Gaia, the Galaxy in One Petabyte

Carme Jordi
University of Barcelona, Spain
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

How many stars are observable?
How does Gaia measure the sky?
Source detection/confirmation in action
Data reception

- New Norcia
- Cebreros
- Malargüe
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 degree since start of nominal operations

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

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

Figure by J. Portell

V1293 Aql
(M5III)
Raw data: 50 GB every day during 5 years → 100 TB

Processing needs

Ingestion
Pre-processing
Data reduction
Classification
Variability analysis
etc.

300 million hours of CPU time

34,000 years !!!!
Six supercomputer centres in place for Gaia: the final data will be available around early 2020s

... and, early releases !!!!!
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

>20 globular clusters
Many thousands of Cepheids and RR Lyrae.

Mass of galaxy from rotation curve at 15 kpc

Sun

30 open clusters within 500 pc

Horizon for proper motions accurate to 1 km/s

Dark matter in disc measured from distances/motions of K giants

Dynamics of disc, spiral arms, and bulge

Horizon for distances accurate to 1.0 per cent

Proper motions in LMC/SMC individually to 2-3 km/s

General relativistic light-bending determined to 1 part in 10^6

1 microarcsec/yr = 300 km/s at z = 0.03
(direct connection to inertial)
1,000,000,000 stars

1,000,000,000 pixels

more than 1,000 people

more than 10,000 scientists

1,000,000,000,000,000,000 bytes
The promises of Gaia

- > 1 billion stars to G = 20
- 106–107 galaxies
- 5 105 quasars
- 3 105 solar system bodies
- tens of thousands of exoplanets
Data collected and processed

Statistics up to June 29

<table>
<thead>
<tr>
<th>Type of Data</th>
<th>Amount</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science telemetry</td>
<td>5.9 TB</td>
<td></td>
</tr>
<tr>
<td>Astrometry transits</td>
<td>$6.5 \times 10^9$</td>
<td>$64.5 \times 10^9$ images</td>
</tr>
<tr>
<td>Photometry transits</td>
<td>$6.1 \times 10^9$</td>
<td>$12 \times 10^9$ images</td>
</tr>
<tr>
<td>Spectroscopy transits</td>
<td>$1 \times 10^9$</td>
<td>$3 \times 10^9$ spectra</td>
</tr>
</tbody>
</table>
Since beginning of nominal mission: 137 days
Raw data received: 3.6 TB (26 GB/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, 64 GB RAM/node)
Horizon 2000+
*Cornerstone mission*

Scientific mission: Astronomy

ESA’s only mission
• 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

Centre d’Operacions Científiques:
ESAC (Villafranca del Castillo, vora Madrid)

- Sistemes d’operacions de càrrega útil
- Equip de Calibració
- Interfície amb el Centre d’Operacions de la Missió (MIT)
- Descompressió de dades (DCS)
- Sistema de processament diari (IDT)
- Sistema de diagnòstic inicial i determinació de calibracions (FL)
Data Processing and Analysis Consortium

[Diagram showing map of Europe with various markers and labels for cities and locations, indicating contributions to the DPAC Consortium.]