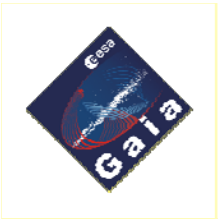


Gaia
Data Processing Framework

Mark ter Linden
Gaia Science Operations Centre
European Space Astronomy Centre (ESAC)
Madrid, Spain

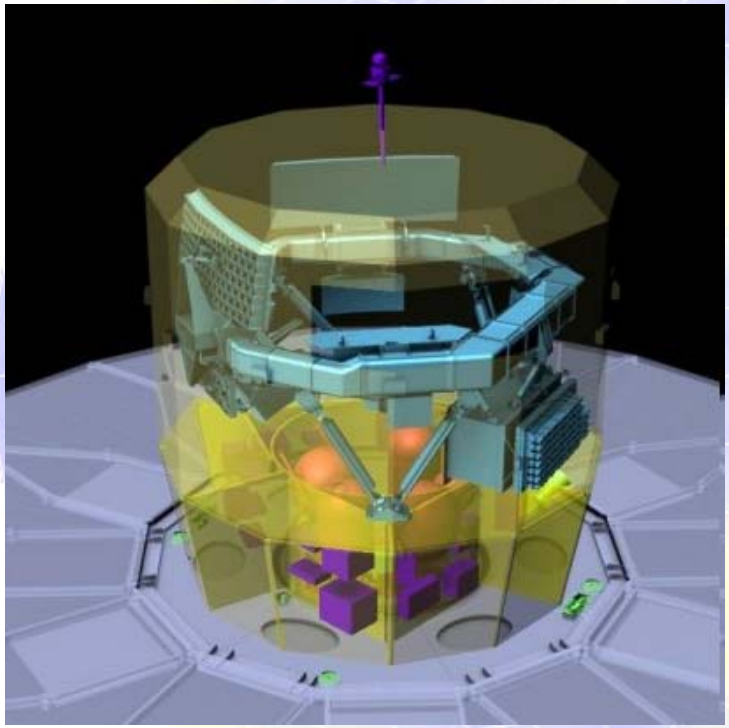
Outline

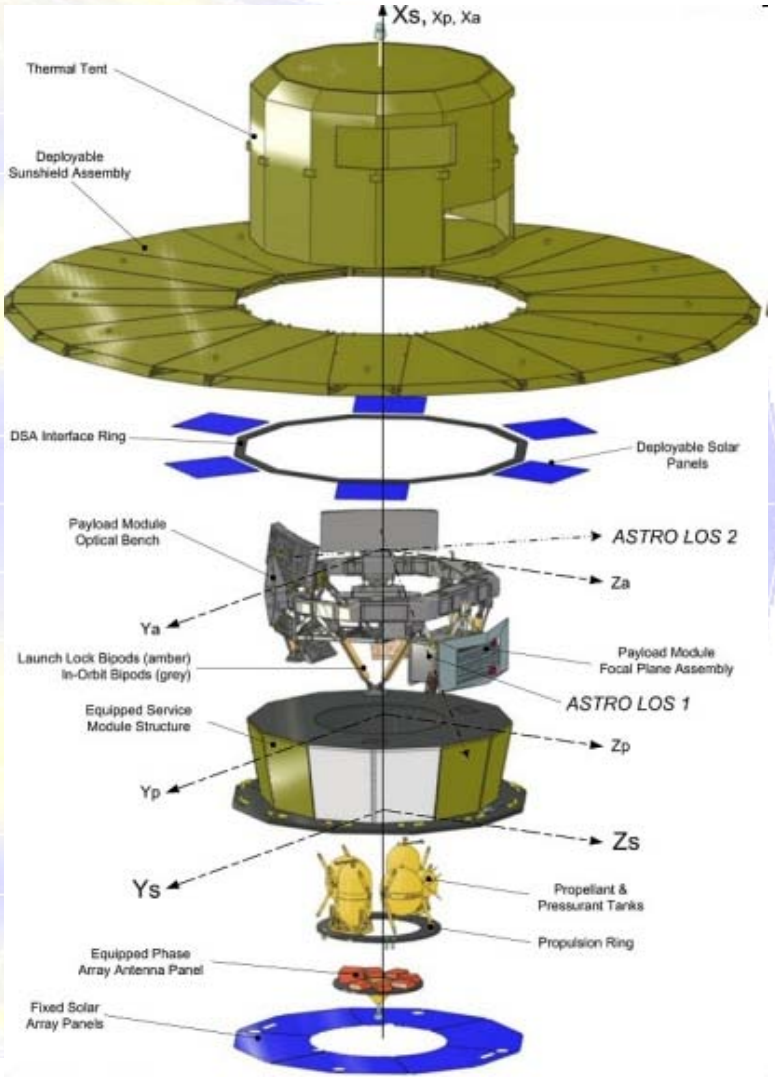
- Overview of mission and hardware
- Ground software development
  - Organisation
  - Standards, methods, tools
  - Architecture

Mission

- Astrometric properties of 1 billion (10^9) sources (G < 20) at microarcsec accuracy
- Photometry (G < 20)
- Radial velocity (G < 16)
- To be lauched in spring 2012
- 5 year survey mission in L2
- Daily about 30 GB download from satellite

Hardware




Xs, Xp, Xa
Thermal Tent
Deployable Sunshield Assembly
DSA Interface Ring
Deployable Solar Panels
Payload Module Optical Bench
ASTRO LOS 2
Za
Ya
Launch Lock Bipods (amber)
In-Orbit Bipods (grey)
Payload Module Focal Plane Assembly
ASTRO LOS 1
Equipped Service Module Structure
Zp
Yp
Zs
Ys
Propellant & Pressurant Tanks
Propulsion Ring
Equipped Phase Array Antenna Panel
Fixed Solar Array Panels

Software

- Ground software to produce final catalogue.
- Large number of observations: 10^{12}
  - 1 millisec per observation → 31 year
  - Obviously parallel processing necessary
- Expected computing power: 10^{21} FLOP
- Contributions from large number of institutes and observatories in Europe
- Funded by community
- To be maintained until 2020

DPAC

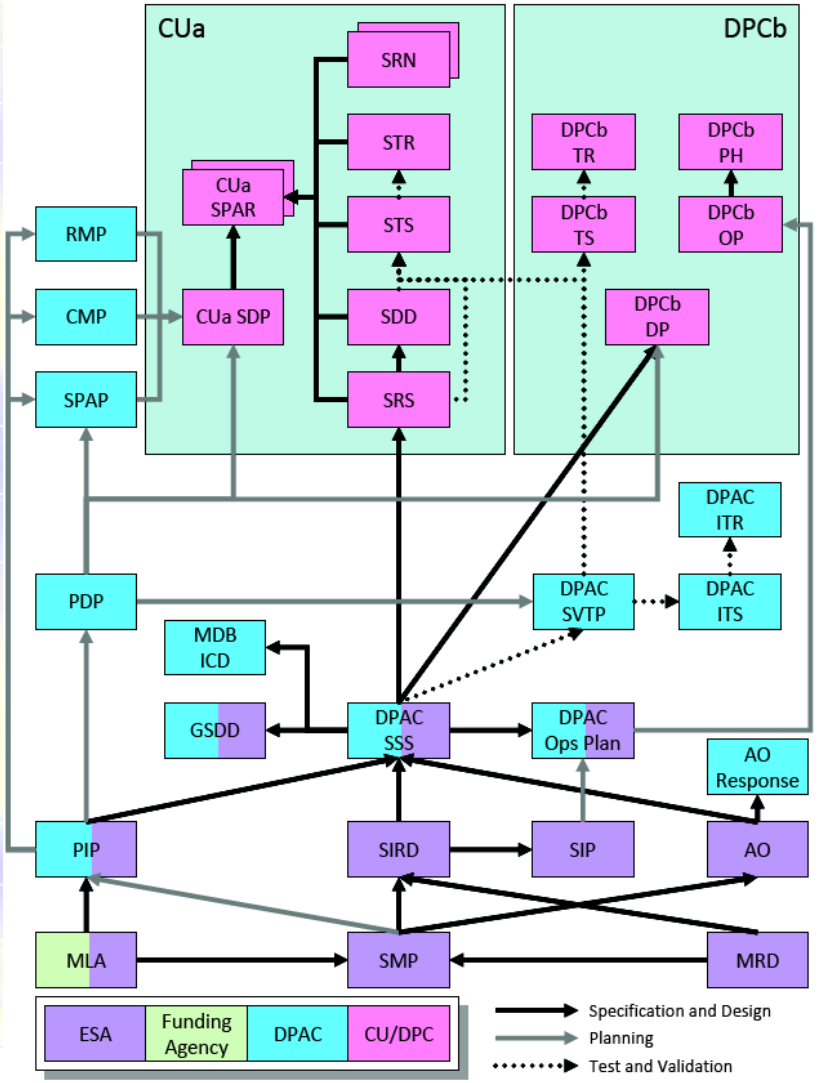
- Data Processing and Analysis Consortium
- Large number of people involved: > 360
- Organized into:
  - 9 Coordination Units (CUs)
  - 6 Data Processing Centres (DPCs)
- ESAC mainly CU1 (System Architecture) and CU3 (Core processing), 20 persons
- 6 Months development cycles, starting from 2006 till launch

Standards

- ESA project – comply with ECSS standard
  (European Cooperation for Space Standardisation)
- Describes processes and documents
- To be tailored for specific projects.
- Many documents to be produced, see document tree (next slide)

Document Tree

- Development plans
- Requirement documents
- Design documents
- Test plans & reports
- Interface Control documents
- Etc.



Collaboration Tools

- Wiki (share information)
- Mantis (issue management)
- Livelink (document repository)
- Subversion (configuration management)

Extreme Programming

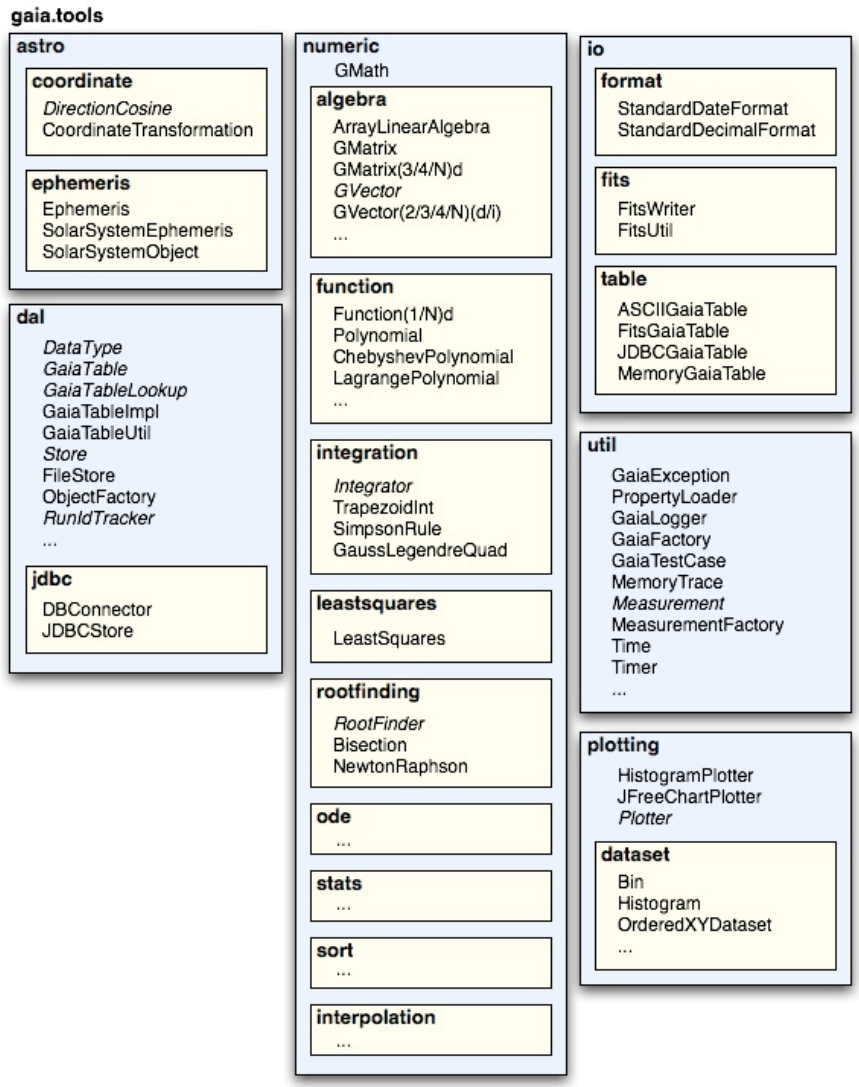
- Agile techniques
- Flexible compared to traditional top-down waterfall model
- Monthly planning meetings
- Stories (activities) with points (1 point is half a day), limited to ~10 points per story
- 6 Months development cycles

Coding

- All programming in the Java programming language
  - Portable (important for 20+ years)
  - Many tools available, such as Integrated Developement Environments (Eclipse)
  - Performance not so bad anymore (just in time compiler, hotspot, improved garbage collection)
- CU1 organizes Java workshops (1 or 2 / year)
- Currently 1 million LOC, expected 3+ million

GaiaTools

- Developed and maintained by CU1
- Reduce duplication
- Tools for:
  - Data access
  - Numerics
  - Infrastructure
  - Plotting
  - Etc.

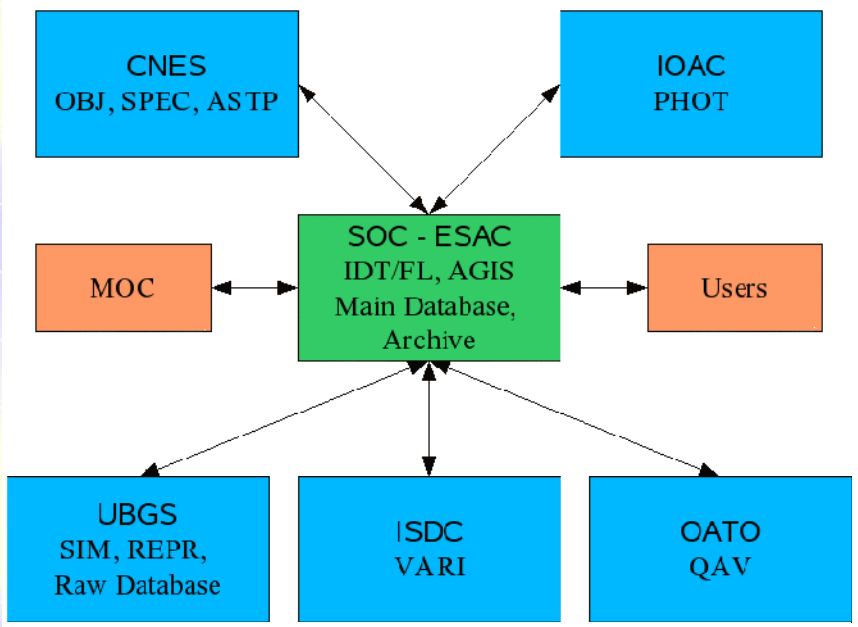


Gaia Parameter Database

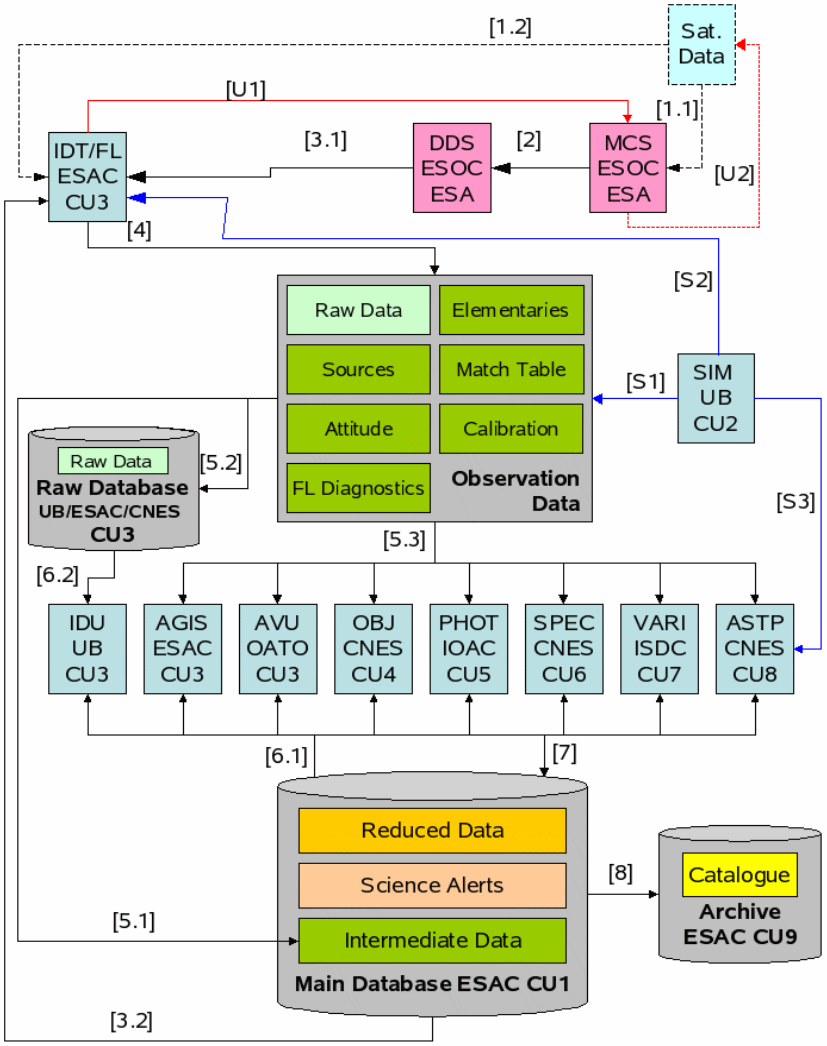
- The Gaia Parameter Database is the central, searchable repository of parameters pertaining to the mission and all its elements.
- It comprises a large variety of entities, ranging from scalar numerical constants to multi-dimensional data sets.
- To be used via web browser or directly in Java code.

Architecture

- Hub-spoke architecture
  - Distributed
  - Limited number of interfaces (reduce dependencies, risks)
  - Independent DPCs



Overall Data Flow



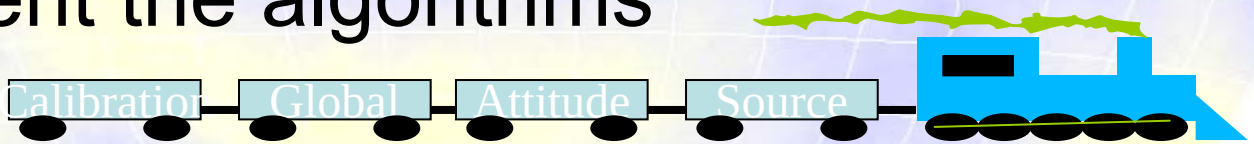
[U1] IDT/FL inputs to MCS – Planning/Calibration
[U2] MOC satellite uplink , TC, Calibrations
[1.1] HK from Gaia to Mission Control System (MCS)
[1.2] TM delivered to Initial Data Treatment and First Look at ESAC directly from ground station
[2] HK data delivered to DDS.
[3.1] IDT/FL Orbit and Aux Data from DDS
[3.2] IDT/FL inputs from Main DB (Obs, Calibration)
[4] Daily observational files created
[5.1] Daily data ingested to Main Database
[5.2] Raw Data Archived in Raw Database
[5.3] Daily data delivered to Subsystems as required.
[6.1] Main Database version delivered to Subsystems
[6.2] Raw Database input to Intermediate Update system
[7] Subsystem updates delivered to main database.
[8] Main Database delivered for Catalogue production
[S1] Possibility to produce observational simulation
[S2] Possibility of TM simulation
[S3] Special Training Set Simulations for CU8

Data Model

- Single data model shared by all CUs.
- Online tool to maintain the model.
  - Keeps data model consist
  - Allow to access & modify model remotely
  - Generates source code and database schemas
- Main database at ESAC, 200 TB in final version.

Infrastructure

- Whiteboard – stores jobs, every node can access it
- Data trains – run on the nodes, retrieve jobs from whiteboard, retrieve data from store and pass data to the takers which implement the algorithms

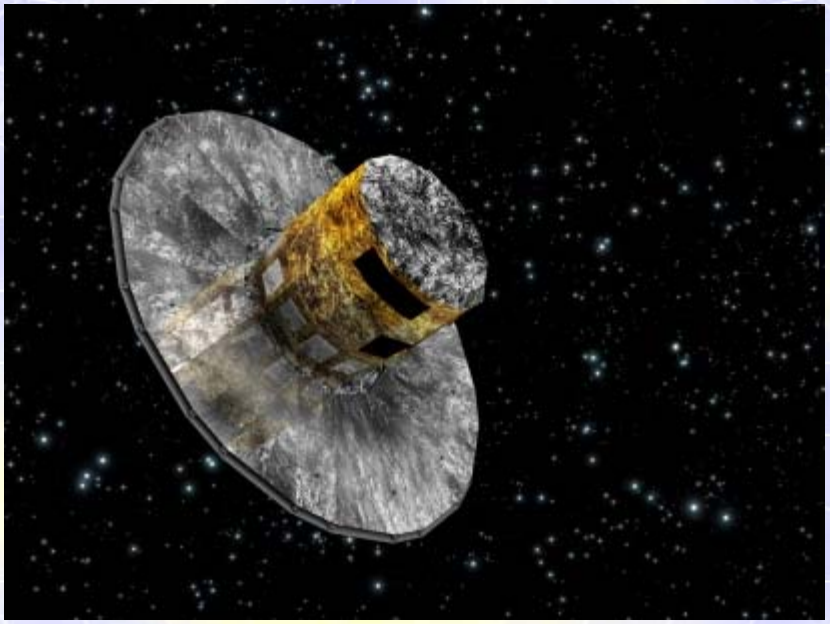


- Optimized for efficient access to data
- RMI servers provide auxiliary data

Clouds

- Experiments with AGIS (Astrometical Global Iterative Solution) on Amazon cloud E2C
- Can provide very large number of nodes
- Easy to port software
- Can be cost effective during operations (4K euro vs 4M euro in 2015 for AGIS)

End



Questions?