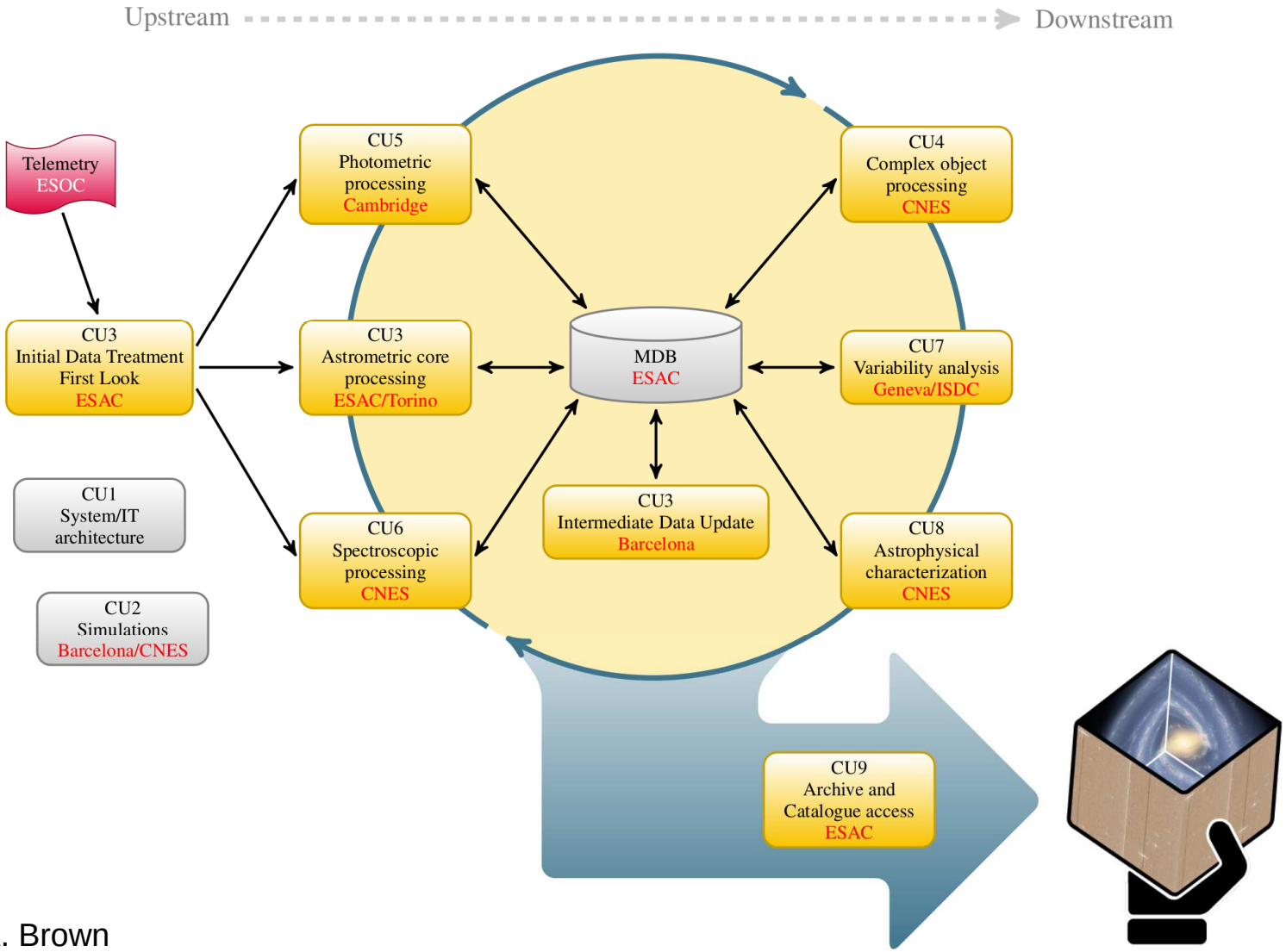


Figure by A. Brown

Sky coverage since mid-July

Number of observations per square degree since start of nominal operations

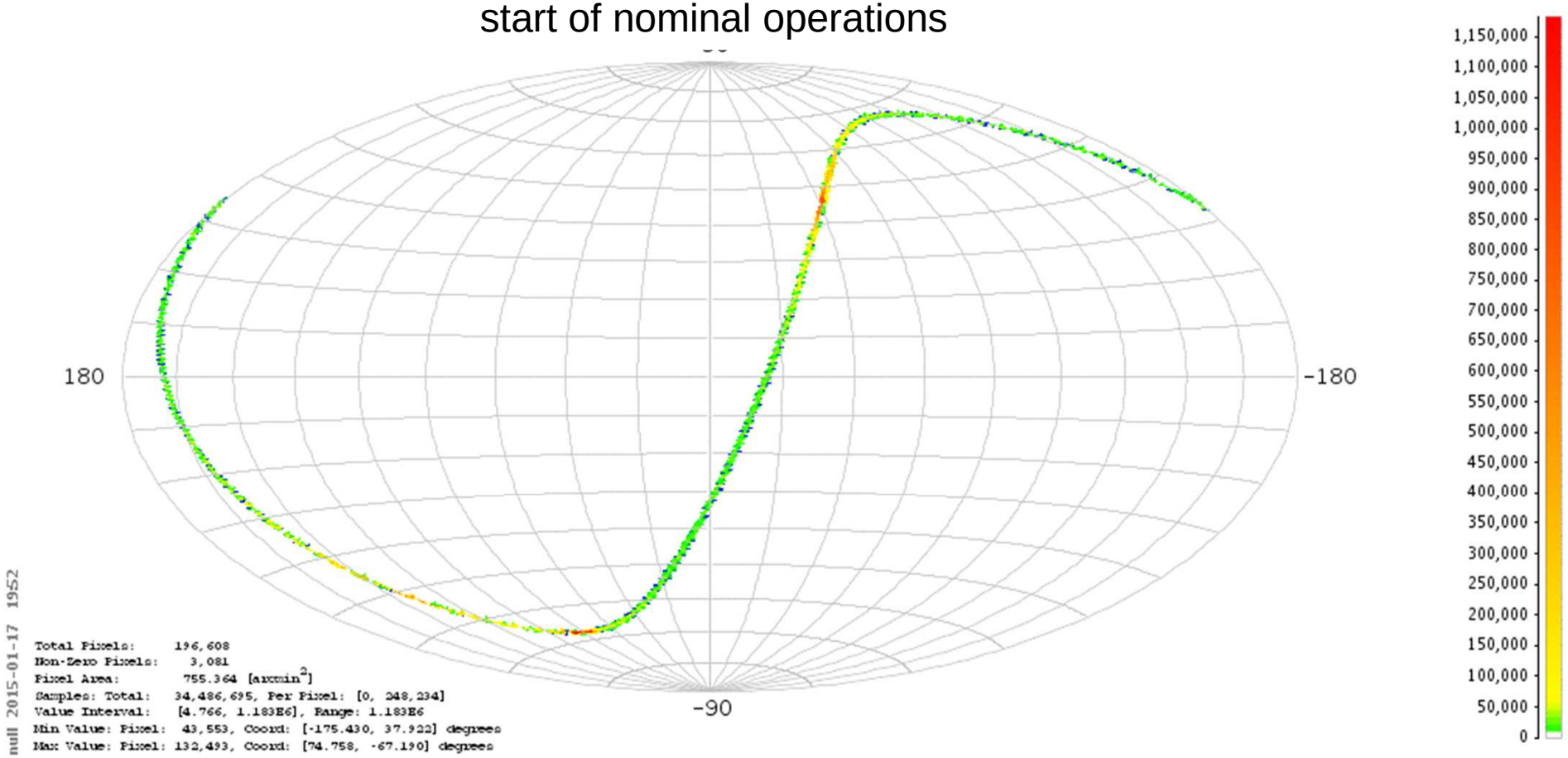
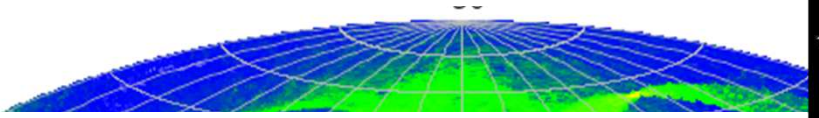
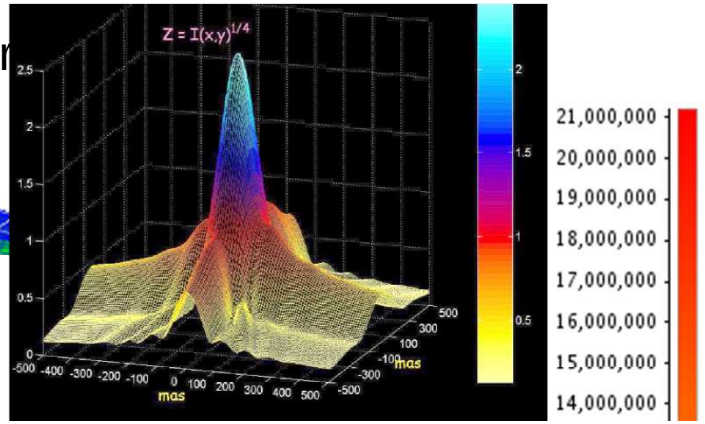


Figure by J. Portell

Sky coverage since mid-July

Number of observations per square degr
start of nominal operations

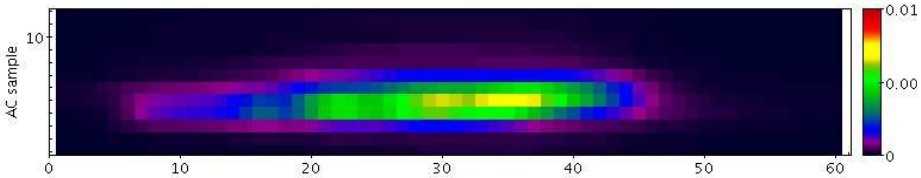




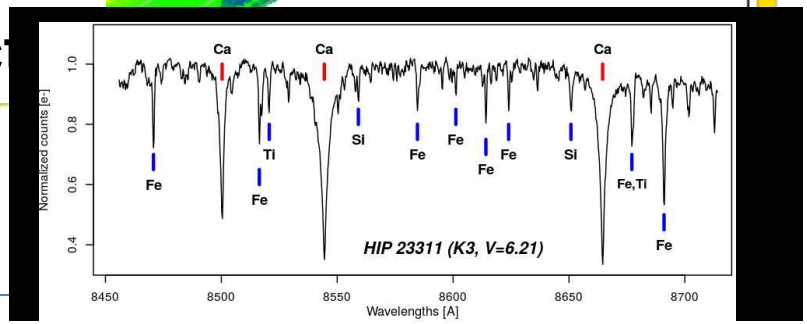
6 months of nominal mission:
full sky coverage
10 billion observations

Gaia-RP spectra

V1293 Aql
(M5III)



90 billion of individual images
20 billion of low-resolution spectra
2.4 billion of high-resolution spec



Processing needs

Raw data:
50 GB every day during 5 years → 100 TB

Ingestion
Pre-processing
Data reduction
Classification
Variability analysis
etc.

} **300 million hours of CPU time**

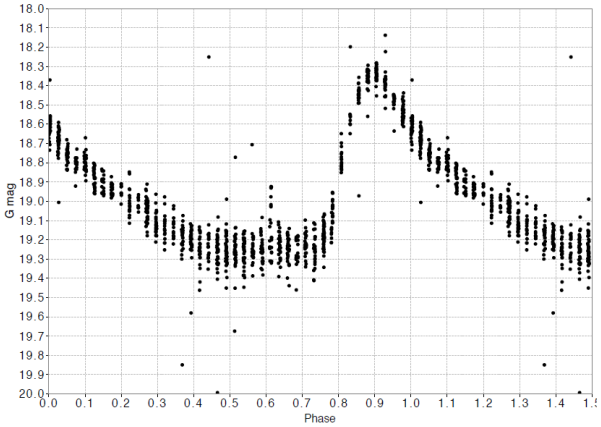
34,000 years !!!

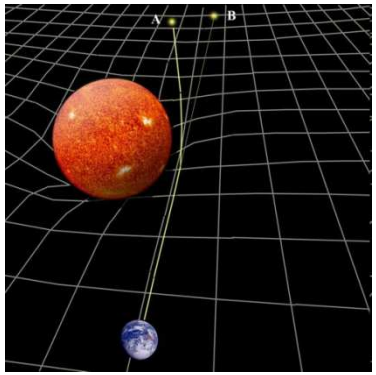
Processing needs

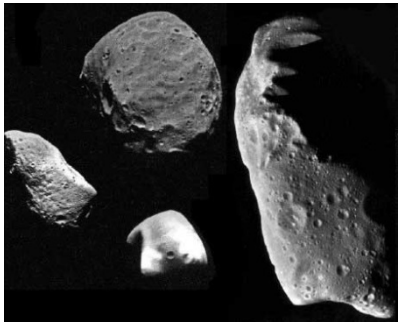
Six supercomputer centres in place for Gaia: the final data will be available around <u>early 2020s</u>

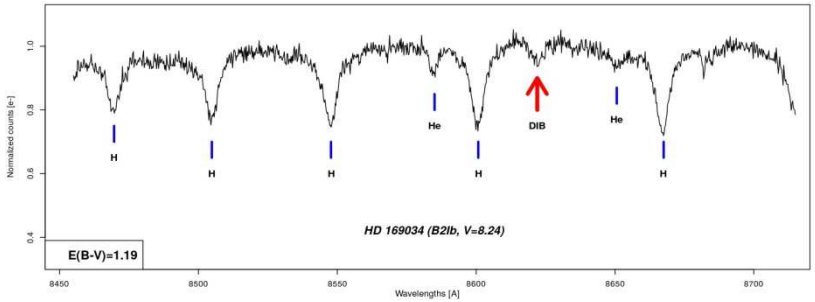
… and, <u>early releases</u> !!!!

Final archive









Archive: DB + Data tools for mining, visualization, analysis, …

Size: 1 Petabyte = 1 million GB
(movie with a run time of 50 yr in HD quality)

Science exploitation

Access to the archive:

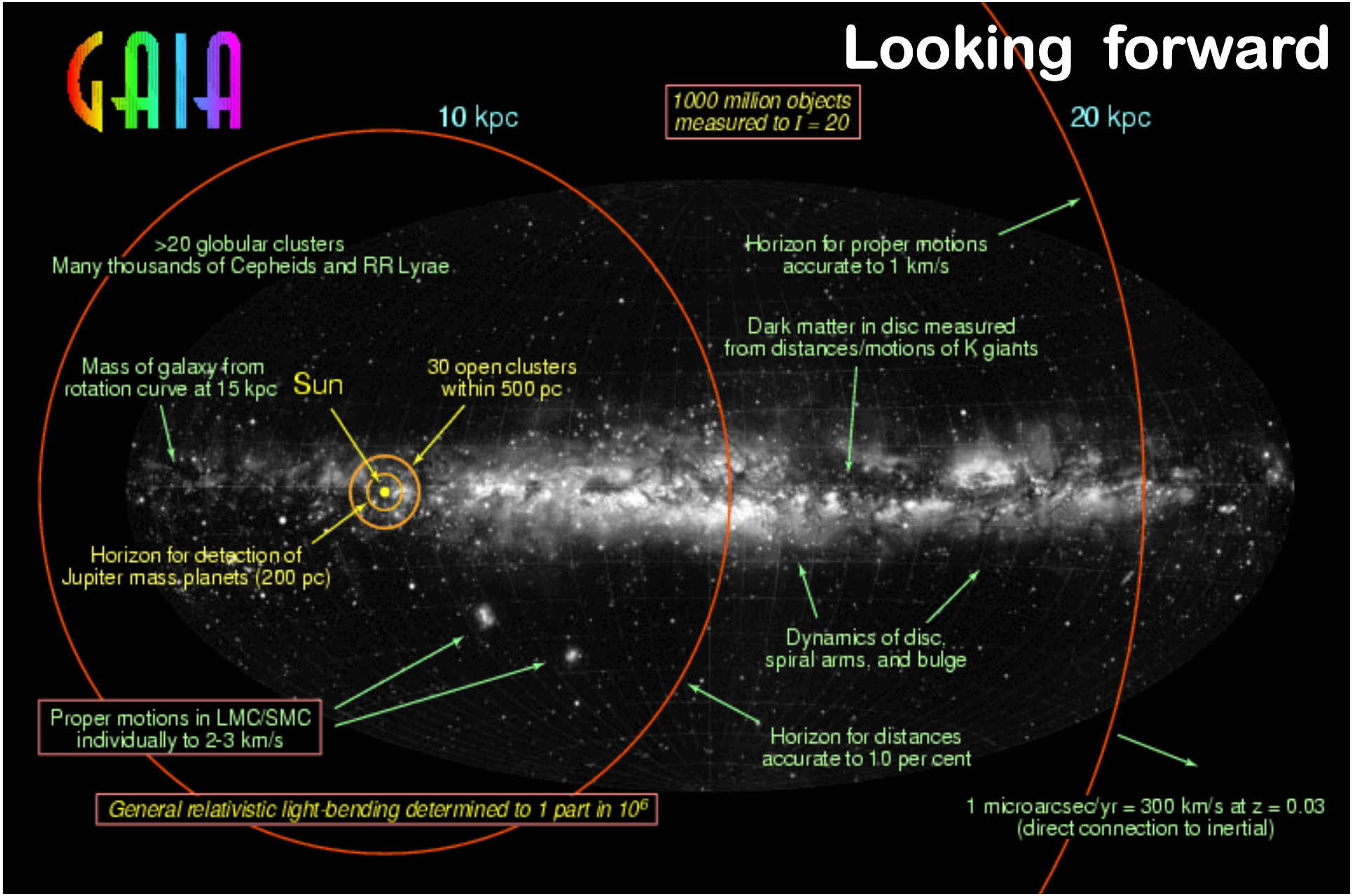
No priority data rights for people involved
Free access to the world wide community

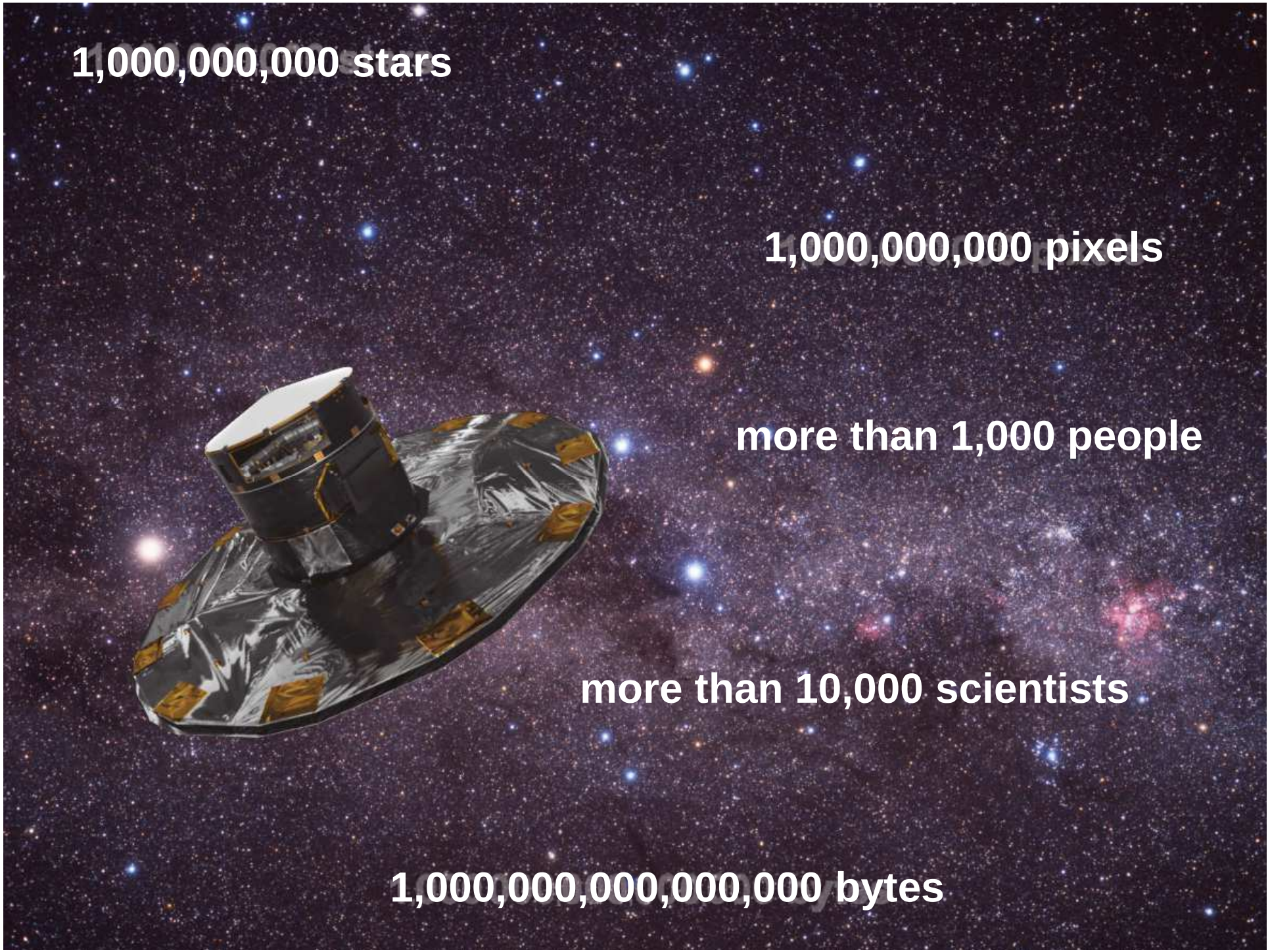
Necessary:

Sharing information
Sharing tools and models
Training young generation of scientists

GREAT: Gaia Research for European Astronomy Training

Looking forward




1,000,000,000 stars
1,000,000,000 pixels
more than 1,000 people
more than 10,000 scientists
1,000,000,000,000,000 bytes

The promises of Gaia

u > 1 billion stars to G = 20
u 106–107 galaxies
u 5 105 quasars
u 3 105 solar system bodies
u tens of thousands of exoplanets

Data collected and processed

Statistics up to June 29

Data collected or processed

Type of Data	Amount	
Science telemetry	5.9 TB	
Astrometry transits	6.5×10^9	64.5×10^9 images
Photometry transits	6.1×10^9	12×10^9 images
Spectroscopy transits	1×10^9	3×10^9 spectra

How many data ?

Since beginning of nominal mission: 137 days

Raw data received: 3.6 TB (26GB/day)

Received & processed measurements: > 6250 million (50-100 million/day)

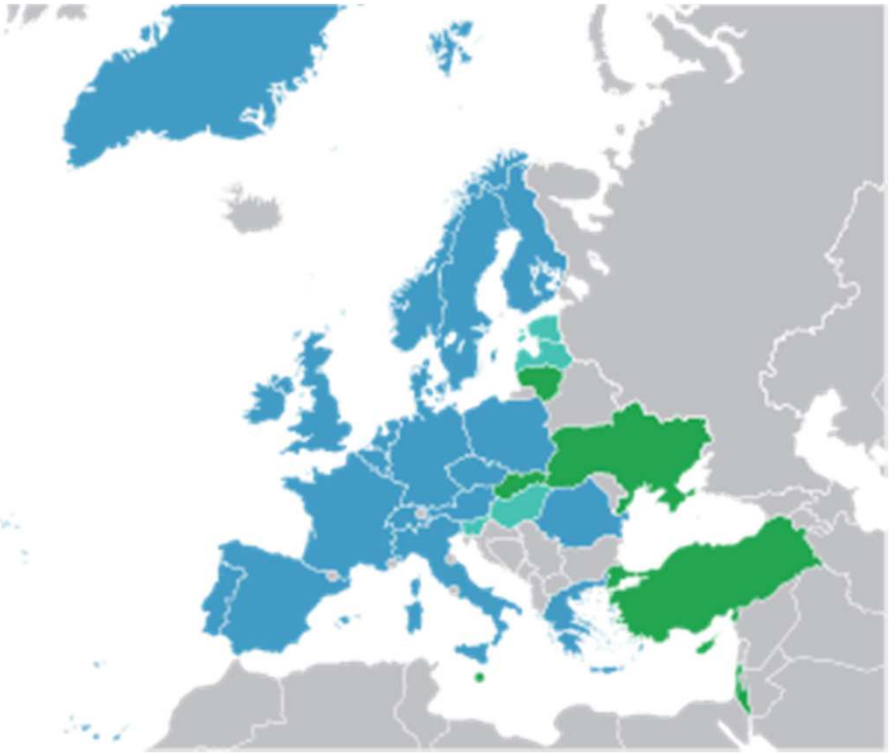
Received & processed images: > 75000 million

Size of the main data base: 30 TB (compressed)

Initial Data Treatment:

Accumulated processing time: ~50 yr

Processing cores used: ~280
(9 nodes x 32 cores, 64GB RAM/node)



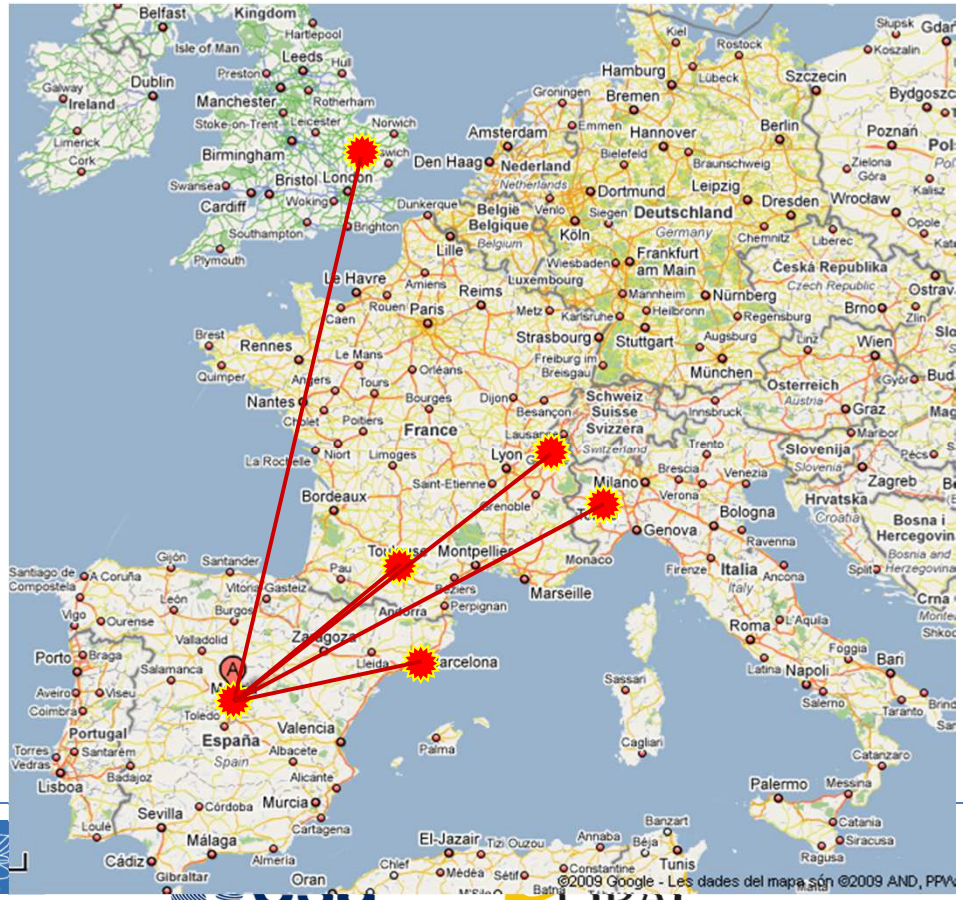
Horizon 2000+
Cornerstone mission

Scientific mission: Astronomy

ESA's only mission

Següents passos

- Seguir amb aquest processament diari fins al final de missió (finals 2019… en principi!)
- Acumulació regular de dades, distribuïnt-ho (com fins ara) als altres centres de processament de dades (incloent MareNostrum)
- Execució dels sistemes iteratius (cíclics) de reducció de dades
- Primera versió del catàleg: **2016**



Processament de dades

gaia

UNIVERSITAT DE BARCELONA

ICCUB

IEEC

esa

Gaia DPAC

Data Processing and Analysis Consortium

gaia

UNIVERSITAT DE BARCELONA

ICCUB

IEEC

esa

Gaia DPAC