Gaia Data Flow System (GDFS) Project: the UK's contribution to Gaia data processing

N. A. Walton[1], M. Cropper[2], G. Gilmore[3], M. Irwin[3], and F. van Leeuwen[3]

[1]Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge, CB24 5JE, UK
email: naw@ast.cam.ac.uk

[2]MSSL, University College London, Dorking, Surrey, UK

[3]Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge, CB24 5JE, UK

Abstract. Gaia is an ESA cornerstone mission which will observe some billion stars in the galaxy enabling micro-arcsec astrometric catalogues to be constructed. In addition Gaia will produce high quality photometric and spectroscopic catalogues.

The data processing tasks are large and complex. A European consortium has been formed - the Gaia Data Processing and Analysis Consortium (DPAC). This paper describes the form of the UK Gaia Data Flow System Project contribution to the DPAC.

Keywords. space vehicles, methods: data analysis

1. Introduction

Gaia (see http://www.rssd.esa.int/gaia) is an approved ESA cornerstone project, which entered its B2 phase in 2006 for launch in late 2011. Gaia will provide photometric, positional, spectroscopic and radial velocity measurements with the accuracies needed to produce a stereoscopic and kinematic census of about one billion stars in both our Galaxy and the Local Group, addressing its core science goals to quantify the formation- and assembly history of a large spiral galaxy, the Milky Way. Gaia will achieve this by obtaining a six-dimensional (spatial & kinematic) phase-space map of the Galaxy. This will be complemented by an optimised high-spatial resolution multi-colour photometric survey, coupled with the largest stellar spectroscopic and radial velocity surveys ever made.

The Gaia data set will be constructed from the many repeat observations of some billion objects. The analysis task is a complex one, involving both real-time and end-of-mission data products.

This paper notes the UK GDFS activities as part of the European wide Gaia Data Analysis and Processing Consortium. We briefly describe the data processing challenges that need to be overcome to meet the heavy demands placed by Gaia.

2. Gaia Data Processing & Analysis Consortium (DPAC)

Gaia is a mission of extremes (see Lindegren, these proceedings). This is reflected in the range of data reduction techniques required: from extreme-accuracy global astrometry to large scale photometry and spectroscopy, all based on data gathered in a constantly

Table 1. DPAC Coordination Units

Unit	Name	Leading Institute	Notes
CU1	System design	ESAC	
CU2	Data simulations	Barcelona	
CU3	Core processing	ESAC	
CU4	Object processing	CNES	
CU5	Photometry	Cambridge	UK Lead
CU6	Spectroscopy	Paris	UK major partner
CU7	Variability analysis	Geneva	
CU8	Astrophysical parameters	Heidelberg	

moving focal plane. The volume of data is large, covering nearly 2×10^{12} individual observations, but more importantly, nearly the entire volume of data needs to be considered in some of the data processing.

The Gaia data processing forms the link between the measurements produced by the satellite payload and the various catalogues with scientific results that will be the products of the mission. Unlike the design and construction of the actual satellite and payload, which is entirely funded by the European Space Agency (ESA), the Gaia data processing is the responsibility, and is being funded by, the wider scientific community. The DPAC (see Mignard, these proceedings) is the pan-European consortium that will construct the data analysis system for Gaia and deliver the science data products after launch.

To manage the various tasks and their interconnectivity, the DPAC data processing is being organised around a small number of coordination units (see Table 1) which cover all data processing and essential analysis software developments.

3. GDFS

The UK Contribution to these efforts is through the Gaia Data Flow System (GDFS) project which is funded by the UK's Science and Technology Facilities Council (`http://www.stfc.ac.uk`). The groups involved in the GDFS are led by the Institute of Astronomy, University of Cambridge, and include, School of Engineering and Design, Brunel University; Institute for Astronomy, University of Edinbugh; Department of Physics and Astronomy, University of Leicester; MSSL, UCL; Rutheford Appleton Laboratory, STFC.

UK participation in the construction of the Gaia Data Flow system is concentrated in the areas of CU5 - Photometric processing (where the IoA, Cambridge is the lead institute). MSSL, UCL lead the significant UK involvement in the CU6 Spectroscopy unit.

The UK GDFS project investment is £10M for all development activities to launch. All staff will be in place by March 2008. The UK team is led by Gilmore at the IoA, and managed by van Leeuwen.

The GDFS project is ensuring that its processing system is compatible with relevant Virtual Observatory standards (see e.g. `http://www.ivoa.net`). Further, members of the DPAC and UK GDFS projects are playing an active role in the further development of the standards process through involvement in appropriate International Virtual Observatory Alliance technical working groups, and interaction with both the UK's AstroGrid (http://www.astrogrid.org) and Euro-VO (http://www.euro-vo.org) projects.