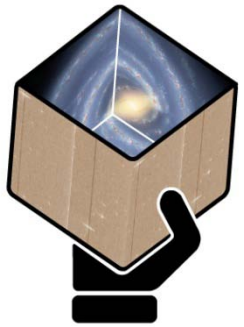
The Gaia Archive (CU9)

REG 2016

gaia

Gaia DPAC

Data Processing & Analysis Consortium

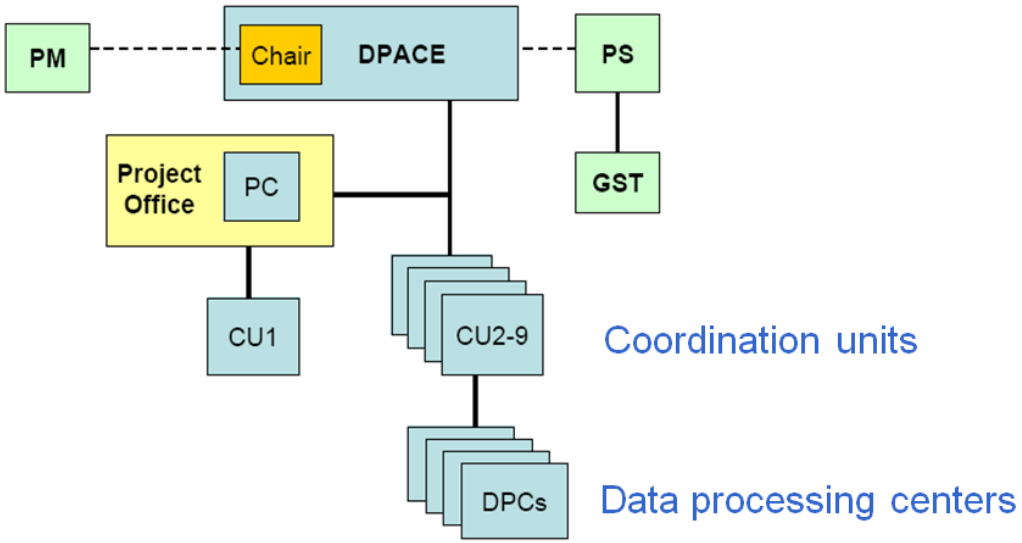


X. Luri

The CU9 and the DPAC

The CU9 is one of the DPAC CUs, but was intentionally left out of its initial structure of in the ESA AO for the creation of the consortium:

- **It was too early to start working on the archive (2007)**
- **ESA wanted to handle the group working on the archive separately**



References for CU9 work

CU9 announcement of opportunities

Defines CU9 responsibilities

- CU9 will be the only responsible for making all Gaia data available except the Science Alerts which are released directly by the respective processing entities. For Science Alerts the CU9 responsibility is limited to recording a full set of all Gaia alerts

- CU9 must be ready to issue the first intermediate data release, currently estimated for launch+22 months. CU9 must also assume 6 months for commissioning and performance verification activities and 5 years of operations. After the nominal 5 year operational period CU9 must assume a three year post operational period for the final data processing and catalogue issue.

- As defined in the SMP, the Gaia mission follows a 'no data rights' policy. […] As stated in the SMP, the GST will define a policy with regard to data access in DPAC.

Gaia Science Management Plan

Basic reference for the Gaia data processing

- Establishes the Gaia Science Team and its responsibilities
- Specifies that **“**In addition, **the Gaia scientific community will contribute, in close cooperation with the SOC, to the design, development and operation of the final Gaia data base, which will contain the intermediate and final mission products.”**
- Sets up the DPAC structure and responsibilities, including ESA commitments to the DPAC

The SMP defines the strategy for the data releases and the data rights

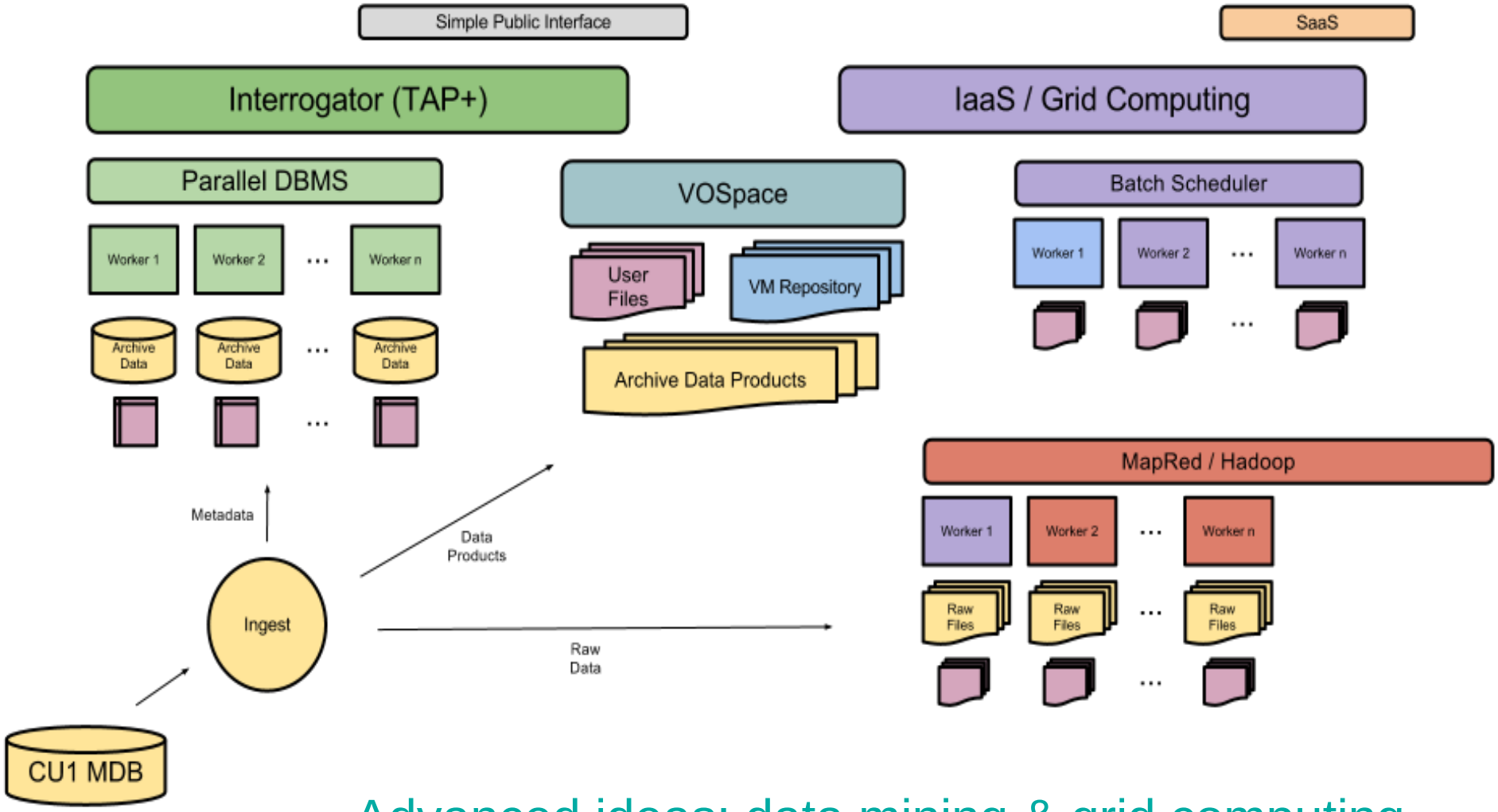
[…] **formal 'data rights […] will not be assigned to any scientists involved in any aspects of the mission, including those participating in the data processing**.

Data will be openly available at the time of each one of the catalogue releases.

Teams participating in the catalogue preparation may have early (pre-official release) access to reduced data (e.g. for validation purposes) in a controlled way.

The archive

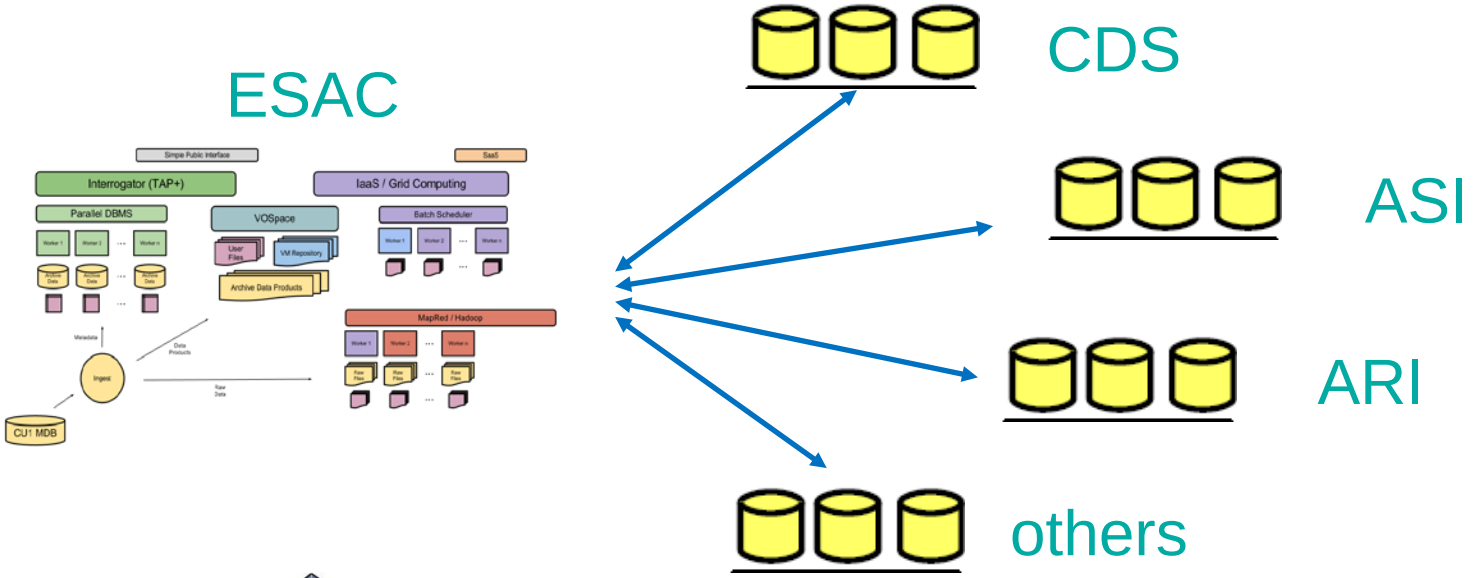
Located at ESAC, operated by ESA



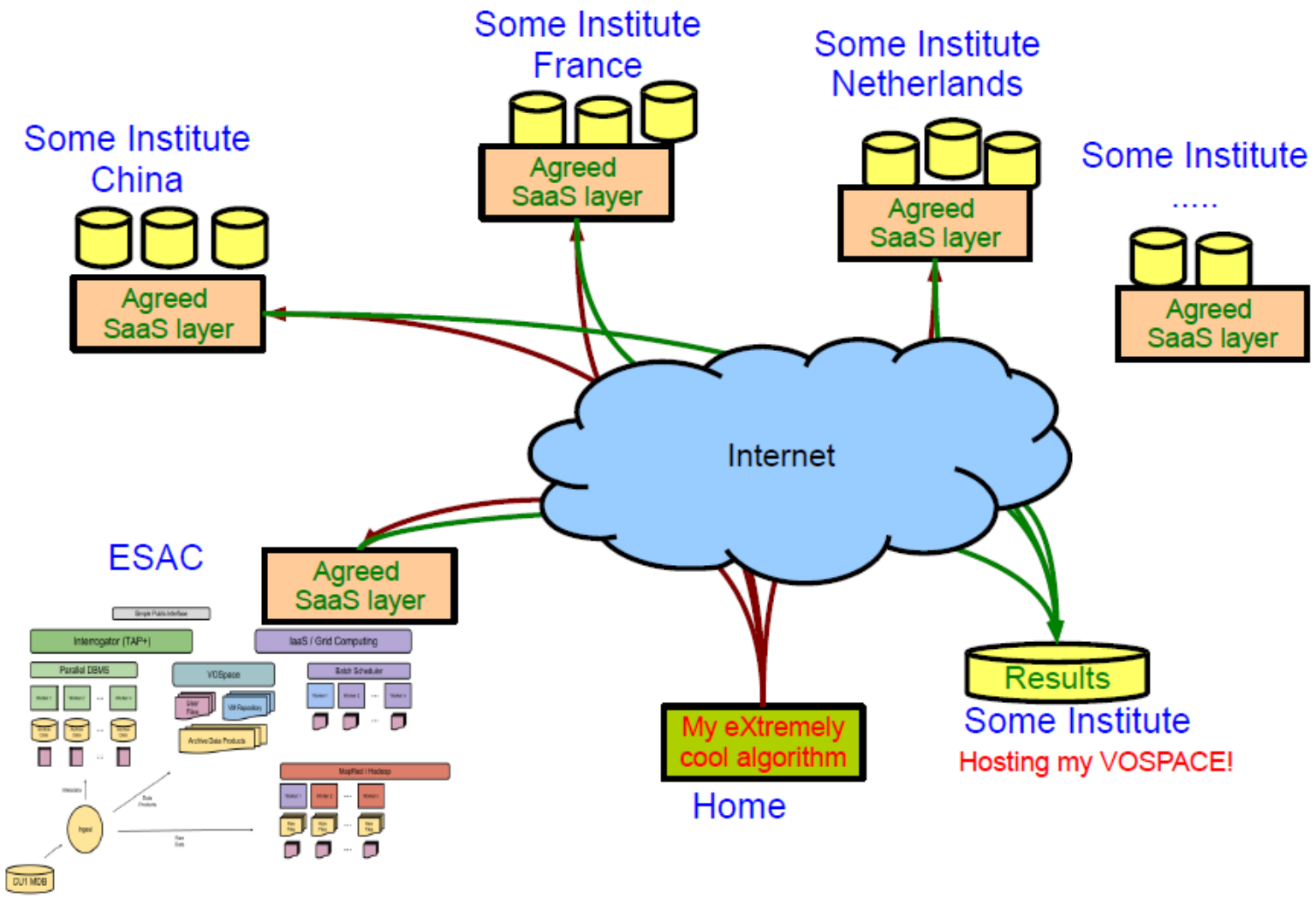
Advanced ideas: data mining & grid computing

Supporting centers

- Partial replication of the Gaia data (only combined data)
- May provide more services to users
  - access through familiar services (e.g. Simbad)
  - value-added services: joint catalogue queries
  - Local data mining
- Closely linked to main archive @ESAC: references point back to full data



Future: cloud computing?



Data and tools

CU9 products

- The data in the catalogue releases

+

- The archive
- Documentation
- Science enabling apps
- Visualization tools
- Outreach activities

User driven development

The CU9 goal is to produce an archive and associated tools **based on the needs of the actual users**. For this an effort has already been done through GREAT to define use scenarios for the archive. This information will be used to drive the CU9 development.

GAP: requirements generation

- [20] GDAS-BR: Browsing and qualitative exploration: including simple usages of first-time users
- [6] GDAS-SA: Science alerts
- [4] GDAS-ED: Early data access: thus access to data from the first releases
- [11] GDAS-EG: Extra Galactic science
- [25] GDAS-GA: Galactic science
- [25] GDAS-ST: Stars and Stellar Physics science
- [4] GDAS-SO: Solar System science
- [1] GDAS-FP: Fundamental Physics science
- [8] GDAS-PR: Public outreach for the non-astronomer/non-scientist
- [19] GDAS-OA: Other and advanced usage scenarios
- **[123] = number of cases submitted as of 2012-06-18**

- Requirements categorised into ten topic areas
- Each described as a user scenario
- Requirements document released 21 June 2012
  - **GAIA-C9-TN-LEI-AB-026-1**

CU9: Top 6 Science Scenarios

Scenario label	Scenario	Rating
GDAS-BR-01	I'm completely new to Gaia. Tell me all about what is contained in the Gaia Archive, and give me some clear examples illustrating how it might be useful to my science.	**48**
GDAS-BR-10	I would like to be able to select objects based on any set of the variables provided in the Gaia catalogue position, parallax, astrophysical parameters, proper motion uncertainties etc. These selections should not be limited to simple "axis-parallel" cuts or cone cuts, but permit a broader array of functions/functional dependencies. An example is selection on fractional parallax error and some relation between G magnitude and extinction. Another is selection on space velocities, which requires a combination of position, parallax and proper motion.	**48**
GDAS-BR-07	I want data of all objects contained in a rectangular/circular region centered on a given sky position.	**48**
GDAS-EG-04	I want the RVS spectra of my favourite source(s)	**48**
GDAS-EG-09	I want astrometric and/or photometric and or/spectroscopic measurements of a specific type of source (GDAS-EG-2), but possibly for each and every epoch of observation.	**48**
GDAS-EG-10	Information about flux variation and position among objects with multi-epoch photometry	**48**

Documentation

Each catalogue release will be fully documented, including details of the archive contents, data processing, validation and error analysis.

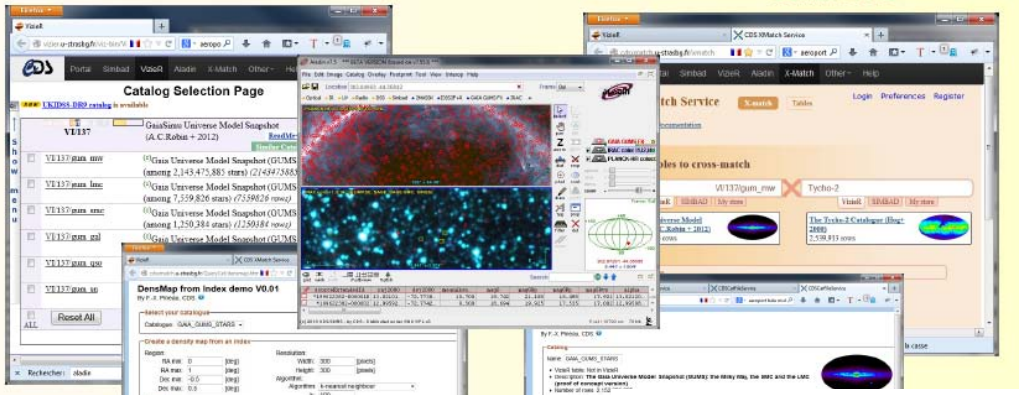
The documentation will be written in cooperation with the CUs doing the actual data processing.

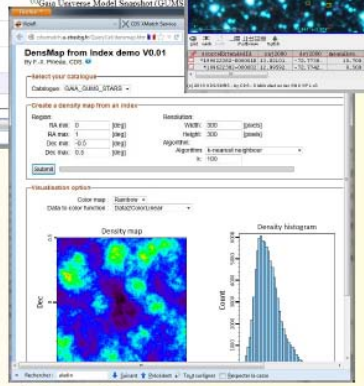
Science enabling applications

The CU9 plans include not only the implementation of a query interface to the archive but also advanced tools allowing a more performing usage of the data:

- Adaptation of existing tools to the Gaia data: Simbad, Vizier, TopCat, VO tools
- Data mining framework allowing massive data exploitation + basic data mining tools
- Pre-computed cross-matching for main existing catalogues
- Integrated auxiliary data
- Integrated science alerts data

VizieR Aladin Xmatch





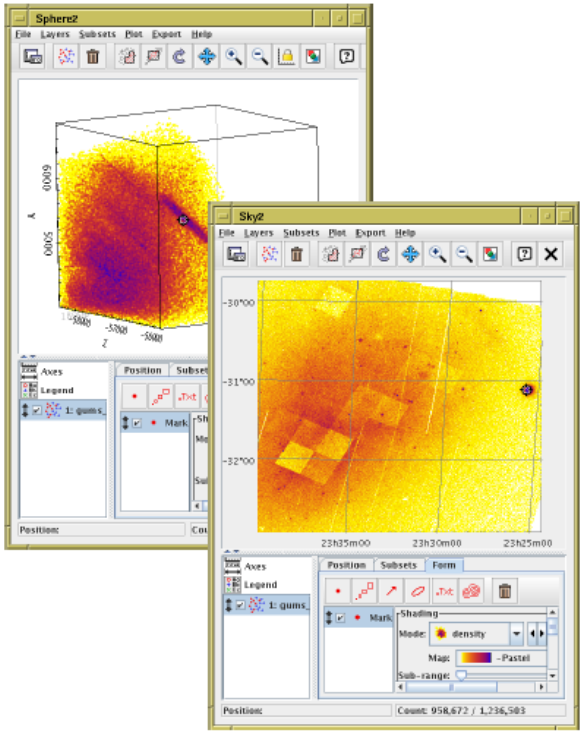
Density map generator

Mul
quer

1.2 million objects by RA, Dec, dist. (SMC region of GUMS)

- Easy navigation:
  - Mouse drag pans (2d) or rotates (3d)
  - Mouse wheel zooms in/out
  - Right click re-centers (3d)
- Easy to see simulation artifacts

Topcat



Visualization

Visualization tools can help improving the scientific exploitation of Gaia data, as well as its outreach. The CU9 plans include the development of several visualization tools for the archive. In a nutshell, WP980 must provide

- 2D browsing and plotting tool(s) with image and graphics overlay support and VO layer.
- 3D navigator/explorer
- Interactivity
- Multiple panels with linked views
- Details on demand [requires pre-computation]
- Visual data selection (and visual queries)

Capable of handling the whole final Gaia catalogue

For GDR1, WP980 will provide:

- 2D browsing and plotting tool(s) with image and graphics overlay support and VO layer. ~10ˆ9 sources
  - Galaxy map; +UCs depending on xmatch
- 3D navigator/explorer. TGAS: ~2.5x10ˆ6
  - Pre-computed Galaxy view; +UCs depending on xmatch
- Interactivity. YES
- Multiple panels with linked views. PARTIAL
- Details on demand. Visualisation server running at ESAC
- Visual data selection (visual queries - not for GDR1)

Outreach

Outreach is considered an integral part of CU9 activity. We will work to ensure the maximum impact of the Gaia products at several levels:

- Academic Outreach
- Education and General Public Outreach
- Media
- Coordination of national activities

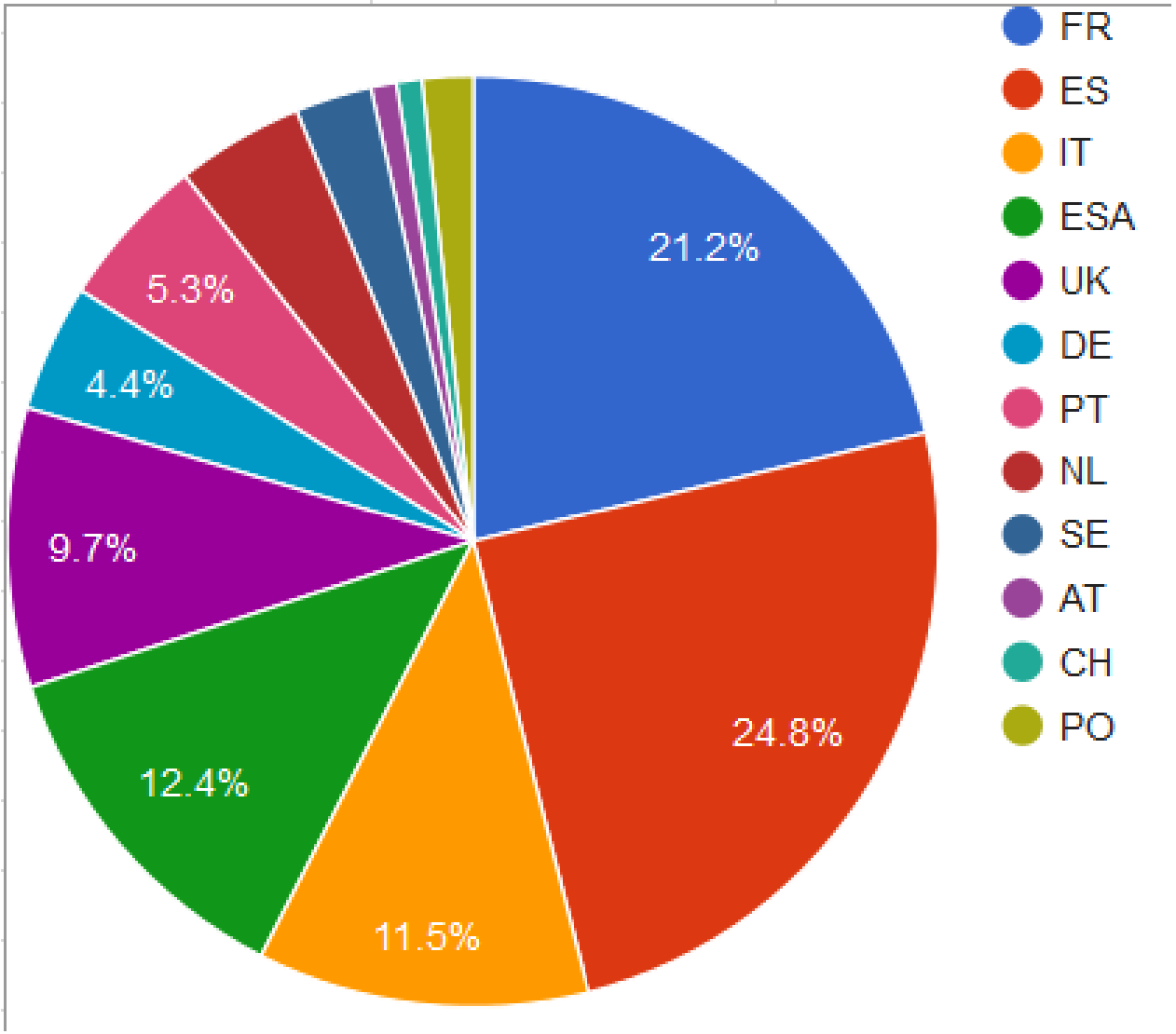
Includes web portals and social media

Status

Current WPs and managers

WP910	Management	X. Luri & W. O'Mullane
WP920	Documentation	F. Van Leeuwen & J. de Bruijne
WP930	Architecture	N. Hambly & J. Salgado
WP940	Validation	F. Arenou & P. di Mateo
WP950	Operations	(G. Gracia) & J. Hernández
WP960	Outreach	S. Jordan & E. Masana
WP970	Science Enabling Applications	X. Luri & P.M. Marrese
WP980	Visualization	A. Moitinho & J. Alves

186 Members (114 full + 71 affiliate)



Effort per country