



## European Space Agency (ESA) + InterSystems

### Crunching Data to Map the Milky Way

Eduardo Anglada of the European Space Agency (ESA) and the European Space Astronomy Centre is currently working on the Gaia mission to map one billion objects in the Milky Way. Anglada and Jordi Calvera Sagué of InterSystems explained how much data is involved in categorizing a billion space objects, and the requirements of the technology that supports this extra-terrestrial task.

The mission of the European Space Agency (ESA) is to explore the universe and it has a fleet of 10 craft out in space to do so. The Gaia space mission began in 1993 and the actual spacecraft was launched in 2013. Its aim is to create a precise 3D map of 10% of the Milky Way, and in doing so characterize up to 1 billion objects.


Eduardo Anglada, computer analyst and grid engineer at ESA, explained that by 'characterize' he meant that he and his team are aiming to chart the position, velocities, physical parameters, and temperature of those objects.

They also want to answer questions relating to the metallicity and age of the objects, as well as whether they are binary systems (with two stars close together). The process will reveal the composition, formation, and evolution of the galaxy.

As one might imagine, the numbers are staggering. The satellite, which costs almost €700 million, is between 700,000 kilometers and 1.5 million kilometers away from Earth. The stakes are high, as the hardware, the software, and the security of the data all have to be impeccable because at that distance, it is impossible to repair or refuel the satellite. Essentially, the data that is downloaded from this satellite is irreplaceable.



*Eduardo Anglada,  
European Space Agency  
(ESA), and European  
Space Astronomy Centre*



**“WE HAVE ANALYZED  
MORE THAN  
ONE TRILLION  
OBSERVATIONS.  
IT IS QUITE A FIGURE  
AND IT IS THANKS  
TO CACHÉ.**

*Eduardo Anglada,  
European Space Agency  
(ESA), and European  
Space Astronomy Centre*

The camera onboard the craft is 938 million pixels and has 106 charge coupled devices (CCDs), the type of image sensor used in most digital cameras. “This is one of the biggest cameras ever built,” said Anglada. The satellite takes six hours to do a full rotation on its axis and sends down between 45 and 100 gigabytes per day, having taken around 70 million images daily. “Once the data is on Earth we have 24 hours to analyze it.”

The data from the camera is downloaded daily and is received by three antennae in Spain, Australia, and Argentina. With the different antennae, ESA have a ‘follow the sun’ way of working so that the satellite is constantly tracked by the satellites in one of those countries. “If we lose it, it’s a disaster. In a few hours, it can be many, many kilometers from where it is supposed to be and it can be very difficult to track and find it again,” Anglada explained.

From those locations, the data is sent to the mission operation center in Germany where it is checked to see if there are any problems with the satellite, and fix them if they are present. Then the compressed data is sent to ESA in Madrid where the observations are decompressed, calibrated, and stored in a cache database.

Then it is sent to the data specialist centers, for among other things, simulations and object, photometric, and spectroscopic processing. It then goes back to Madrid, the location of the central database.

The system was set up in a hub and spokes model so that all researchers could access the data in order to fulfill the requirements. The spoke locations are dotted around Europe including Cambridge, Turin, and Toulouse.

There are four stages to the data processing; daily and cyclic operations, the main database, calibration activities and payload commanding, and finally development.

Anglada said that the daily database is quite big now. “It is almost 30 terabytes. It’s a single instance. We have a big server with 1.5 terabytes of RAM and seven terabytes of solid state disk.”

The cyclic operations are completed every four to six months. In that time the data centers have to finish their processes of the data and start sending it round to other data centers in order for them to refine the results.

InterSystems Caché is part of the main database and the daily processing. The space agency got in touch with Jordi Calvera Sagué, regional managing director at InterSystems, to say that it faced the challenge of inserting and processing a large volume of data in a short period of time. ESA provided the company with some dummy data and the software engineers did a configuration of the database cache in three days. “We changed the architecture after the proof of concept, but it was relatively quick and just did the configuration,” he said.

In addition to InterSystems Caché, ESA also uses Aspera for distribution across the data processing centers and Atlassian Jira for bug tracking. The ESA team comprises 400 people in Europe and 26 in ESAC with different areas of expertise from calibration to management to daily download.

So what was the attraction to InterSystems? Anglada said that it is extremely reliable and the company offers comprehensive support. “In the last 11 months we have had no problems at all. I am a very technical person, I am part of the daily team and InterSystems has the best support that we have had by far.” Anglada added that it is very easy to get in touch with the account manager and find out if the tool can feasibly meet a new requirement. “It is not so common that a company, once the project is mature, wants to continue working with you like that.” He added that that Caché is so robust that daily checks can be completed in minutes.

As previously mentioned, there are some big numbers involved in this project and for Anglada, one is particularly impressive. “We have analyzed more than one trillion observations. It is quite a figure and it is thanks to Caché. It handles 30 terabytes of database without problems.” InterSystems has even increased the shelf life of the project. Calvera Sagué said that the satellite is so optimized that the mission can keep going for two to three years longer than anticipated.

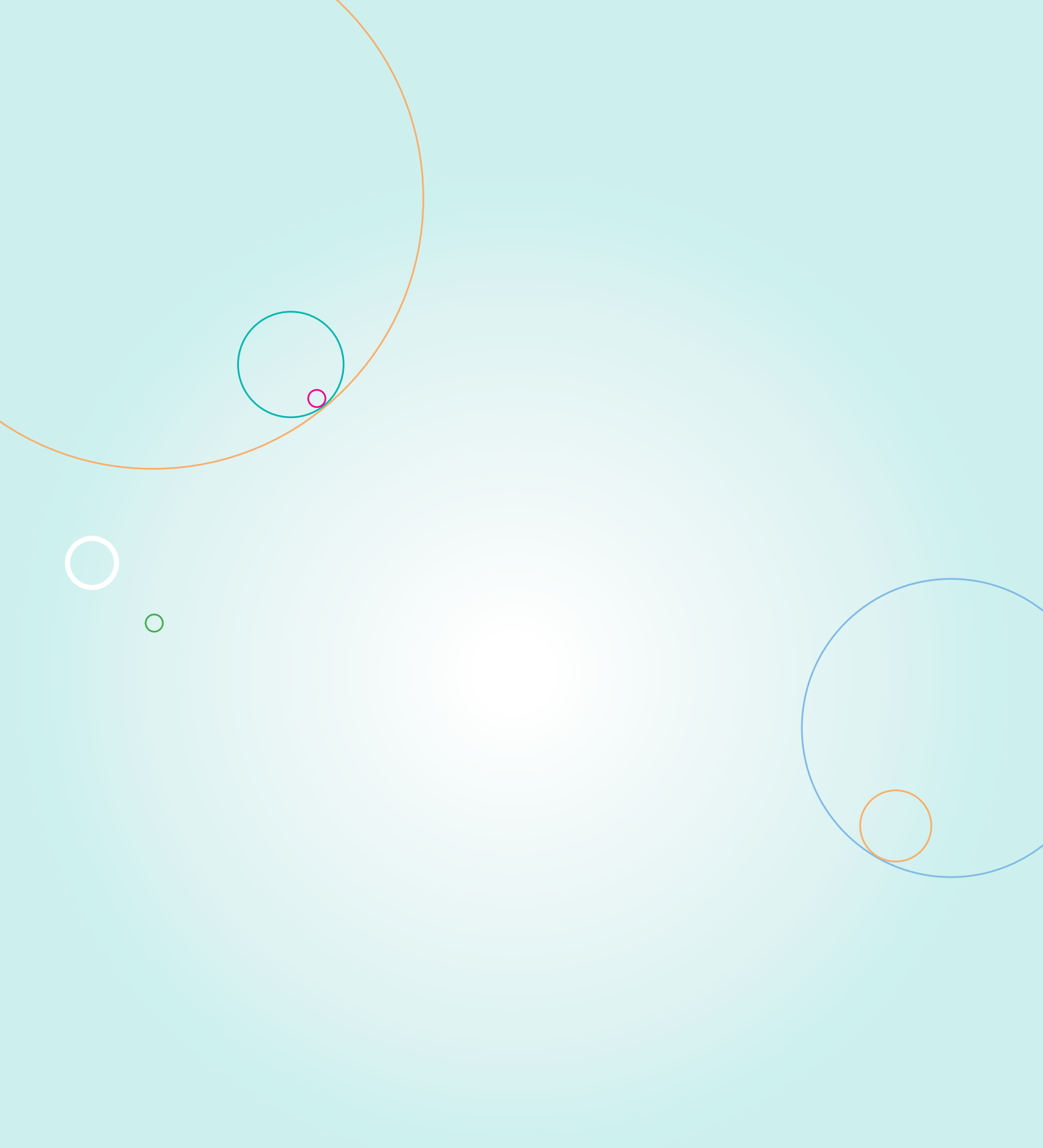
The results of this mission were published in a catalog for the first time in 2016 and mapped 300 light years, just part of the Milky Way. The second release April 2018 has mapped about 8,000 light years. The Gaia team was able to document the brightness of 1.7 billion stars and the surface temperature of 1.4 billion.

The level of detail meant that there were some surprising discoveries. “Having the velocities meant we could study the kinematics of the galaxy. For example the Milky Way clashed with another galaxy many years ago and we have seen the last stages of how both galaxies are merging. It has been known for years but so far nobody could measure it. With this data it has been possible.”

Furthermore, last year, the first interstellar comet, Oumuamua, was detected. “It is the first object detected in the Solar System which is not part of the solar system. This comet was travelling through space for about 6,000 years and it came into the Solar System so fast that it wasn’t trapped by the Sun but the Sun’s gravity changed its trajectory.”

The third release is likely to take place in the first half of 2021, and the date of the final release, which will consist of full astrometric, photometric, and radial-velocity catalogs, is yet to be decided. Anglada said the astronomy community is extremely happy with the results of this mission so far and the average number of scientific articles citing the catalog is three or four per day “which is amazing for science.”

“Everything is quite material now,” said Anglada. “Each time a scientist realizes that we can measure something new, new ideas for the corresponding development takes place. With this catalog the scientific community will be able to study the physics they want about these stars for the next 30 or 40 years.”



The power behind what matters.

