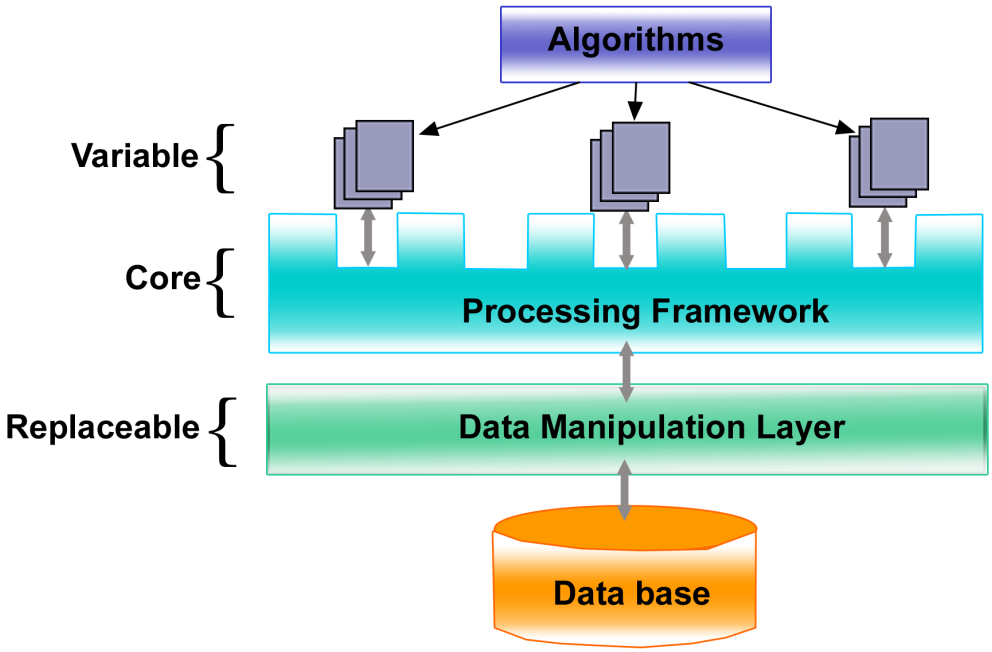

The data processing and analysis architecture chosen for Gaia is highly adaptable. While algorithms may vary for different tasks, the processing and data handling layers remain fixed within the core. The data base managing the large amount of data can be replaced, providing the ability to adapt to evolution in data base technology.

The data processing requirements for Gaia are amongst the most challenging of any scientific endeavour to date. Due to the immense volume of data that will be collected, for 1 billion stars, it will be a major challenge, even by the standards of computational power in the next decade, to process, manage and extract the scientific results necessary to build a 3-dimensional view of our Galaxy, the Milky Way. Gaia is in some senses the astronomical equivalent of the Human Genome Project, and is a pioneering undertaking being led by ESA.

A total of some 100 Terabytes of science data will be collected during Gaia's lifetime. The total data archive will surpass 1 Petabyte – at current state-of-the-art disc access rates, this would require 40 days simply to read. The required numerical processing is colossal – of order 10^{21} flops. Using the world's fastest computer (as of early 2006, IMB/DOE's BlueGene 136 TFlop system), would require 85 days of processor power. The size of the problem, and the requirements for rigorous testing and optimisation, can be emphasised by noting that just 1 second of CPU time devoted to any specific task, per object (e.g. variability classification), would require 30 years of CPU time for the entire data set, per task.

In order to assess the complex requirements for Gaia's data analysis, ESA initiated a Gaia Data Access and Analysis Study in 2001. The aim was to undertake a first implementation of the Global Iterative Solution, and to establish some basic assumptions concerning the data base technology required to manage 1 billion objects, each observed 100 times during the five-year mission lifetime. Led by the Spanish software company GMV, supported by scientists from the University of Barcelona, and making use of the university's supercomputer, the results showed that the iterative adjustment of merely a million stars would consume months of computer time.

A deep revision of the access and processing framework necessary for the Global Iterative Solution was clearly needed. An ESA team, in place at the European Space Astronomy Centre (ESAC) in Madrid since September 2005, completed a totally revised implementation of this core processing in December 2005. Totally in Java, and using Oracle 10g RAC as the data base management system, the new implementation uses a dedicated Dell system comprising 6 PowerEdge 1850 dual-processor nodes, interconnected by Gigabit Ethernet, and each equipped with two 3.6 GHz Intel Xeon processors, in total delivering about 45 GFlop. With this system, an 'outer iteration' (source, attitude and calibration) of 1 million stars over 18 months of mission data now runs in just under 3 hours. A major advance in the demonstration of the core global iterative solution for Gaia has therefore now been achieved.