

# Data Processing and Analysis Consortium

In the course of Gaia's five year astronomical survey, around 100 terabytes of raw data concerning our galaxy, the Milky Way, will be harvested and transmitted to Earth. Sophisticated data processing will be necessary to distil this immense quantity of complex data into the final Gaia Catalogue, which will comprise astrometric, photometric and spectroscopic information for around one billion celestial objects. When the Announcement of Opportunity for this formidable task was issued by ESA, a group of about 400 European scientists and developers rose to the challenge and responded. This group, called the Data Processing and Analysis Consortium, is already working in preparation for Gaia's anticipated early 2012 launch.



In late 2006, ESA's Announcement of Opportunity for Gaia's data processing was released, calling for a proposal to build and operate Gaia's ground segment data processing, a single processing pipeline leading to the intermediate and final mission products. The announcement expected the system to be developed as a collaboration between ESA's Gaia Science Operations Centre and a broad scientific community supported by national funding agencies.

In response to the Announcement of Opportunity, a large pan-European team of expert scientists and software developers submitted their proposal for a comprehensive system capable of handling the full size and complexity of the Gaia data. In May of 2007, ESA's Science Programme Committee approved the proposal put forward by the Data Processing and Analysis Consortium (DPAC). At this point DPAC became officially responsible for Gaia data processing and analysis.

The nature of the Gaia mission leads to the acquisition of an enormous quantity of complex, extremely precise data, representing the multiple observations of a billion diverse objects by a 'double vision' instrument that is spinning and precessing. The Gaia data challenge - processing raw satellite telemetry into valuable science products - is therefore a huge task in terms of expertise, effort and dedicated computing power.

DPAC is a collaboration that draws its membership from all over Europe, including a diverse community of about 400 scientists and software engineers, spread throughout 23 countries, and six large Data Processing Centres. The consortium brings together skills and expertise from across the continent; its international nature and cooperative spirit reflects that of ESA itself.



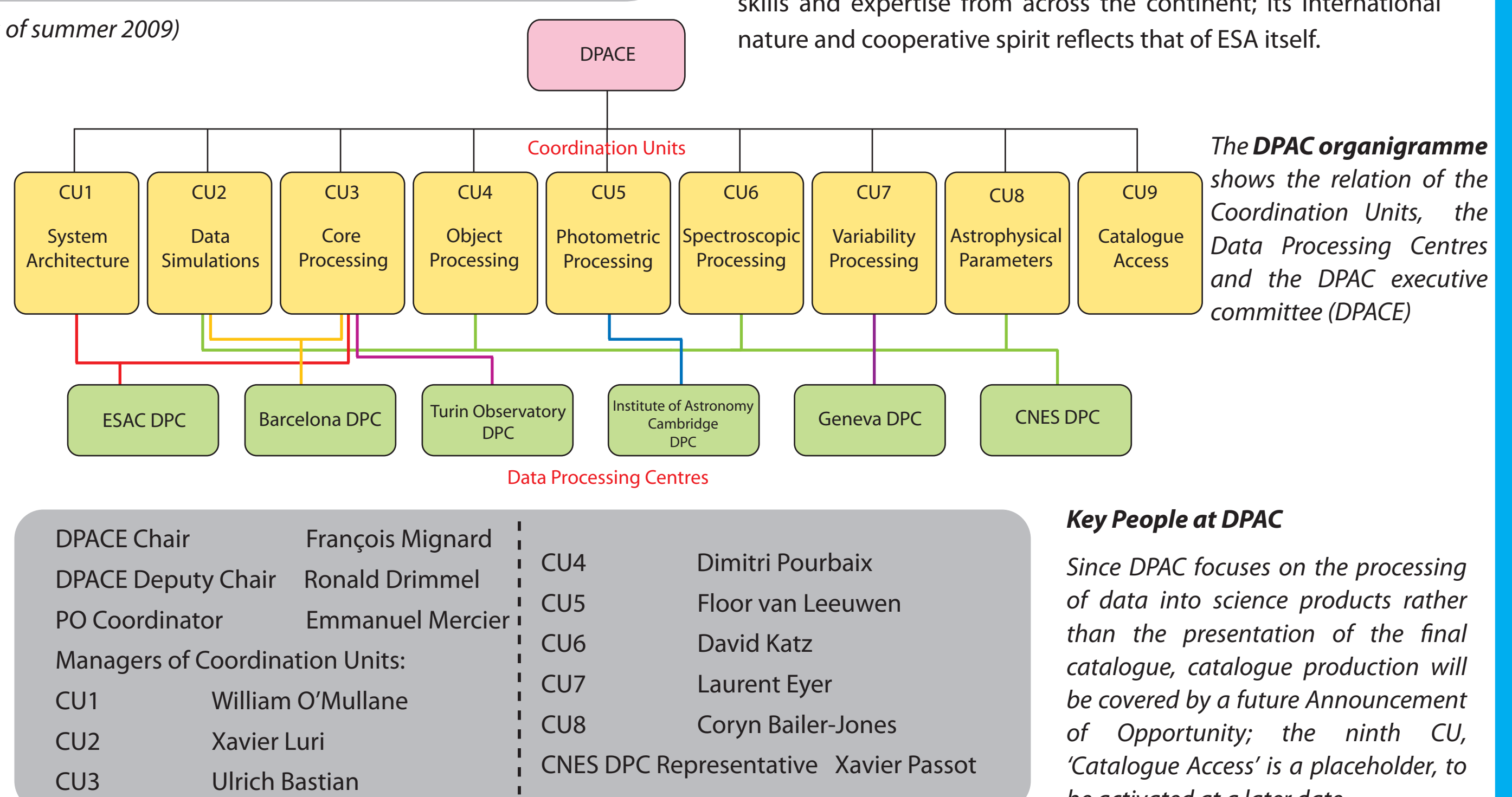
DPAC membership is distributed across 23 countries (map accurate as of summer 2009)

The consortium is sub-divided into smaller, specialist units known as Coordination Units, or CUs. These are the building blocks of DPAC, with each unit being assigned a unique set of data processing tasks. The CUs are supported by the Data Processing Centres, or DPCs: the centres at which the actual computer hardware is available for the processing. Whilst the CUs are primarily structured specifically for the development of software, each is then closely associated with at least one DPC, where their processing algorithms are actually implemented.

Besides the clear technical challenge faced by DPAC, the sociological challenge of effectively coordinating the efforts of about 400 people undertaking an unprecedented challenge in data processing for astronomy is not to be underestimated. Effective communication, organisation and maintaining motivation for the large consortium is recognised as essential.

The day-to-day management of the overall DPAC development and operations is delegated to the DPAC Project Office (PO). The Project Office monitors all technical, managerial and administrative activities of its own and, where appropriate, of the CU/DPCs, and provides the directive necessary to accomplish the project, expedite resolution of problems and interface with ESA Gaia Project.

For more on DPAC: [www.rssd.esa.int/gaia/dpac](http://www.rssd.esa.int/gaia/dpac)  
For more on Gaia or to download this poster: [www.rssd.esa.int/gaia](http://www.rssd.esa.int/gaia)



The **DPAC organigramme** shows the relation of the Coordination Units, the Data Processing Centres and the DPAC executive committee (DPAC)

## Key People at DPAC

Since DPAC focuses on the processing of data into science products rather than the presentation of the final catalogue, catalogue production will be covered by a future Announcement of Opportunity; the ninth CU, 'Catalogue Access' is a placeholder, to be activated at a later date.

DPAC plays a pivotal role in ensuring Gaia's potential is fulfilled. In the course of the processing, intermediate data releases with valuable scientific content are expected. When the Gaia Catalogue is finally published in around 2020, DPAC's work will be complete, and Gaia's processed data immediately made freely available for investigation by the world's entire scientific community.



# Gaia: Surveying the Galaxy