

Organisation of the Data Processing

F. Mignard

From GST and DACC inputs

Objectives and challenges

- Gaia will observe over 5 years with three instruments
- Gaia will produce ~ 200 TB of raw data
 - At the lowest level they are primarily CCD counts
- The DP must cope with
 - A large treatment both in volume and computation
 - A complex and iterative treatment
 - Interconnected data and tasks
- It is a community task, neither organised, nor funded by ESA
- It must be ready by the launch time

Objectives and challenges

- It requires specific expertises and commitment :
 - Overall analysis of a complex project
 - Professional S/W developers
 - Quality control
 - Strong commitment of teams over ~ 10 to 15 years
 - Ability to work with large team under pressure with tight schedule
- The basic organisation does not exist yet
 - This is very different from the WG structure
 - The level of coordination is much more demanding
 - Science will not be the primary activity in the DP
 - The organisation must have clearly defined responsibilities

- There are three obvious main steps in the DP
 - The initial data treatment of the CCD images up to centroiding
 - The astrometric, photometric and spectroscopic iterative solution
 - The special non iterative processes
- The global structure should be
 - Must be as simple as possible
 - Clear hierarchy with non overlapping responsibilities
 - GST must have the full visibility and retain control of the DP
- The current workplan
 - Global system called Gaia Data Analysis Consortium (DAC)
 - Central concept around Coordination Units (CU)
 - Within each CU there are Development Units (DU)
 - The lower level is the Work Package (WP)

Coordination Units

- Small in number (< 10)
- Each must be in charge of a significant fraction of the DP
- Interfaces between CU's are limited
- Each CU coordinates the developments of many tasks
 - It defines the structure in DU
 - It follows the development, testing and documentation of the S/W
 - It defines and implement the interfaces between the DUs
- A CU is placed under the responsibility of a manager and his deputy
 - They manage their CU
 - They provide contact and interfaces with the other CUs

Identified Cus (provisional)

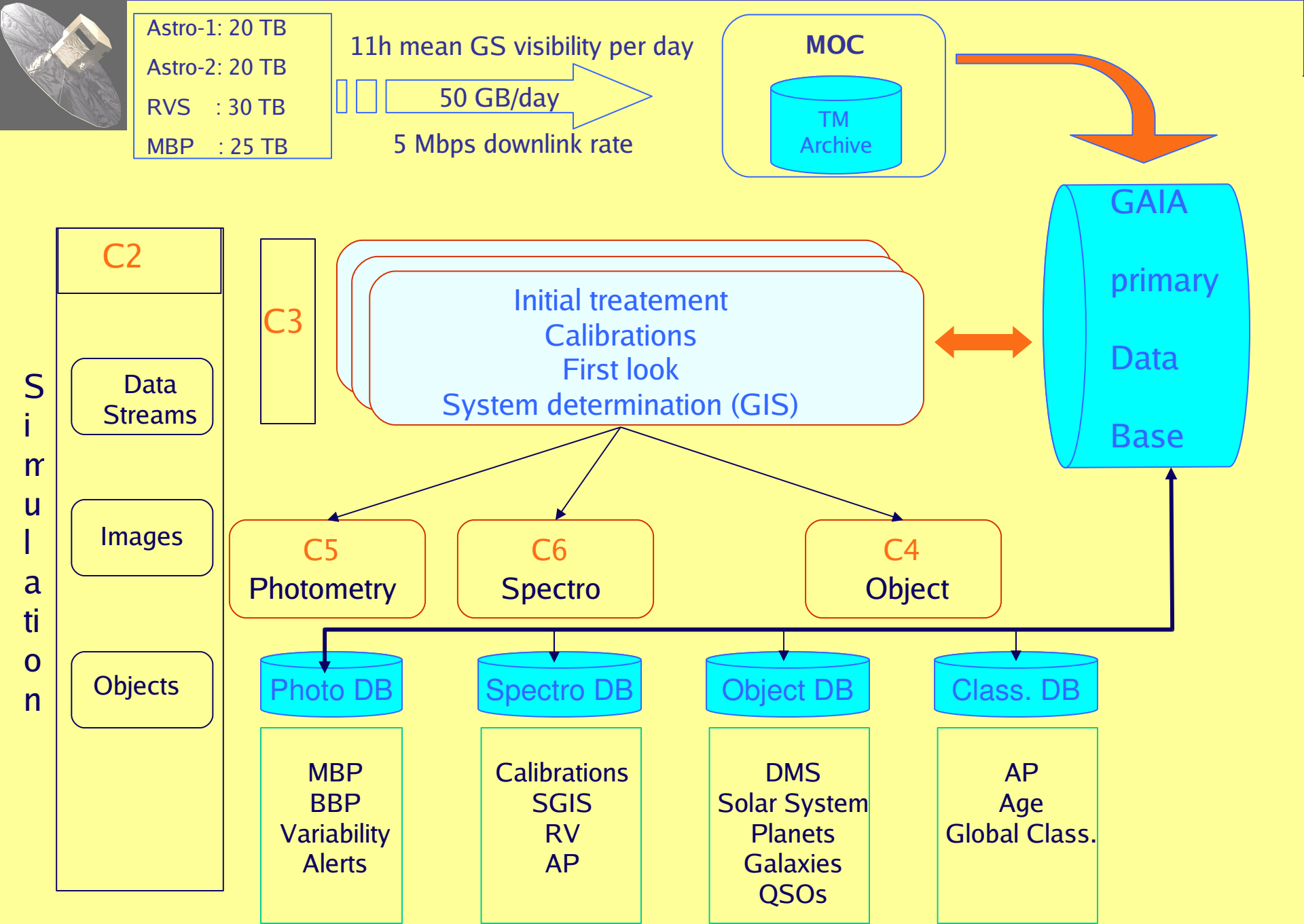
- C1 : Overall system architecture and central DB
- C2 : Data simulation (three levels)
- C3 : Astrometric core processing
- C4 : Object processing
- C5 : Photometric reduction
- C6 : Spectroscopic reduction
- C7 : [*Scientific exploration and Catalog access*]

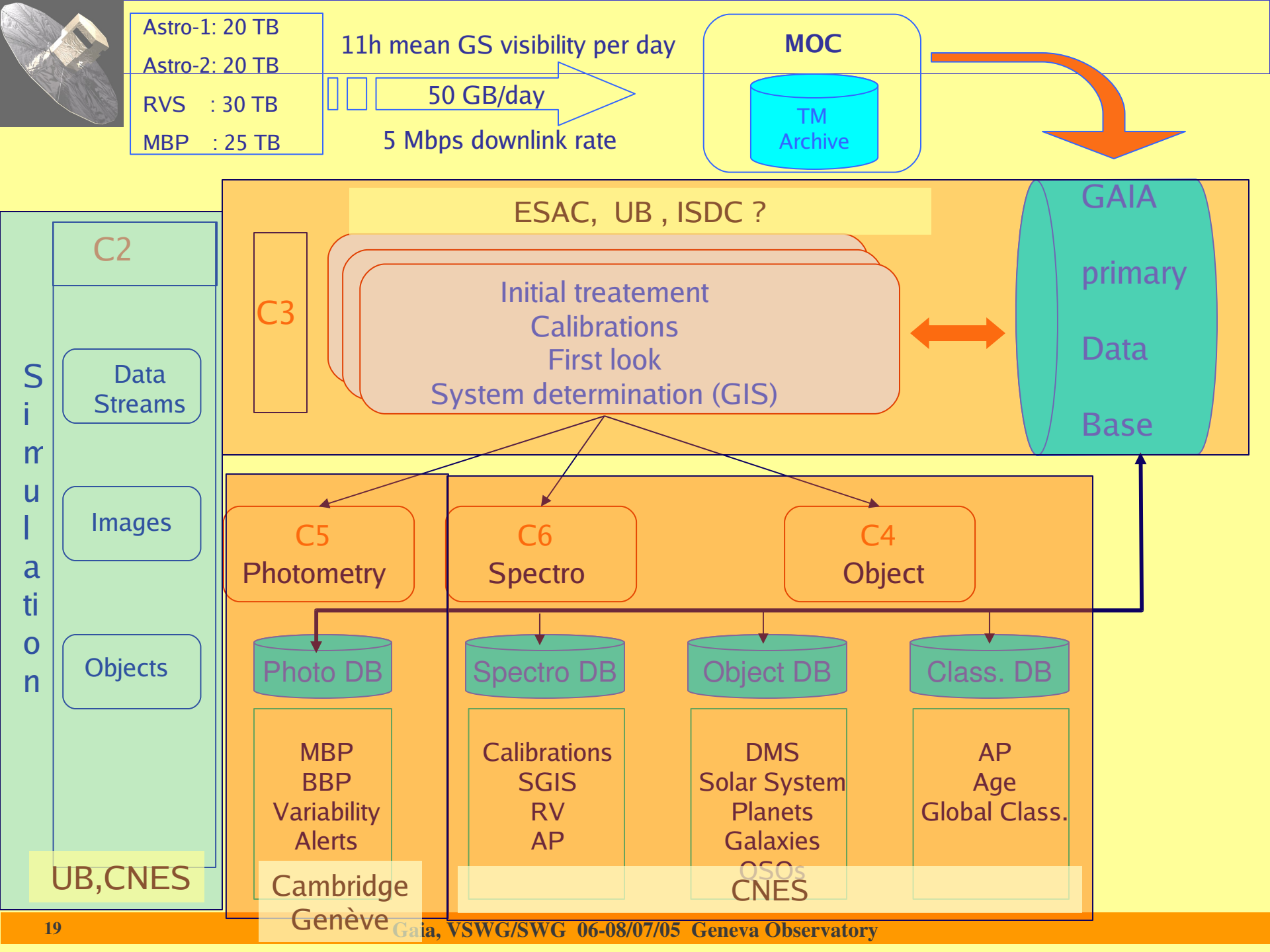
Identified Cus leaders and deputies

- C1 : Overall system architecture
W. O'Mullane, U. Lammers
- C2 : Data simulation
X. Luri, C. Babusiaux, F. Mignard,
U. Bastian, M. Lattanzi, J. Torra
- C3 : Astrometric core processing
D. Pourbaix, C. Bailer Jones, P. Tanga
- C4 : Object processing
F. van Leeuwen, A. Brown, C. Jordi
- C5 : Photometric reduction
D. Katz, M. Cropper, U. Munari
- C6 : Spectroscopic reduction

- Responsibility for the development of S/W
 - Conceptual development
 - Algorithm testing on simulated data
 - Preparation of the implementation of the operational versions
 - Process the Gaia data and provide outputs for the DB
- A DU should be based on a small number of well connected groups
 - Typically 5 people in each place
- Examples of Dus :
 - PSF calibration, RV determination, orbit determination, taxonomic classification

- Overall architecture not yet defined
- Several approaches possible :
 - All computational H/W and DB at one place serious funding problem
 - All core processing at one place + distributed shell tasks data transfer
 - Centralised data base and core tasks ; separate location for integration of shell tasks and centralised or distributed execution
 - Local data bases with shell tasks processing and coordination units





How to form the Consortium ?

- Setting of a DACC (Data Analysis Coordination Committee)
- Must establish the structure of the consortium :
 - Breakdown into CU compatible with data flow
 - List of DUs within each CU
 - Examine the funding issues, data flow problem, system architecture
 - Identify institutes, teams or individuals for the WPs
- DACC established for one year
 - Will be replaced by the DACE (Data Analysis Consortium Executive)
- An AO will be issued to formalise the structure
 - no competitive AO is foreseen

Composition of the DACC

- Chair : F. Mignard
- Co-chair : C. Bailer-Jones
- CU managers :
 - W. O'Mullane, X. Luri, U. Bastian, D. Pourbaix, F. Van Leeuwen, D. Katz
- Project scientists : M. Perryman
- + Experts : A. Brown, U. Lammers, L. Lindegren, J. Torra

Pyramidal structure

- GST (includes all the CU managers)
- Data Analysis Consortium Executive (DACE)
 - Consists of the CU managers
- Coordination Units
 - Manager, deputy manager, S/W engineers, DUs
- Development Units
 - Leader, includes several work packages
- Work Packages
 - From individual to small groups within one location

- List of work-packages of each CU
 - Compilation document in preparation
- Construct the data flow constraints
 - Functional analysis
 - List of processes and dependencies
- CU managers build their architecture of DUs
 - Should reflect the answers to the Lol
 - Large contact with the community needed
- Next meeting of the DACC : 6-7/10 in Heidelberg

