

# Gaia Catalogue and Archive Plans and Status

**29 June 2009**

**AS Gaia, Besançon**

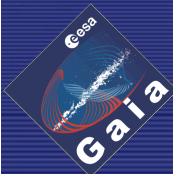
**William O'Mullane**

**Gaia Science Operations Development Manager**

**European Space Astronomy Centre**

**Madrid**

AS Gaia Besançon, France  
June 29th 2009



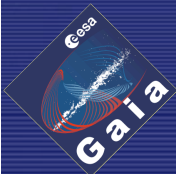
# A little background

- Already heard about the Satellite from Jos
  - That is ESA responsibility
- Data processing is done by the Data Processing and Analysis Consortium (DPAC)
  - That is a community effort (>300 people)
  - Some ESA involvement
- The Project Scientist Oversees all this
  - Advised by the Gaia Science Team



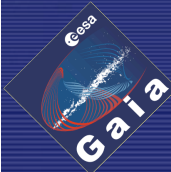
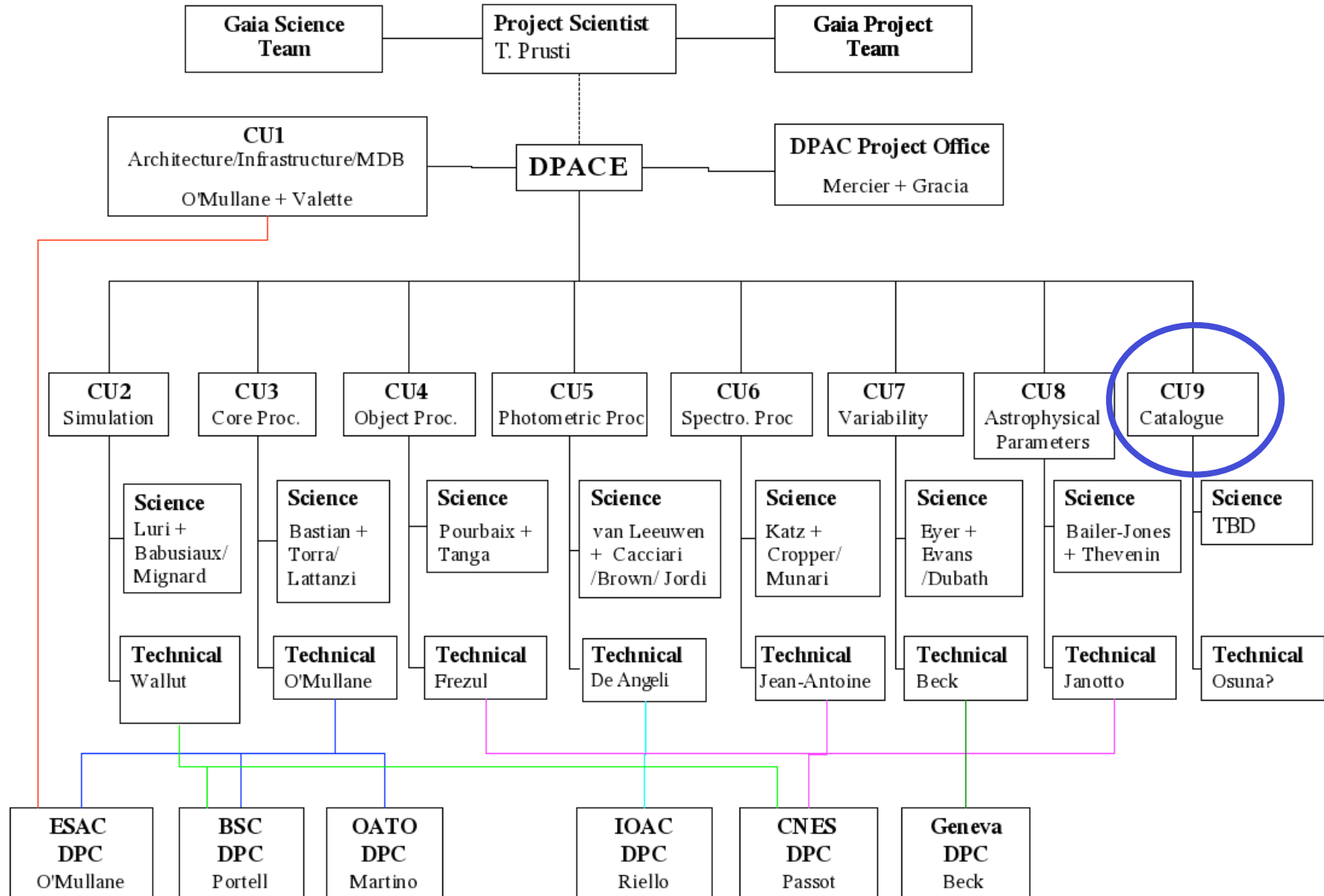
# Background

- DPAC officially created in answer to ESA's AO for Gaia Data Processing 2006.
- DPAC is composed of
  - Eight Coordination Units (CU)
    - area of competence/software production
  - Six Data Processing Centres (DPC)
    - Hardware to run processing
- AO explicitly said DPAC not to include Archive/Catalogue production
  - At that time
- To be integrated in DPAC later
  - A ninth Coordination Unit CU9
  - Possible new AO



# The Organisation Chart

AS Gaia Besaçon, France  
June 29th 2009



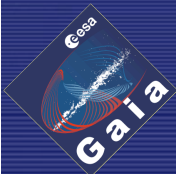
# Why Me ?

- CU1 leader in DPAC
  - System architect
- Also Science Operations for ESA
- End up at DPACE and GST meetings
- Scribe for thoughts on CU9 past twelve months in GST and DPACE
- Experience (SDSS, GSC, Hipparcos)
- But probably Schultz has it
  - "Nothing personal, your name just happened to come up."



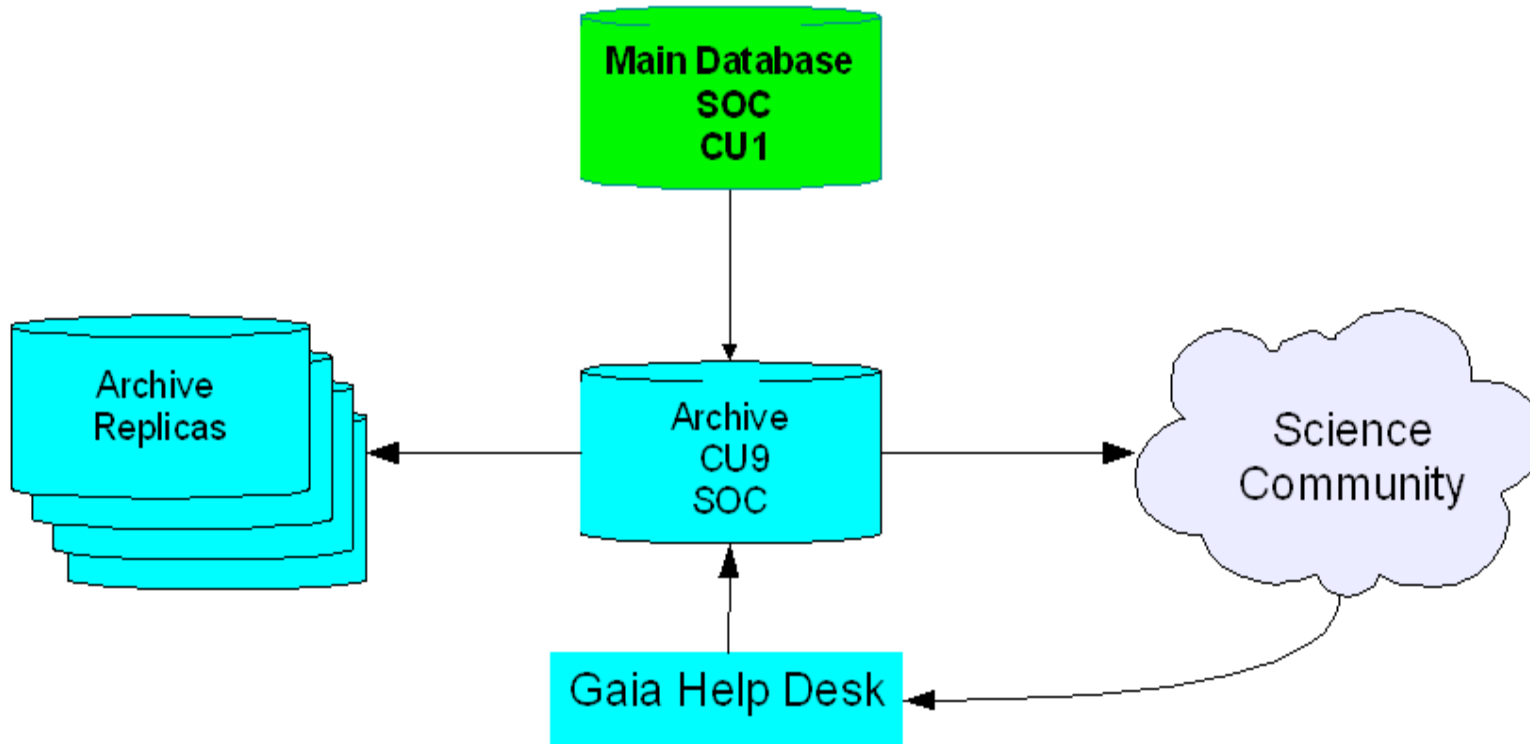
AS Gaia Besaçon, France  
June 29th 2009

# Proposed Architecture



# Archive in context

AS Gaia Besaçon, France  
June 29th 2009



# Archive is NOT simply MDB

- - M
  - D
  - B is CU oriented and contains redundancy
- Archive should be more cohesive and present single coherent data set (No more CUs visible)
- Mode of access is radically different
  - Arbitrary queries
  - Data mining
  - Download in various formats etc..





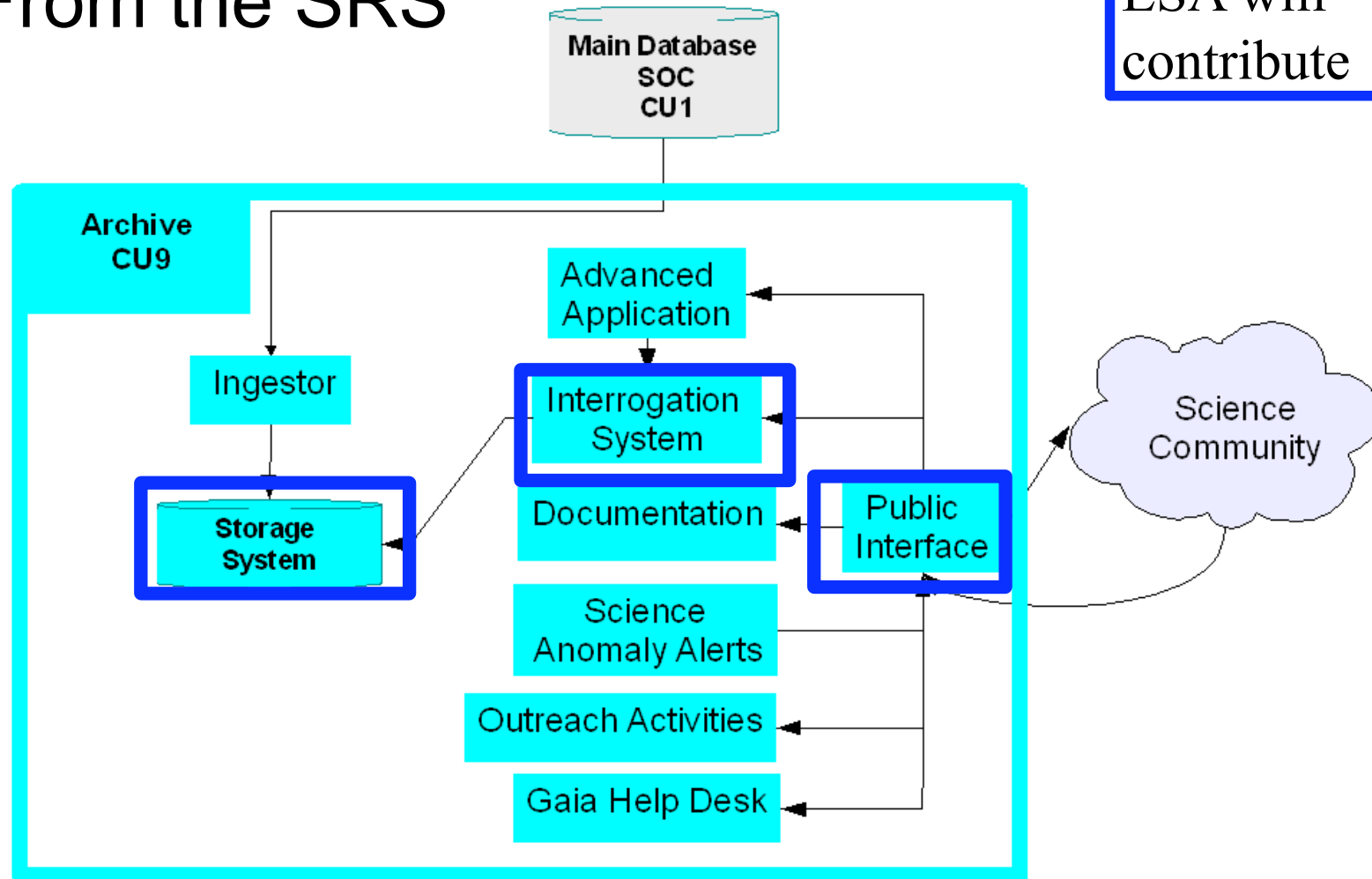
- ESA must have a copy of the Archive (ESAC mandated to be repository of all ESA space science data)
- Does not preclude others also having a copy
- SDSS do this quite successfully.
- Intention is that any institute involved should be allowed a replica – if they are “**SERIOUS**” about it
  - Purchasing machine, disk, licenses etc.



# CU9 decomposition

- From the SRS

ESA will contribute



CU9 may decide on another decomposition

# General Requirements

- Part of DPAC i.e. CU9
  - Reporting etc. similar to other CUs
  - CU9-L in DPAC
- Some DPAC members will have to be in CU9
  - No one understands the data and algorithms better
- Portability – to make easy replicas

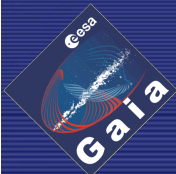


# Community Interface

---

AS Gaia Besaçon, France  
June 29th 2009

- Should be Public oriented
  - i.e. no separate site for “professionals”
- Basics with no login
- Easy self-registration for advanced features
- Some sort of Sky Browser (ala SDSS, GoogleSky)
- Multi lingual of course ....



- MDB ICD already exists
  - Contains ALL data – CU9 needs to use it
- Lots of data
  - Probably needs to be distributed
  - Should not take too long – one week



# Interrogator

- Whatever the interface , VO, Sky Browser etc.  
Need fast engine to answer queries
- Putting data in some DBMS will not be sufficient
  - Tuning needed !!! LOTS OF IT.
- Should allow powerful queries to user (SQL and/or ADQL)
- Local space for registered users to upload data and store query results
- Extraction in multiple formats (csv, FITS, VOTable)
- VO access of course .. TAP, SIAP .....
- AND FAST !!!!!!!



# Advanced Applications

- Assume we will have “added value” apps
- 3D visualisation for all or part of sky (i.e. globular cluster)
  - Animated with proper motions of course
- Light curve tools
- ..... your idea here ...
- Interrogator should provide an API to build these apps on.



- Alerts – bit of a grey area for now
- Flux based alerts will come straight from CU5
- NEO orbits from CU4 go direct to IMCCE
- Others will come up – DPAC is concentrating on processing first !
- CU9 should also provide a facility for DPAC to publish alerts
- However alerts are sent should be archived by CU9
- A follow up should be gathered..





- Usual problem reporting, tracking , FAQ etc..
- Some real people somewhere..
- Should be ready to politely answer all questions



- Probably most important CU9 task
- Uniform documentation on all data and algorithms used to derive
- MDB dictionary provides a lot on the data
- Algorithm documentation to be extracted from CUs
- Project History ...
- Bunch of statistics e.g. maps source, observation, histograms of sources per x time period .....



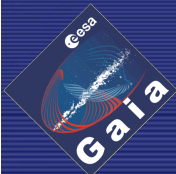
- Derived plots such as
  - Hertzsprung Russell Diagrams
  - galactic-kinematics diagrams
  - Hess diagrams and colour magnitude Plots for multiple stars and for non-stellar sources
  - Where possible/sensible both a print quality copy and an interactive
- Printed volume – would be nice !
- At least Source Catalogue on Media..



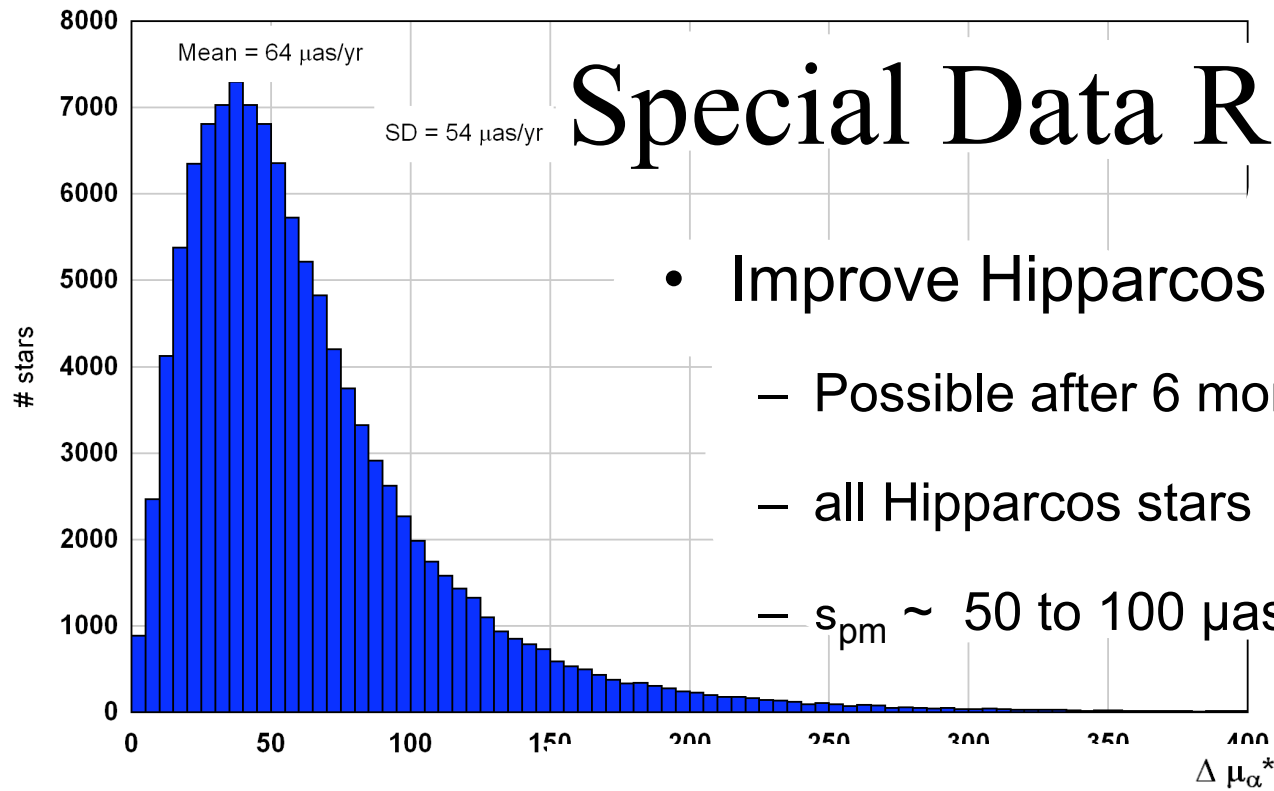
# Releases very tentative ...

AS Gaia Besaçon, France  
June 29th 2009

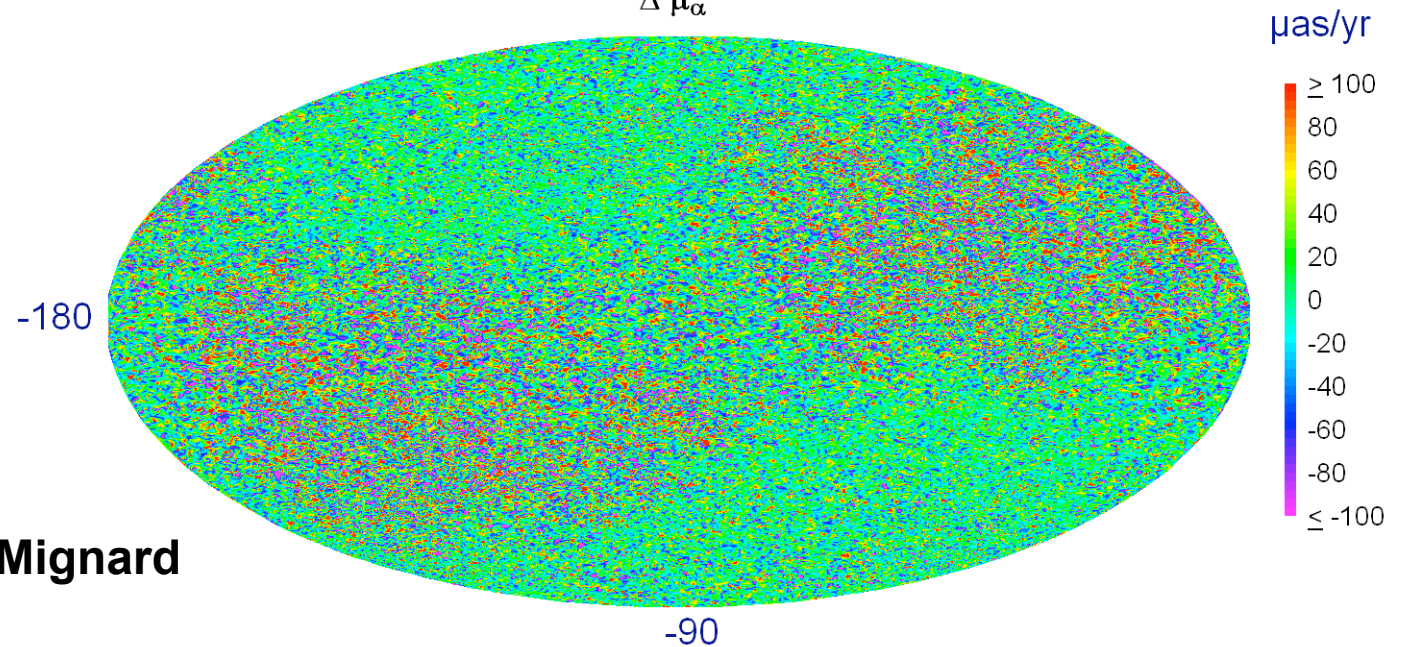
- Yes more than one – we wont make you wait till 2020. But all remains tentative
- Probably one after 2-3 years
  - Some Astrometry and some Photometry
- One more after 5 years
  - Full astrometry better photometry, spectra
- Then the final one 8 years after launch
  - Variable stars , astroparams ...



# Special Data Release



- Improve Hipparcos proper motions
  - Possible after 6 months
  - all Hipparcos stars  $V > 6$
  - $s_{\text{pm}} \sim 50$  to  $100 \mu\text{as/yr}$  with 21 years timespan



From Francois Mignard

# Open areas

- Living archive (Anthony Brown)
  - Can we/should we allow additions to the Archive?
  - e.g. improved solutions for binaries using follow observations
- Implications for maintenance, quality and security
- But there will be no printed catalogue so why not a new type of astronomical archive ?





- Archive as a model (James Binney)
  - How can we compare models of the Galaxy to Gaia Data ?
  - How can the Archive facilitate that ?
- David Hogg (<http://arxiv.org/abs/0810.3851>) goes one further:
  - We should try to encode the archive in a Model



# Open areas

- Cross matching other Catalogues
- A job for the Virtual Observatory?
- Perhaps but VO almost by definition has to give Lowest Common Denominator of Catalogues.
- May merit cross matching a few major surveys and including the matches in the Archive.





# Open areas

- Virtualisation (William O'Mullane)
- Alex Szalay (SDSS) has said for years we need to bring the processing to the data for large Archives.
  - Hence rather sophisticated CASJOBS
- Virtualisation could allow users a virtual machine in the Data Centre with the Archive
  - Code what you want how you want run it there
- Many more things for sure ...



# When does it start ...

---

- No real possibility to have initial catalogue before 18 -24 months after launch
- Hence CU9 could start just after launch
- It is probably enough time to make an initial archive with VO layers
- Would still require some CU9 building to start soon ..
- My personal idea – get CU9-L onboard now!



# Questions ??

AS Gaia Besaçon, France  
June 29th 2009



Ariane V188 carrying Herchel and Planck (May 14 2009)

