Gaia: Processing to Archive

William O'Mullane, Uwe Lammers, Jose Hernandez
European Space Astronomy Centre, Madrid, Spain

Abstract

Gaia is ESA's ambitious space astrometry mission with a foreseen launch date in late 2012. Its main objective is to perform a stellar census of the 1000 Million brightest objects in our galaxy (completeness to V=20 mag) from which an astrometric catalog of micro-arcsec level accuracy will be constructed.
We update the viewer briefly on the status of the Astrometric Global Iterative Solution for Gaia. The results of AGIS feed in to the Main Database (MDB) will be described here also. All results from Gaia processing in fact are in the MDB which is governed by a strict ICD. We describe the Distributed Data Model tool developed for Gaia - the Data Dictionary. Finally we mention public access to Gaia data - the archive. We present current plans and thinking on the archive from the ESA/DPAC perspective.

The Astrometric Global Iterative Solution (AGIS)

- AGIS is the mathematical and numerical scheme that shall be used to generate the core astrometric mission products of Gaia: The five standard astrometric parameters, position (2x), parallax (1x), proper motion (2x), for all observed celestial objects brighter than G=20mag with targeted accuracies (e.g. <10 uas [G<10mag], <25 uas [G=15mag], <300uas [G=20]).
- The best global match between all measurement data and an observational model involving the unknown source, attitude, calibration, and global parameters is sought in a least-square sense a block-iterative solution scheme
- The system is developed in DPAC's CU3 by ESAC (Spain) and Lund Observatory (Sweden)
- Distributed all-in-Java system – about 120.000 lines of code
- Shall run about every 6 months on an ever increasing raw data volume
- The operational system is baselined to run on dedicated processing hardware at ESAC – a 10 TFLOP/s machine is envisaged – an alternative "Cloud" computing solution is being studied

Unknowns: source+attitude+calibration

Residual (O-C) of observation

Downweighting function

$$\chi^2(x) = \sum_i \frac{R_i(x)^2}{\sigma_i^2 + \varepsilon_i^2} \cdot w\left(\frac{R_i(x)}{\sqrt{\sigma_i^2 + \varepsilon_i^2}}\right)$$

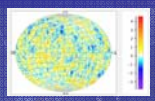
Sum over all observations

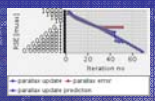
Merit function

Observation noise

Excess noise

- The above minimization problem is cast into an algebraic form with Normal equations and solved **iteratively** with Simple Iterations (SI) or Conjugate Gradients (CG) using a Gauss-Seidel pre-conditioner
- Recently, all four blocks (source, attitude, calibration, global) have been brought into the CG scheme
- CG is about 2 times better in convergence rate





Parallax convergence in recent all-block CG run

- CG is better but needs fixed excess noise values ε_i which are determined as part of a robust (observation) outlier rejection algorithm
- Observation weights and excess noise values must be determined with "Simple Iterations"
- In operations a "hybrid" solution scheme will be used:
  1. Iterate with SI thereby determine observation weights and excess noises
  2. Continue iterating with CG and **frozen** weights
  3. Perhaps go back to step 1 …
- Implementation of the hybrid solver and the outlier rejection scheme is well underway – it is believed to be the last major algorithmic building block of AGIS!!

AGIS and NanoJASMINE

- NanoJASMINE is a small Japanese astrometry mission that launches in 2011
- A Gaia-similar design and operating principles makes it feasible to use AGIS for this mission also
- This collaboration is well-undeway, for details see Poster **P057** by Yamada et. al

For more information contact Uwe.Lammers@sciops.esa.int

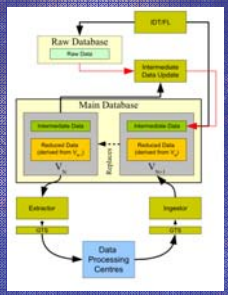
Gaia's Main Database (MDB)

Gaia's repository of Data

- Six Data Processing Centres across Europe will be performing different parts of the reduction of Gaia's Data.
- The Main Database will be the hub of the Gaia Processing architecture, it will collect and integrate data received from the Data Processing Centres.
- There are many interdependencies within the processing tasks running at different locations. The data reduction will be iterative and there will be new versions of the Gaia Main Database regularly.
- The estimated size of the Main Database after the final data reduction is 1 Petabyte.

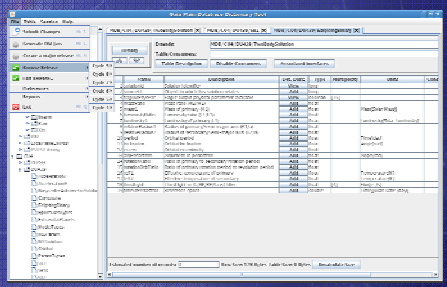
Collaborative Data Modeling

- Defining and agreeing on a common Data Model across the processing consortium is a challenge.
- We've developed an on-line MDB Dictionary Tool
- Each coordination unit defines their part of the model
- From the Dictionary Tool we make DM releases, generate the Java entity classes, database schemas, ICDs…
- Full Object Oriented modeling is supported
- Dependencies and interfaces are tracked



Gaia's Data Model facts

- Ten releases in the period 2006-2010
- 100 DPAC contributors
- More than 1000 tables
- Over 8000 parameters
- CCB in place reviewing and approving releases



For more information contact Jose.Hernandez@sciops.esa.int

The Gaia Archive

Timeline

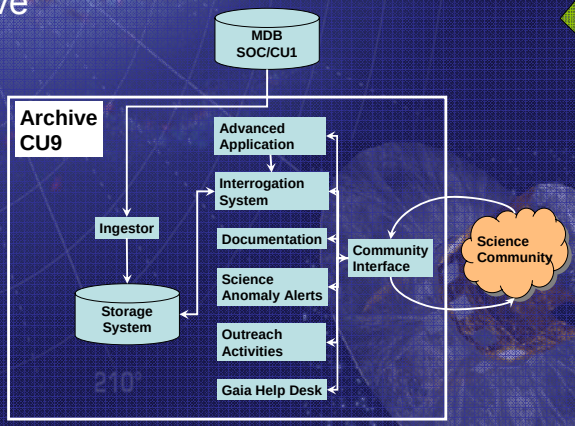
2009: First archive requirements draft
2010: Gaia Archive Preparation Working Group (GAP) forming
2014: First data to public about 2 years after launch
… : Other releases follow
2021:Final catalogue release
… :Figure out if we can reprocess on our laptop !

Gaia Archive Preparation Working Group (GAP)

Charged with organising Gaia archive work namely:

- Coordinating a response to the ESA AO for the Gaia Archive
- Definition of and Archive plan for activities in each CU
- Soliciting interest and intent from parties currently external to DPAC
- Definition of a mode of work between existing CUs and possible new partners
- Coordinating any externally funded activities/investigations considered as potentially useful to the Gaia Archive

If you think your institute should be involved contact womullan@sciops.esa.int.



Possible Archive Architecture