Gaia DR3

Un reto científico y tecnológico

Xavier Luri – ICCUB/IEEC, Universitat de Barcelona

Triangulando el cielo

01

Contexto histórico

Los catálogos

La herramienta del astrónomo

Tres estrellas cada uno

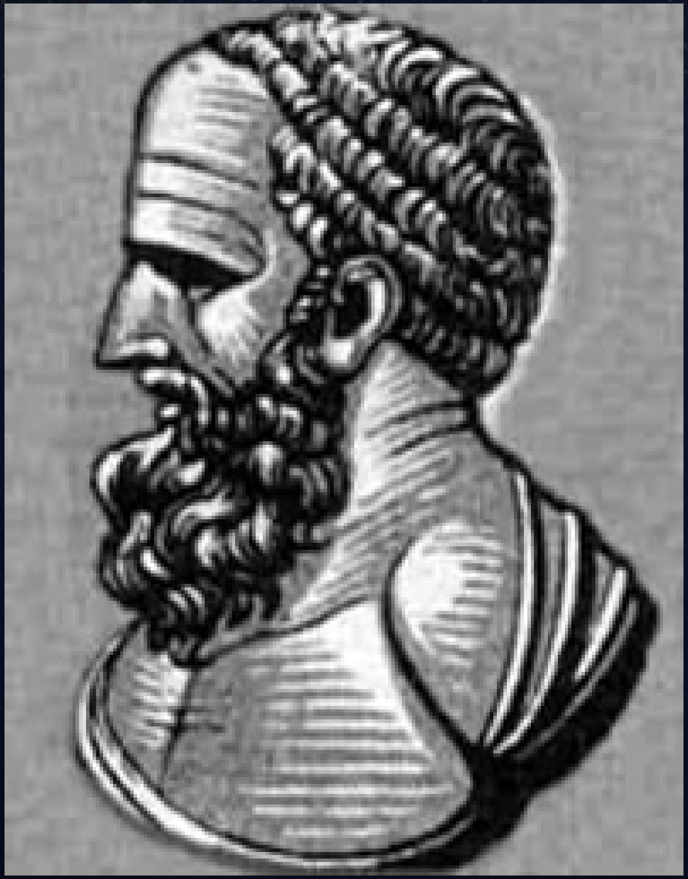
- El catálogo más antiguo conocido (*ca*. 1531 AC to *ca*. 1155 AC)
- 36 estrellas

MUL.APIN

- 71 estrellas

Los catálogos griegos

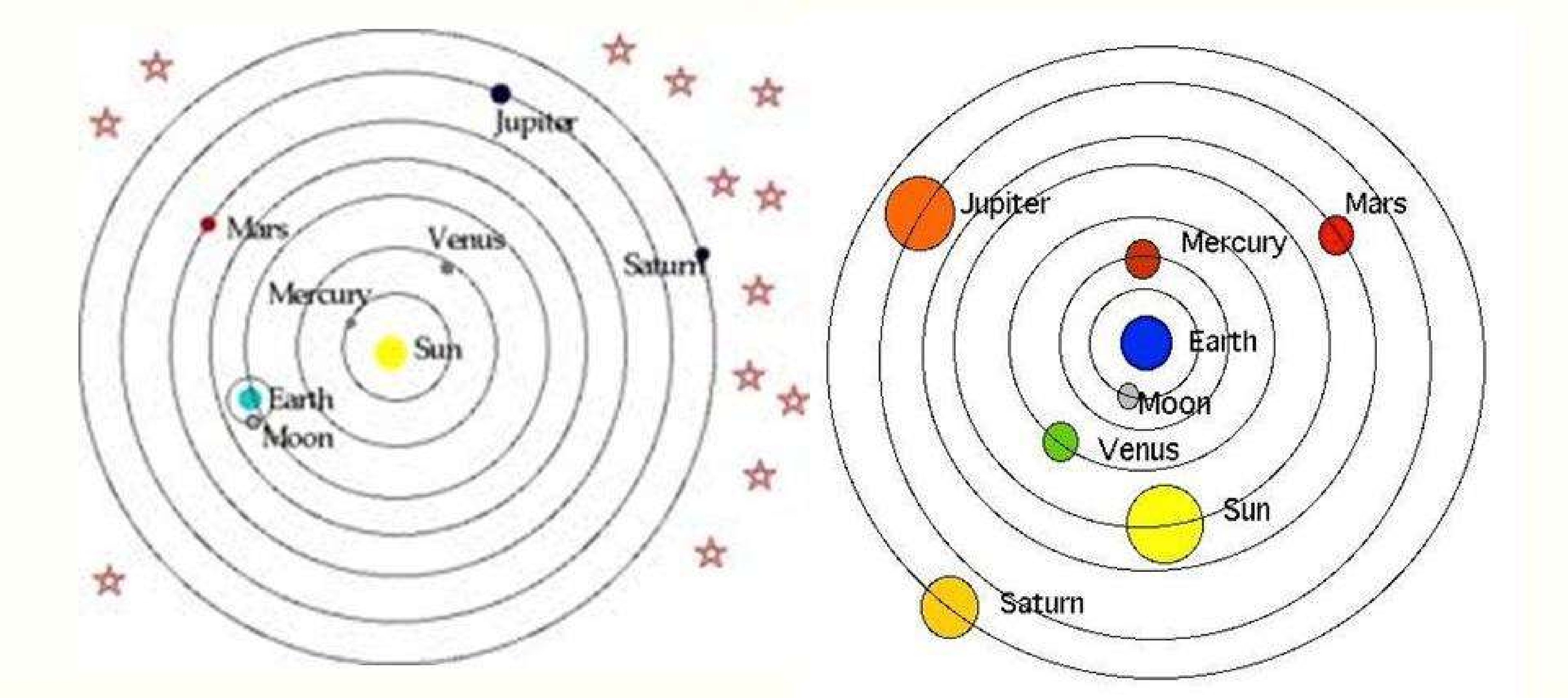
Los clásicos



- Eudoxo de Cnidos (370 AC)
- Timocharis (~300 AC)
- **Hiparco de Nicea (129 AC)**
  - ~1000 estrellas
  - Descubre de la precesión de los equinocios
  - Primera nova
  - Primer modelo del movimento solar y lunar
  - Las mejores observaciones planetarias del momento

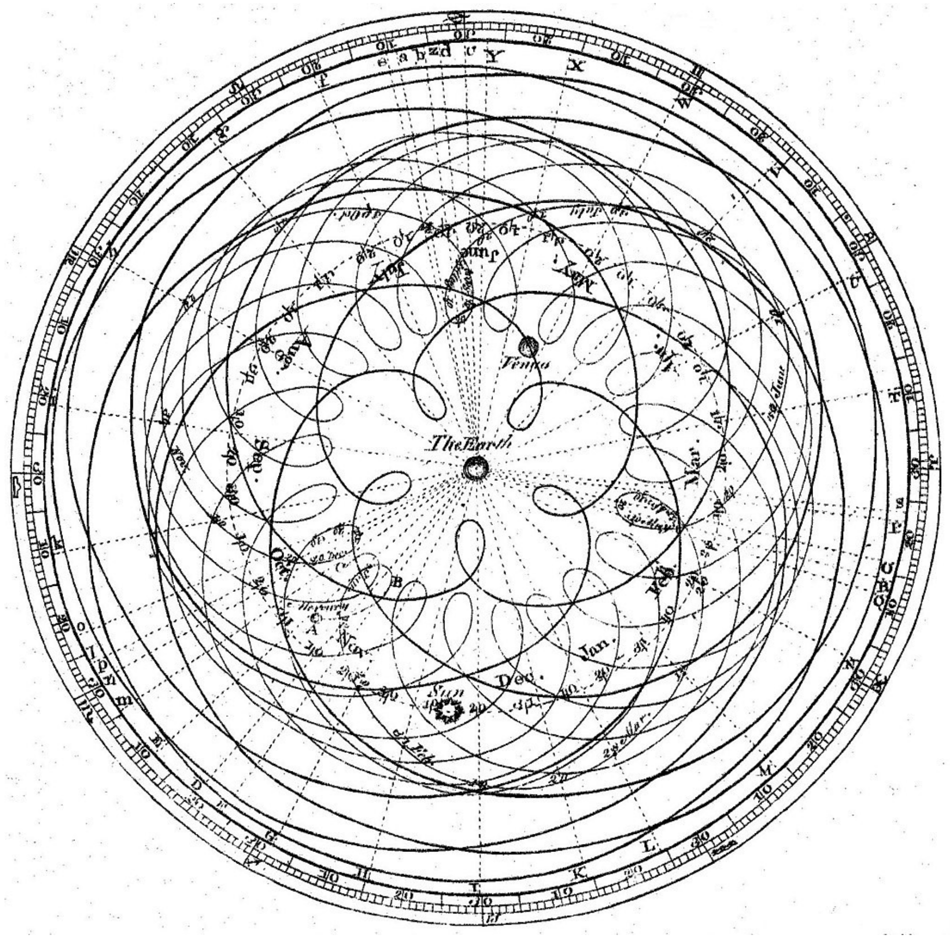
Copernico vs. Tolomeo

¿Cual es el “correcto”?

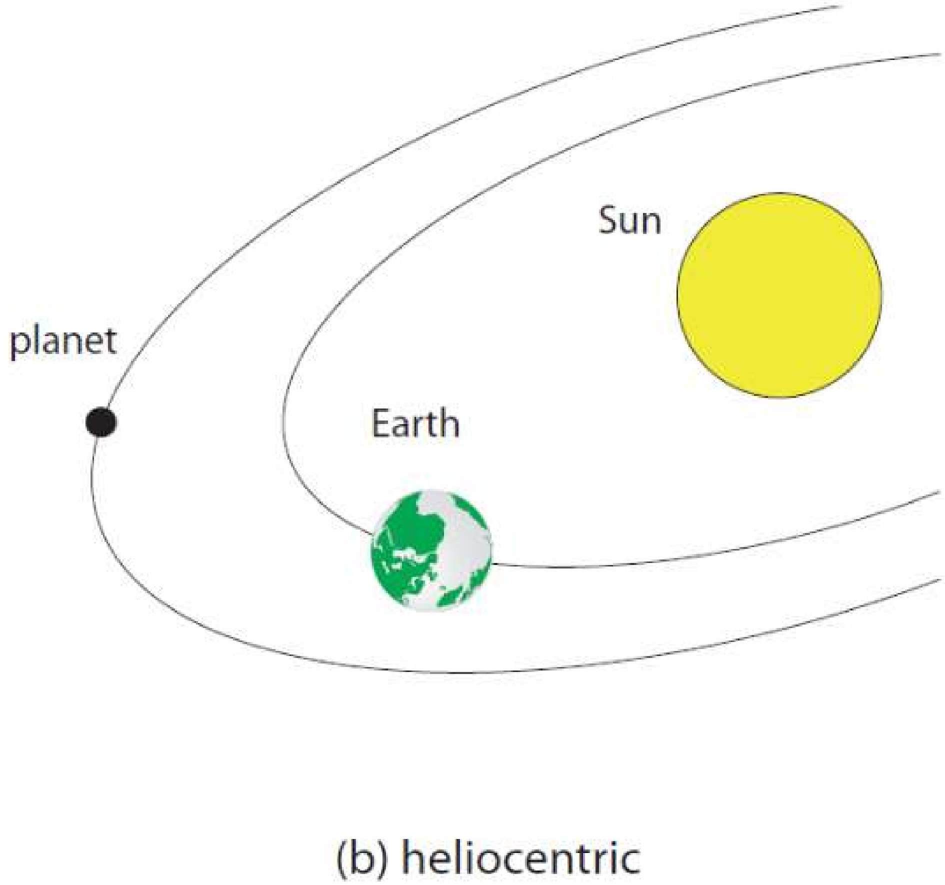

Jupiter
Mars
Venus
Saturn
Mercury
Sun
Earth
Moon
Jupiter
Mars
Mercury
Earth
Moon
Venus
Sun
Saturn

Predicción de posiciones

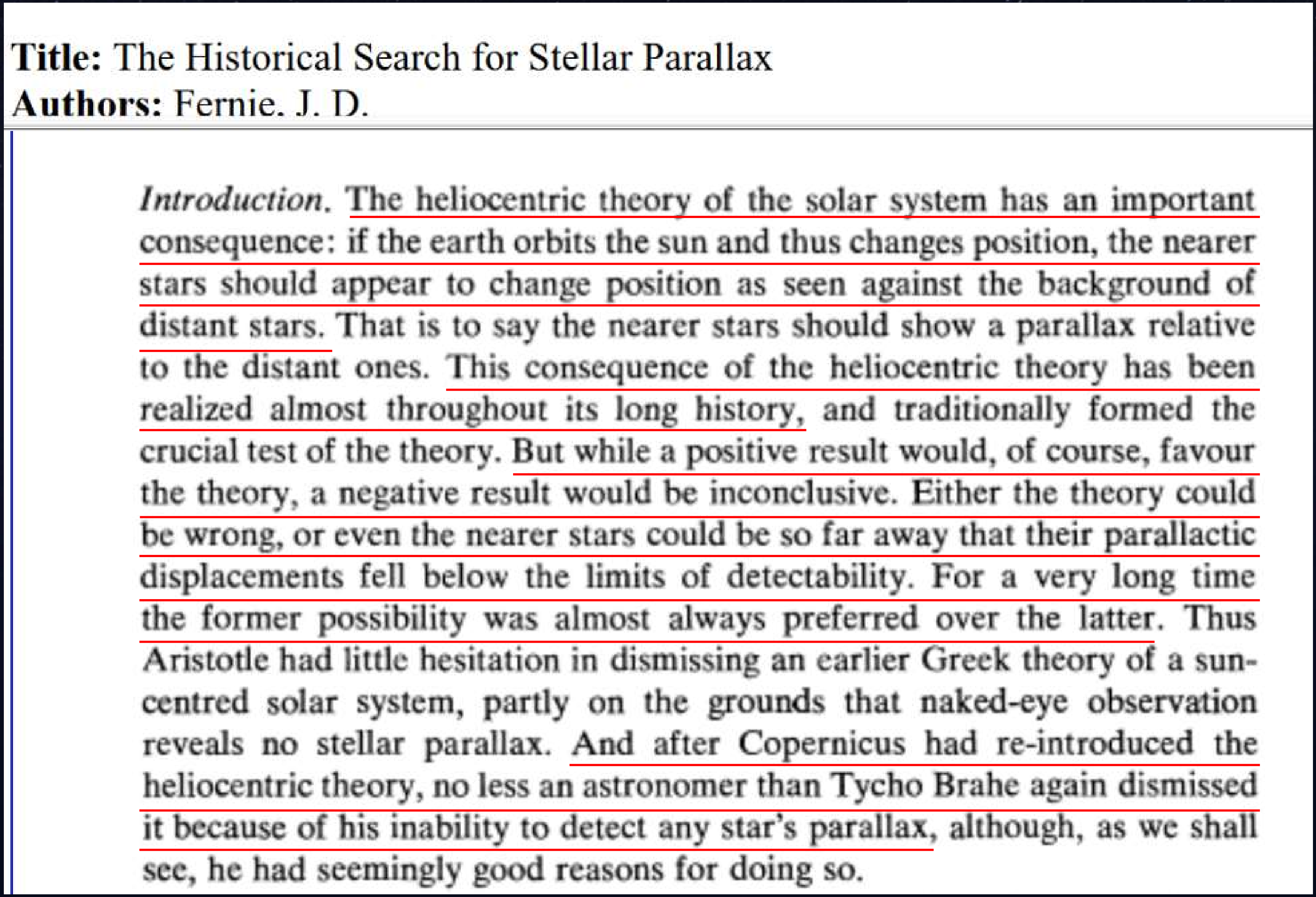
Las dos teorias pueden hacerlo correctamente


The Earth
Venus
Sun
Jan.
Dec.

Epiciclos y eferentes

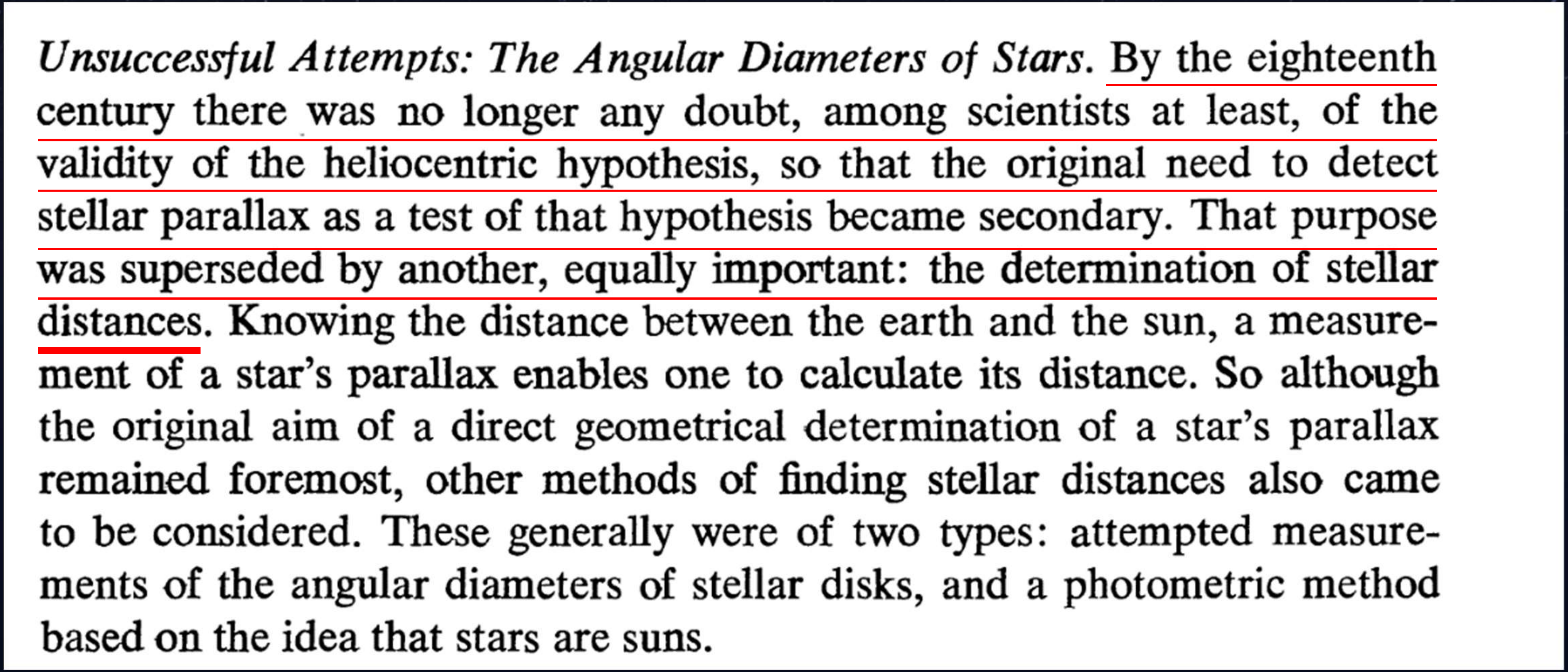

Sun
planet
Earth
(b) heliocentric

(figura Perryman 2012)



Title: The Historical Search for Stellar Parallax
Authors: Fernie, J. D.

Introduction. The heliocentric theory of the solar system has an important consequence: if the earth orbits the sun and thus changes position, the nearer stars should appear to change position as seen against the background of distant stars. That is to say the nearer stars should show a parallax relative to the distant ones. This consequence of the heliocentric theory has been realized almost throughout its long history, and traditionally formed the crucial test of the theory. But while a positive result would, of course, favour the theory, a negative result would be inconclusive. Either the theory could be wrong, or even the nearer stars could be so far away that their parallactic displacements fell below the limits of detectability. For a very long time the former possibility was almost always preferred over the latter. Thus Aristotle had little hesitation in dismissing an earlier Greek theory of a sun-centred solar system, partly on the grounds that naked-eye observation reveals no stellar parallax. And after Copernicus had re-introduced the heliocentric theory, no less an astronomer than Tycho Brahe again dismissed it because of his inability to detect any star's parallax, although, as we shall see, he had seemingly good reasons for doing so.



Unsuccessful Attempts: The Angular Diameters of Stars. By the eighteenth century there was no longer any doubt, among scientists at least, of the validity of the heliocentric hypothesis, so that the original need to detect stellar parallax as a test of that hypothesis became secondary. That purpose was superseded by another, equally important: the determination of stellar distances. Knowing the distance between the earth and the sun, a measurement of a star's parallax enables one to calculate its distance. So although the original aim of a direct geometrical determination of a star's parallax remained foremost, other methods of finding stellar distances also came to be considered. These generally were of two types: attempted measurements of the angular diameters of stellar disks, and a photometric method based on the idea that stars are suns.

Paralajes estelares

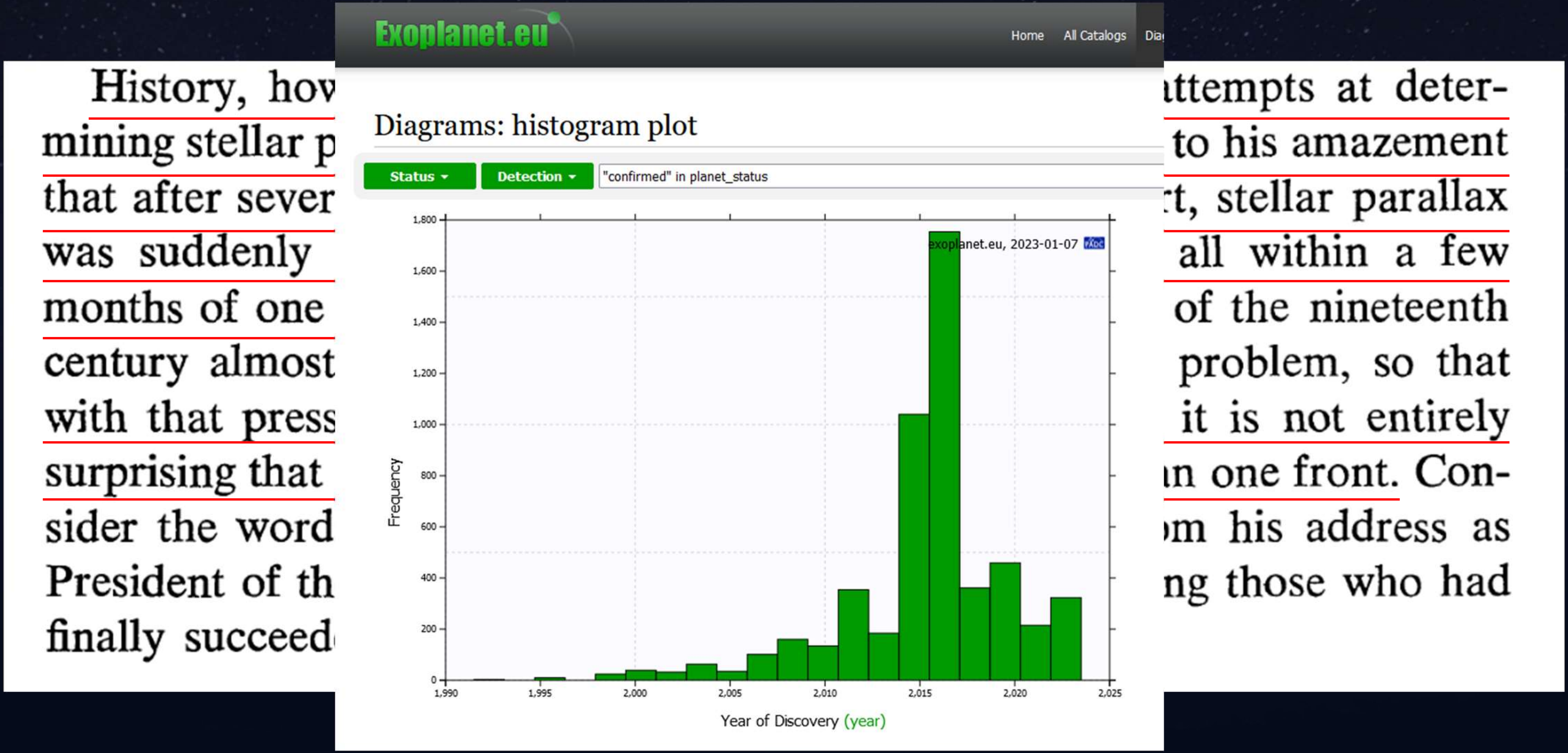
Con el telescopio pueden por fin observarse

En el siglo XVII se produjo una revolución en la astronomía: la invención del telescopio y el reconocimiento del sistema heliocéntrico desencadenaron una carrera entre astrónomos para medir la paralajes de las estrellas.

Pero no fue hasta finales de la década de 1830 que estas medidas permitieron a los astrónomos determinar por primera vez las distancias a un puñado de estrellas.

Bessel, Struve and Henderson

Un esfuerzo colectivo



History, hov
mining stellar p
that after sever
was suddenly
months of one
century almost
with that press
surprising that
sider the word
President of th
finally succeed

ıttempts at deter-
to his amazement
:t, stellar parallax
all within a few
of the nineteenth
problem, so that
it is not entirely
ın one front. Con-
›m his address as
ng those who had

Exoplanet.eu
Home
All Catalogs
Dia

Diagrams: histogram plot

Status
Detection

"confirmed" in planet_status

exoplanet.eu, 2023-01-07
Frequency
1,800
1,600
1,400
1,200
1,000
800
600
400
200
0
1,990
1,995
2,000
2,005
2,010
2,015
2,020
2,025
Year of Discovery (year)

Friedrich Wilhelm Bessel

La primera paralaje estelar

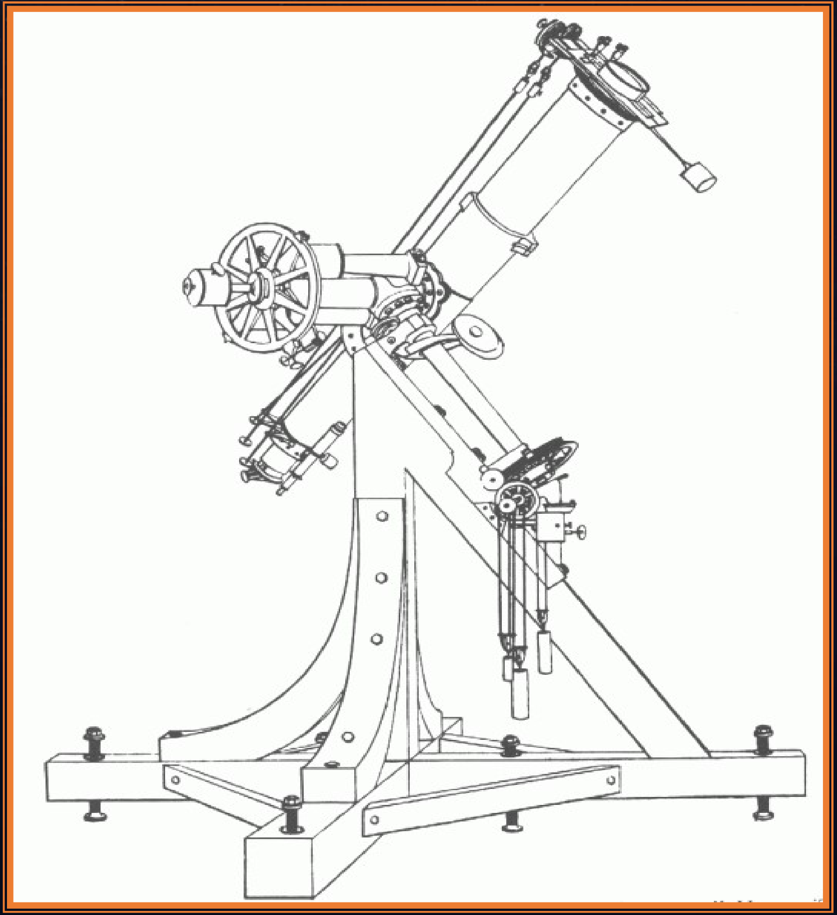
Nacido en Minden (Alemania) en 1784

Autodidacta, trabajaba como contable en una empresa hasta que pasó a dedicarse a la astronomía gracias a W. Olbers

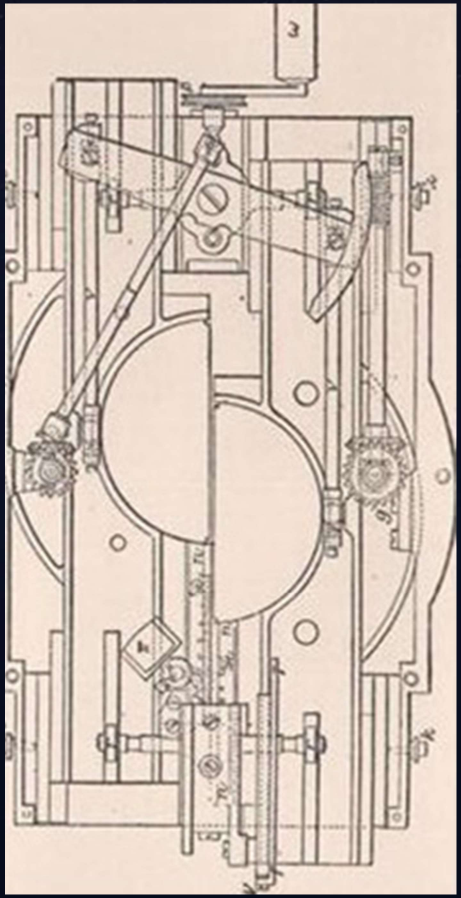
Obtuvo la primera paralaje estelar fiable en 1838

Franhoufer, el constructor de instrumentos

Inicialmente diseñado para la medida del diámetro de los astros, principalmente Sol y Luna



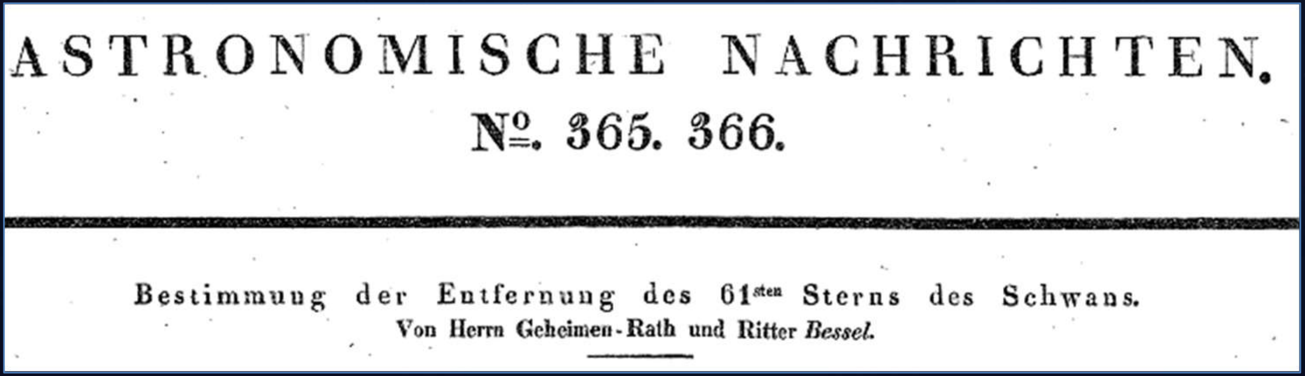
Heliómetro de Fraunhofer



Resultado (1838)

La primera paralje estelar fiable

- Medidas relativas durante 15 meses: 85 Sa, 98 Sb
- Ajuste de movemento propio y paralaxis por minimos cuadrados



ASTRONOMISCHE NACHRICHTEN.

№. 365. 366.

Bestimmung der Entfernung des 61sten Sterns des Schwans.

Von Herrn Geheimen-Rath und Ritter *Bessel.*

$$\varpi = 0.314 \pm 0.02'' \quad (1838)\; D = 10.4\ \mathrm{ly}$$

Más allà de la astronomia terrestre

02

Astrometría desde el espacio

Los límites de la observación terrestre

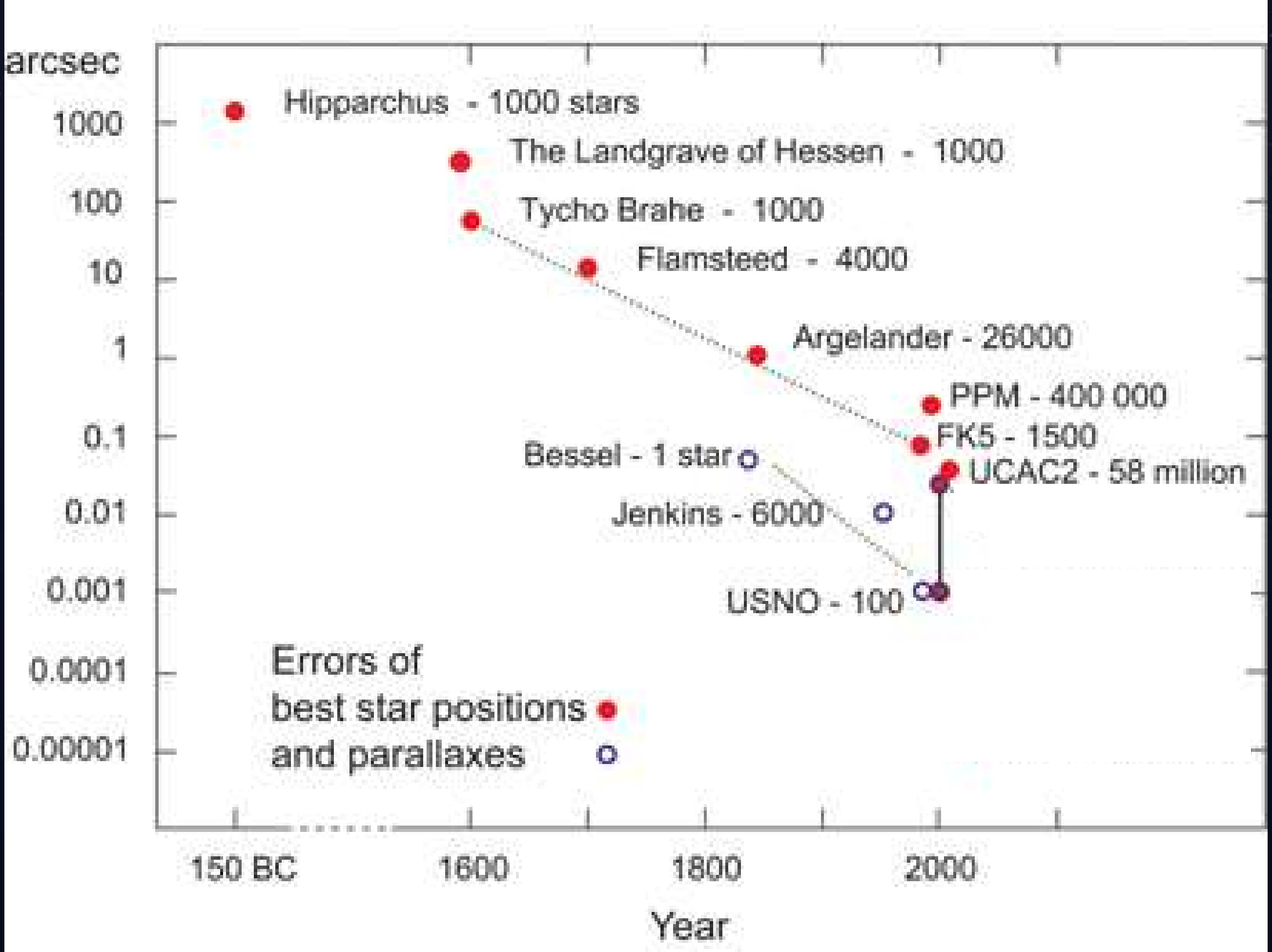
¿Hasta dónde podemos mejorar desde tierra?

A mediados del siglo XX la astrometría había llegado a un cierto estancamiento debido a las limitaciones de las observaciones desde tierra, principalmente:

- Efectos atmosféricos (seeing): límite de precisión
- Observaciones de campo reducido: medidas locales
- Medidas angulares relativas (fondos de estrellas débiles): efectos sistemáticos

La era de las paralajes

Disponibilidad de distancias estelares



La respuesta: astrometría desde el espacio

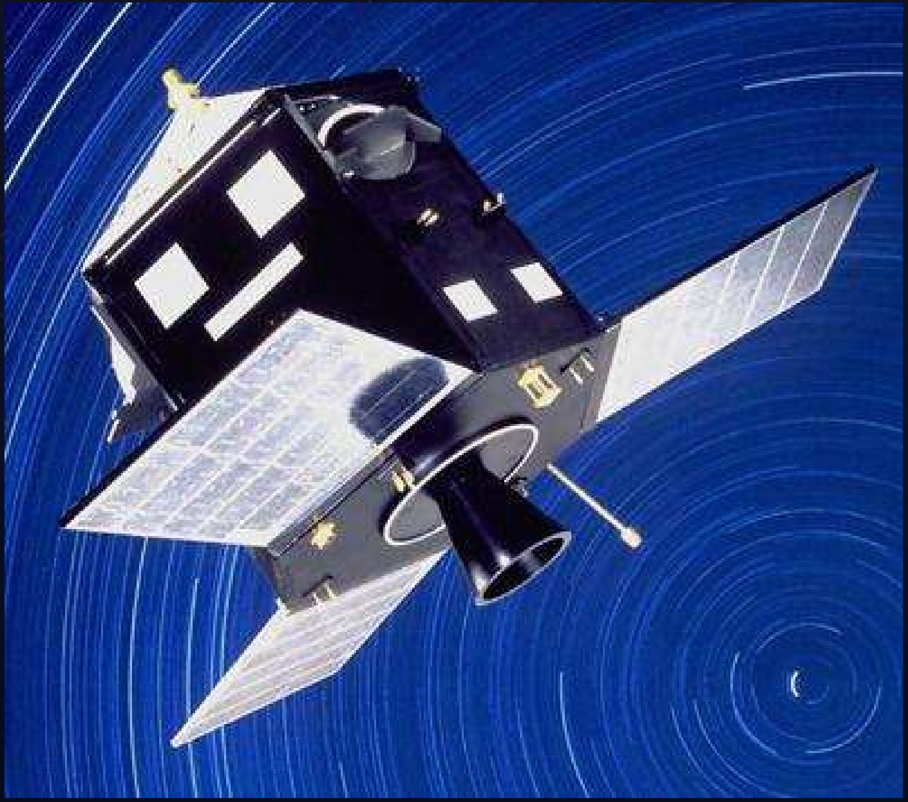
Hipparcos y Gaia, las misiones de la agencia europea del espacio

Hipparcos (1989-1993)

Gaia (2013-2020+)

Los catálogos Hipparcos y Tycho

Los primeros catálogos de astrometría espacial



- Catálogo Hipparcos
  - 118.218 estellas
  - Precisión 1mas
  - Primeras paralajes absolutas!
  - Movimentos propios

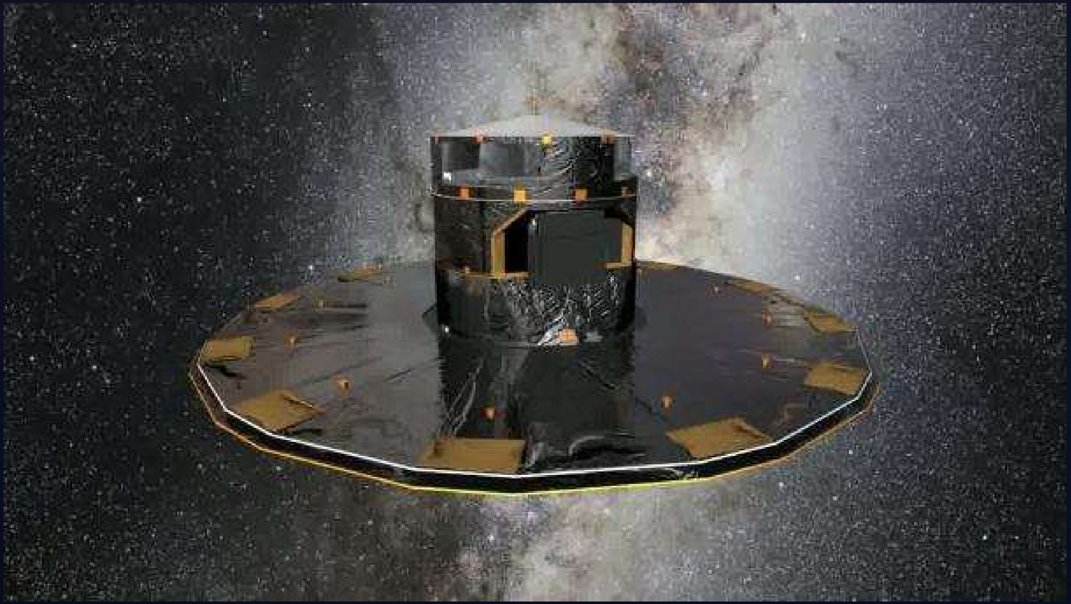
- Catálogo Tycho-2
  - 2,5 millones de estrellas
  - Precisión 2.5mas
  - Posiciones y movimientos propios
  -

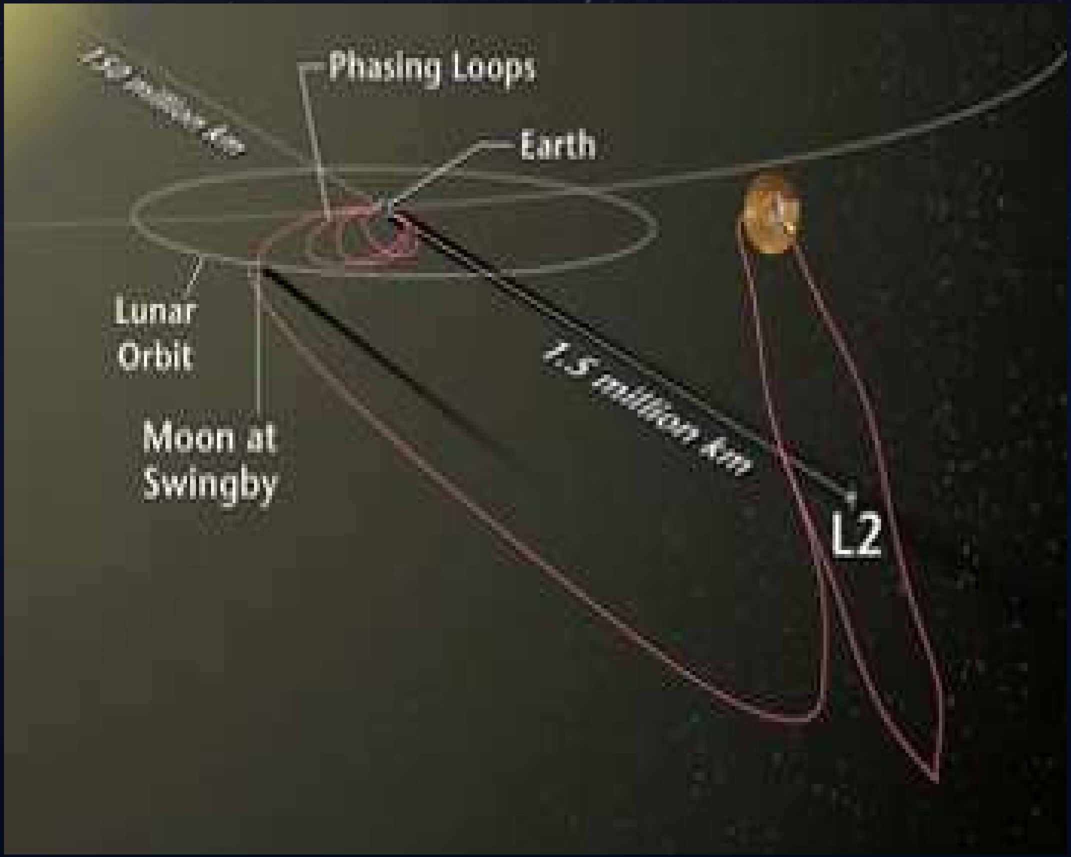
Contexto **03**

La misión Gaia

La misión Gaia

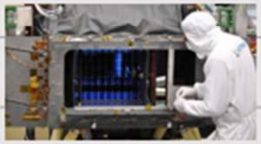
DR1, DR2, EDR3, DR3, DR4, y DR5




150 million km
Phasing Loops
Earth
Lunar Orbit
Moon at Swingby
1.5 million km
L2

A brief history of Gaia

1993 1997 2000 2005-6



2013 2016 2018 2020 **2022** 2025 2030

Initial concept

Gaia Science
Advisory Group

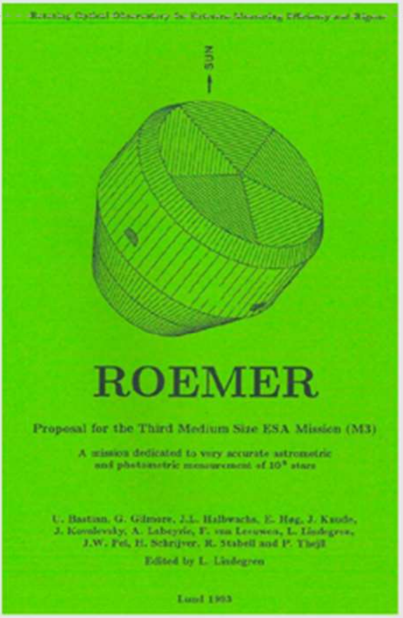
Gaia “Red Book”

Prime contractor selected,
construction starts

Launch! 🥳

Start of nominal operations

~70M transits and ~37GB every day (on average)



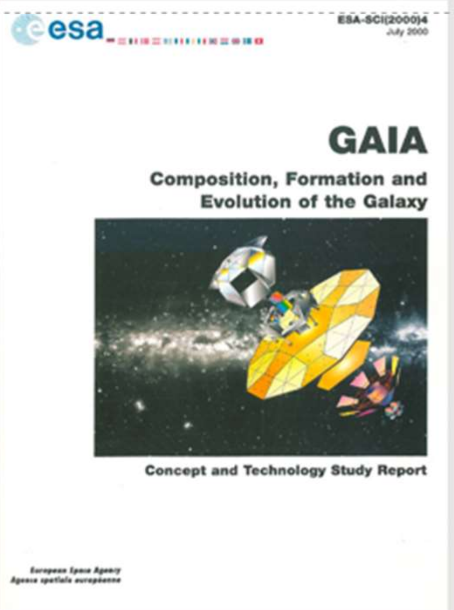
ROEMER

Proposal for the Third Medium Size ESA Mission (M3)

A mission dedicated to very accurate astrometric
and photometric measurement of 10^8 stars

U. Bastian, G. Gilmore, J.L. Halbwachs, E. Høg, J. Knude,
J. Kovalevsky, A. Labeyrie, F. van Leeuwen, L. Lindegren,
J.W. Pel, H. Schrijver, R. Stabell and P. Thejll
Edited by L. Lindegren

Lund 1993



esa

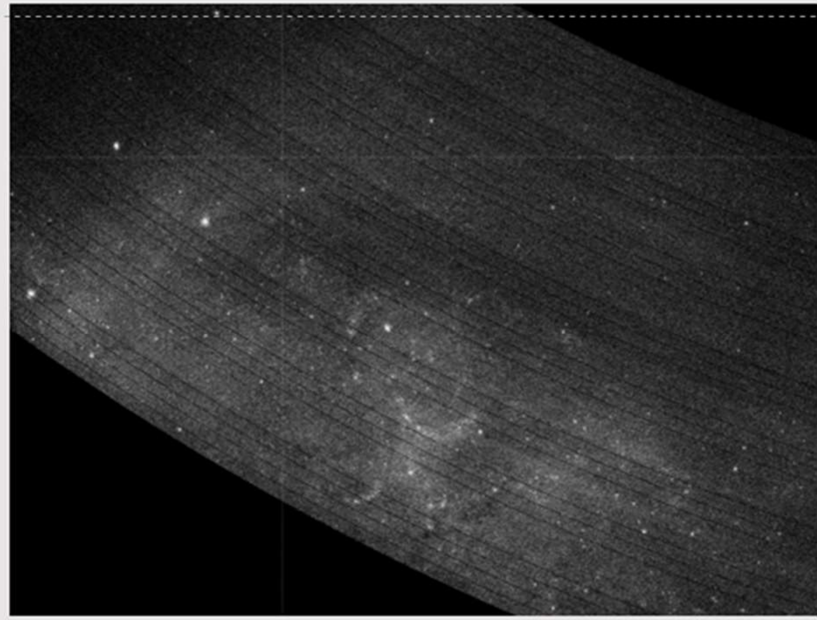
ESA-SCI(2000)4
July 2000

GAIA

Composition, Formation and
Evolution of the Galaxy

Concept and Technology Study Report

European Space Agency
Agence spatiale européenne

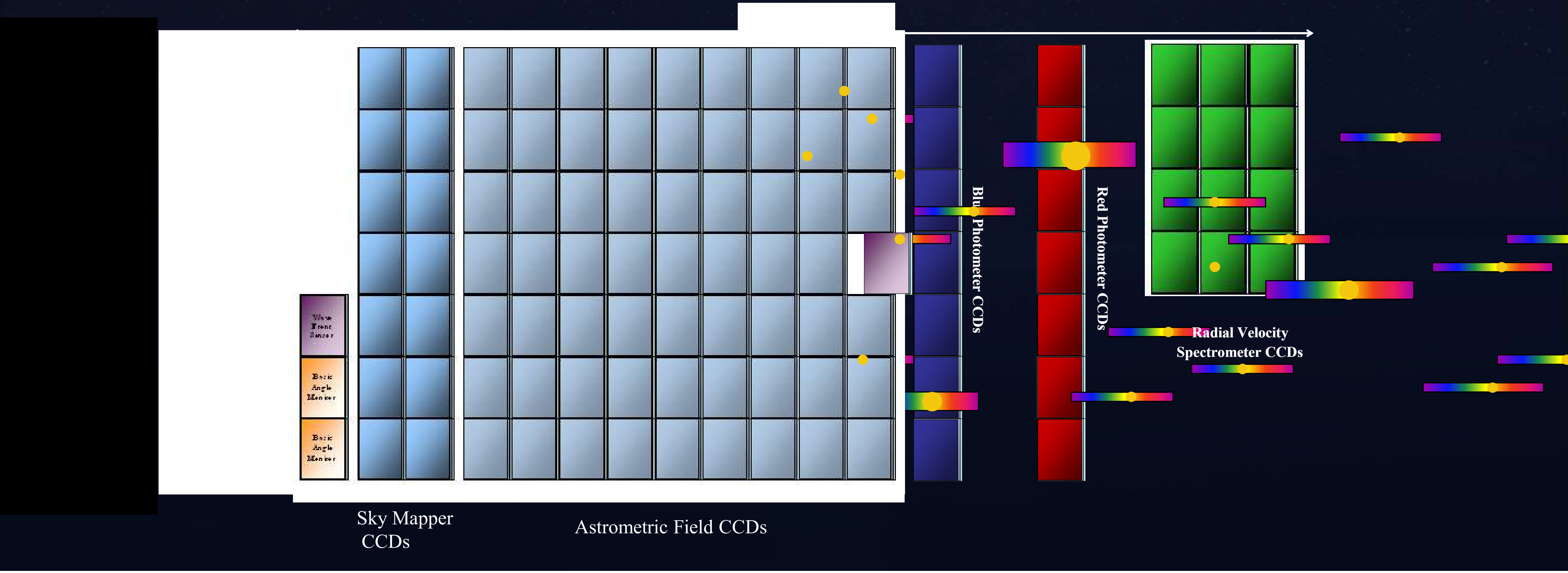


Instrumentos y plano focal

Plano focal

106 CCDs , 938 million pixels, 2800 cm^2

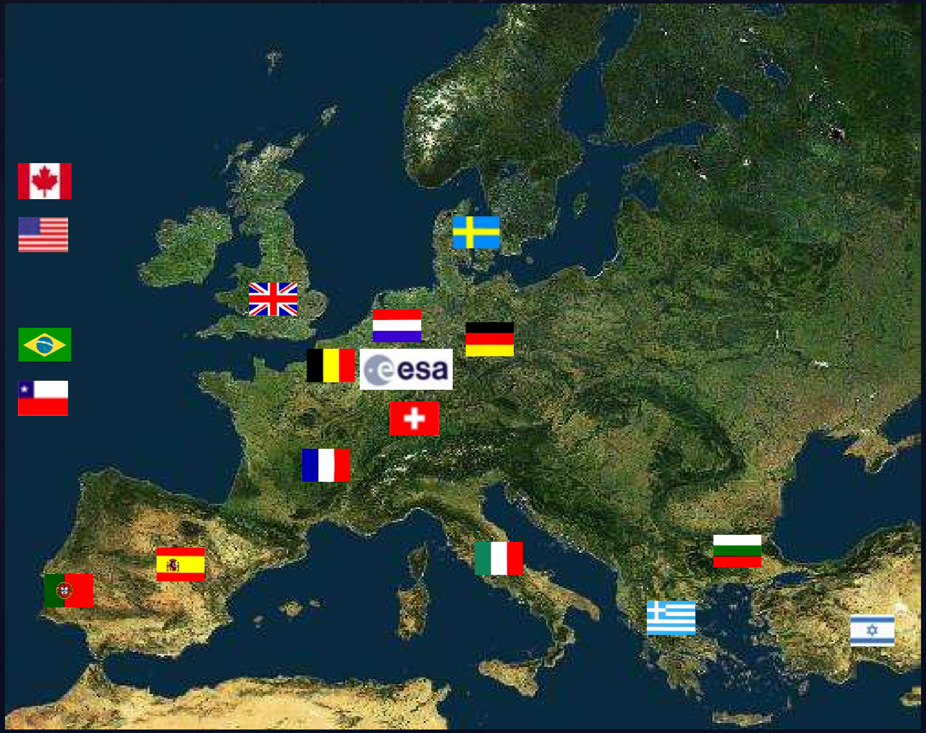
2


Wave Front Sensor
Basic Angle Monitor
Basic Angle Monitor
Blue Photometer CCDs
Red Photometer CCDs
Radial Velocity Spectrometer CCDs
Sky Mapper CCDs
Astrometric Field CCDs

El consorcio DPAC

Gaia Data Processing and Analysis Consortium

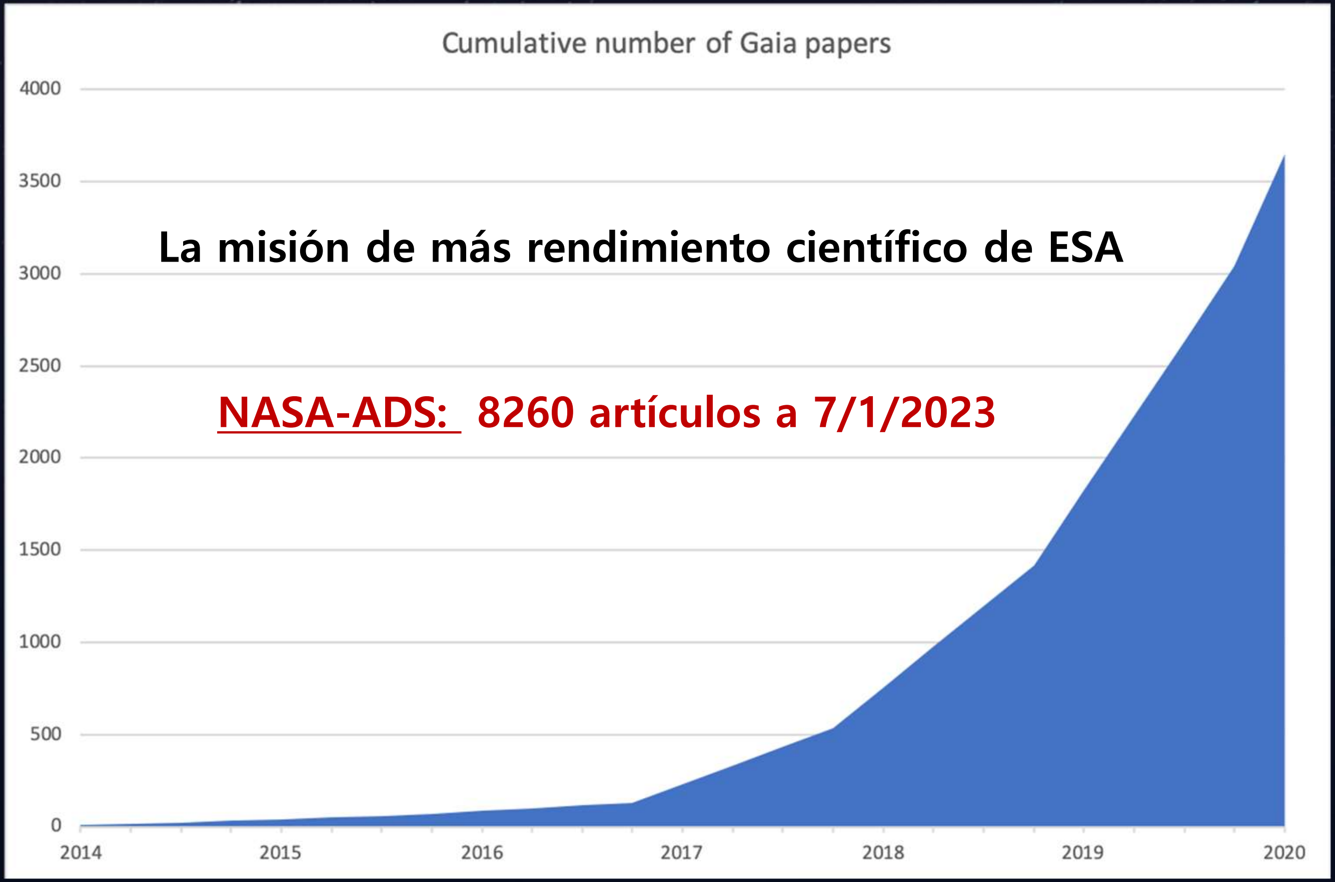
- ~500 miembros
- 24 agencias de financiación
- 93% en las 10 mayores



Duración MLA: 2007-2022+

Coste medio anual: ~30M€
Coste global: ~500M€

El éxito de Gaia

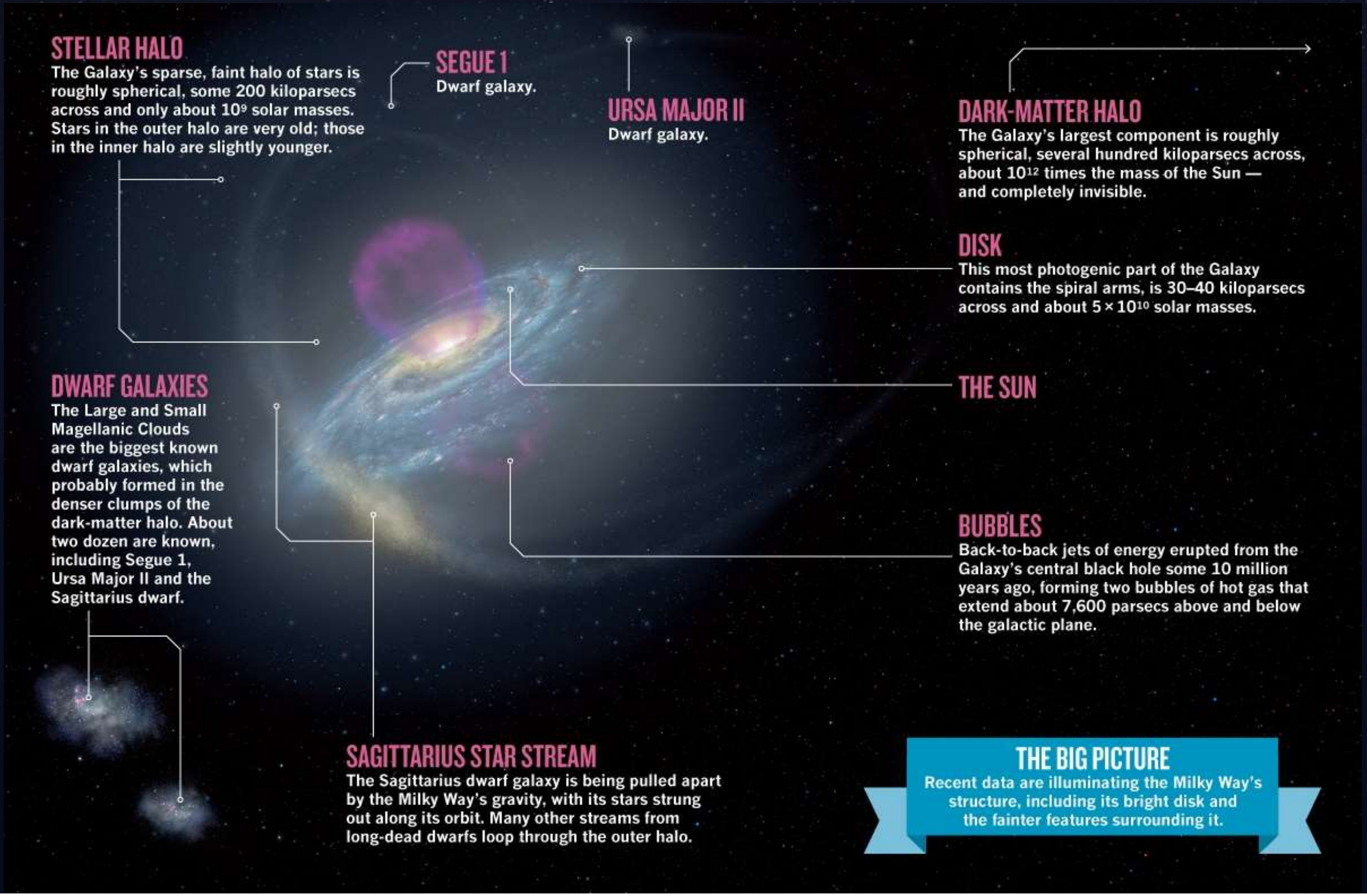


La misión de más rendimiento científico de ESA

NASA-ADS: 8260 artículos a 7/1/2023

Objetivo científico: nuestra Galaxia

Composición, evolución, dinàmica y estructura de la Vía Láctea



STELLAR HALO
The Galaxy's sparse, faint halo of stars is roughly spherical, some 200 kiloparsecs across and only about 10^9 solar masses. Stars in the outer halo are very old; those in the inner halo are slightly younger.

SEGUE 1
Dwarf galaxy.

URSA MAJOR II
Dwarf galaxy.

DARK-MATTER HALO
The Galaxy's largest component is roughly spherical, several hundred kiloparsecs across, about 10^{12} times the mass of the Sun — and completely invisible.

DISK
This most photogenic part of the Galaxy contains the spiral arms, is 30–40 kiloparsecs across and about 5×10^{10} solar masses.

THE SUN

DWARF GALAXIES
The Large and Small Magellanic Clouds are the biggest known dwarf galaxies, which probably formed in the denser clumps of the dark-matter halo. About two dozen are known, including Segue 1, Ursa Major II and the Sagittarius dwarf.

BUBBLES
Back-to-back jets of energy erupted from the Galaxy's central black hole some 10 million years ago, forming two bubbles of hot gas that extend about 7,600 parsecs above and below the galactic plane.

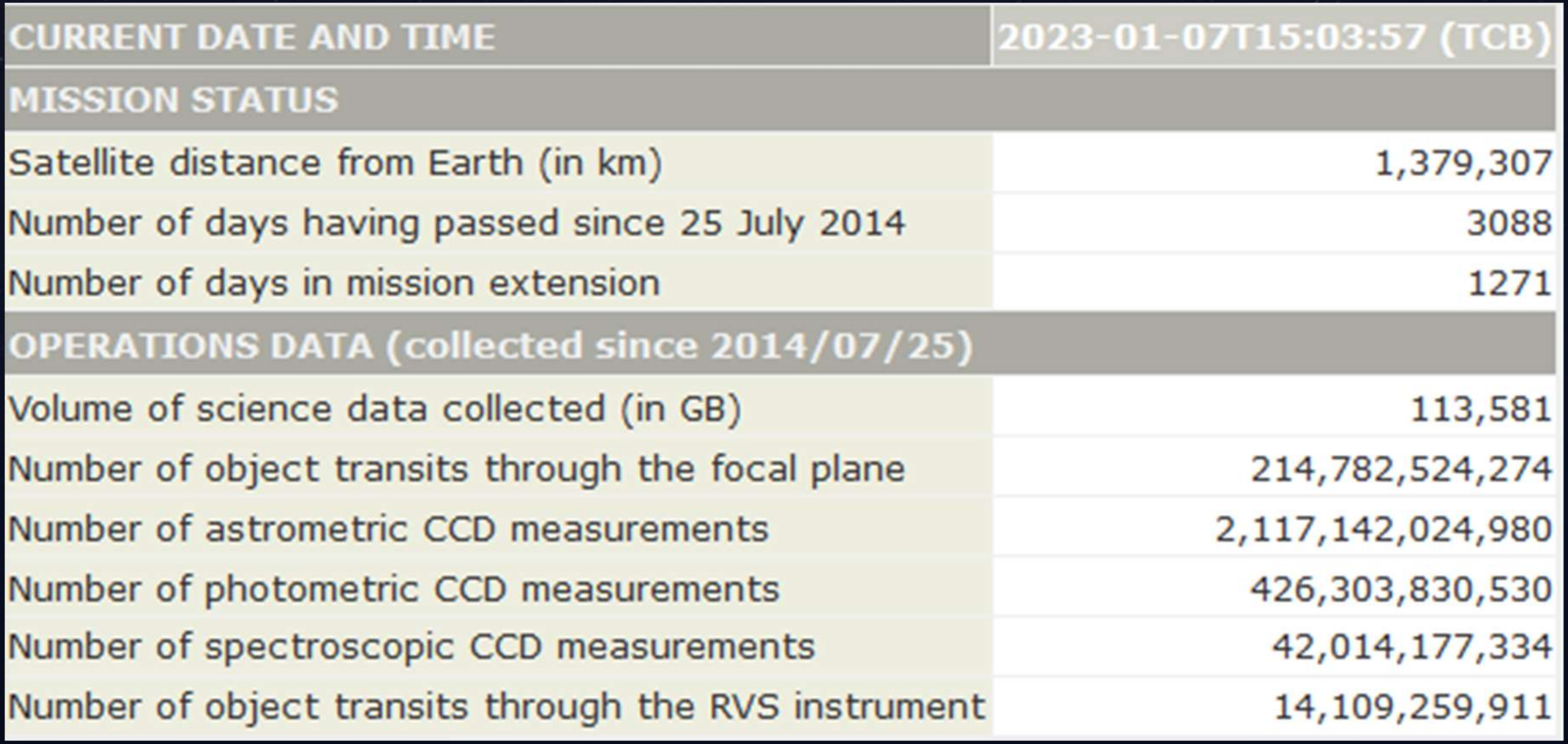
SAGITTARIUS STAR STREAM
The Sagittarius dwarf galaxy is being pulled apart by the Milky Way's gravity, with its stars strung out along its orbit. Many other streams from long-dead dwarfs loop through the outer halo.

THE BIG PICTURE
Recent data are illuminating the Milky Way's structure, including its bright disk and the fainter features surrounding it.

Nuestra galaxia, una visión actualizada

Gaia en números

Big data



CURRENT DATE AND TIME	2023-01-07T15:03:57 (TCB)
MISSION STATUS	
Satellite distance from Earth (in km)	1,379,307
Number of days having passed since 25 July 2014	3088
Number of days in mission extension	1271
OPERATIONS DATA (collected since 2014/07/25)	
Volume of science data collected (in GB)	113,581
Number of object transits through the focal plane	214,782,524,274
Number of astrometric CCD measurements	2,117,142,024,980
Number of photometric CCD measurements	426,303,830,530
Number of spectroscopic CCD measurements	42,014,177,334
Number of object transits through the RVS instrument	14,109,259,911

https://www.cosmos.esa.int/web/gaia/mission-numbers

Contexto **04**

El grupo Gaia en la UB

Gaia@UB

http://gaia.ub.edu

Equipo Gaia en la UB

IPs: X. Luri, C. Jordi, F. Figueras

~25 miembros
Científicos/Ingenieros

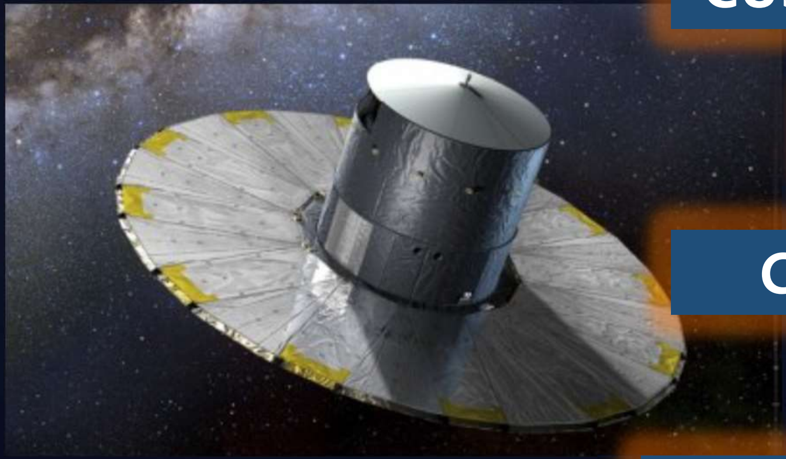
Equipo multidisciplinar
Ciencia y Tecnología

El proyecto científico Gaia@UB

Historia y evolución de la Galaxia
- historia formación estelar
- función de masa inicial
- disco y su ondulación
- barra central
- brazos espirales
- acreción de galaxias
- poblaciones estelares
- cúmulos estelares
- determinación de distancias

Desarrollo tecnológico
- compresión
- emparejamiento de observaciones
- minería de datos
- inteligencia artificial

Gaia



CU2 - Simuladores

CU3 - Astrometria

CU5 - Fotometria

CU9 - Archivo

Big data

Minería de datos

HPC/HTC

Cloud computing

Metodología

La misión **05**

Procesado de datos

Las entregas de datos

Data Releases

https://www.cosmos.esa.int/web/gaia/release

Lanzamiento 2013

Fin de misión~2025

Aprendiendo sobre los datos y el procesado / más recursos requeridos

DR1
Sep. 2016

DR2
Apr. 2018

EDR3
Dec. 2020

DR3
2022

DR4

Final
catalogue
~2030

Más tránsitos, más base temporal, mejores calibraciones → más precision, menos sistemàtica

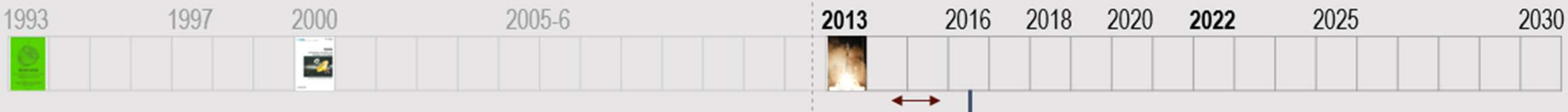
La elección de este modelo de procesado no fue una decisión técnica sino política:

- ESA no quería pasar tanto tiempo sin publicar datos
- El equipo científico prefería un modelo con una entrega única, basado en la experiencia de Hipparcos (no publicar hasta que se consiga la máxima precisión)
- Ganó ESA

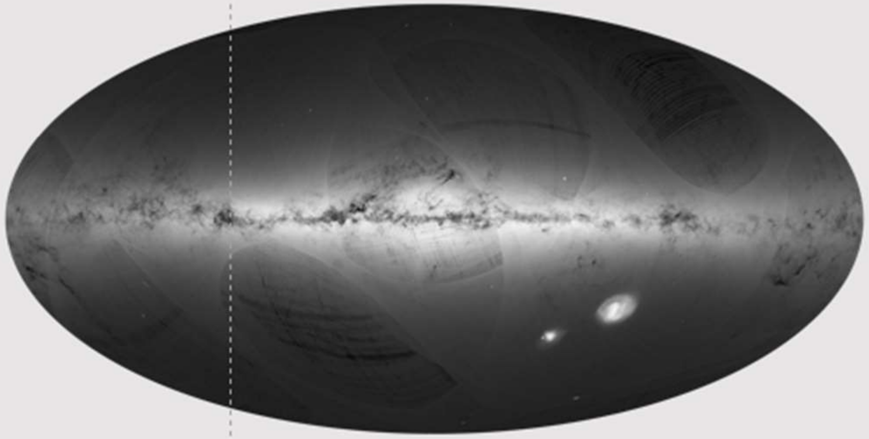
En mi opinión ha sido positivo

- Modelo Waterfall developing vs. Cyclic Prototyping; nos permite probar, aprender, repetir y mejorar
- Pero tiene más coste, cada entrega de datos es un esfuerzo enorme

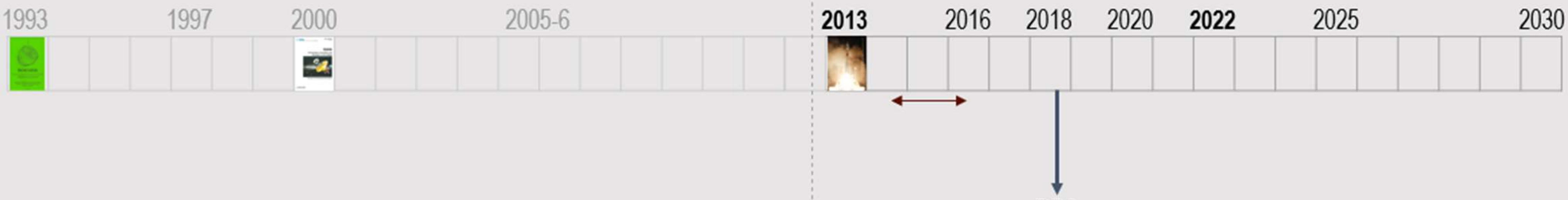
A brief history of Gaia


1993
1997
2000
2005-6
2013
2016
2018
2020
2022
2025
2030

DR1
13.7 months
29B transits used
2M parallaxes +
motions
1.1B positions



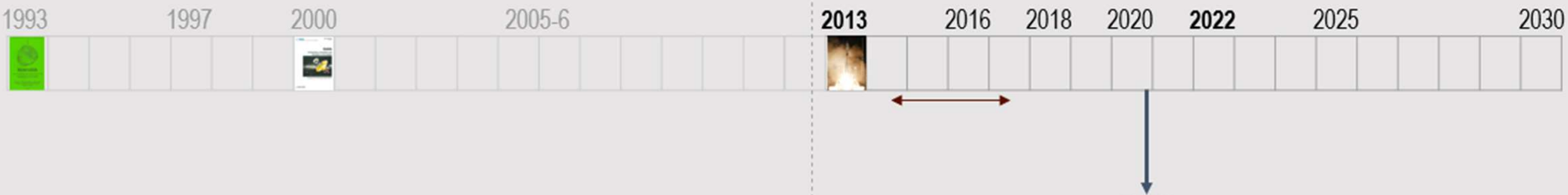
A brief history of Gaia


1993
1997
2000
2005-6
2013
2016
2018
2020
2022
2025
2030

DR2

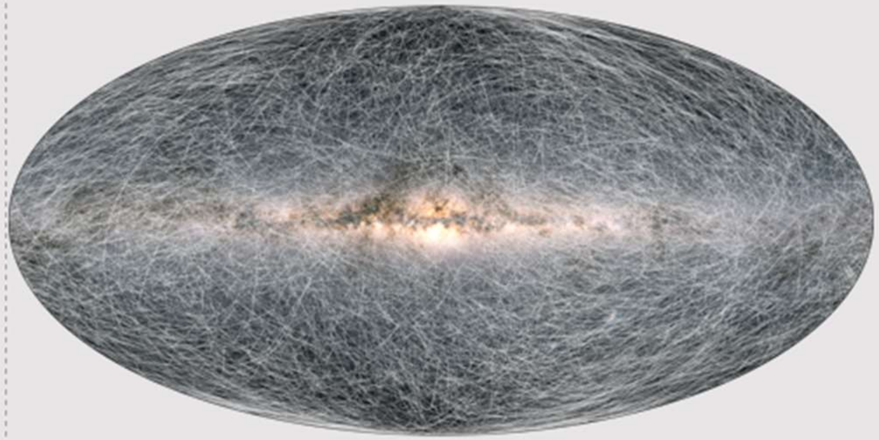
22 months
57B transits used (3.7B RVS)
1.3B parallaxes, motions and colours
7M radial velocities
Variables, asteroids, astrophys. params…

A brief history of Gaia


1993
1997
2000
2005-6
2013
2016
2018
2020
2022
2025
2030

EDR3

34 months
78B transits used (4.8B RVS)
1.4B parallaxes, motions and colours
7M radial velocities (DR2)



A brief history of Gaia

1993
1997
2000
2005-6
2013
2016
2018
2020
2022
2025
2030

DR3
See next slides!

Looking ahead

Spacecraft end-of-life estimated for ~H1 2025

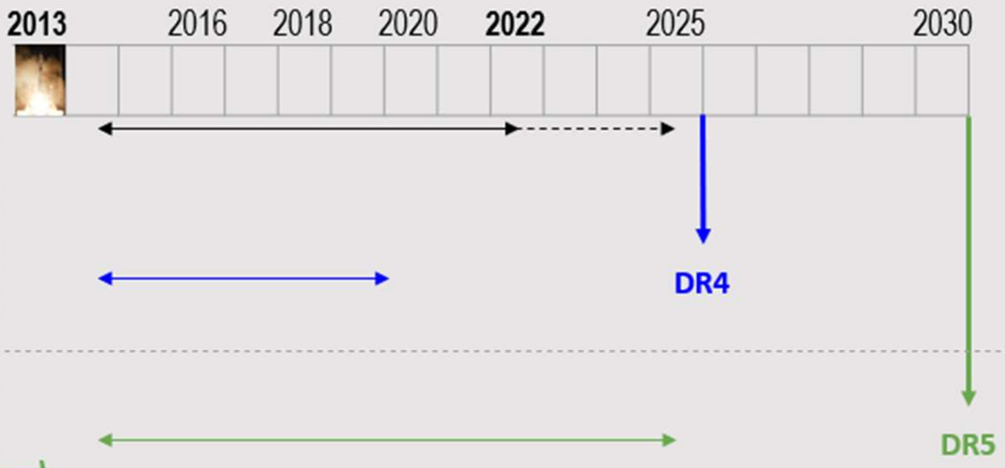
Final mission extension to be approved next year

DR4 expected ~end 2025

All data from nominal mission (5.5 years)

DR5 expected ~end 2030

All data from extended mission (~10.5 years)


2013
2016
2018
2020
2022
2025
2030
DR4
DR5

Some cyclic DPAC systems (mainly astrometry) well advanced towards DR4

Daily DPAC systems already processing DR5 data since 2020

La última entrega **06**

Gaia Data Release 3

GAIA EARLY DATA RELEASE 3

esa

1 811 709 771
stellar positions

1 806 254 432
brightness
in white light

1 542 033 472
brightness
in blue light

1 540 770 489
colour

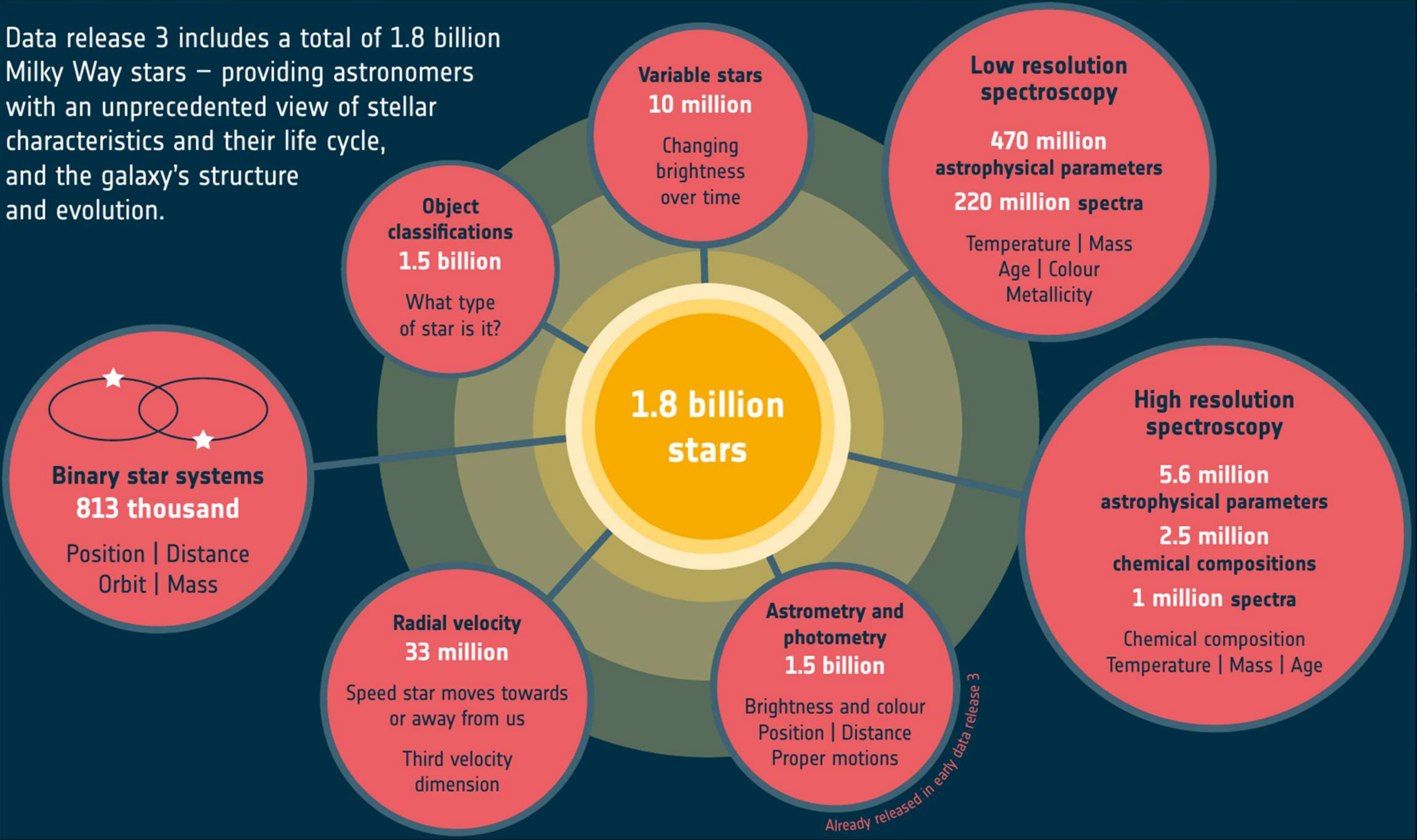
1 467 744 818
parallax and
proper motions

1 614 173
extragalactic
sources

1 554 997 939
brightness
in red light

#SpaceCare #ExploreFarther

Data release 3 includes a total of 1.8 billion Milky Way stars – providing astronomers with an unprecedented view of stellar characteristics and their life cycle, and the galaxy's structure and evolution.



1.8 billion stars

Variable stars
10 million
Changing brightness over time

Low resolution spectroscopy
470 million astrophysical parameters
220 million spectra
Temperature | Mass
Age | Colour
Metallicity

Object classifications
1.5 billion
What type of star is it?

Binary star systems
813 thousand
Position | Distance
Orbit | Mass

High resolution spectroscopy
5.6 million astrophysical parameters
2.5 million chemical compositions
1 million spectra
Chemical composition
Temperature | Mass | Age

Radial velocity
33 million
Speed star moves towards or away from us
Third velocity dimension

Astrometry and photometry
1.5 billion
Brightness and colour
Position | Distance
Proper motions

Already released in early data release 3

Rompiendo récords

Más allá del estudio astrométrico y fotométrico más grande y preciso hasta la fecha:

- El estudio espectrofotométrico más grande de la historia
- El estudio de velocidad radial más grande de la historia
- Primer estudio espacial de todo el cielo de QSO y galaxias en el universo local
- Estudio espectrofotométrico-dinámico de mayor precisión de asteroides
- El survey más grande para muchas clases de estrellas variables
- La mayor colección de datos astrofísicos para estrellas MW
- Estudio de estrellas no individuales que supera todo el trabajo de NSS de los últimos dos siglos

DR3 en números

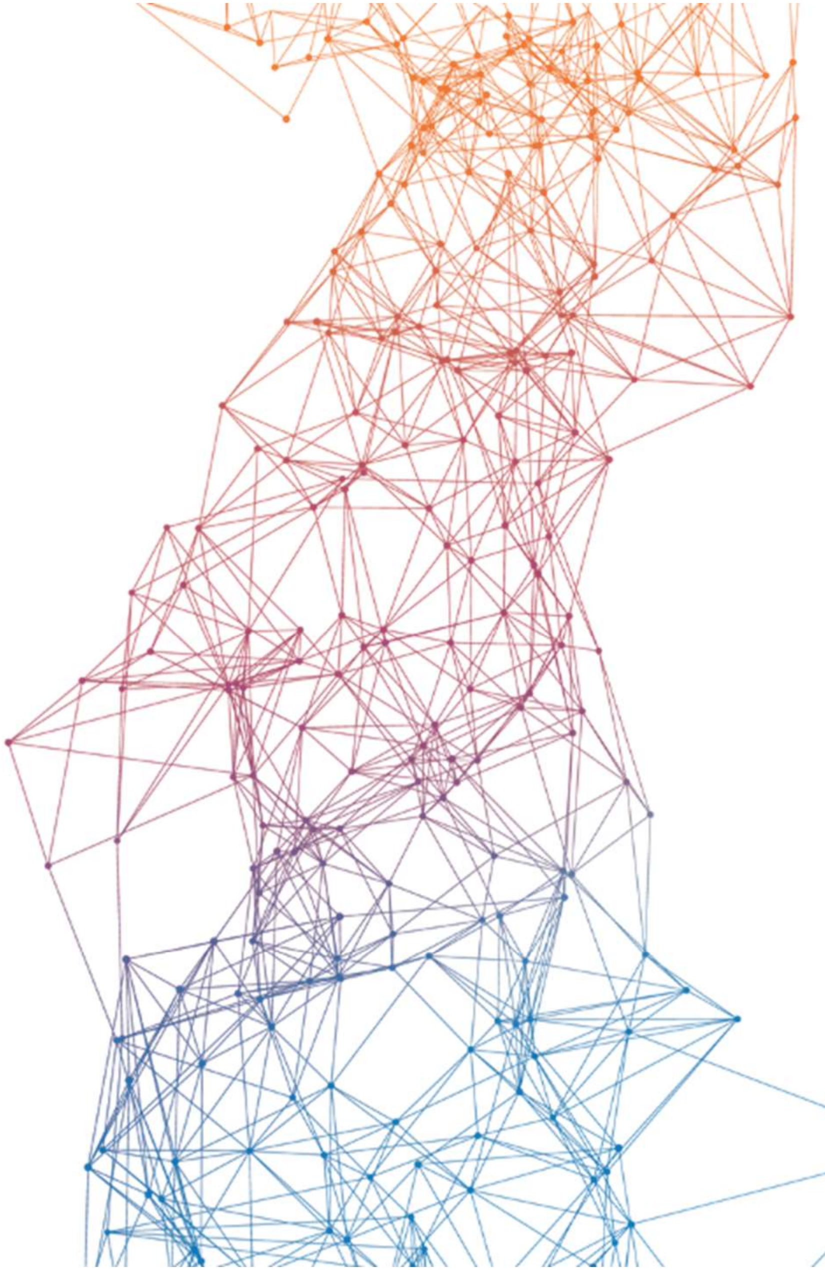
- Descarga masiva del catálogo DR3: 9 TB
- Más de 90 tablas publicadas
- 78 mil millones de tránsitos utilizados:
  - 770 mil millones de mediciones astrométricas + fotometría G
  - 156 mil millones de mediciones espectrofotométricas (BP/RP)
  - 14 mil millones de espectros de alta resolución (RVS)
- 41 TB de datos de la sonda
- Más de 1 ZettaFLOP consumido (10^{21} operaciones de coma flotante)

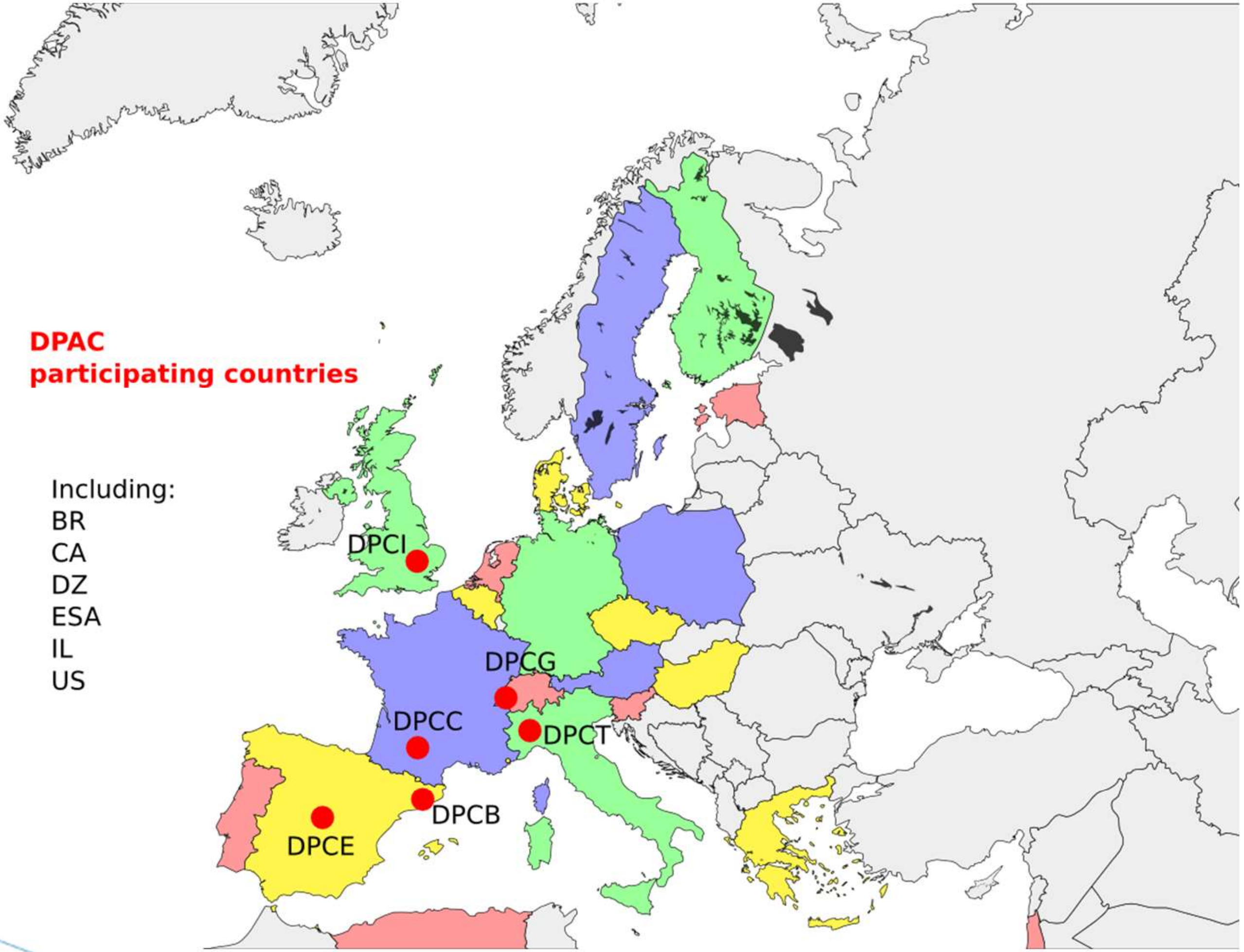
El procesado de datos

06

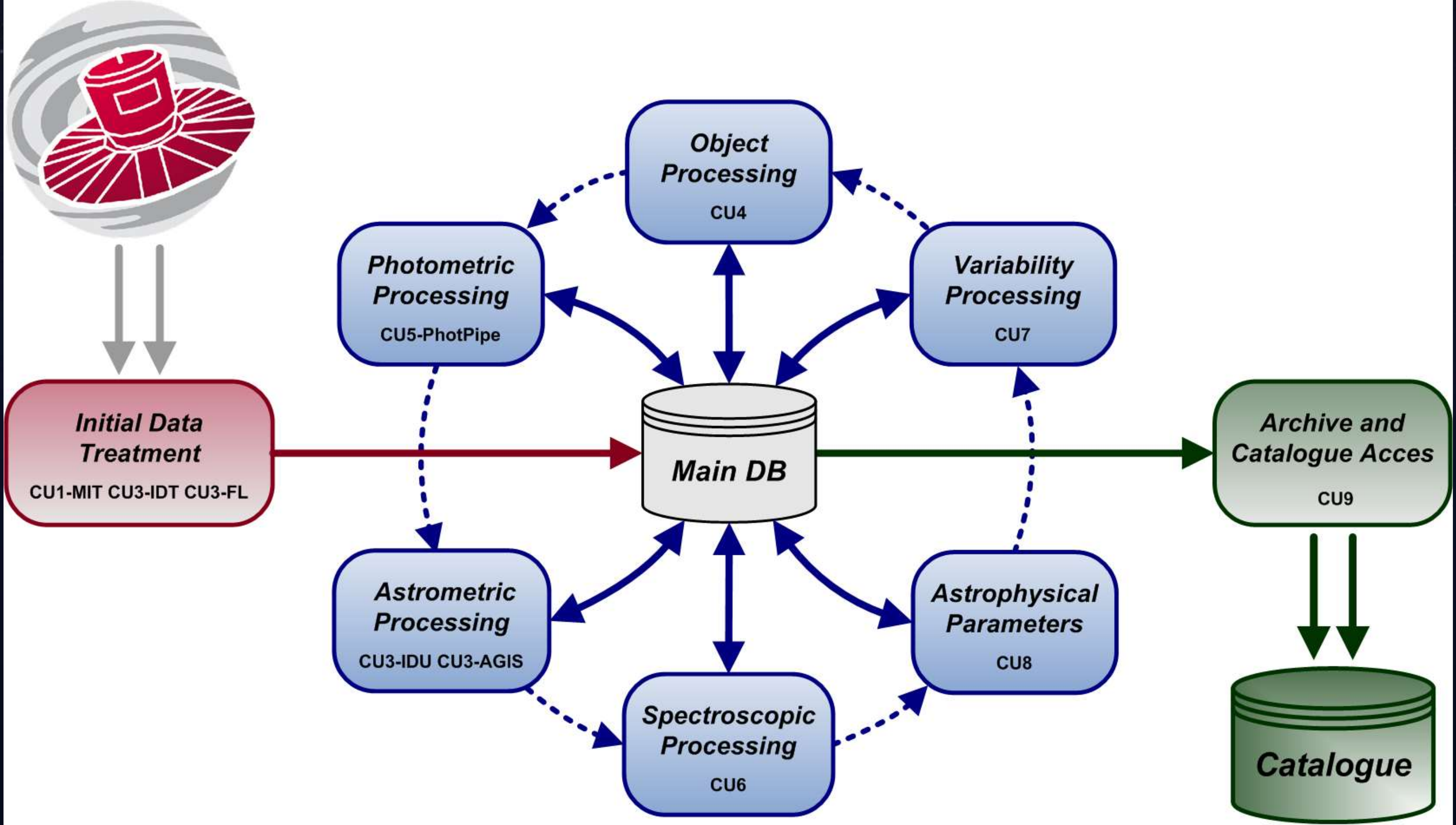
Los centros de datos de DPAC

Los centros de datos de Gaia-DPAC




DPAC
participating countries
Including:
BR
CA
DZ
ESA
IL
US
DPCI
DPCG
DPCC
DPCT
DPCB
DPCE

Concepto de procesado: “hub and spokes”


Object Processing
CU4
Photometric Processing
CU5-PhotPipe
Variability Processing
CU7
Initial Data Treatment
CU1-MIT CU3-IDT CU3-FL
Main DB
Archive and Catalogue Acces
CU9
Astrometric Processing
CU3-IDU CU3-AGIS
Astrophysical Parameters
CU8
Spectroscopic Processing
CU6
Catalogue

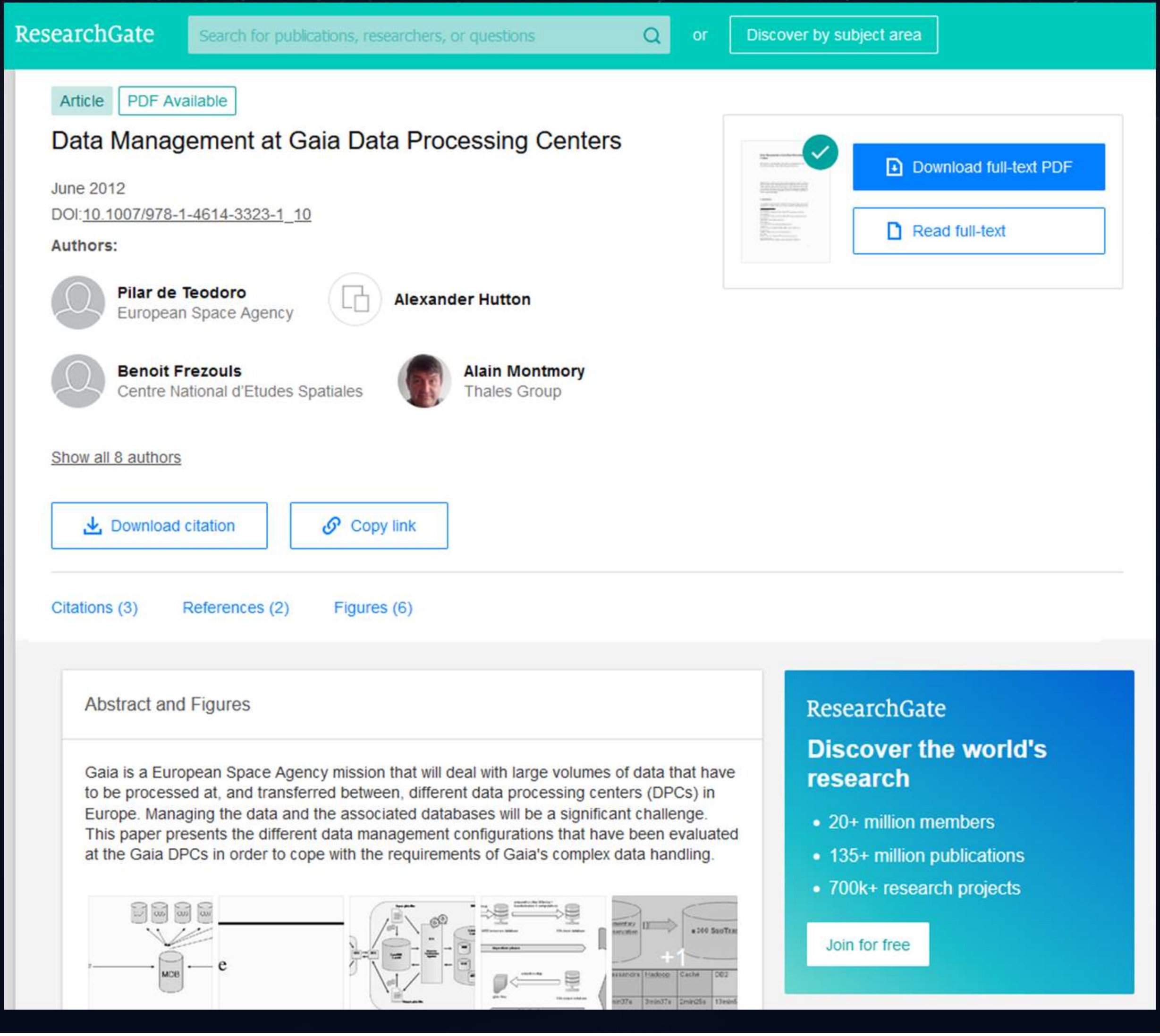
La elección de este modelo de centros de datos no fue una decisión técnica sino política:

- Técnicamente hubiese sido preferible centralizar el procesado en un único centro de datos
- Políticamente era imposible que las agencias de financiación lo aceptasen. Cada una quería financiar el procesado en su territorio.

En mi opinión ha sido un "mixed blessing"

- Cada centro se ha organizado alrededor de los científicos expertos en un tema, con ESA como eje central de datos
- La coordinación del procesado es más complicada; la gestión del calendario de procesado por el PO es una pesadilla, pero en DR3 empezamos a tenerlo controlado ☺

https://www.researchgate.net/publication/258724991_Data_Management_at_Gaia_Data_Processing_Centers


ResearchGate
Search for publications, researchers, or questions
or
Discover by subject area
Article
PDF Available
Download full-text PDF
Read full-text
Download citation
Copy link
Citations (3)
References (2)
Figures (6)
ResearchGate
Discover the world's research
20+ million members
135+ million publications
700k+ research projects
Join for free

Data Management at Gaia Data Processing Centers

June 2012

DOI:10.1007/978-1-4614-3323-1_10

Authors:

Pilar de Teodoro European Space Agency

Alexander Hutton

Benoit Frezouls Centre National d'Etudes Spatiales

Alain Montmory Thales Group

Show all 8 authors

Abstract and Figures

Gaia is a European Space Agency mission that will deal with large volumes of data that have to be processed at, and transferred between, different data processing centers (DPCs) in Europe. Managing the data and the associated databases will be a significant challenge. This paper presents the different data management configurations that have been evaluated at the Gaia DPCs in order to cope with the requirements of Gaia's complex data handling.

MDB + “Hub and spokes”

- La MDB (ESAC) es el repositorio de datos de referencia
- Control estricto de versiones de los datos
- Los datos viajan entre MDB y los centros de datos: se mandan datos para procesado y se reciben resultados
- Centros productores / centros consumidores
- Posible gracias a la red europea de investigación (RedIris/GEANT)
- Uso de herramientas especializadas (ASPERA)

Ingenieria de software: organizando el consorcio

Una unidad especializada (CU1) y un grupo de supervisión (Project Office) se encargan de asegurar que el desarrollo de software de DPAC se haga bajo estándares de calidad.

Choque de culturas:

- Ingenieros de SW vs. Científicos: estándares vs “que funcione”
- Lucha por la documentación (como con el cinturón de seguridad se aprende mucho cuando te estrellas, si sobrevives)
- Después de años de funcionamiento se ha conseguido que los equipos adquieran una cultura de desarrollo de software común.
- Es una cultura flexible, se adapta a los distintos caso (e.g. ESA vs. Validadores)

Algunos elementos

Modelo de datos de la MBD
- Crucial para el desarrollo conjunto
- Herramienta especializada para su gestión

Decisión temprana del uso de un lenguaje común: Java
- Polémica inicial por su rendimiento (interpretado vs. compilado)
- Lenguaje con buen control formal, ventajoso (C nos hubiese matado; FORTRAN ni te cuento)
- Cuidado con la planificación a muy largo plazo (compra de Sun por Oracle)
- Millones de líneas de código

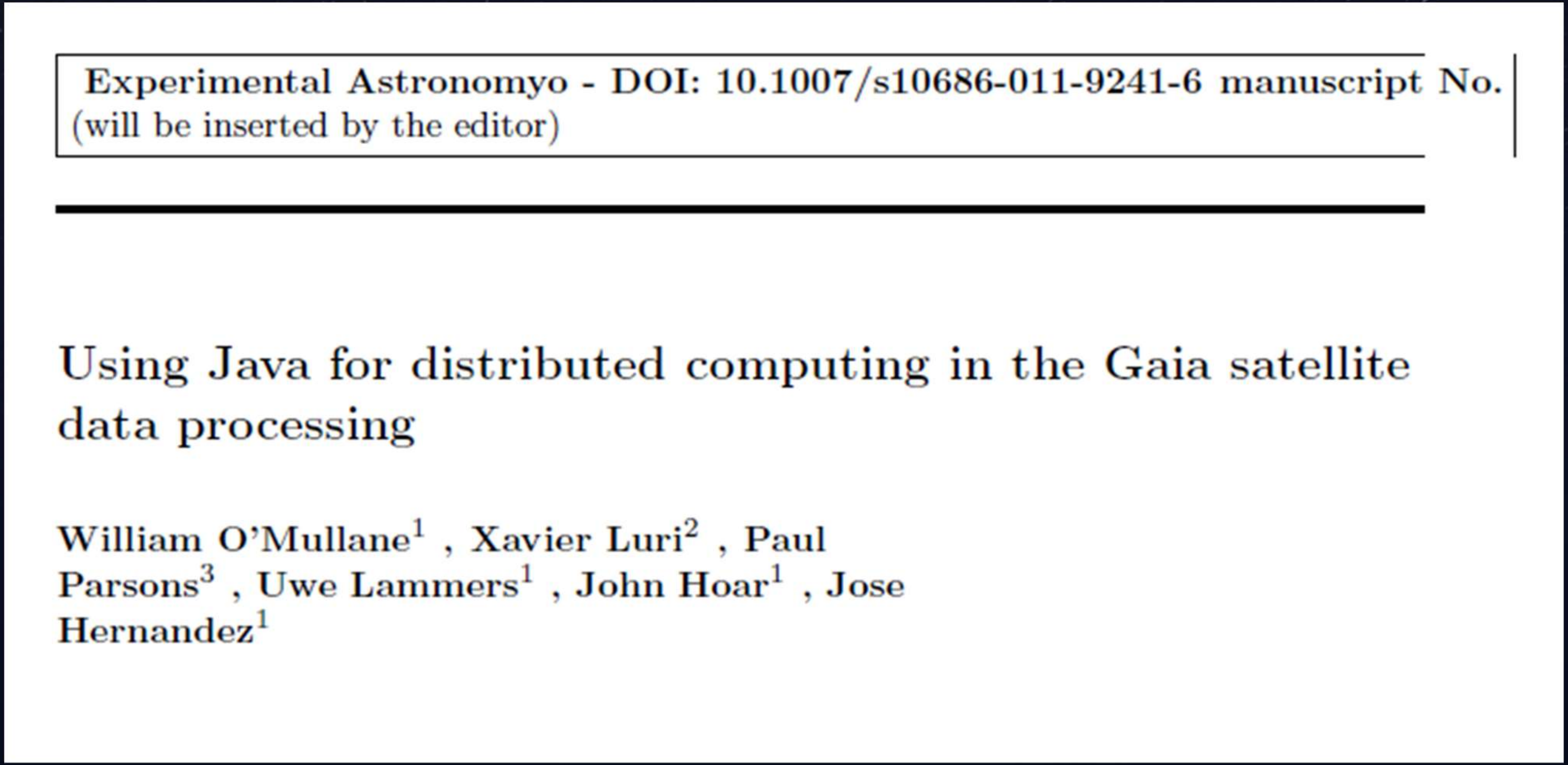
Librerias desarrolladas en Java:
- GaiaTools: herramientas comunes, de uso obligatorio (ej.. cambios de coordenadas)
- GaiaParamDB: parámetros comunes, de uso obligatorio (ej.. constantes físicas)

Aplicación flexible de ECSS (European Cooperation for Space Standardization) en desarrollo y documentación

JIRA, Jenkins, SVN

Cuidado con los temas legales: licencias!!!

2011



Experimental Astronomyo - DOI: 10.1007/s10686-011-9241-6 manuscript No.
(will be inserted by the editor)

Using Java for distributed computing in the Gaia satellite data processing

William O'Mullane[1] , Xavier Luri[2] , Paul Parsons[3] , Uwe Lammers[1] , John Hoar[1] , Jose Hernandez[1]

El procesado de datos **07**

DPCE (ESAC, España)

IDT → Daily monitoring System (desarrollado en BCN)

AGIS → Implement the astrometric solution

Main Database → ¡Todos los datos!

Archivo principal → Acceso público

Processing nodes

180 servers of type:

Dell PowerEdge C6420/R610/R630/R740/R820/R840

IBM NeXtScale nx360

IBM System X iDataPlex dx360

Lenovo NeXtScale nx360

With 3600 CPU cores, 11TB of memory in total (providing ~6TFLOPS).

AGIS Core Processing

$t_1, t_2, \ldots, t_{20}$

100 million “well-behaved” sources

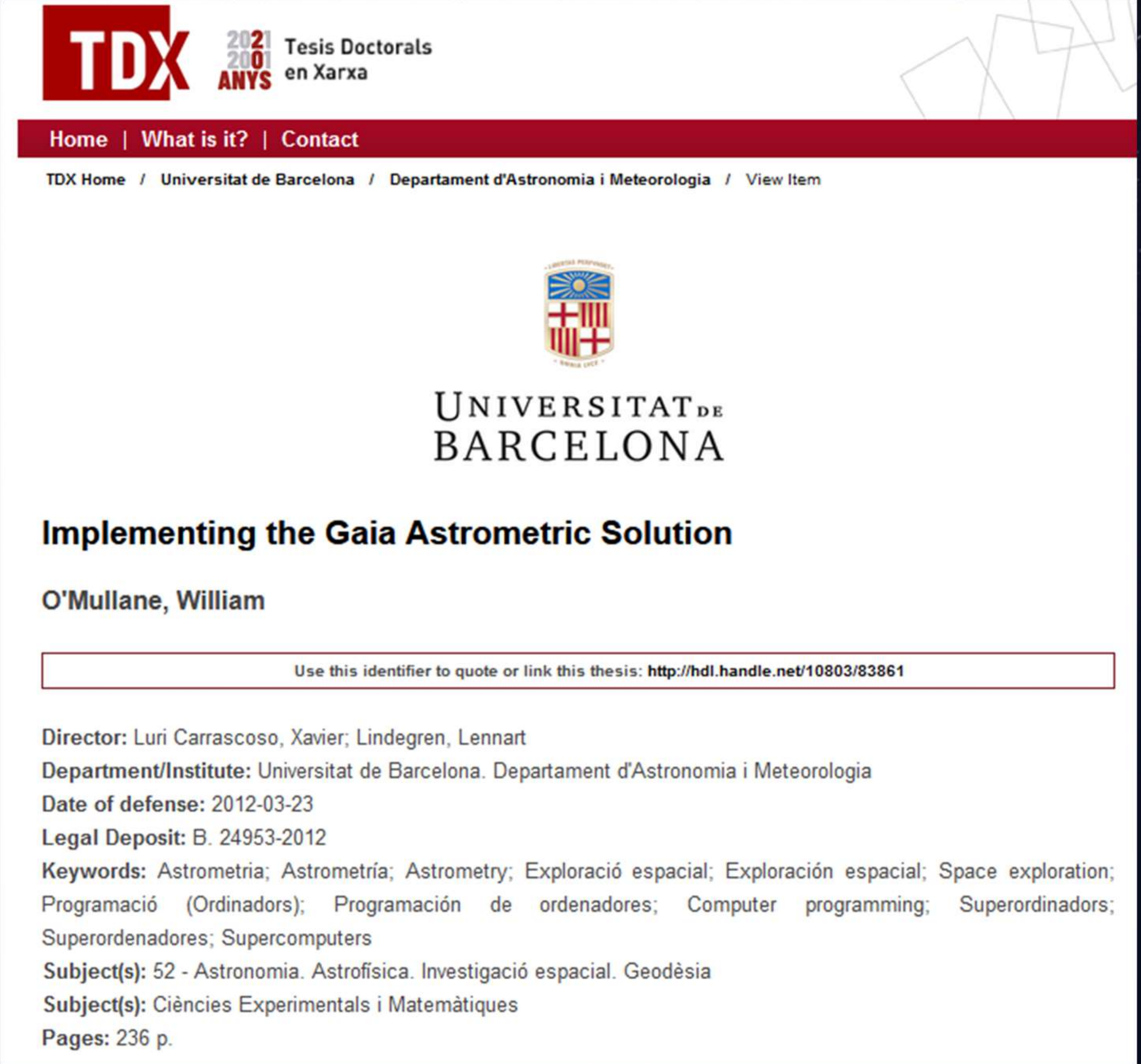
Astrometric solution

Estimates parameters of 4 models:

(S) Source	5 x 10^9 param
(A) Attitude	~ 10^8 param
(C) Calibration	~ 10^6 param
(G) Global	< 10^2 param

$$\min_{\boldsymbol{s},\boldsymbol{a},\boldsymbol{c},\boldsymbol{g}} \sum_l \left[\frac{t_l - f_l(\boldsymbol{s},\boldsymbol{a},\boldsymbol{c},\boldsymbol{g})}{\sigma_l}\right]^2$$

- Least squares solution:
  10^{10} parameters using 10^{12} observations,
- direct solution unfeasible
  (Bombrun et al. 2011),
- use Astrometric Global Iterative Solution:
  AGIS (Lindegren et al. 2012)


TDX
2021 2001 ANYS
Tesis Doctorals en Xarxa
Home | What is it? | Contact
TDX Home / Universitat de Barcelona / Departament d'Astronomia i Meteorologia / View Item
UNIVERSITAT DE BARCELONA

Implementing the Gaia Astrometric Solution

O'Mullane, William

Use this identifier to quote or link this thesis: **http://hdl.handle.net/10803/83861**

Director: Luri Carrascoso, Xavier; Lindegren, Lennart

Department/Institute: Universitat de Barcelona. Departament d'Astronomia i Meteorologia

Date of defense: 2012-03-23

Legal Deposit: B. 24953-2012

Keywords: Astrometria; Astrometría; Astrometry; Exploració espacial; Exploración espacial; Space exploration; Programació (Ordinadors); Programación de ordenadores; Computer programming; Superordinadors; Superordenadores; Supercomputers

Subject(s): 52 - Astronomia. Astrofísica. Investigació espacial. Geodèsia

Subject(s): Ciències Experimentals i Matemàtiques

Pages: 236 p.

El procesado de datos

08

DPCB (Barcelona, España)

- Run CU3 Intermediate Data Updating (IDU)
- IDU products delivery downstream
- Run CU2 GAia System-level Simulator (GASS)
- Run CU2 Gaia Object Generator (GOG)

- Data Reception from DPCE on daily basis:
    - Raw Data from the spacecraft and intermediate data, calibrations and reference data from other downstream system
- Backup of the full Telemetry archive and DPAC data (more than 1PB for the full mission)

Processing hardware

DPCB uses the resources of the Barcelona Supercomputing Center (BSC, strategic project, ~10Mhoras of CPU per year)

Mainly the MareNostrum supercomputer; we have gone through 4 updates. The current one, MareNostrum 4, has a maximum yield of 1.1 Petaflops. It consists of 48,896 Intel Sandy Bridge processors on 3,056 nodes, including 84 Xeon Phi 5110P on 42 nodes, with more than 115 TB of main memory and 2 PB of GPFS disk storage.

MareNostrum provides massive processing resources, but also has some limitations:

- No exclusive processing resources; Provides a shared environment with other users: jobs run through a queue
- Need to plan and reserve resources well in advance (2-3 months)
- There is no database, so file I/O is key: tasks must be equalized for good performance

Latest activities

- For the final execution of IDU for eDR3/DR3, up to 300 ExaFLOPs (300E18 FLOPs (3E20)) have been required… ~1 ZettaFLOP if we take into account the bootstrap execution and test campaign
- For DR4, IDU will process twice as much data and we will perform several iterations with more complex and sophisticated algorithms and require up to 3E21 FLOPs to complete all activities.
- The overall Flops estimate for the full mission will easily reach 1e22 FLOPs (100 ZettaFLOPs)

El procesado de datos

09

DPCC (Toulouse, Francia)

DPCC scientific tasks

DPCC hosts 7 scientific tasks, done by 3 CUs

[CU4]

- Detection of Solar System Objects
- Characterization of Non Single Stars (binaries at least), quasars, galaxies and gravitational lenses
- Characterization of quasars
- Characterization of galaxies
- Characterization of gravitational lenses

[CU6]

- Calibration of the Spectrometer and computation of radial and rotational velocities

[CU8]

- Classification of objects and determination of the astrophysical parameters

DPCC and big data solutions

- **A proven solution from web companies to deal with huge data volume**
  - Horizontal linear scalability
    - For storage and computing power growth
  - Parallelization with process localization
    - The process goes to the data
  - Map Reduce
    - complex data manipulation (joins, sort)
  - HDFS: distributed data storage
  - Computing resources management

- To simplify the data manipulation queries on top of Hadoop !

❖ **As an intermediate filesystem to manage gbin files**

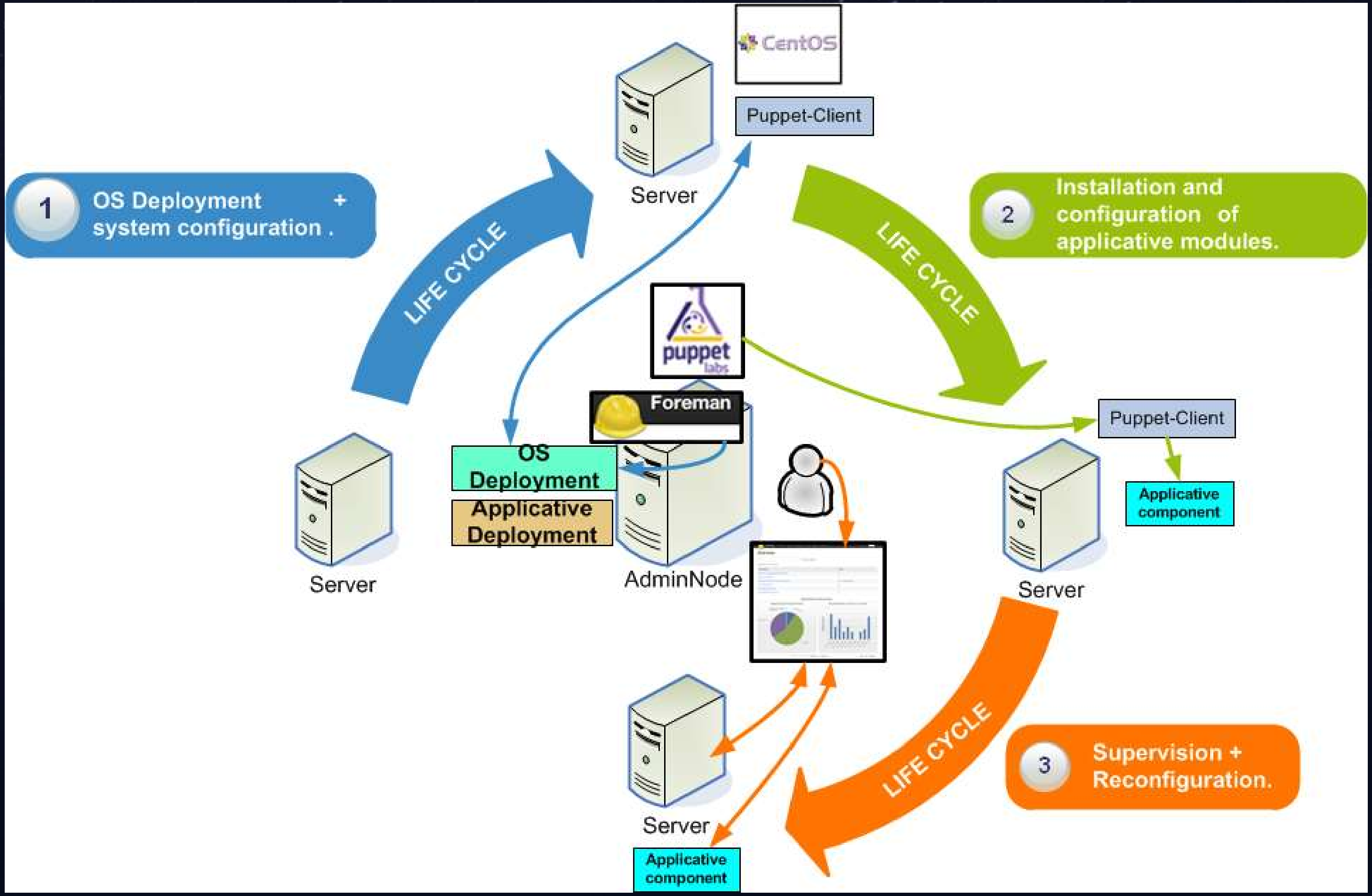
❖ **To manage ephemeris from iMCCE**

❖ **To manage a huge log volume, generated by big computes**

DPCC Operational Platform

- Based on standard servers with
- Servers bought progressively, according to computing and storage needs
  - 32 Datanodes (calculus nodes) for the cycle 0
  - 68 Datanodes after next month for the beginning of the cycle 1, with up to date technologies (cpu,disk…)
- And now for the cycle 4…
  - 225 Datanodes . 6 generations of servers (2013 ⇨ 2021), more than 5000 cores
  - High density (24 cpu/ 18 To /128 Go ⇨ 96 cpu / 192 To / 1024 Go)
  - 15 Management servers
  - 14 switchs and a network throughput = 10 Gbps
  - Storage = 1000 disks
  - High availability : 24/24 hours and 7/7 days

DPCC platform management

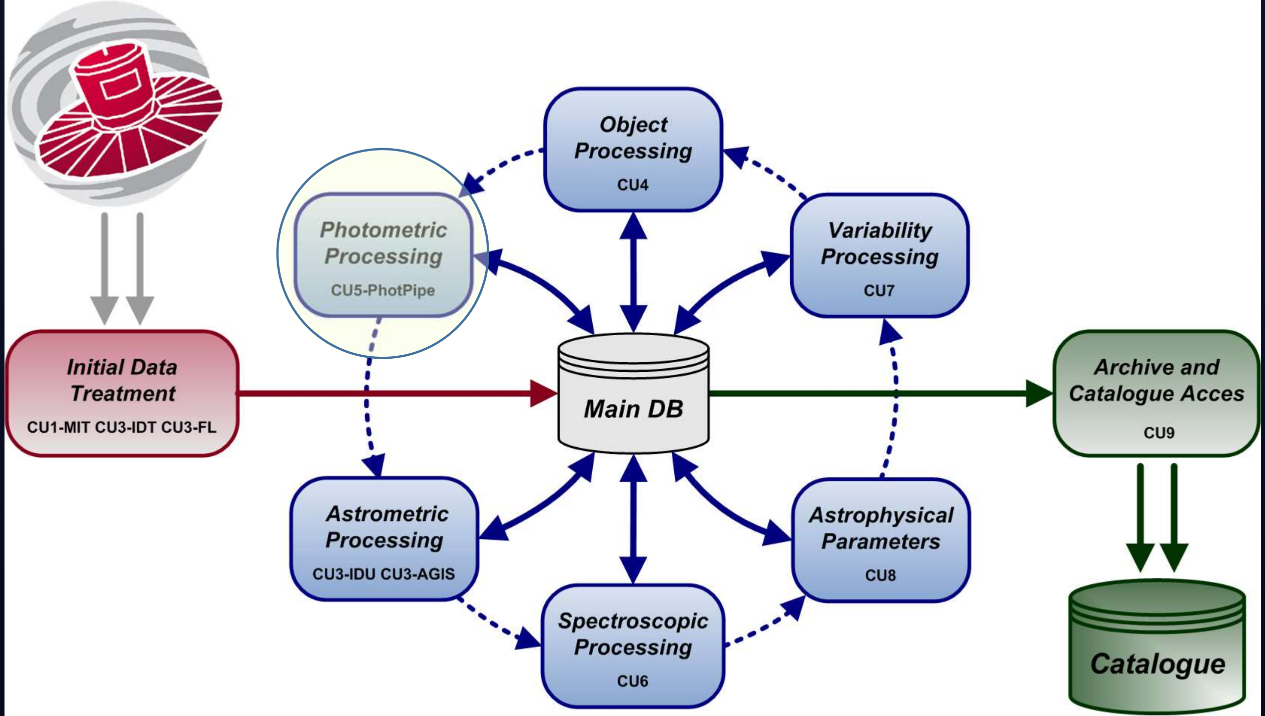

CentOS
Puppet-Client
Server
1 OS Deployment + system configuration .
LIFE CYCLE
2 Installation and configuration of applicative modules.
LIFE CYCLE
puppet labs
Foreman
OS Deployment
Applicative Deployment
Server
AdminNode
Puppet-Client
Applicative component
Server
LIFE CYCLE
3 Supervision + Reconfiguration.
Server
Applicative component

El procesado de datos

10

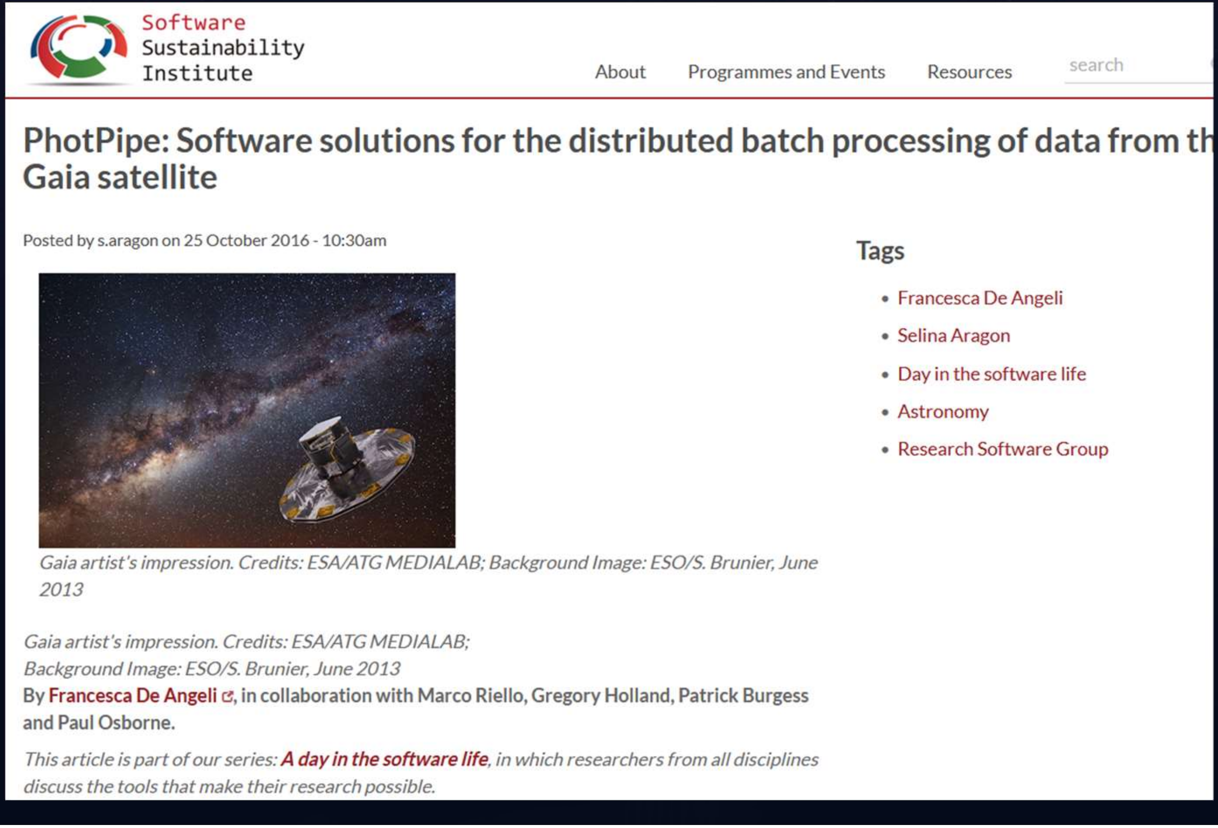
DPCI (Cambridge, Reino Unido)

DPCI implements the photometric processing


Object Processing
CU4
Photometric Processing
CU5-PhotPipe
Variability Processing
CU7
Initial Data Treatment
CU1-MIT CU3-IDT CU3-FL
Main DB
Archive and Catalogue Acces
CU9
Astrometric Processing
CU3-IDU CU3-AGIS
Astrophysical Parameters
CU8
Spectroscopic Processing
CU6
Catalogue

PhotPipe is the main processing system. Currently using a mixture of Hadoop/Spark, but moving to pure spark for future processing. A basic description is provided in this article:

https://www.software.ac.uk/blog/2016-11-23-photpipe-software-solutions-distributed-batch-processing-data-gaia-satellite



PhotPipe: Software solutions for the distributed batch processing of data from th Gaia satellite

Posted by s.aragon on 25 October 2016 - 10:30am

Gaia artist's impression. Credits: ESA/ATG MEDIALAB; Background Image: ESO/S. Brunier, June 2013

Gaia artist's impression. Credits: ESA/ATG MEDIALAB;
Background Image: ESO/S. Brunier, June 2013

By Francesca De Angeli, in collaboration with Marco Riello, Gregory Holland, Patrick Burgess and Paul Osborne.

This article is part of our series: ***A day in the software life****, in which researchers from all disciplines discuss the tools that make their research possible.*

Tags

- Francesca De Angeli
- Selina Aragon
- Day in the software life
- Astronomy
- Research Software Group

New DPCI hardware has been delivered, installed; aim to transition all DPCI functions (including Science Alerts) over next 5 months.

Processing cluster:

- Total cluster memory 55TB
- Total cluster cores 5280 T
- otal cluster local hard drive storage 2.34PB
- Total off-cluster storage : 1.1PB
- Network Full Data Rate Infiniband - 56Gb/s
- Hardware housed in purpose-built University data center with generator power backup.

El procesado de datos

11

DPCG (Ginebra, Suiza)

- Variable stars processing (CU7)
- 3 x enclosures x 45 blades = 135 * 8 cores = 1080 pure CPu nodes, plus 8*40+4*48 = 512 DB cores.
- Gaia is quite special with a dedicated parallel Database Management System deployed on the cluster of 12 nodes and plus number of auxiliary nodes for logging, visualisation and interactive work.

El procesado de datos

12

DPCI (Torino, Italia)

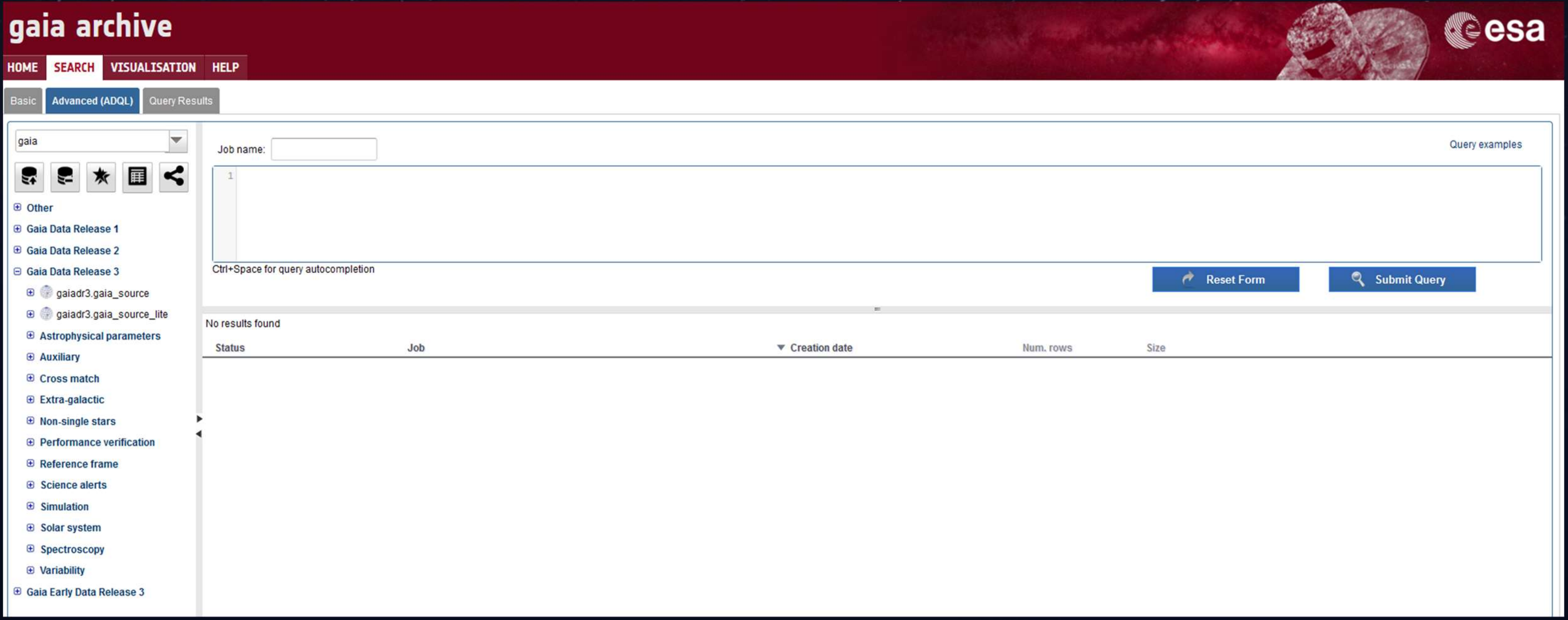
- Alternative astrometric reduction and basic angle analysis (CU3)
- **STORAGE CAPACITY**: 1.5 PB overall raw disk space distributed between two HPE P7400 storage units and one P8400.
- **PROCESSING CAPACITY**: 14 servers HP DL580 G7/G9 with a total of about 600 CPU cores and 4.5TB RAM.
- **DB SERVERS**: 3 HPE DL580 G7 dedicated to the database cluster based on Oracle RAC technology.
- **INFRASTRUCTURE MONITORING AND MANAGEMENT**: services based on VMWare virtual environment configured with two HPE DL 580 G7 servers clustered and managed by vCenter Server.

Publicando los datos

13

El archivo de Gaia en ESAC

A public database interface


gaia archive
esa
HOME
SEARCH
VISUALISATION
HELP
Basic
Advanced (ADQL)
Query Results
gaia
Other
Gaia Data Release 1
Gaia Data Release 2
Gaia Data Release 3
gaiadr3.gaia_source
gaiadr3.gaia_source_lite
Astrophysical parameters
Auxiliary
Cross match
Extra-galactic
Non-single stars
Performance verification
Reference frame
Science alerts
Simulation
Solar system
Spectroscopy
Variability
Gaia Early Data Release 3
Job name:
Query examples
Ctrl+Space for query autocompletion
Reset Form
Submit Query
No results found
Status
Job
Creation date
Num. rows
Size

OPERATIONAL GAIA DB characteristics

MAIN DB Server:

- Model: Dell Power Edge R840
- 64 Cores
- CPU: Intel(R) Xeon(R) Gold 5218 CPU @ 2.30GHz
- RAM: 1.5 TB Physical Memory (RAM DDR4-2666 MHz)
- 2 x 480GB Local SSDs
- 10Gbit network card
- OS: RedHat Enterprise Linux Server 7

BACKUP DB Server:

- Model: Dell Power Edge R840
- 64 Cores
- CPU: Intel(R) Xeon(R) CPU E5-4640 0 @ 2.40GHz
- RAM: 1.2 TB Physical Memory (RAM DDR3-1600 MHz)
- 1 x 400GB Local SSDs
- 10Gbit network card
- OS: RedHat Enterprise Linux Server 7

Postgres database engine

Content Delivery Network for bulk download

Storage

Gaia Uses Hybrid Storage:

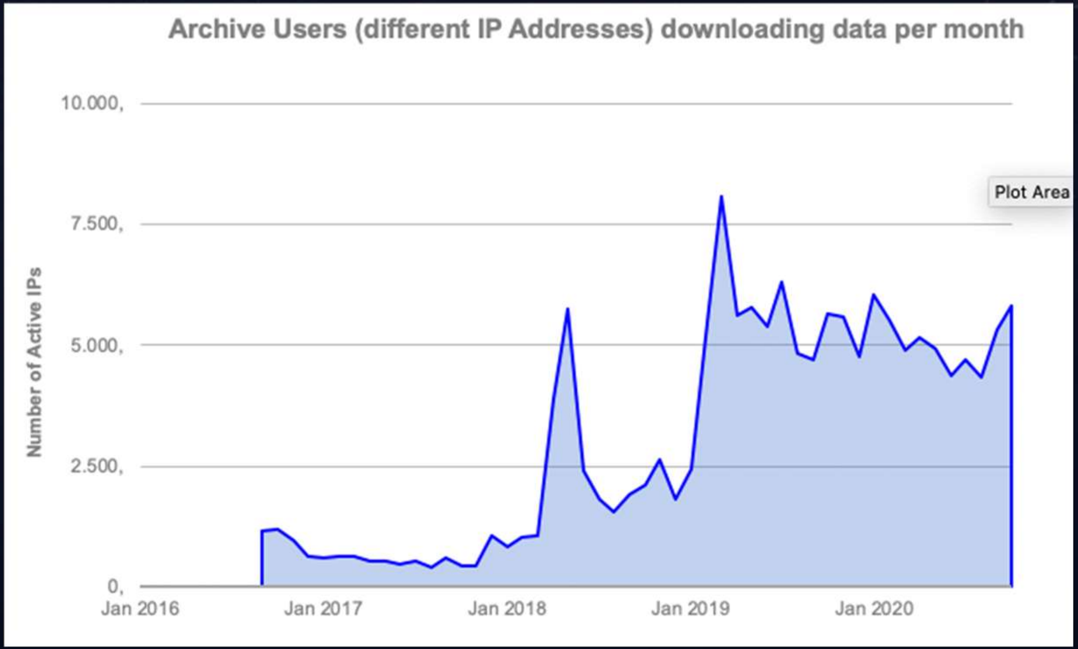
- High Performance Storage (Netapp) for the most demanding applications, the core database technology requiring typical latencies ~<10 ms
- Medium Performance Storage (Isilon Dell) for the rest (not so demanding) applications, requiring typical latencies ~<1000ms

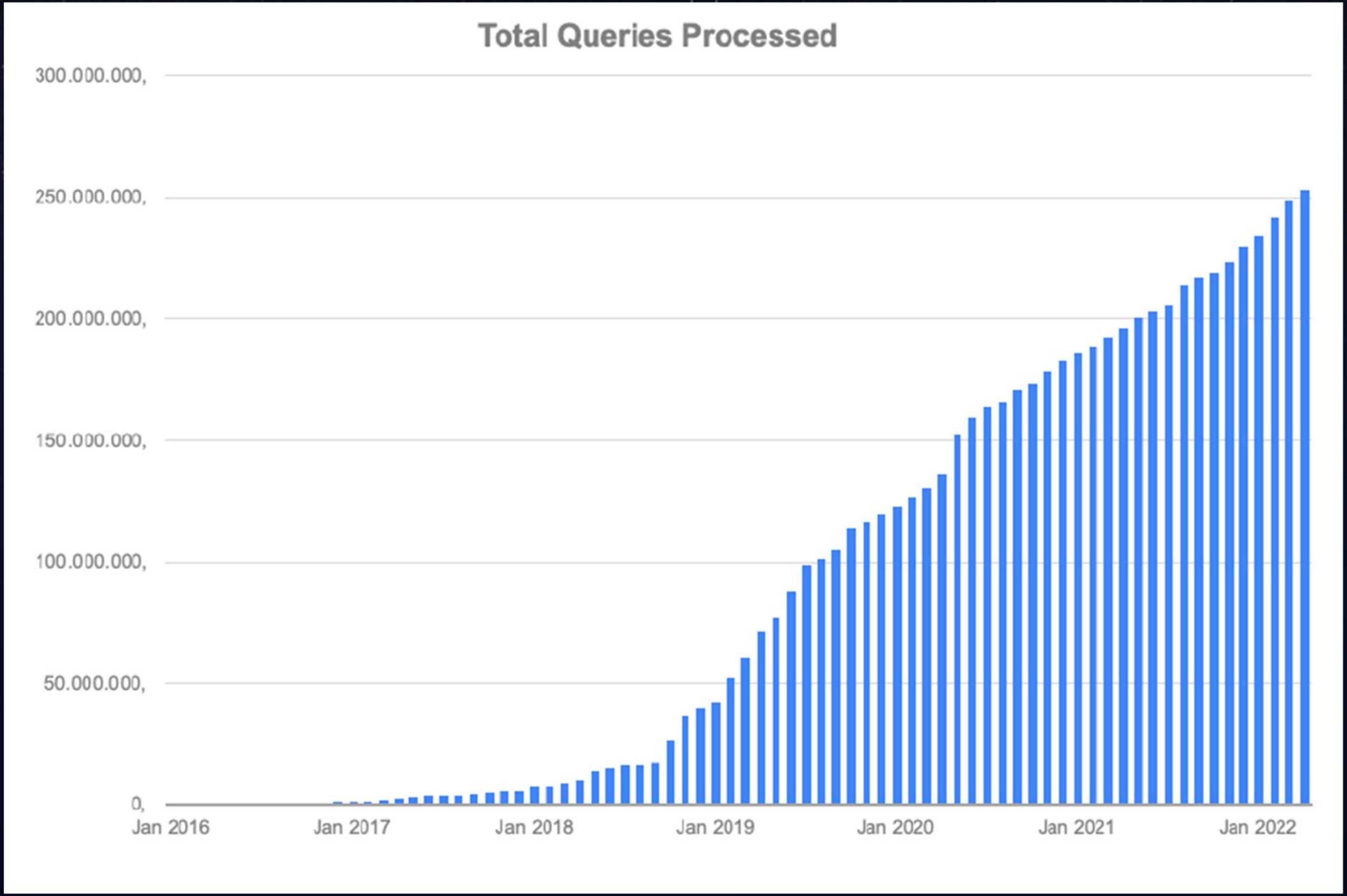
Current 2022 storage: 2.5 PB

Envisaged 2023-2030 storage: ~6PB

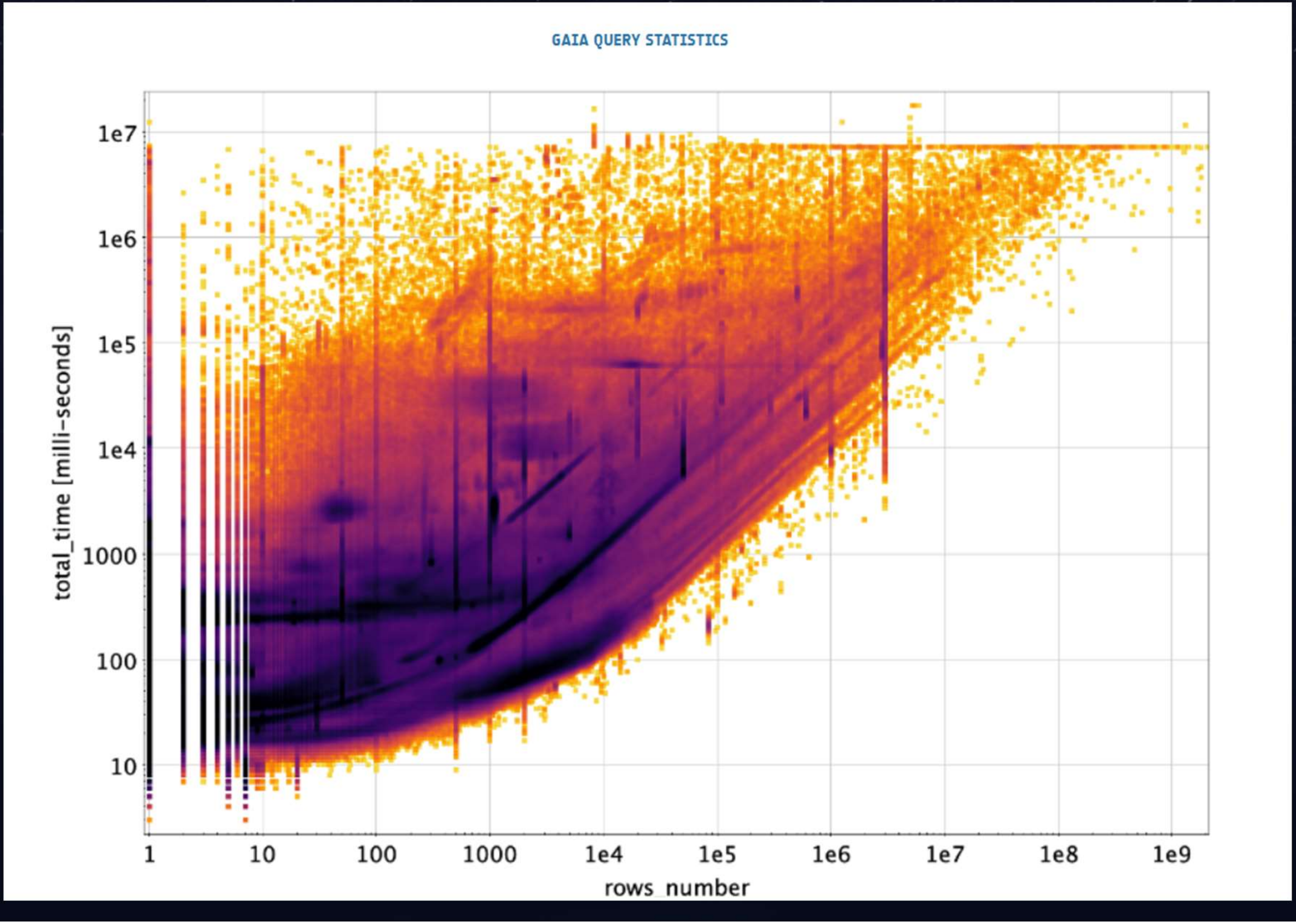
Gaia data usage and publications

- Use of the ESA Gaia archive remains at high level
    - ~5,000 monthly users downloading data
    - Typically 2.5 to 5 million monthly queries
- Similar level of usage from the partner data centres also serving Gaia data




Total Queries Processed
300.000.000,
250.000.000,
200.000.000,
150.000.000,
100.000.000,
50.000.000,
0,
Jan 2016
Jan 2017
Jan 2018
Jan 2019
Jan 2020
Jan 2021
Jan 2022

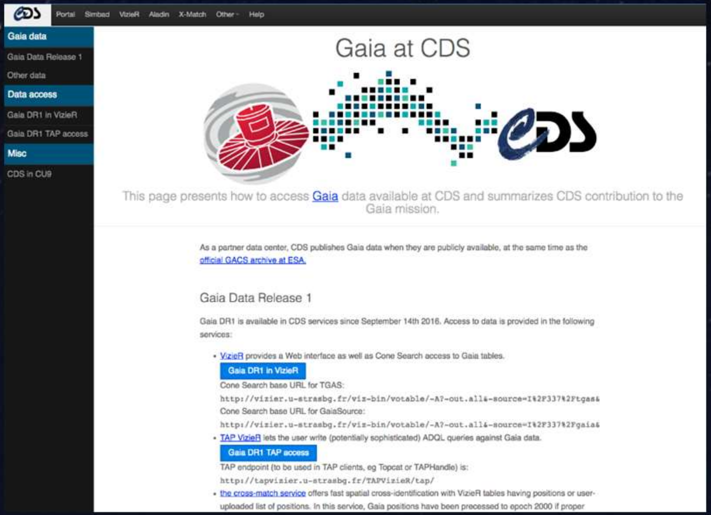
https://www.cosmos.esa.int/web/gaia/iow_20220328



Publicando los datos

13

Otros archivos de datos de Gaia



Gaia at CDS

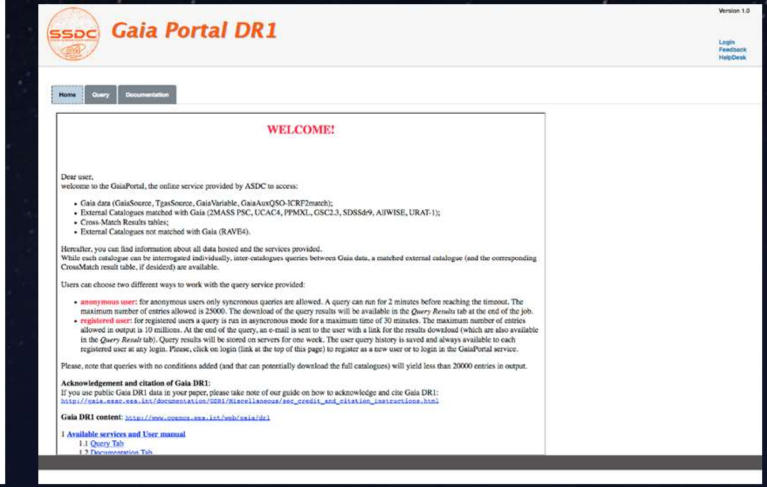
This page presents how to access Gaia data available at CDS and summarizes CDS contribution to the Gaia mission.

As a partner data center, CDS publishes Gaia data when they are publicly available, at the same time as the official GACS archive at ESA.

Gaia Data Release 1

Gaia DR1 is available in CDS services since September 14th 2016. Access to data is provided in the following services:

- VizieR provides a Web interface as well as Cone Search access to Gaia tables.
  Gaia DR1 in VizieR
  Cone Search base URL for TGAS:
  `http://vizier.u-strasbg.fr/viz-bin/votable/-A?-out.all&-source=I%2F337%2Ftgas&`
  Cone Search base URL for GaiaSource:
  `http://vizier.u-strasbg.fr/viz-bin/votable/-A?-out.all&-source=I%2F337%2Fgaia&`
- TAP VizieR lets the user write (potentially sophisticated) ADQL queries against Gaia data.
  Gaia DR1 TAP access
  TAP endpoint (to be used in TAP clients, eg Topcat or TAPHandle) is:
  `http://tapvizier.u-strasbg.fr/TAPVizieR/tap/`
- the cross-match service offers fast spatial cross-identification with VizieR tables having positions or user-uploaded list of positions. In this service, Gaia positions have been processed to epoch 2000 if proper



WELCOME!

Dear user,
welcome to the GaiaPortal, the online service provided by ASDC to access:

- Gaia data (GaiaSource, TgasSource, GaiaVariable, GaiaAuxQSO-ICRF2match);
- External Catalogues matched with Gaia (2MASS PSC, UCAC4, PPMXL, GSC2.3, SDSSdr9, AllWISE, URAT-1);
- Cross-Match Results tables;
- External Catalogues not matched with Gaia (RAVE4).

Hereafter, you can find information about all data hosted and the services provided.
While each catalogue can be interrogated individually, inter-catalogues queries between Gaia data, a matched external catalogue (and the corresponding CrossMatch result table, if desiderd) are available.

Users can choose two different ways to work with the query service provided:

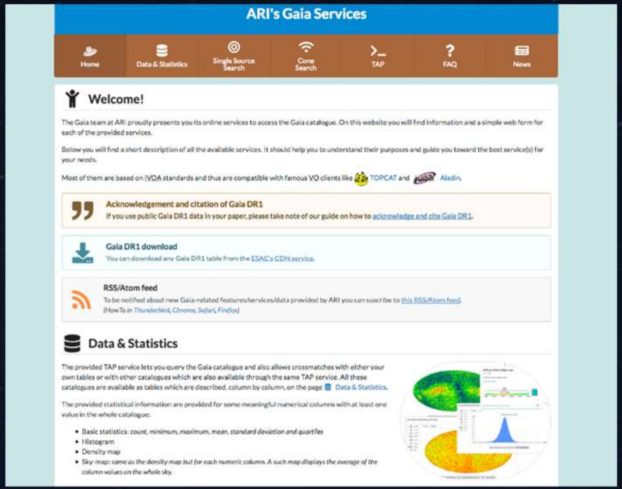
- anonymous user: for anonymous users only syncronous queries are allowed. A query can run for 2 minutes before reaching the timeout. The maximum number of entries allowed is 25000. The download of the query results will be available in the *Query Results* tab at the end of the job.
- registered user: for registered users a query is run in asyncronous mode for a maximum time of 30 minutes. The maximum number of entries allowed in output is 10 millions. At the end of the query, an e-mail is sent to the user with a link for the results download (which are also available in the *Query Result* tab). Query results will be stored on servers for one week. The user query history is saved and always available to each registered user at any login. Please, click on login (link at the top of this page) to register as a new user or to login in the GaiaPortal service.

Please, note that queries with no conditions added (and that can potentially download the full catalogues) will yield less than 20000 entries in output.

Acknowledgement and citation of Gaia DR1:
If you use public Gaia DR1 data in your paper, please take note of our guide on how to acknowledge and cite Gaia DR1:
`http://gaia.esac.esa.int/documentation/GDR1/Miscellaneous/sec_credit_and_citation_instructions.html`

Gaia DR1 content: `http://www.cosmos.esa.int/web/gaia/dr1`

1 Available services and User manual
  1.1 Query Tab
  1.2 Documentation Tab



ARI's Gaia Services

Welcome!

The Gaia team at ARI proudly presents you its online services to access the Gaia catalogue. On this website you will find information and a simple web form for each of the provided services.

Below you will find a short description of all the available services. It should help you to understand their purposes and guide you toward the best service(s) for your needs.

Most of them are based on IVOA standards and thus are compatible with famous VO clients like TOPCAT and Aladin.

Acknowledgement and citation of Gaia DR1
If you use public Gaia DR1 data in your paper, please take note of our guide on how to acknowledge and cite Gaia DR1.

Gaia DR1 download
You can download any Gaia DR1 table from the ESAC's CDN service.

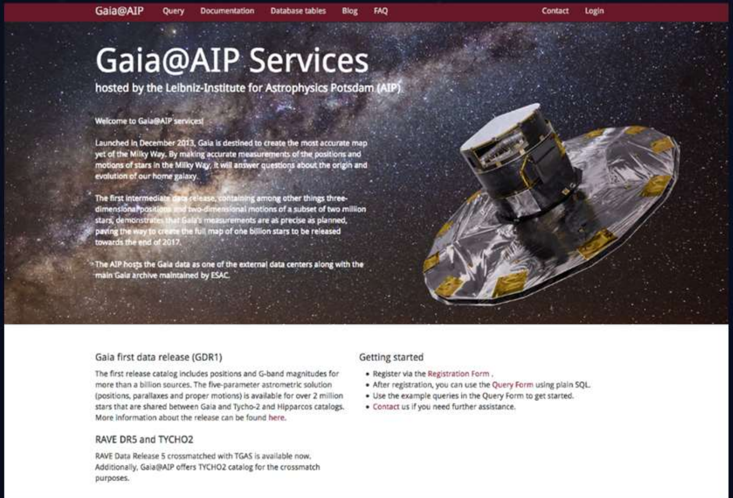
RSS/Atom feed
To be notified about new Gaia-related features/services/data provided by ARI you can suscribe to this RSS/Atom feed.
(HowTo in Thunderbird, Chrome, Safari, Firefox)

Data & Statistics

The provided TAP service lets you query the Gaia catalogue and also allows crossmatches with either your own tables or with other catalogues which are also available through the same TAP service. All these catalogues are available as tables which are described, column by column, on the page Data & Statistics.

The provided statistical information are provided for some meaningful numerical columns with at least one value in the whole catalogue:

- Basic statistics: *count, minimum, maximum, mean, standard deviation and quartiles*
- Histogram
- Density map
- Sky-map: *same as the density map but for each numeric column. A such map displays the average of the column values on the whole sky.*



Gaia@AIP Services

hosted by the Leibniz-Institute for Astrophysics Potsdam (AIP)

Welcome to Gaia@AIP services!

Launched in December 2013, Gaia is destined to create the most accurate map yet of the Milky Way. By making accurate measurements of the positions and motions of stars in the Milky Way, it will answer questions about the origin and evolution of our home galaxy.

The first intermediate data release, containing among other things three-dimensional positions and two-dimensional motions of a subset of two million stars, demonstrates that Gaia's measurements are as precise as planned, paving the way to create the full map of one billion stars to be released towards the end of 2017.

The AIP hosts the Gaia data as one of the external data centers along with the main Gaia archive maintained by ESAC.

Gaia first data release (GDR1)

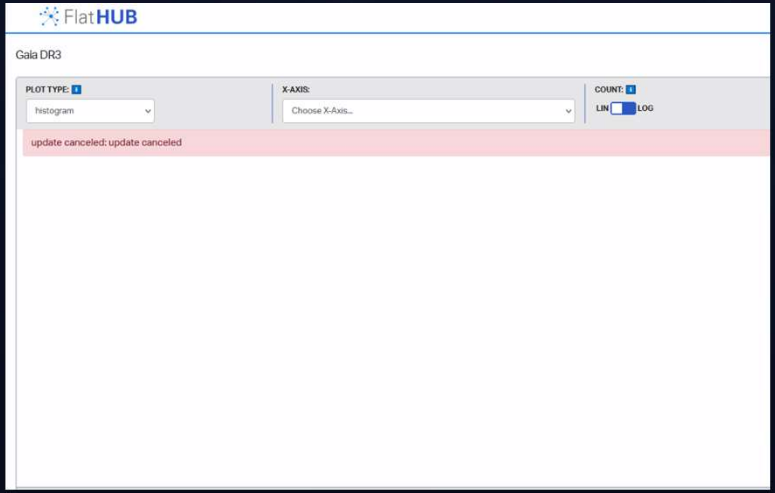
The first release catalog includes positions and G-band magnitudes for more than a billion sources. The five-parameter astrometric solution (positions, parallaxes and proper motions) is available for over 2 million stars that are shared between Gaia and Tycho-2 and Hipparcos catalogs. More information about the release can be found here.

RAVE DR5 and TYCHO2

RAVE Data Release 5 crossmatched with TGAS is available now. Additionally, Gaia@AIP offers TYCHO2 catalog for the crossmatch purposes.

Getting started

- Register via the Registration Form ,
- After registration, you can use the Query Form using plain SQL.
- Use the example queries in the Query Form to get started.
- Contact us if you need further assistance.



PLOT TYPE: histogram

X-AXIS: Choose X-Axis...

COUNT: LIN LOG

update canceled: update canceled

OBSERVATOIRE DE PARIS-MEUDON / GEPI (FRANCE)

NATIONAL ASTRONOMICAL OBSERVATORY (NAO) OF JAPAN

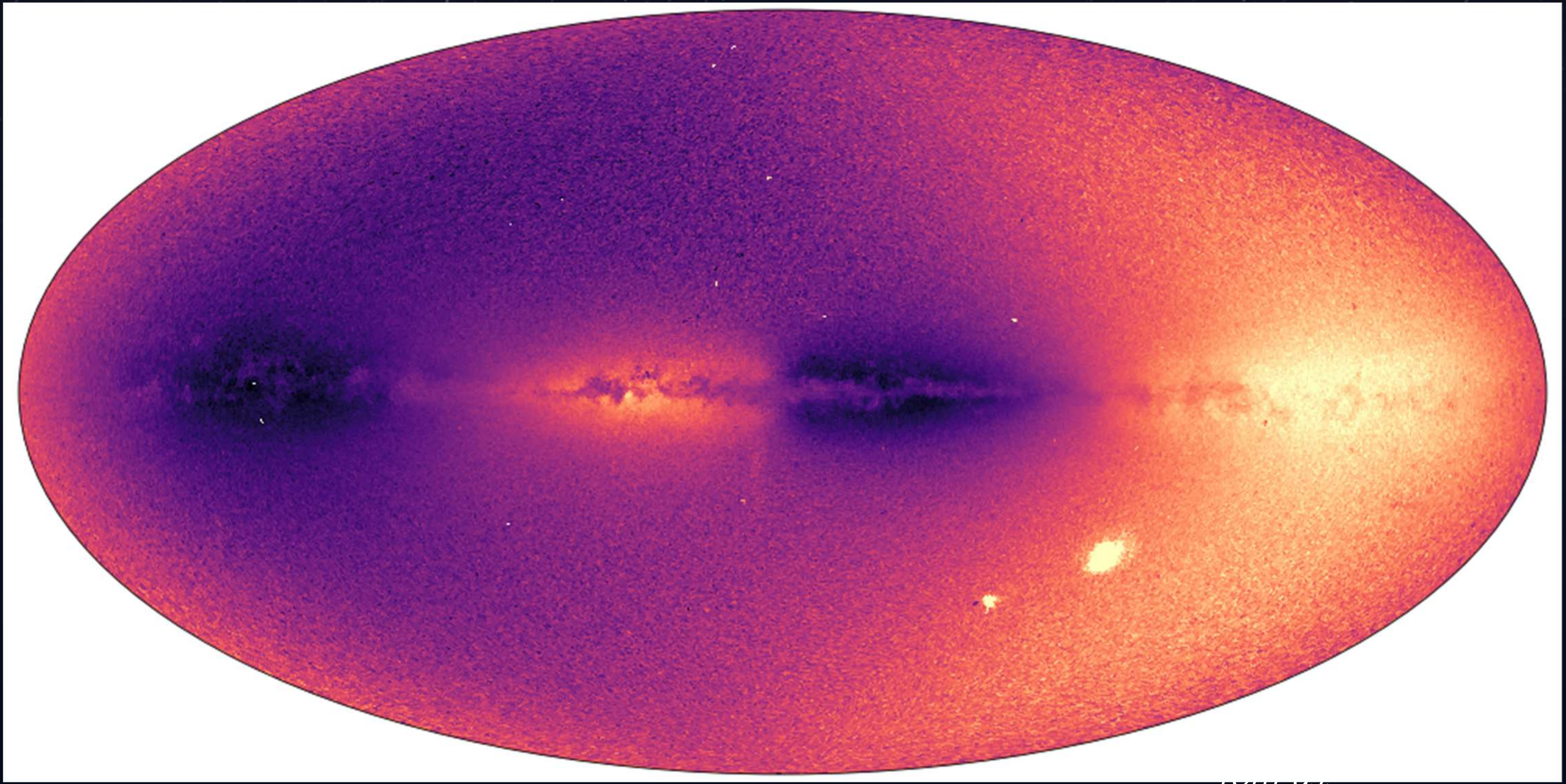
PORT D'INFORMACIÓ CIENTÍFICA (PIC) BARCELONA IN SPAIN

KAPTEYN ASTRONOMICAL INSTITUTE IN THE NETHERLANDS

GAIA SKY
3.2.0
Gaia DR3 very large catalog
466 million stars

Gracias por vuestra atención

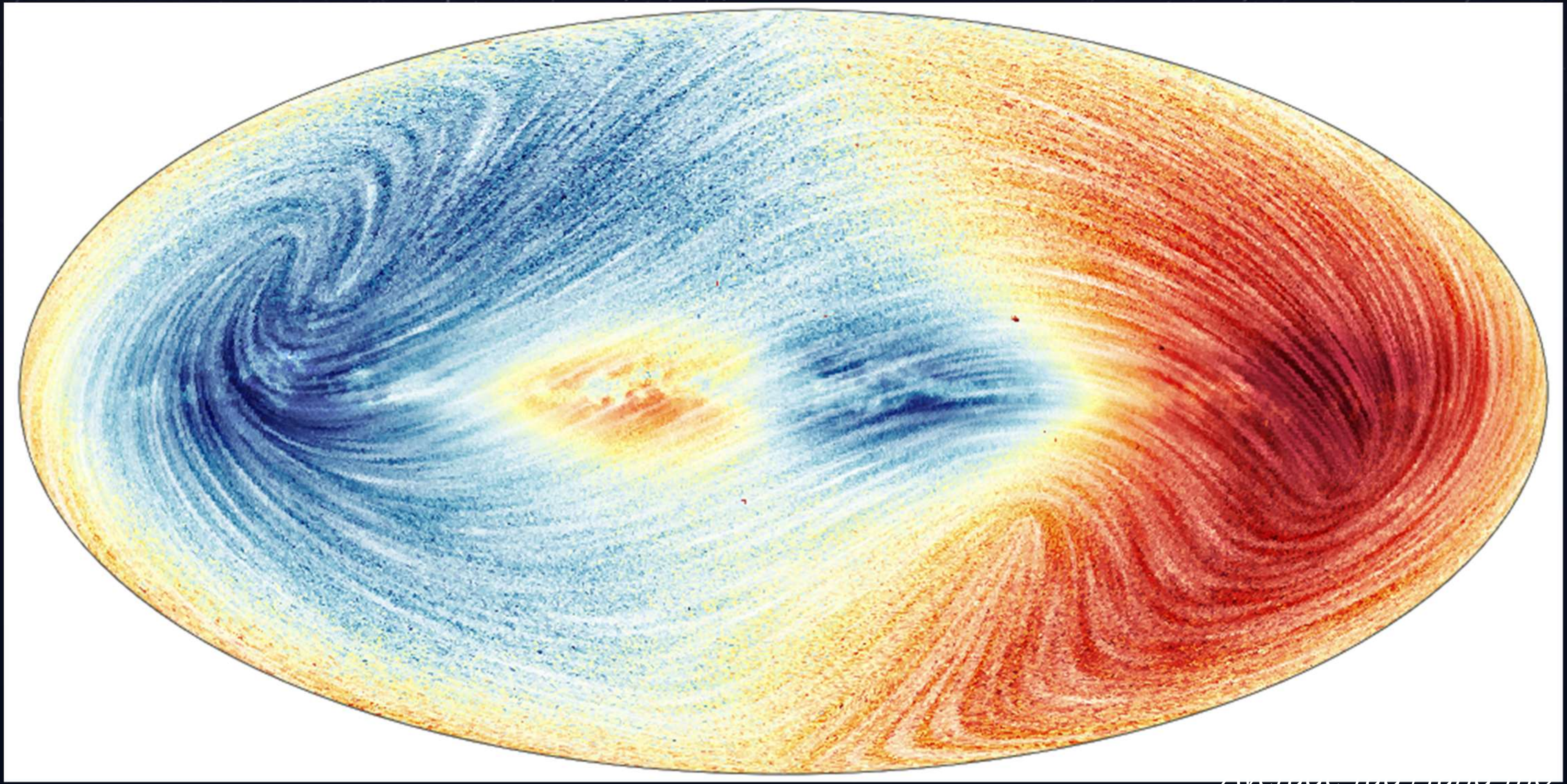
Gaia DR3: radial velocities



rom us.

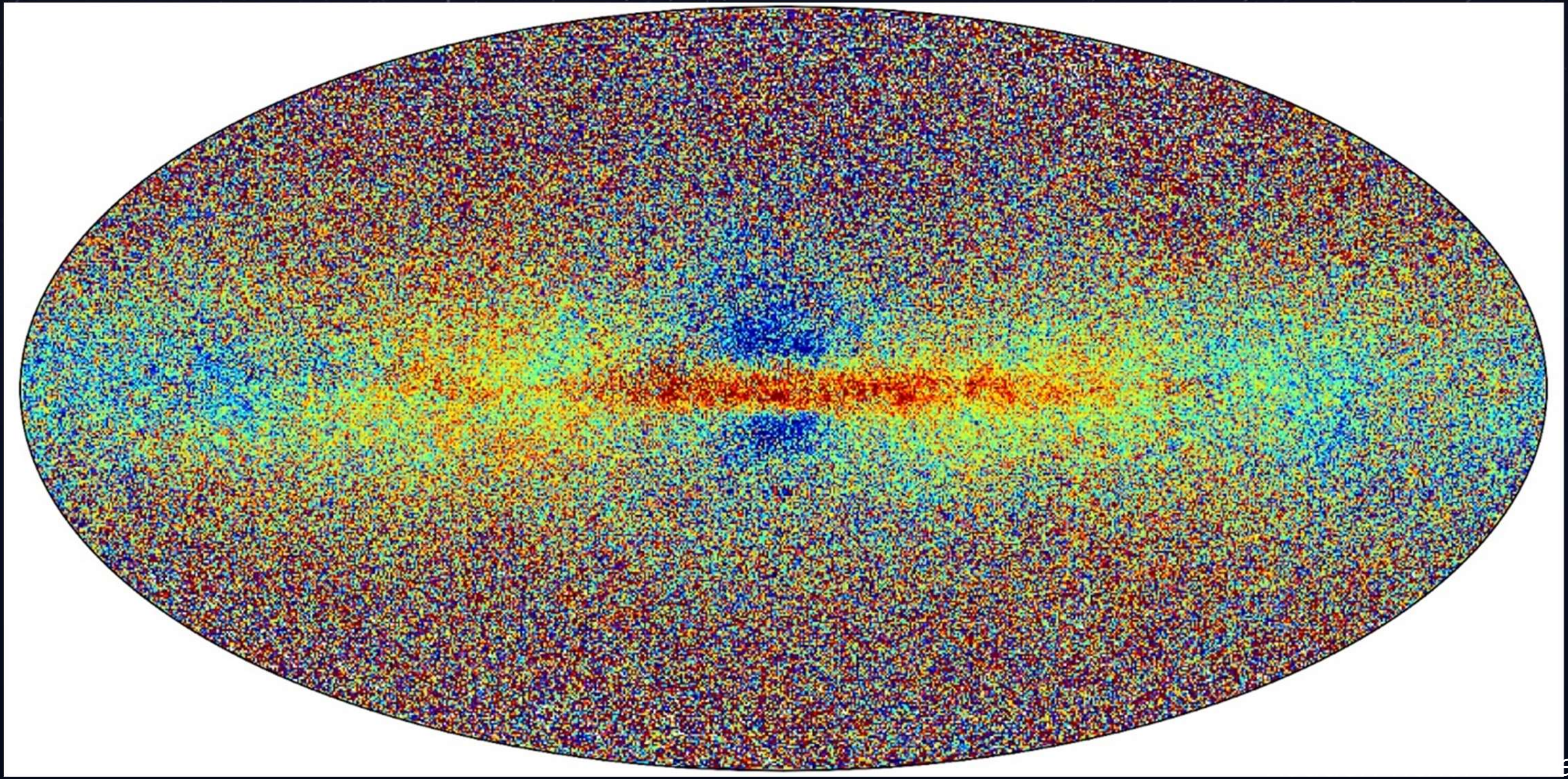
Dark: moving towards

Gaia DR3: radial velocities + proper motions



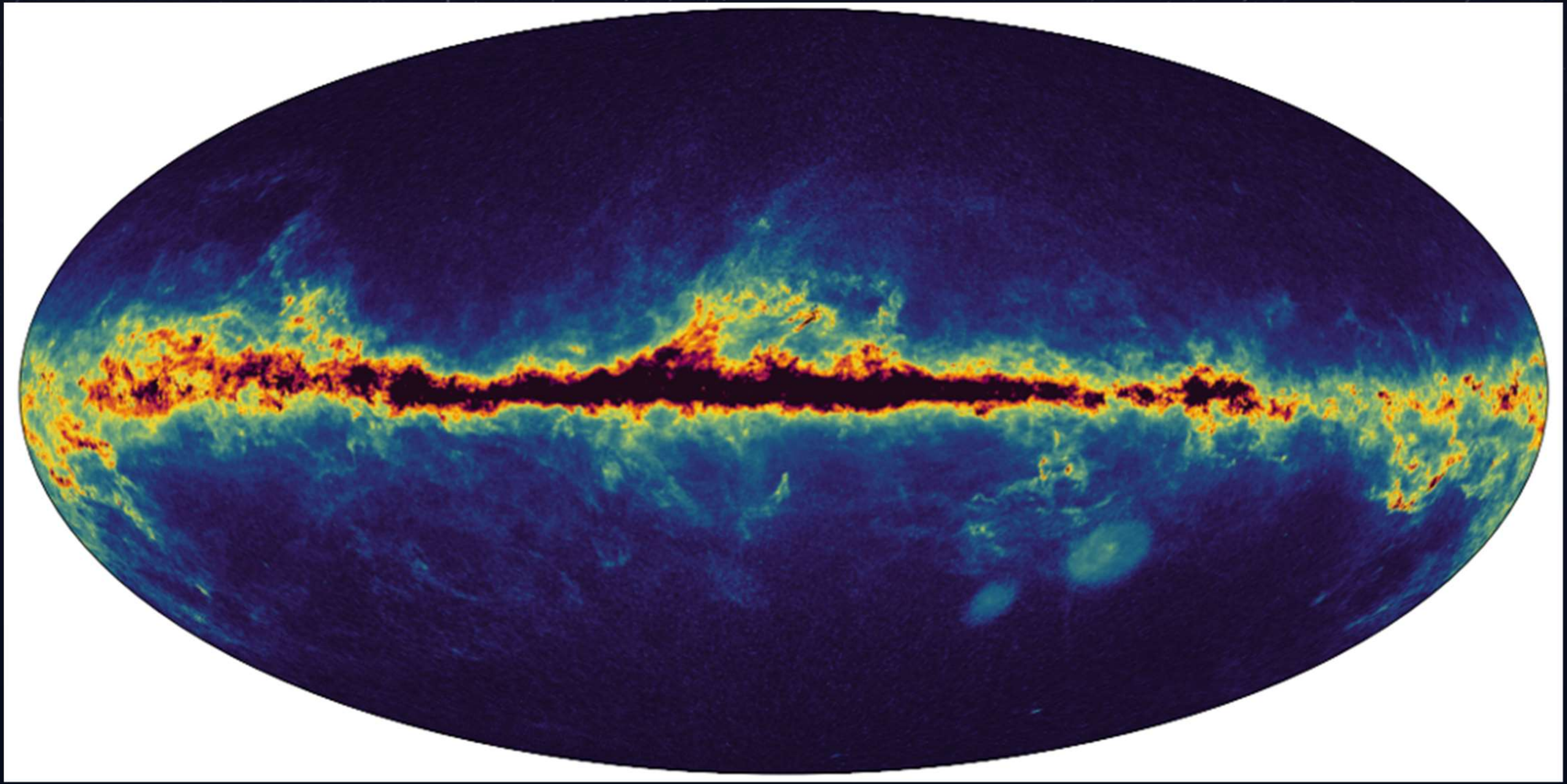
Average rad+tang motions

Gaia DR3: chemical map



eta

ls.

Gaia DR3: interstellar dust