



L'Universo di Gaia

(Scienza e Tecnologia)

Mario G. Lattanzi (INAF-OATo)

per la partecipazione Italiana alla missione Gaia



PREMESSA: Questa presentazione sulle attività Italiane per la realizzazione della missione Gaia non può che partire dai sinceri e sentiti ringraziamenti ai ~ 70 collaboratori INAF o assoc. INAF (non tutti censiti qui) e gli oltre 20 colleghi di aziende che operano nel campo delle Scienze Spaziali e Tecnologie correlate e dell'Informatica avanzata e che hanno permesso, con il loro lavoro di quasi 15 anni, alla Dott.ssa Busonero e me di mettere insieme le schede GaiUniverse-0 e la figlia "G_DPCT" (GaiaUniverse-1). Queste sono dedicate, nel contesto specifico del PT 2021-2023 che fa da sfondo a queste audizioni, alle attività di fase operativa della missione.

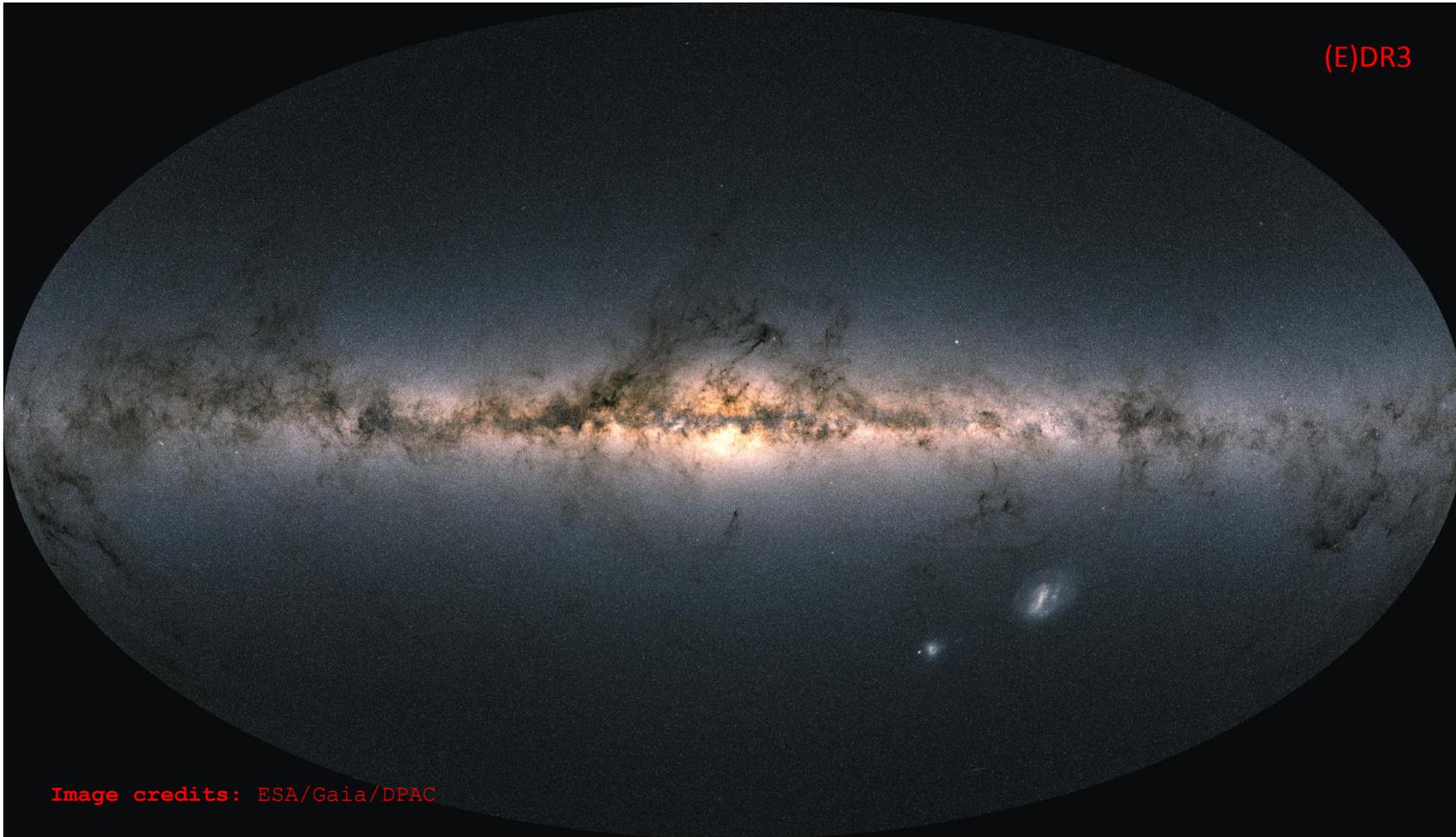
Pertanto, questa presentazione non riguarderà gli ultimi risultati della missione Gaia. L'enfasi è sul contributo Italiano (INAF), la sua qualità e quantità, il suo impatto nel triennio di interesse.

Infine, diversi i riferimenti alla missione Gaia da tutte le RSN, anche se non sempre via schede "collegate". Un elenco dei "collegamenti" di cui sono a conoscenza segue:

- Schede Programma : CARPA, **EXO-Arch**, **EXO-Star**, **EXO-Young**, Galassia.To, GraviMetrA, STARK,
- Schede Progetto: ASTRA, FOTOSS, PLATO, **SHOT**, TASSEL, **WEAVE**
- Schede Didattica di Alta Formazione: MAA, GMAC

Si noti l'importanza delle schede Programma che si "implementano" in Gaia e che dal suo "caso scientifico", da cui sono derivati i requisiti di missione oltre 25 anni fa, hanno tratto ispirazione per il futuro.

L'Universo di Gaia: Introduzione



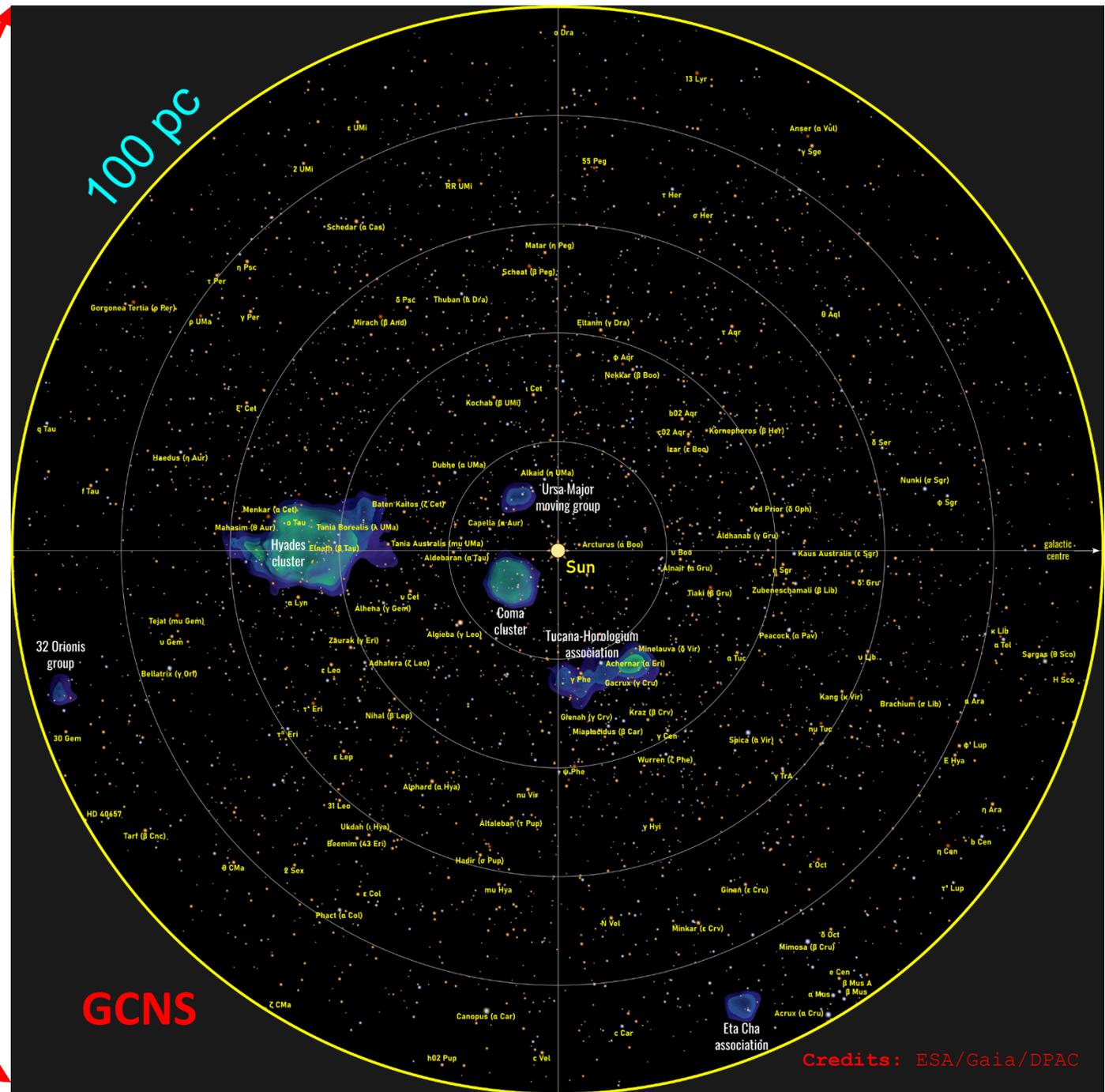
Le coordinate di circa due miliardi di oggetti sono riportate su questa mappa del cielo in coordinate Galattiche.

Come non confonderla con una vera e propria foto digitale di tutto il cielo astronomico profondo!

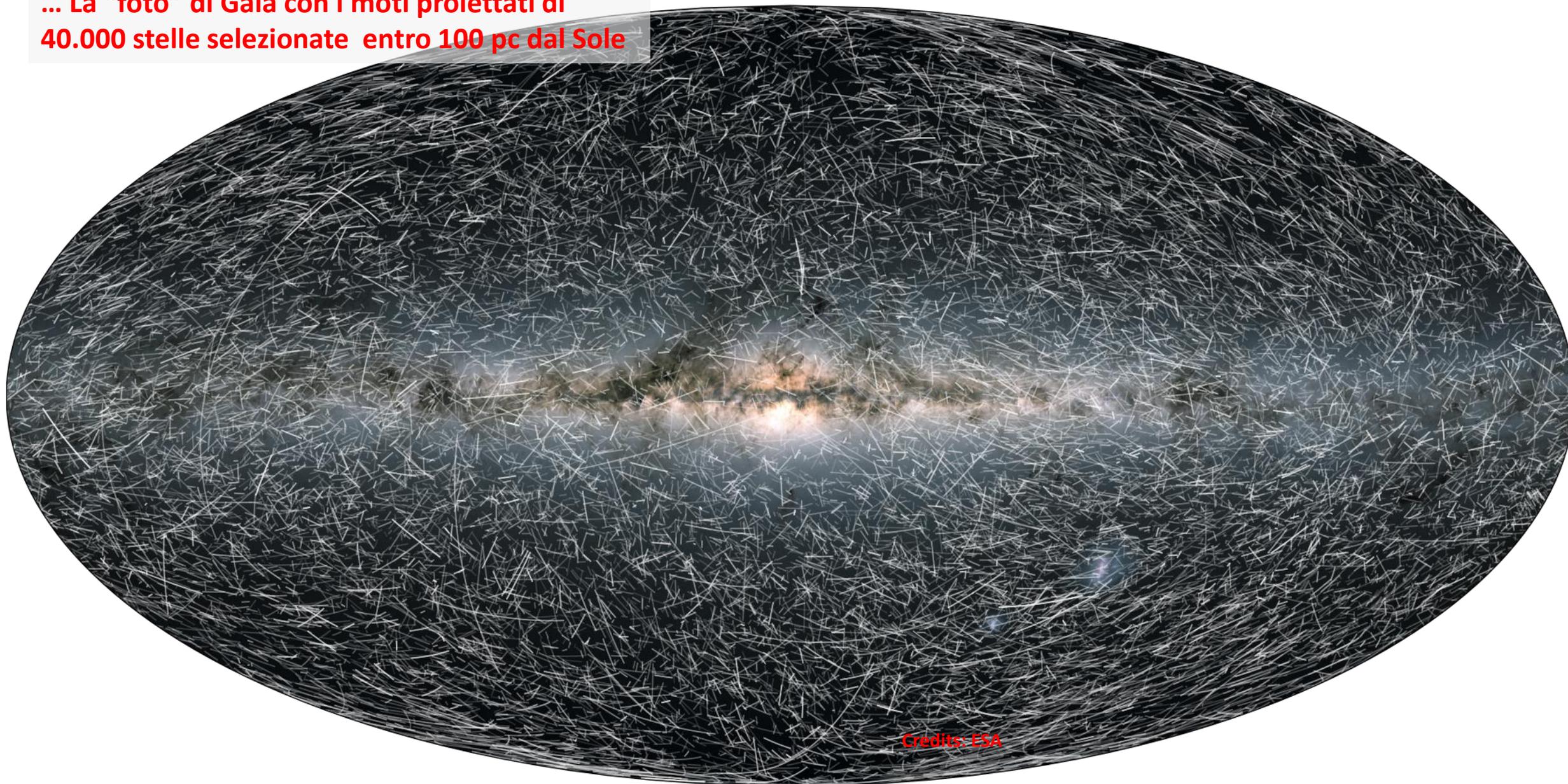
E' quanto Gaia ci ha consegnato lo scorso 3 dicembre (rilascio del catalogo (E)DR3).

E' forse questo il modo più efficace per presentare la portata del lavoro della missione Gaia: una immagine che nessuna camera su nessun satellite è, o sarà, in grado di realizzare.

... E limitandoci ai dintorni del Sole
(ad una distanza di circa 300
anni luce da esso) e
selezionando solo 40.000 delle
stelle censite in questo volume
per cui sono state misurate
distanze e velocità di ottima
precisione, otteniamo.....



... La "foto" di Gaia con i moti proiettati di
40.000 stelle selezionate entro 100 pc dal Sole



Credits: ESA

.....Ovvero, una “istantanea” globale che, attraverso parallassi, moti propri e velocità radiali, fornisce una visione stereoscopica in grado di “propagarsi” avanti ed indietro nel tempo.

E la “dimensionalità stereoscopica” aumenterà con le prossime “Data Releases” con la pubblicazione di parametri astrofisici, classificazioni e serie temporali della variabilità sia astrometrica che spettro-fotometrica.

Questa è dunque l’unicità della missione Gaia:

una straordinaria macchina da censimento spaziale (census machine) che sta costituendo un laboratorio dell’Universo Locale senza precedenti e che tale rimarrà per molti decenni a venire. Il «laboratorio» per la cosmologia di precisione su cui dovranno confrontarsi sia i risultati di missioni spaziali e programmi da Terra multifrequenza, sia le più avanzate simulazioni dell’astrofisica Galattica come risultato dell’evoluzione cosmologica.

Indice

A. L'Universo di Gaia: Aspetti Scientifici

- Obiettivi e Risultati
- Organizzazione delle Attività (di Progetto - overall)
- Programmazione (di Progetto - overall)
- Team – Organizzazione, Programmazione, Produzione, Leadership
- Fondi
- Criticità

B. L'Universo di Gaia: Aspetti Tecnologici

- Team – Organizzazione, Programmazione, Risultati, Leadership
- Criticità

Obiettivi

La missione Gaia dell'ESA è uno dei progetti principali della comunità astronomica Europea nell'ambito del programma «mandatory» dell'Agenzia Spaziale Europea.

Lanciata nel dicembre 2013 iniziava le operazioni scientifiche nel luglio del 2014.

Gli obiettivi principali, originari, della missione Gaia, così come definiti dal Gaia Science Team che portò all'approvazione della missione (Perryman et al, A&A, 2001,369, 339), è quello di **rivoluzionare Fisica ed Evoluzione stellare e la conoscenza della nostra Galassia come prodotto Cosmologico per una conoscenza senza precedenti del nostro posto nell'Universo.**

Tale obiettivo da raggiungere con osservazioni astrometriche e spettro-fotometriche ripetute nel tempo di tutti gli oggetti della volta celeste rivelabili fino alla magnitudine G(aia) di 20.7 (vicina alla R di Johnson-Cousin). La precisione senza precedenti delle misure angolari di Gaia permettono la misura assoluta di coordinate, parallassi e moti propri su scala Galattica. In particolare, l'errore sulla parallasse a fine missione, ovvero dopo l'elaborazione globale della totalità dei dati che saranno accumulati (**circa 2000 miliardi di osservazioni sugli 11.5 anni!!**), sarà di **9-25 μ as a R=15 (a seconda del colore)**. Ciò implica **distanze al meglio di 0.1% entro 100 pc e del 10% fino a 10 kpc**, permettendo calibrazioni dinamiche e di energia radiante di qualità senza precedenti. Inoltre, le misure spettroscopiche e spettrofotometriche prese in contemporanea forniscono in aggiunta **velocità radiali (al meglio delle centinaia di m/sec) e fotometria multi-banda calibrata (~ 0.001 mag)** di altissima qualità per una mappatura statisticamente completa della dinamica e della chimica Galattica: una banca dati celeste dalle caratteristiche, appunto, rivoluzionarie.

Un formidabile laboratorio di Cosmologia Locale per testare, come il Sole per la fisica stellare, le previsioni delle più avanzate simulazioni cosmologiche a red-shift zero.

Se la maggior parte degli oggetti censiti da Gaia è composta da stelle della Via Lattea (VL), alcune **centinaia di milioni di questi sono extragalattici** (es., stelle delle galassie nane nell'alone esterno della VL e delle Nubi di Magellano, oltre a 1.5 milioni di QSO fino a $z=2-3$) **e milioni di nuovi asteroidi del Sistema Solare.**

Per questo, si poteva già intravedere, fin dal concepimento della missione, la rivoluzione che Gaia avrebbe apportato anche nei campi:

- a) **della Fisica Fondamentale e dell' astronomia gravitazionale (Sistemi di Riferimento, test a tutte le scale (locale e Galattica) delle Teorie della Gravitazione, onde gravitazionali);**
- b) **della Verifica Locale delle Teorie Cosmologiche (bulge, disco spesso ed alone vicino in contesto Λ CDM);**
- c) **della Scienza dei Pianeti Extrasolari;**
- d) **dei corpi minori del Sistema Solare;**
- e) **l'astronomia extra-galattica (Galassie attive, Blazars).**

[Molteplici i riferimenti a questa scheda Gaia, GaiaUniverse-0 (ed a quella figlia, G_DPCT/GaiaUniverse-1), da parte di molte delle schede relative a Progetti e Programmi di Ricerca dei nostri ricercatori e tecnologi INAF anche non direttamente coinvolti.]

Risultati Gaia entra in fase operativa solo nel luglio 2014, dopo una prolungata messa a punto in orbita (“science verification”) determinata da alcune importanti anomalie riscontrate sul payload. Nonostante queste difficoltà iniziali, il consorzio per il processamento e l’analisi dati Gaia (DPAC, si veda più avanti), grazie alle straordinarie capacità dei suoi astronomi e tecnologi di adattare algoritmi e strategie di riduzione, è riuscito a riportare la qualità dei risultati in linea con le aspettative originali (rif. GaiaUniverse-1).

Questo straordinario impegno è stato ripagato dall’enorme riscontro che i due cataloghi Gaia fin qui pubblicati, il DR2 (aprile 2018) e la prima parte della terzo (Early Gaia Data Release 3 - EDR3, rilasciato il 3 dicembre 2020; si veda

<https://www.cosmos.esa.int/web/gaia/release>) hanno avuto tra le comunità dei fisici e degli astronomi di tutto il mondo.

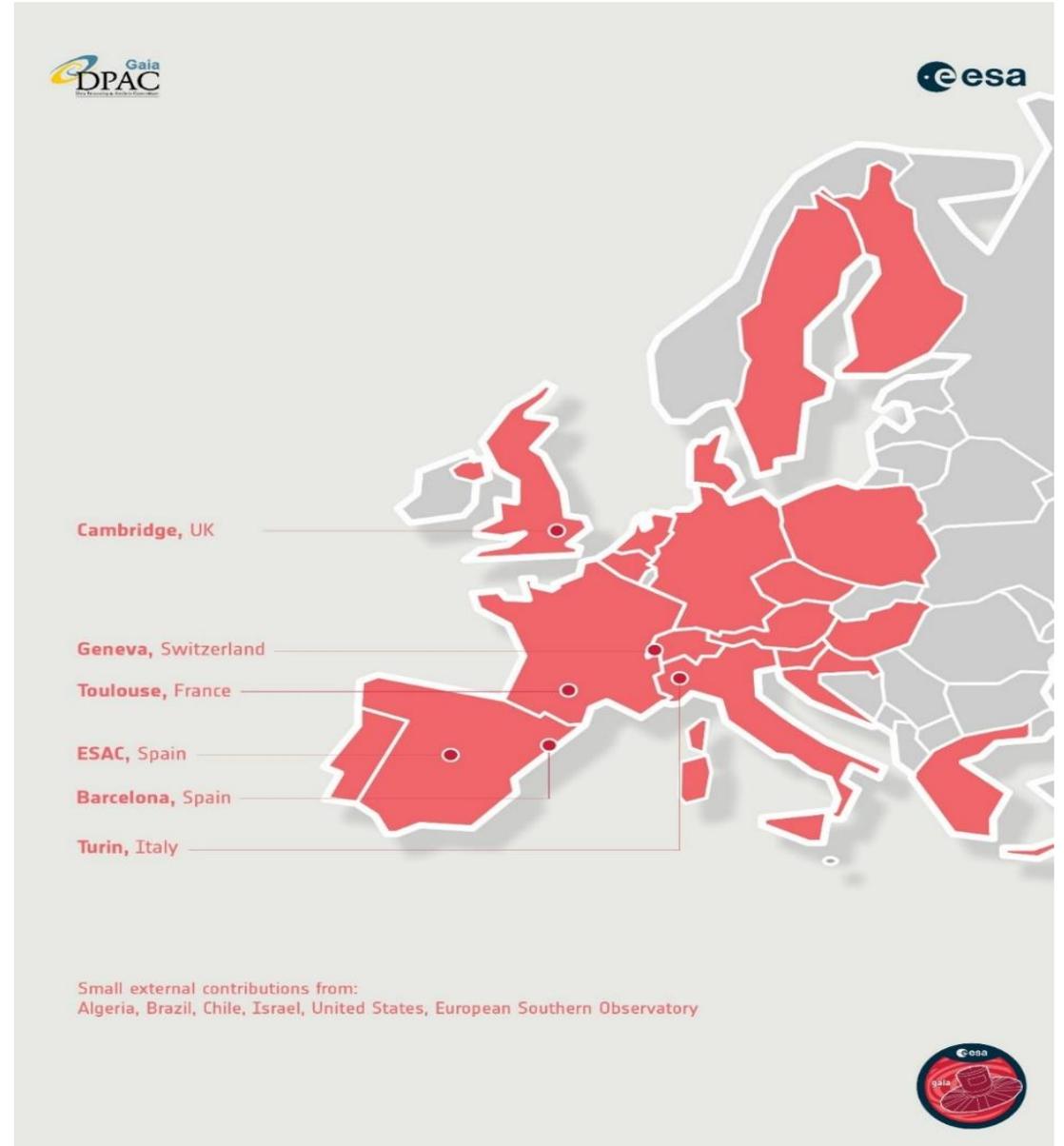
Infatti, nonostante il DR2 consideri solo 22 mesi di dati, l’impatto sulla comunità astronomica mondiale è tale che nell’anno passato (dati 2020) sono apparsi ~250 articoli/mese su riviste internazionali referate che lo utilizzano in modo significativo. [Aggiornamenti riguardo ai cataloghi disponibili e futuri all’indirizzo web di cui sopra.]

Organizzazione delle Attività (overall)

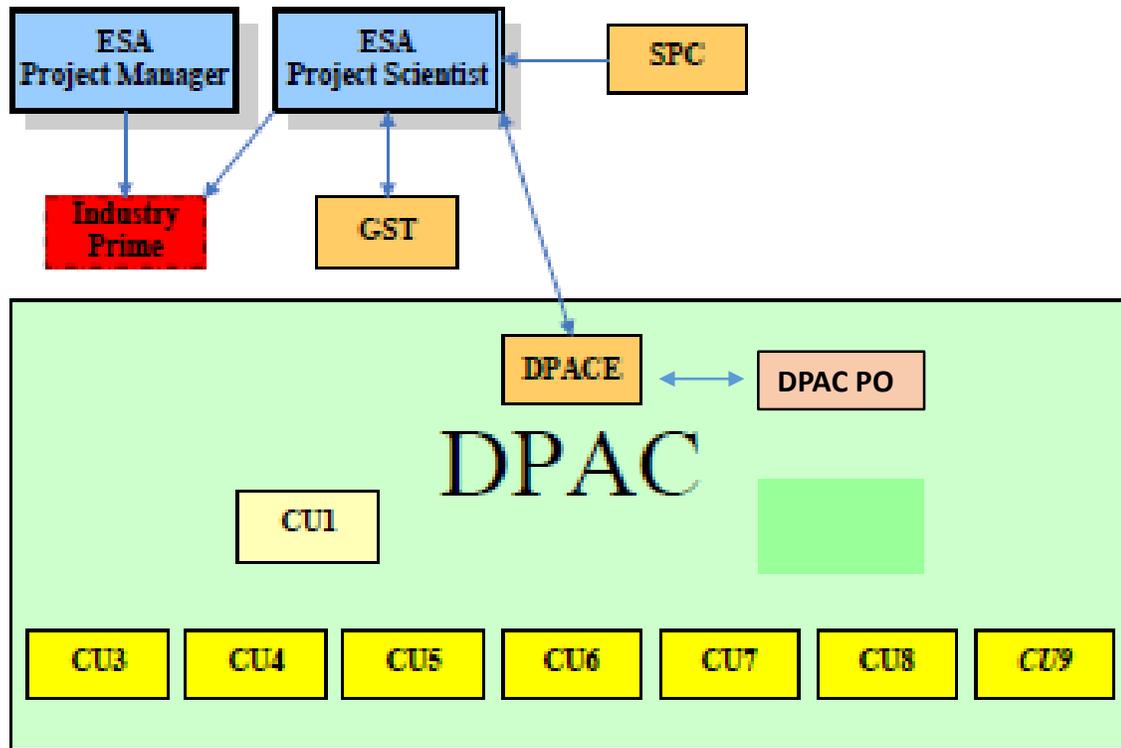
Gaia è una missione con ‘payload’ realizzato direttamente da **ESA**, ovvero non acquisito via contributi nazionali diretti. Responsabilità **esclusiva della comunità scientifica è invece la riduzione ed analisi dati.**

Nel **2007**, a conclusione di un processo di valutazione e revisione da parte di un Comitato di esperti Internazionale, **l’ESA approvava la formazione del Consorzio Europeo DPAC** (Data Processing and Analysis Consortium; si veda la figura a fianco) **formalizzando l’operazione attraverso la stipula di un MLA** tra ESA e le Agenzie Partecipanti (rif. «Obiettivi e Risultati») e che ha richiesto la costituzione di PI-ship nazionali per la sua realizzazione.

La Figura 3 riassume la struttura della «**Governance**» del **Programma Gaia** e del DPAC in particolare (<https://www.cosmos.esa.int/web/gaia/dpac/consortium>).



The DPAC for the Gaia mission



CU1: Consortium wide SW system architecture and repository (via SubVerioN-SVN); ECSS Documentation, Scheduling and WBS repository via Livelink/Wiki (@DPCE).

CU3	Core Processing: Absolute Astrometry and Reference Frame (DPCB, DPCE, DPCT)
CU4	Non-single objects (Solar System Asteroids, binaries, Exoplanets,...) (DPCC)
CU5	Spectro-photometry (instrumental and absolute) (DPCI)
CU6	Radial velocities and HiRes spectroscopy (DPCC)
CU7	Variability (DPCG)
CU8	Stellar (object) parameters (DPCE)
CU9	(Science) Performance validation

6 Data Processing Centers (DPCs):

- Univ. Barcelona (DPCB)
- IoA Cambridge (DPCI)
- Univ. Geneve (DPCG)
- Gaia SOC Madrid (DPCE)
- OATo/ALTEC Torino (DPCT)
- CNES Toulouse (DPCC)

Figura 3 Scheda GaiaUniverse-0. Organizzazione schematica del consorzio Europeo DPAC, compresi i sei Data Processing Centers (DPCs).

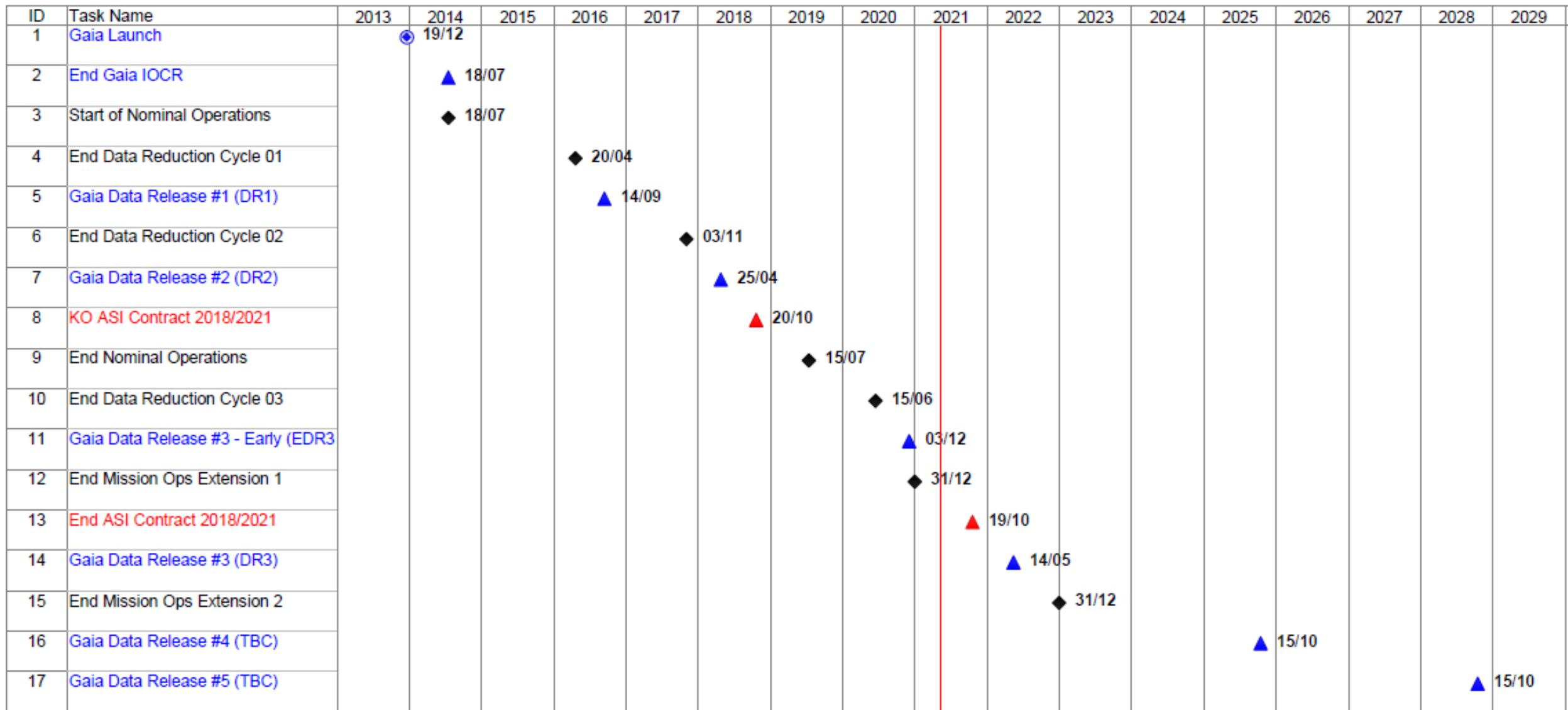
Il Progetto SW del DPAC è stato sviluppato in Java, un **linguaggio piuttosto insolito per una impresa astronomica** e che ha richiesto uno sforzo iniziale particolarmente pesante per agli astronomi e tecnologi di tutta Europa, italiani compresi.

Il SW sviluppato (su piattaforma SubVersion-SVN) è rilasciato periodicamente (“SW releases”) dopo attenta review da parte di ingegneri informatici di processo e controllo di qualità (PA and QA engineers) durante appositi meeting dei Configuration Control Board (CCB). Anche tutta la documentazione prodotta e richiesta per le varie accettazioni e qualificazioni, su piattaforma Livelink/Wiki, deve soddisfare gli **standard ECSS** mentre tutta la **parte economica**, in particolare quella con l’ASI, deve essere **gestita via ECOS**.

Programmazione (overall)

La missione Gaia avrebbe dovuto terminare il suo censimento del cielo, come da durata nominale, dopo 5 anni dall'inizio di fase operativa. Tuttavia, la qualità e l'impatto dei risultati parziali fin qui elaborati e la resilienza di payload e satellite in orbita sono tali che ESA intende continuare la presa dati fino alla fine del 2025, con una data di fine "operations" ora prevista per il 2028. Infatti, le "operations" hanno bisogno di altri 3 anni, dopo il "fine vita" orbitale (previsto per il dic 2025), per completare il ciclo di elaborazione da parte del Gaia «Science Ground Segment» (SGS) che fa capo al DPAC e di cui fa parte il DPCT.

L'evoluzione programmatica del progetto è elaborata mantenuta ed implementata dal DPAC e procede secondo gli accordi sottoscritti nell'MLA. La programmazione («Gantt chart») di alto livello che illustra l'avanzamento programmatico delle attività è illustrata nel diagramma che segue.



DPAC main schedule with integrated ESA reviews (including Data Releases).

La complessa partecipazione Italiana impatta, con compiti altamente qualificati, la maggior parte delle CU ed è illustrata nella **Figura 4**. Questo comporta che l'Italia contribuisce SW e supporto alla sua integrazione, non solo al DPCT, ma anche ai di DPC di Cambridge (CU5), Ginevra (CU7), Tolosa e Madrid (CU3, CU4).

IL contributo Italiano è presentato ad alto livello con il Product Tree/WBS in **Figura 5** (direttamente dal Technical and Management Proposal approvato dall'ASI con il Kick-Off dell'ultimo contratto stipulato con INAF (n. 2018-24-HH.0) ed attualmente in corso). Tutti WP che iniziano con la lettera “G” (per Gaia) provengono dalla WBS delle “task” di elaborazione di primo livello ed analisi dati Gaia del DPAC. I WP che iniziano con “I” (per “Italian”) si riferiscono ai WP specifici/aggiuntivi dei contratti con ASI, compreso l'IWP-M-900 a supporto della partecipazione degli scienziati INAF alla CU9 (escluse le FTE per il team Gaia di SSDC).

**Major products
and systems
(article 5 of Gaia
MLA: [Italian
Contribution](#))**

Contribution to CU3

Core Processing@ESAC& DPCT

- Initial Gaia Source Catalog (**IGSL via GSC2 DB**)
- DPCT
- Astrometric Verification Unit (AIM, BAM,GSR) (**OATo, OACT**)

Contribution to CU4

Object Processing@CNES

- Minor planets identification (**OAFi**) and Phot. Classification (**OATo**)
- Exoplanets identifications and characterization (**OATo**)

Contribution to CU7

Variability Processing @ Geneva

- Cepheids, RR Lyrae (**OABo, OANa**)
- Solar-Like variability (**OACT**)
- QSO variability & classification (**OATo**)

Contribution to CU5

Photpipe@Cambridge

- Absolute flux calibration, monitoring & external verification (**OABo**)
- SpectroPhotometricStandardStars-SPSS (**OAFi**)
- Crowded fields (**OARm, OATe**)
- High Background fields (**OARm, OATe**)
- Science alerts (**OASBo, OAPd**)

Contribution to CU9

Data distribution and exploitation@SOC

- External Cross Matching (**Gaia@SSDC, OARm**)
- Scientific validation tasks (e.g., Open clusters) (**OAPd, Gaia@SSDC**)
- Support scientific exploitation (**Gaia@SSDC**)

Contribution to CU8

Astrophysical characterisation@CNES

- Synthetic libraries & stellar parameters (**OACT, OAPd**)
- Interstellar absorption (**OATo**)

Figura 4 Scheda GaiaUniverse-0. Dettaglio dei Contributi INAF alla missione Gaia (come da MLA ESA/Agenzie/Istituzioni nazionali).

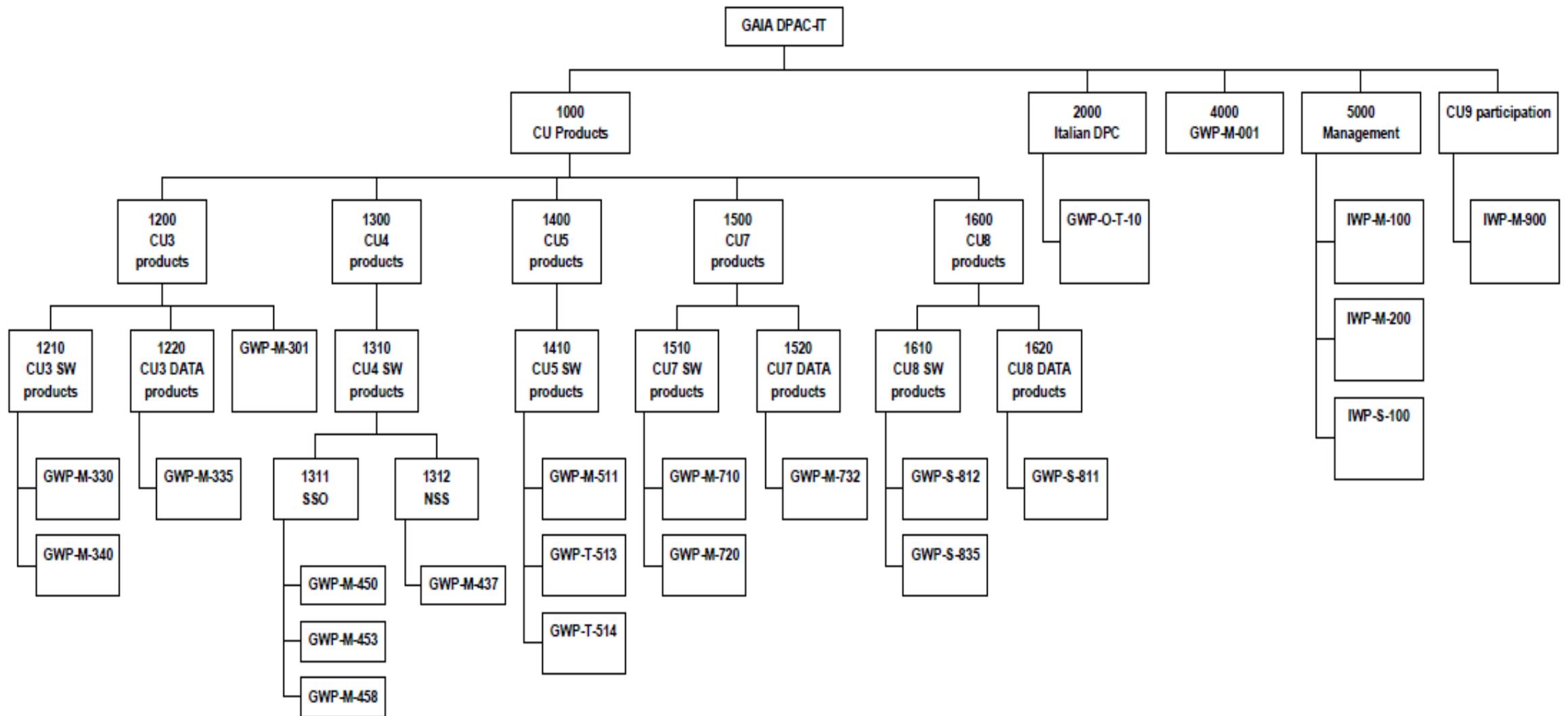


Figura 5 Scheda GaiaUniverse-0. WBS (and Product Tree) of the Italian Participation (Gaia DPAC-IT) as in the Technical and Management proposal approved for the current ASI contract to INAF (n. 2018-24-HH.0).

La «governance» della partecipazione Italiana è realizzata via un contratto ASI-INAF primario: allo sforzo finanziario dell'ASI corrisponde un contributo equivalente INAF con FTE di proprio personale.

A questo si aggiungono un contratto industriale (alla ALTEC, To) per la realizzazione e le operazioni dell'Infrastruttura (HW e SW) del Data Processing Center Italiano (DPCT) e altro accordo ASI-INAF per il Gaia Team presso l'SSDC. La Figura 1 allegata per questa sezione riassume l'organizzazione attraverso la quale opera il nostro Team. Un agile Ufficio di Progetto Italiano (OATo), supportato dal Gruppo di Coordinamento, assicura sia il coordinamento Nazionale (in particolare con ASI) che quello con il Project Office del DPAC).

The INAF contribution (organization & work)

The Italian contribution to DPAC is (numerically) first with that of France at approx. 19 % of the total FTE effort.

~ 30 FTE/yr of INAF staff and contract personnel (spread over just above 60 colleagues!!).

S. Randich is Member of the GST, A. Vallenari is DPAC Deputy Chair

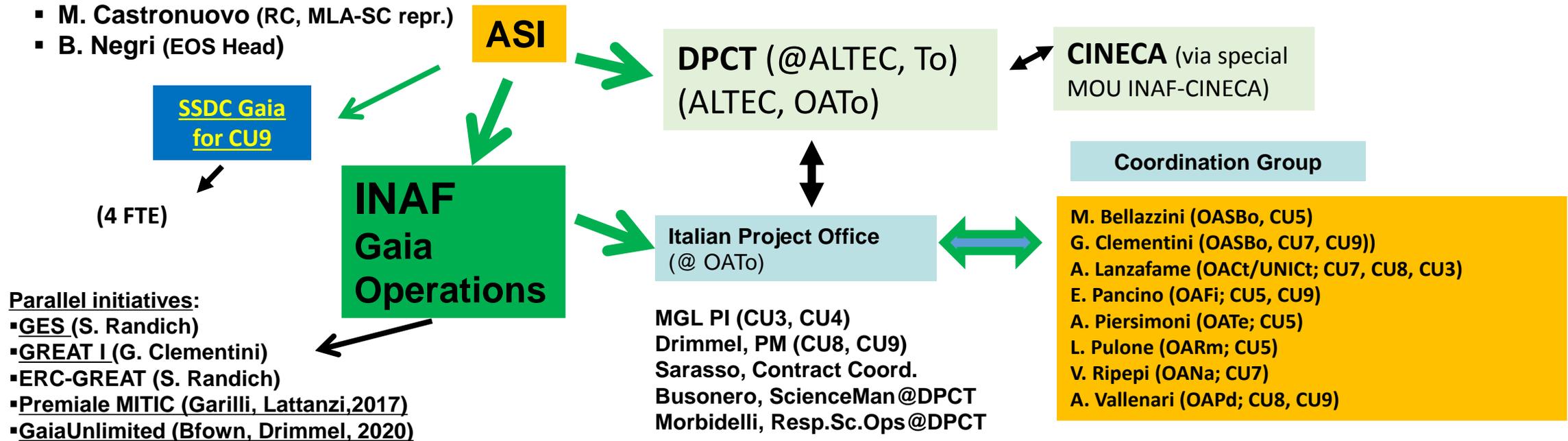


Figura 1 Scheda GaiaUniverse-0. Organizzazione del Team INAF



Otto strutture INAF

contribuiscono alla fase operativa della missione: **OASBo, OACt, OANa, OARm, OATe, OAFi, OAPd e OATo**. Il totale del contributo Italiano, quasi interamente INAF, è di circa **30 FTE/anno**, distribuite su circa **70 unità di personale** tra staff permanente ed a contratto (totale GaiaUNiverse-0 e GaiaUNiverse-1).

La figura illustra la distribuzione geografica della partecipazione INAF.

La frazione media di FTE del personale TI INAF nell'arco di durata del progetto è del 70% mentre si mantiene attorno al 50% nei 3 anni oggetto di questo PTA.

Grazie agli accordi/contratti di cui sopra, **l'INAF e ALTEC S.p.A hanno realizzato ed operano il DPCT**, uno dei 6 Centri di Processamento Dati che costituiscono il Gaia Science Ground Segment (GaiaUNiverse-1). Inoltre, ha realizzato ed opera presso **l'SSDC uno dei 4 GaiaPortal** per la distribuzione ed il 'mining' dei cataloghi rilasciati dal DPAC, così da supportare efficacemente lo sfruttamento scientifico dei dati Gaia dalla comunità Nazionale.

Infine, partecipiamo fin dal "commissioning" iniziale al monitoraggio dello stato del Payload Astrometrico tramite il coinvolgimento di tre tecnologi OATo nel gruppo dei Payload Experts (PEs) del DPAC.

Cambridge, UK

Geneva, Switzerland

Toulouse, France

ESAC, Spain

Barcelona, Spain

Turin, Italy

Small external contributions from:
Algeria, Brazil, Chile, Israel, United States, European Southern Observatory



STORAGE CAPACITY: 2.5 PB
overall raw disk space
distributed between two HPE
P7400 storage units and one
P8400.

PROCESSING CAPACITY: 14
servers HP DL580 G7/G9 with
a total of about 600 CPU cores
and 4.5TB RAM.

DB SERVERS: 3 HPE DL580 G7
dedicated to the database
cluster based on Oracle RAC
technology.



Collegamento diretto con
sistemi HPC CINECA via
MOU dedicato

La **Figura 6** illustra il diagramma di Gantt di alto livello delle attività DPAC con integrate le date delle ESA «reviews» (comprese le date previste per il rilascio pubblico dei futuri cataloghi) e con inserite la date di inizio e fine del **contratto ASI-INAF in corso**.

La figura proviene, come dal «Technical and Management Proposal» INAF approvato da ASI nel 2018 e in vigore fino a fine 2021. D'altra parte, come detto in precedenza, tutto il crono-programma delle attività DPAC è mantenuto aggiornato dal PO del DPAC e costantemente recepito dal PO Italiano e presentato all'ASI alle riunioni di avanzamento.

ID	Task Name	Start	Finish	Duration	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
1	Gaia Launch	Thu 12/19/13	Thu 12/19/13	0 mons		12/19									
2	End Gaia IOCR	Fri 7/18/14	Fri 7/18/14	0 mons		7/18									
3	Start of Nominal Operations	Fri 7/18/14	Fri 7/18/14	0 mons		7/18									
4	End Data Reduction Cycle 01	Wed 4/20/16	Wed 4/20/16	0 mons				4/20							
5	Gaia Data Release #1	Wed 9/14/16	Wed 9/14/16	0 mons				9/14							
6	End Data Reduction Cycle 02	Fri 11/3/17	Fri 11/3/17	0 mons					11/3						
7	Gaia Data Release #2	Wed 4/25/18	Wed 4/25/18	0 mons						4/25					
8	KO ASI Contract 2018/2021	Tue 7/24/18	Tue 7/24/18	0 mons						7/24					
9	End Nominal Operations	Mon 7/15/19	Mon 7/15/19	0 mons							7/15				
10	End Data Reduction Cycle 03	Mon 6/15/20	Mon 6/15/20	0 mons								6/15			
11	Gaia Data Release #3	Tue 12/15/20	Tue 12/15/20	0 mons									12/15		
12	End Mission Ops Extension 1	Thu 12/31/20	Thu 12/31/20	0 mons									12/31		
13	End ASI Contract 2018/2021	Fri 7/23/21	Fri 7/23/21	0 mons										7/23	
14	Gaia Data Release #4	Thu 12/15/22	Thu 12/15/22	0 mons											12/15
15	End Mission Ops Extension 2 (TBC)	Sat 12/31/22	Sat 12/31/22	0 mons											12/31

Figura 6 Scheda GaiaUniverse-0. Example of DPAC main schedule with integrated ESA reviews and time coverage of ASI current contract to INAF (n. 2018-24-HH.0) after approval of INAF Technical and Management proposal.

Produzione scientifica e tecnologica

Oltre ai **20 articoli scientifici e tecnologici rappresentativi allegati** alle schede GaiaUniverse-0 e GaiaUniverse-1, è importante ricordare le oltre 22 pubblicazioni/anno che dal 2008 al 2020 sono uscite su riviste referate e legate strettamente ad attività scientifica di fase operativa (algoritmica, calcolo HPC, verifica scientifica della qualità dei dati prodotti).

Per la parte più tecnologica, nello stesso periodo, sono state pubblicate sul Livelink, repository mantenuto al DPCE-ESAC, **759 note tecniche (TN) di livello DPAC relativi a prodotti dalla nostra partecipazione**. A queste TN DPAC, vanno aggiunti **oltre 5000 rapporti tecnici** dalle nostre pipeline AVU in esecuzione dal 2014 al DPCT ed archiviati sul locale DB (si veda GaiaUniverse-1). Tutta questa produzione è stata/è naturalmente parte consistente delle deliverables dei contratti ASI.

Lo stesso DPCT e il GaiaPortal presso l'SSDC devono essere considerati prodotti di altissima tecnologia e unici per le scienze dell'Universo non solo in Italia ma internazionalmente.

Infine ci sono le importanti ricadute, e quindi il valore aggiunto, che le abilità tecnologiche sviluppate negli anni per la missione Gaia hanno per INAF. Dal punto di vista accademico va sottolineato il programma GraviMetrA, i **progetti TLS (premiare MITIC), ASTRA** (finanziato dal MAECI in collaborazione con il Ministero della Ricerca Cinese per **astrometria sub- μ as**) ed il progetto EU "Gaia Unlimited"; **i contratti industriali con la TAS-I**, come sub-contractors ESA, **per le missioni Euclid e Spica**. In particolare, per **Euclid** facciamo parte del **team che ha costruito il Fine Guidance Sensor per il puntamento ed assetto fine della piattaforma Euclid** fornendo il catalogo profondo (fino alla mag 19) per il calcolo dei quaternioni di assetto derivato dal materiale Gaia. E sempre per Euclid, il contributo già dato alla partecipazione italiana, in particolare all'avvio della realizzazione del Science Data Center-I.

La leadership della partecipazione nazionale a Gaia è dell'INAF, infatti, nonostante alcuni rilevanti contributi Universitari (realizzati attraverso il meccanismo dell'associatura) si può dire la partecipazione in Gaia è «all-INAF».

Sul piano internazionale, lo sforzo Italiano (INAF) è secondo solo, con un minimo scarto, alla partecipazione Francese. Come detto in precedenza ed illustrato nel dettaglio della scheda GaiaUniverse-0, le responsabilità Italiane a tutti i livelli della «governance» della missione (sia al livello ESA che DPAC) è di alto livello.

Fondi

Come illustrato in scheda, per il 2022 e 2023, dopo il prenegoziato con ASI per la definizione delle necessità Gaia da inserire nel PTA ASI per il periodo 2022-2024 (compreso), siamo in attesa del consolidamento del budget della nostra agenzia spaziale (previsto per giugno) per l'avvio della procedura per l'apertura del Tavolo Negoziale che speriamo possa accadere nell'autunno.

Criticità

Non dovrebbe esserci dubbio alcuno che le figure altamente interdisciplinari formate con tenacia e “vision” nel corso di oltre 20 anni di sviluppo del Progetto Gaia in Italia e che INAF ha, in parte, acquisito attraverso posizioni TI di tecnologo e ricercatore, daranno all’INAF una marcia in più per continuare a sostenere, con le giuste competenze ed abilità (anche manageriali e di processo), progetti avanzati in sviluppo e/o futuri (spaziali e non), garantendo al contempo ruoli di leadership sia nazionale che internazionale.

D’altra parte la crescita delle potenzialità tecnologiche del **DPCT** durante il triennio coperto da questo PTA e l’enorme potenziale di sfruttamento dei Big Data del suo DB (sia per l’astronomia non prevista dal “core program” scientifico di Gaia, ma soprattutto per le scienze e le tecnologie spaziali), un unicum senza precedenti in Italia, dovrebbe convincere l’INAF ad acquisirlo definitivamente **come facility dell’Ente**. Ponendo così anche le basi per **“The Living Sky”**, l’idea di **Cosmo-teca Universale**, più volte presentata all’Ente, recentemente sperimentata con successo nella sua forma prototipale nell’ambito del progetto Premiale MITIC (PI B. Garilli) e **che solo l’INAF Ente può pensare di realizzare e mantenere nel tempo**.

Questa acquisizione richiede però che l’INAF consideri prioritario, e continui nella sua acquisizione, personale TI le cui ragioni e caratteristiche sono meglio discusse nella scheda “figlia” GaiaUNiverse-1.

E' una criticità assoluta, la necessità di completare nel breve-medio periodo coperto da questo PTA INAF, 2021-2023, la stabilizzazione di personale di altissimo livello che opera da anni in Gaia e che fa riferimento ai settori ex MA 2 e MA 5.

In sostanza, per garantire all'INAF il giusto ritorno scientifico che il suo staff si è conquistato negli anni con una partecipazione intensa e di alto profilo è indispensabile procedere al reclutamento del personale a contratto Gaia che risulta nelle graduatorie:

- i) III RIC comma2 riservato – MA 2.1 – Popolazioni e ammassi stellari galattici ed extragalattici
- ii) III RIC comma2 libero – MA 5.Inf (esperimenti GR e modellistica GR)
- iii) III RIC Comma2 riservati – MA 2.3 – sistemi planetari

G-DPCT

(GaiaUniverse-1)

The Italian Gaia Data Processing Center and beyond: AVU, OPS4, TLS

Deborah Busonero (INAF-OATo)

DPCT OVERVIEW

- Il DPCT è uno dei 6 centri di processamento dati (DPC) del Gaia Science Ground Segment, ubicato in ALTEC a Torino. La sua realizzazione ed operatività è il risultato del lavoro di un team integrato INAF-OATo/ALTEC.
- Ideazione e progettazione del DPCT e dei sistemi AVU nel lontano 2005.
- Il personale coinvolto nelle attività afferisce principalmente a **INAF-OATo** in quanto leader per l'ideazione e realizzazione di tecniche avanzate di calibrizioni e riduzione dati per l'Astrometria a campo largo o survey-oriented a livello del microarcsec e oltre e risoluzione della sfera e **INAF-OACT** per codici HPC e porting su exa-scale
- I sistemi sw interessati fanno capo all'Astrometric Verification Unit e costituiscono 3 dei 7 sistemi scientifici del Core Processing di Gaia che riproduce tutta la catena di riduzione astrometrica, dalla telemetria alla produzione dei 5 parametri astrometrici di catalogo.
- **OBBIETTIVO A LUNGO TERMINE:** Il progetto vuole guardare oltre l'operatività della missione Gaia, con l'obiettivo di fornire alla comunità INAF e non, un Centro Dati con capacità di data analysis e di riprocessamento unico nel suo genere per l'astrofisica e le scienze e tecnologie spaziali in genere.
(Progetto OPS4-TLS per la realizzare di uno dei più avanzati sistemi di archiviazione, distribuzione, processing, analisi e sfruttamento per Big Data dedicato all'indagine dell'Universo vicino e lontano)

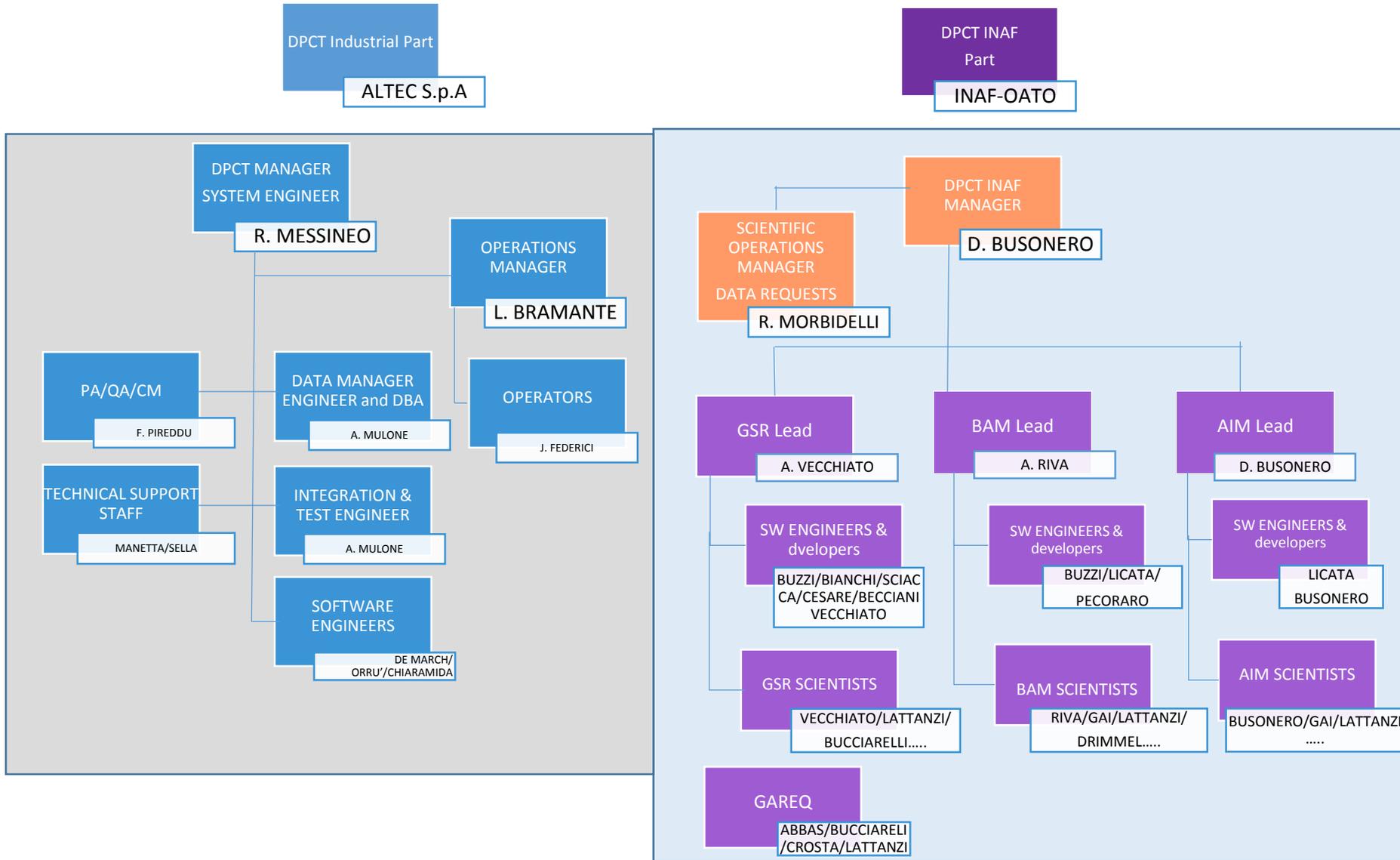
ORGANIZZAZIONE

- INAF-OATo è responsabile della definizione, progettazione, sviluppo, test e validazione dei 3 sistemi AVU e di tutte le attività di coordinamento scientifico necessarie per lo sviluppo e l'operatività del DPCT
- ALTEC è responsabile della definizione, progettazione, implementazione, test, validazione e operazioni dell'infrastruttura HW e SW del DPCT e integrazione, test e operazioni di tutto il sw fornito e di tutte le attività di coordinamento tecnico necessarie per lo sviluppo e l'operatività del DPCT

Il gruppo coinvolto presenta competenze che spaziano da:

- Software system engineering,
- gestione di DataBase relazionali e non
- realizzazione, implementazione e operabilità di sistemi di processamento dati tipici di un segmento di terra di una missione spaziale,
- Quality e product assurance,
- data science and data analytics,
- porting su pre-Exascale e ingegnerizzazione di codici HPC.

INAF-OATo / ALTEC joint effort: THE TEAM



FTE INAF dedicati al progetto

	TI	TD
FTE INAF certi 21-23	12.45	7.2
FTE INAF inviluppo	115	80

**IMPORTANTE: dal 2008 al 2018 l'80% delle attività è stata portata avanti e sotto responsabilità di personale non staff
Attualmente siamo al 60%**

LEADERSHIP

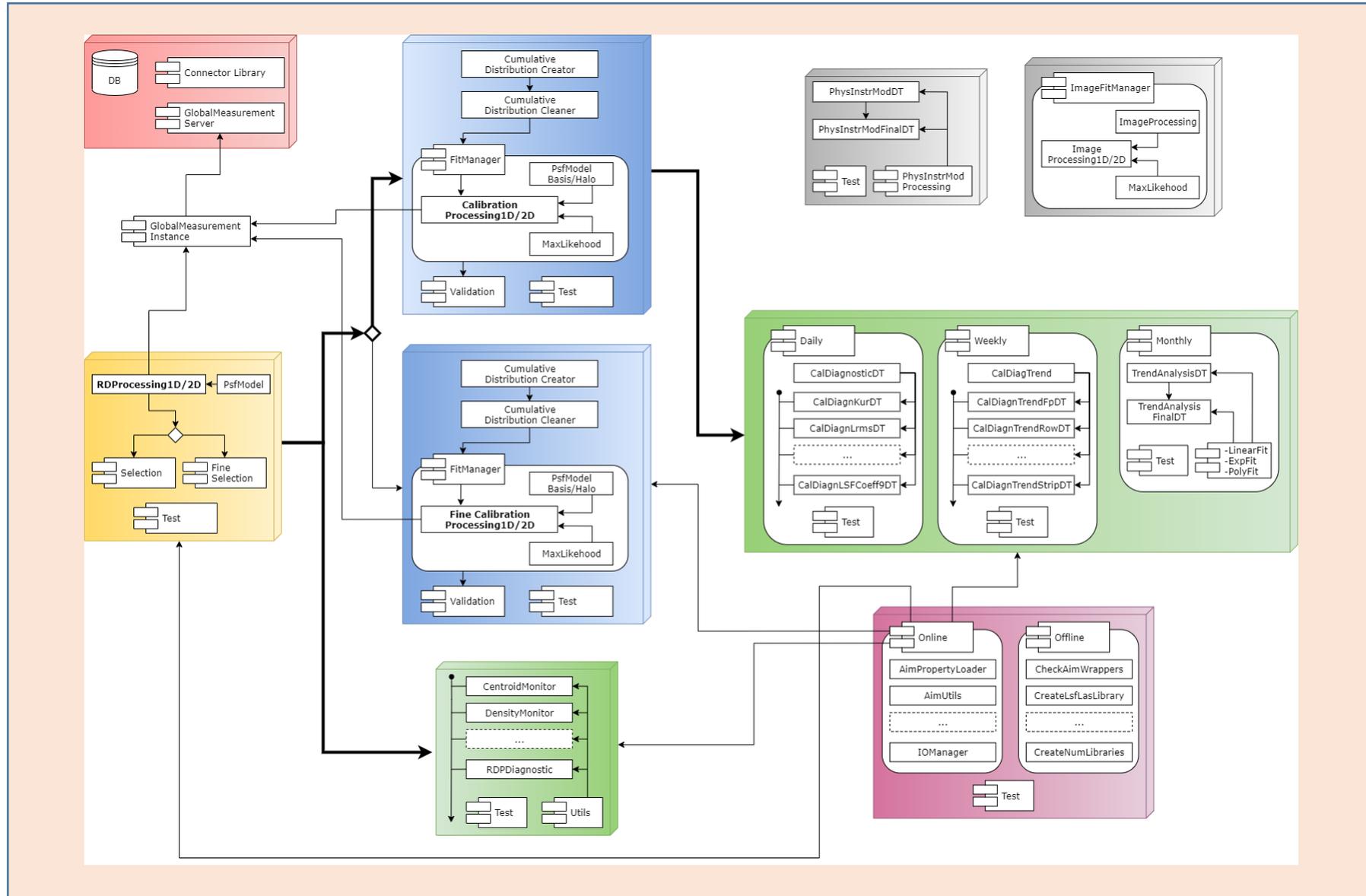
- Il gruppo INAF coinvolto nel progetto ha all'interno del Science Ground Segment di Gaia **il ruolo di leadership di tutta la catena indipendente di validation e verification del core processing (AVU)** a cui è dedicato proprio il DPC Italiano.
- **Conduzione insieme ad ALTEC del DPCT:** INAF DPCT Science Manager and AVU lead (Deborah Busonero); CU3 (core processing) deputy (Mario G. Lattanzi)
- **Leadership nel monitoraggio dello stato del BAM e del piano focale astrometrico** all'interno del gruppo dei PE che insieme al SOC è responsabile del monitoraggio dello stato del Payload e di definire le necessarie procedure di calibrazione a bordo.

AIM AND BAM PIPELINES

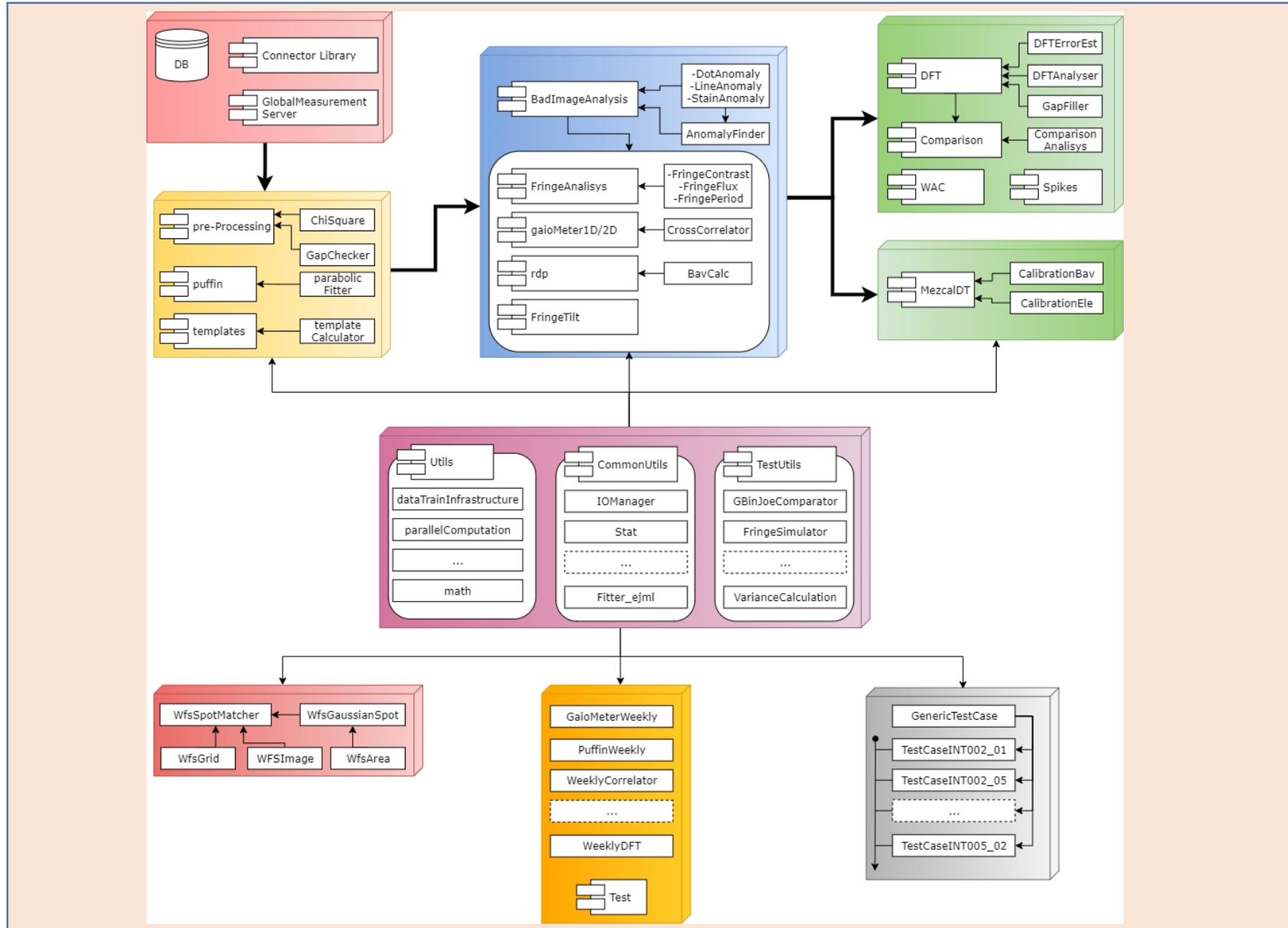
- **AIM pipeline:** raw data processing, image parameter determination, LSF/PSF modelization and calibration, astrometric instrument monitoring and diagnostics throughout the mission lifetime
 - **Versione daily 100.000 righe di codice**
 - **Versione ciclica circa 100.000 righe di codice**
 - 24 hours of raw data each run: **from 2×10^6 to 15×10^6 raw images**
 - Complex structure of the pipeline: **10 sw modules managed by a coordinator in an automatic way**, the output of one run become the input of the next one
 - **6 hours** of time execution on the DPCT Operation platform **for each run**

- **AVU/BAM pipeline:** raw data coming from the Basic Angle Monitoring (BAM) instrument, i.e. fringes, for monitoring and analyzing the instrument behaviour throughout the mission and performing the BAV calibration.
 - **Versione daily 100.000 righe di codice**
 - **Versione ciclica circa 50.000 righe di codice**
 - 24 hours of raw data each run: almost **8×10^4 images**
 - **1-2 hours for each run**
 - The pipeline output is sent to DPCE and ingested into the MDB
 - AVU/BAM cyclic version of the software aims to fringes reprocessing for calibration improvement

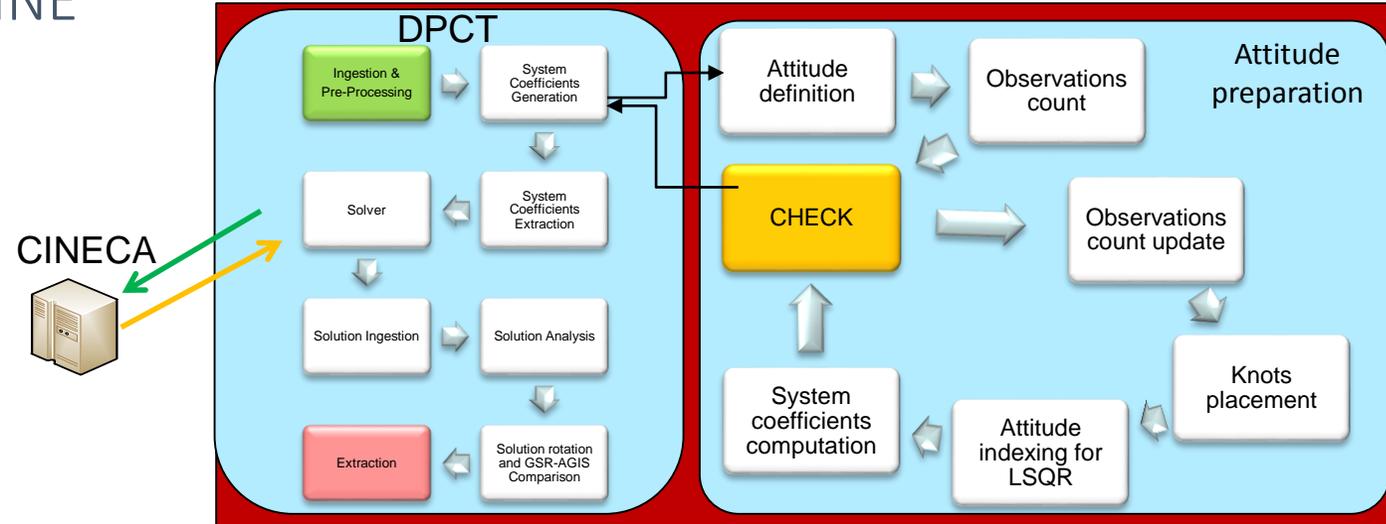
HIGH LEVEL ACTIVITY DIAGRAM of AIM SYSTEM:



HIGH LEVEL ACTIVITY DIAGRAM of BAM SYSTEM:



GSR PIPELINE



- **The Global Sphere Reconstruction** (GSR) solves a linearized system of equations whose result gives the global astrometric reference system (position, parallax, proper motions). This solution is compared with that of AGIS
 - **Starting from 10^7 to 10^8 objects for each run**
 - Very complex pipeline structure
 - **130.000 code lines in Java + 30.000 code lines in C, C++ for the Solver running at CINECA**
 - Final GSR output sent to DPCE in the MDB
 - The Solver module run at CINECA which is managed as one processing node of the DPCT
 - The whole process could be iterated for Non-Linearity
 - **One run takes from 3 to 6 days** on 10^7 objects.

HW INFRASTRUCTURE

- Operation and test & validation platform committed to the DPCT project
- Procurement performed incrementally according to mission needs

INTERNET LINK : 1Gbps (300 Mbps guaranteed) via GARR

STORAGE CAPACITY: **2.5 PB overall raw disk space** distributed between **two HP P7400** storage units and **one P8400**.

COMPUTING : **14 servers** HP DL580 G7/G9 with a total of about **600 CPU cores and 4.5TB RAM**.

DEV & TEST: 7 servers HP

DB SERVERS: **3 servers** HP DL580 G7 (**32 cores**, 256MB RAM each) based on Oracle RAC technology (**DBMS Oracle 12.2c to 18.5**).

NETWORK CONNECTION: LAN network up to 10 Gbps. SAN network redundant at 8 Gbps.

SECURITY SERVICE: redundant firewall based on pfSense, enabling secure remote access via VPN.

INFRA MONITORING AND MANAGEMENT: services based on VMWare virtual environment configured with two HP DL 580 G7 servers clustered and managed by vCenter Server.

BACKUP SERVERS: HP DL580 G7 dedicated to DB and filesystem backups from data volume snapshots.

3 LEVELS BACKUP : L1 on primary storage array, L2 on disks (StoreOnce 6600) and L3 on tape libraries (HP ESL G3).

HPC INTERCONNECTION: access to HPC super computer at CINECA for dedicated processing.



QUALCHE NUMERO.....

Attività operative compiute dai sistemi AVU al DPCT da inizio fase operativa fino ad oggi:

322.300 workflows
35.068.170 job processati
~12 GB di dati di telemetria da Gaia scaricati ogni giorno (SOC),
~16 GB di files ricevuti ogni giorno al DPCT,
~ 114.5×10^9 di Osservazioni ricevute,
2.469 processi AVU/BAM completati,
2.534 processi AIM completati,
21.000 Mission Log entries

GSR DRC attivato alla fine del ciclo 2 e attualmente con 42 processi all'attivo.

L'occupazione totale del DB al momento ha raggiunto ~**614** TB di cui

~400 TB nel database Repository (RDB),
~110 TB nel LOCAL database (LDB),
~104 TB nel GSR database (GSRDB).
per un totale di **1.5 PB** di dati archiviati.

PRODOTTI:

AIM e BAM attivi fin dalla prima fase di Commissioning;
Output utilizzati dai Payload Expert italiani per l'individuazione e risoluzione delle problematiche riscontrate sul satellite.

GSR attivo dalla fine del ciclo 2

GAREQ esperimento per la misura del quadrupolo di Giove

- 5500 rapporti tecnico-scientifici giornalieri,
- 100 rapporti tecnico-scientifici settimanali,
- 100 documenti ECSS quali SRN, STR, SRS, STS, SDD, SUM, ICD, tra cui il DPCT development plan, DPCT operations plan, DPCT Procedure Handbook DPCT design document
- una ventina di note tecniche DPAC descrittive le soluzioni architetture e algoritmiche
- una ventina di pubblicazioni referate e non (SPIE) specifiche reperibili su ADS, oltre alla documentazione accompagnante le 3 data release avvenute fino ad oggi (GDR1, GDR2, GEDR3).
- A queste si aggiungono gli articoli della Gaia Collaboration pubblicati in occasione del rilascio di ogni Data Release.
- Per il progetto TLS sono state prodotte 3 note tecniche INAF, mentre per l'espansione OPS4 è stata prodotta a inizio 2021 una nota tecnica dedicata

Criticità 1

- **Possibili perdite di competenze del personale a contatto** coinvolto nel progetto in particolare quello coinvolto nelle attività di sw engineering, data scientist e data analyst: **posizioni di CTER e Tecnologo III livello**
- **Investire in percorsi formativi** (dottorati, post-doc, contatti TD, tenure-track) per personale specializzato che abbia competenze sia dal lato scientifico che informatico e di supporto alla realizzazione di SGS dalla fase di messa a punto fino alla fase di post-operation
- **Adeguati profili interni INAF** per garantire la gestione e lo sfruttamento scientifico di enormi moli di dati quali quelle di un Data Base tipo Gaia che è unico attualmente nel suo genere per complessità.
- **Policy DPAC** molto rigida riguardo al rilascio di documentazione o paper riguardanti i sistemi di processing e i rispettivi risultati

Criticità 2

Un progetto scientifico comporta una forma mentis e un approccio alla gestione delle varie fasi di progetto che può essere in alcune fasi non compatibile con le scelte aziendali perdendo in flessibilità e motivazione, comportando anche un aggravio di costi per certe professionalità.

Necessario mantenere il giusto equilibrio nel rapporto con l'Azienda.

Il DPCT, come i centri di processamento dati previsti in futuro, dovrebbero essere considerati infrastrutture INAF indipendentemente da dove risiedono, considerando l'interesse scientifico e tecnologico della comunità tutta nello sfruttamento dei dati, ben oltre la conclusione della fase operativa di missione. Per poter garantire ciò e non aver speso sforzi invano dobbiamo **coordinare in modo più organico i gruppi in INAF con tale expertise e pensare a un piano a lungo termine sia per lo sviluppo di tali centri di PROCESSAMENTO ED ANALISI dati (science ground segments per missioni spaziali), che per il loro mantenimento e sfruttamento.**