

Gaia: la Galàxia en un petabyte

Eduard Masana, Xavier Luri, Jordi Portell

Departament d'Astronomia i Meteorologia, Institut de Ciències del Cosmos, Universitat de Barcelona (IEEC-UB)

RESUM

La missió Gaia de l'Agència Espacial Europea (ESA) es va llançar des del Centre Espacial de la Guaiana, a la Guaiana Francesa, el 19 de desembre de 2013 amb l'objectiu de determinar les posicions, moviments propis i distàncies de més de mil milions d'estrelles amb una precisió sense precedents. Aquestes dades permetran en els propers anys avenços significatius en moltes àrees de l'astrofísica, i en particular en el coneixement de la nostra galàxia.

A més de constituir un repte tecnològic en si mateix, Gaia també suposa un repte en el tractament de les dades que generarà: uns 150TB enviats a terra, que es convertiran en 1PB en la base de dades resultant del seu processat. Per tal de processar aquestes dades s'ha format un consorci (DPAC) format per uns 450 científics i enginyers d'arreu d'Europa. Aquest consorci porta a terme diverses tasques, incloent el desenvolupament d'un simulador de missió, el processat massiu de dades i la implementació de l'arxiu que allotjarà les dades finals de la missió. Aquestes tasques són un bon exemple de com el processat de grans quantitats de dades científiques (Big Data) està esdevenint un repte tan significatiu com el de la construcció dels mateixos instruments.

Doi: <http://dx.doi.org/10.2436/20.2001.01.2>

1 Introducció

La missió Gaia de l'Agència Espacial Europea (ESA) es va llançar des del Centre Espacial de la Guaiana, a la Guaiana Francesa, el 19 de desembre de 2013 amb l'objectiu principal de determinar les posicions, moviments propis i distàncies de més de mil milions d'estrelles, aproximadament l'1 % del total que formen la Galàxia. Gaia permetrà, entre altres coses, ampliar el nostre coneixement sobre l'origen, l'estructura i l'evolució de la Galàxia gràcies no només al gran nombre d'estrelles observades sinó també a la precisió de les seves mesures, de l'ordre del microsegon d'arc, o el que és equivalent, a l'angle amb el qual veuríem des de la Terra una moneda d'un euro situada a la Lluna. Al llarg dels propers cinc anys cadascun dels mil milions d'objectes serà observat una mitjana de 70 vegades, i això donarà lloc a un catàleg final amb un contingut estimat d'un petabyte de dades, equivalent a uns 200.000 DVD.

Gaia suposarà una revolució en tots els àmbits de l'astrofísica. Les seves observacions abasten tot tipus d'estrelles en els seus diferents estats evolutius, així s'obté informació sobre els seu procés de formació i evolució. Podrem conèixer la natura dels braços espirals de la Galàxia i com s'han format, i la distribució de matèria fosca en de la Via Làctia. Pel que fa al Sistema Solar es podrà determinar amb una precisió sense precedents l'òrbita de més de 200.000 asteroides i detectar més d'un miler d'objectes que es podrien atansar perillosament a la Terra. També es detectaran més de 7000 nous planetes extrasolars. Les observacions de Gaia suposaran també una nova i molt estricta verificació de la teoria de la relativitat general.

El repte plantejat per Gaia és doble. D'una banda hi ha el repte tecnològic que ha suposat construir un instrument capaç d'assolir les precisions esmentades. D'altra banda, a partir d'ara es planteja el repte de processar, emmagatzemar i accedir a l'ingent volum de dades generat per la missió. Per exemple, el procés de reducció de les dades en brut enviades pel satèl·lit fins a l'obtenció del catàleg final requerirà més de 10^{21} FLOPs i òbviament només serà possible fent servir eines de supercomputació, com ara les disponibles al CESCA/CSUC o al Barcelona Supercomputing Center (BSC).

Els grans supercomputadors han estat també imprescindibles en la preparació de la missió. En particular s'han utilitzat per simular de forma realista dades de la missió que han permès posar a punt el procés de reducció de dades, des de la primera ingestió a la base de dades fins a l'obtenció dels resultats finals de la missió.

2 La missió Gaia

Gaia és una missió de l'Agència Espacial Europea (ESA) dissenyada per fer astrometria d'alta precisió de mil milions d'estrelles amb una precisió sense precedents. A més, proporcionarà informació dels seus paràmetres físics (temperatura, massa, edat i composició química) i determinarà la velocitat radial (la velocitat a la qual una estrella es mou al llarg de la línia de visió) de les estrelles més brillants [1].

L'astrometria és la branca de l'astronomia que s'encarrega de la determinació de les posicions, velocitats i distàncies dels cossos celestes. L'única manera directa que hi

ha per determinar la distància a una estrella és mesurar la seva paral·laxi (vegeu la figura 1 i la secció «Gaia: astrometria global»). Fins i tot per a les estrelles més properes l'angle de paral·laxi és molt petit. Per exemple, Alfa Centauri, l'estrella més propera a nosaltres, situada a 1,3 parsecs (4,3 anys llum), té una paral·laxi de només 0,765". No va ser fins a l'any 1838 que Friedrich Bessel va aconseguir mesurar per primer cop la paral·laxi d'una estrella, en concret 61 Cyg, i fins a l'any 1989 s'havien mesurat alguns milers de paral·laxis, amb precisions en el millor dels casos de centèsimes de segon d'arc. Aquell any l'ESA va posar en òrbita el satèl·lit HIPPARCOS, que va elaborar un catàleg de 118.000 estrelles, amb la seva posició, el seu moviment propi i la seva paral·laxi amb una precisió del mil·lisegon d'arc. Va ser el primer cop que es feia astrometria des de l'espai.

Gaia és l'hereva d'HIPPARCOS, però multiplica per 10.000 el nombre d'objectes observats i per 100 la precisió de les mesures. A diferència d'HIPPARCOS, que orbitava la Terra, Gaia observa des del punt de Lagrange L2 del sistema Terra-Sol. En aquest punt, situat a 1,5 milions de quilòmetres de la Terra en direcció oposada al Sol, les forces gravitatòries de la Terra i el Sol es combinen de tal manera que és possible mantenir un satèl·lit al seu veïnatge sense, pràcticament, necessitat de propulsors. A més, L2 ofereix un entorn tèrmic molt estable, imprescindible per a les mesures d'alta precisió de Gaia.

Tots els cossos celestes més brillants que la magnitud $G = 20$ (unes 400.000 vegades menys brillant que l'estrella més feble visible a simple vista) són detectats i observats per Gaia. Això inclou no només estrelles de la nostra Galàxia, sinó també estrelles de galàxies properes com ara els núvols de Magallanes i Andròmeda, així com quàsars i galàxies llunyanes que no poden ser resoltes en les seves estrelles individuals i asteroides del Sistema Solar.

Gaia consta de dos telescopis que operen conjuntament focalitzant la llum de dues direccions diferents en

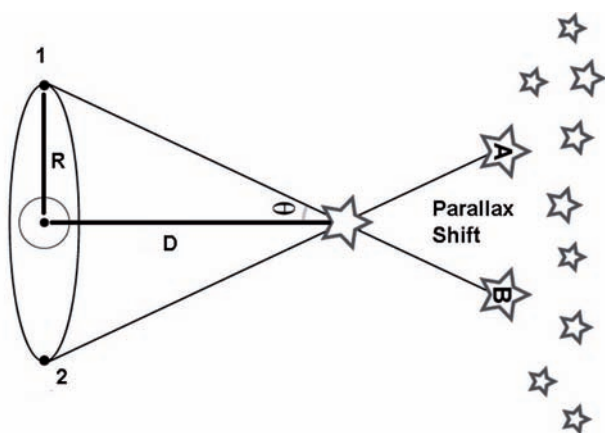


Figura 1: Paral·laxi estel·lar

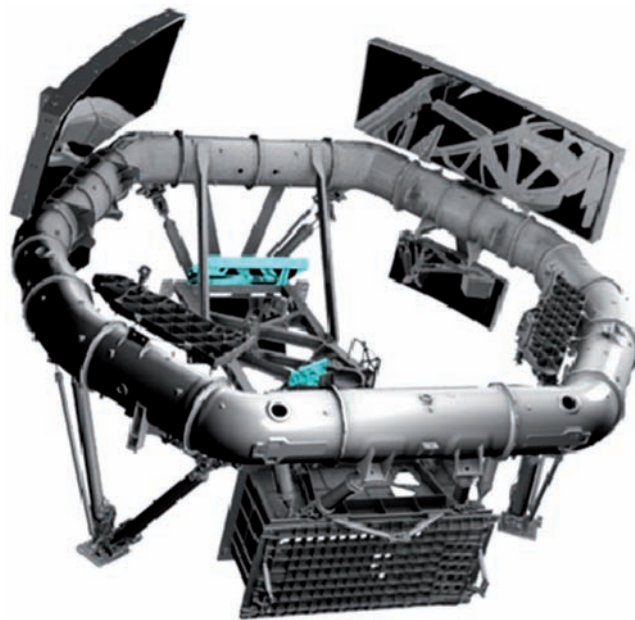


Figura 2: Estructura interna del satèl·lit

un únic pla focal. Els dos telescopis estan separats un angle de $106,5^\circ$ (anomenat *angle bàsic*) i tenen una distància focal de 35 m. Tant els telescopis com l'estructura del satèl·lit on van muntats els miralls i el pla focal estan fets de carbur de silici, un material lleuger i altament resistent amb unes propietats tèrmiques i mecàniques òptimes. Per fer-se una idea de la precisió amb què han estat construïts els diferents elements que formen el satèl·lit, podem mencionar que si els miralls tinguessin la superfície de l'oceà Atlàntic, les màximes rugositats no excedirien el metre d'alçada.

Per observar la totalitat de la volta celeste, Gaia fa una rotació sobre el seu eix cada sis hores. A la vegada aquest eix fa un moviment de precessió al voltant de la línia Sol-satèl·lit de manera que els telescopis van escanrant tot el cel. Aquesta llei d'escaneig permet observar la totalitat del cel cada sis mesos i fa que, en acabar els cinc anys de missió, cada objecte s'hagi observat de mitjana unes setanta vegades (figura 3).

El pla focal està format per 106 CCD que conformen tres instruments (figura 4). Per un costat tenim l'instrument astromètric, encarregat de mesurar les posicions de les estrelles. L'instrument fotomètric consta de dos prismes de baixa dispersió (un al blau i un al vermell) que permeten d'obtenir espectres que donaran informació sobre els paràmetres físics de les estrelles. Finalment, l'espectrògraf de velocitats radials mesura la velocitat radial de les estrelles a partir de l'efecte Doppler de les línies d'absorció del seu espectre. El total de píxels del pla focal és de gairebé 1000 milions, cadascun d'una mida de $10 \times 30 \mu\text{m}$. És el pla focal més gran que mai s'ha enviat a l'espai.

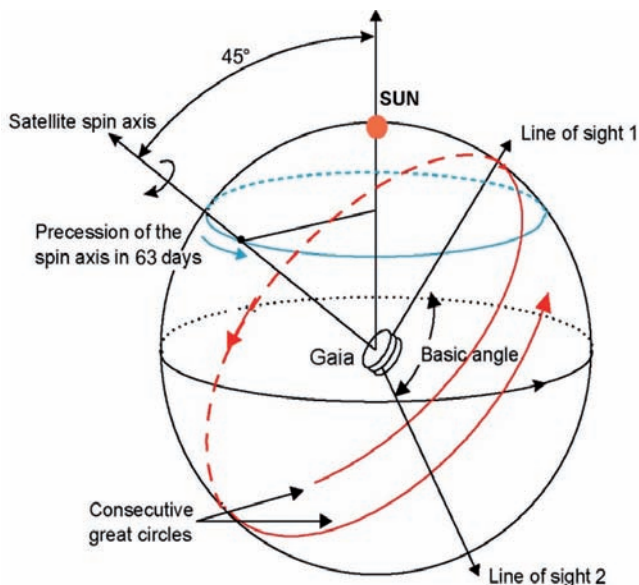


Figura 3: Llei d'escaneig de Gaia

El fet que Gaia estigui girant constantment sobre el seu eix fa que la imatge de les estrelles es vagi desplaçant sobre el pla focal; triga aproximadament un minut a creuar-lo. Les CCD de Gaia funcionen en l'anomenat mode TDI (Time Delay Integration). En aquest mode la càrrega dins de la CCD es va transmetent de píxel a píxel a la mateixa velocitat a la qual la imatge de l'estrella passa per sobre de la CCD, fins a arribar al registre de lectura. El resultat és equivalent a haver integrat la llum un temps igual al del trànsit de l'estrella per la CCD (gairebé 5 segons). Per limitacions de l'ample de banda de les comunicacions amb el satèl·lit, de cada observació només s'envia a terra una petita finestra de píxels al voltant de l'estrella, que, juntament amb la mesura del temps en què ha estat llegida i la posició i l'orientació del satèl·lit en aquell instant, formen les observacions bàsiques que han de ser processades per tal d'obtenir finalment les dades astromètriques de les estrelles. En total, cada vegada que una estrella és observada s'envien a terra 15 d'aques-

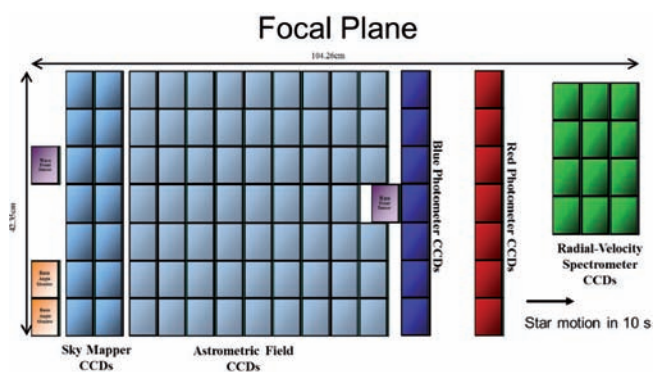


Figura 4: Pla focal de Gaia

tes finestres. Això fa que, de mitjana, cada dia s'enviïn més de 750 milions de finestres, equivalents a 50 Gb, que han de ser emmagatzemades i pre-processades en temps quasi real. Les dades enviades pel satèl·lit són recollides per tres antenes de 35 m de diàmetre, situades a Cebreros (Espanya), New Norcia (Austràlia) i Malargüe (Argentina).

3 El Data Processing and Analysis Consortium (DPAC)

Els estatuts de l'ESA, i a partir d'ells la definició de les tasques relacionades amb una missió, comporten que aquesta s'ocupi de la construcció, el llançament i l'operació dels satèl·lits del seu programa científic, però no de l'explotació científica dels resultats. Aquesta tasca es delega en la comunitat científica europea de diverses maneres i, en qualsevol cas, les dades són posades finalment a disposició de la comunitat científica internacional.

En el cas de Gaia, l'Agència va delegar la tasca del processament de les dades en brut del satèl·lit per convertir-les en dades científicament utilitzables i la seva posterior publicació a un consorci europeu, el Gaia Data Processing and Analysis Consortium (DPAC). Aquest consorci es va formar el 2006 en resposta a una oferta oberta de l'ESA (Announcement of Opportunity) i des d'aleshores es responsabilitza de:

- La preparació dels algorismes d'anàlisi de dades per a la reducció de les dades astromètriques, fotomètriques i espectroscòpiques en un marc de processament coherent i integrat; inclou el tractament d'objectes especials com ara els estels múltiples i els petits planetes, entre d'altres.

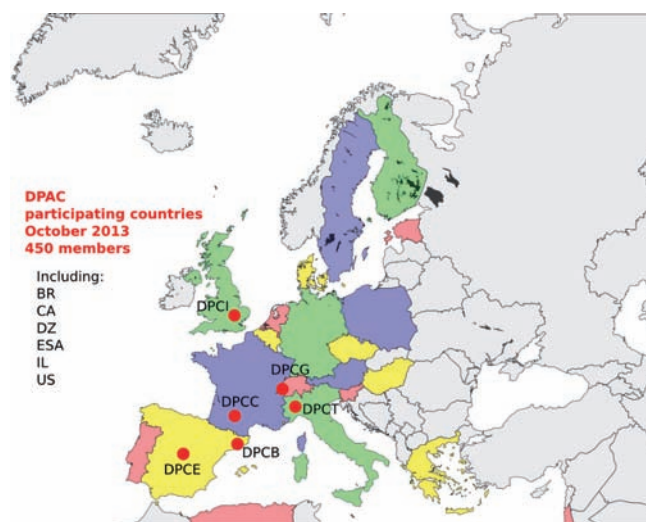


Figura 5: Països participants en el DPAC i els seus centres de computació

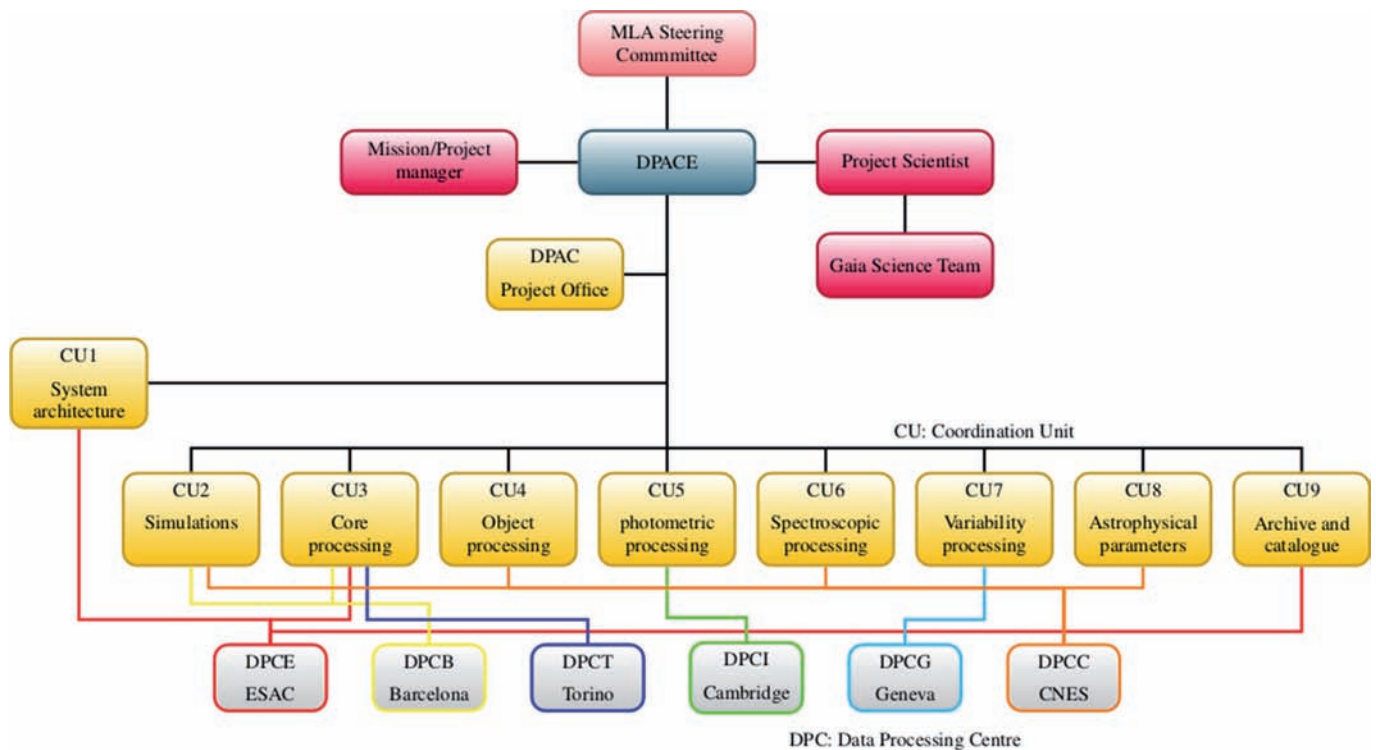


Figura 6: Organització del consorci DPAC en unitats i centres

- La generació i provisió de dades simulades per utilitzar-les en el disseny, el desenvolupament i la verificació del sistema complet de processament.
- El disseny, el desenvolupament, l'adquisició i l'operació de tots els aspectes del programari i maquinari necessaris per processar les dades de la missió a les fases de simulació, operacions de la missió i producció del catàleg final.
- El disseny, el desenvolupament i l'operació de la base de dades i arxiu de Gaia, que contindrà els productes intermedis i finals de la missió d'interès per a la comunitat científica en general.

El DPAC està format actualment per uns 450 científics i enginyers de 25 institucions d'arreu d'Europa, amb un pressupost global d'uns 500 milions d'euros al període 2006-2022.

EL DPAC s'estructura en nou subunitats anomenades *coordination units* (CU). Cadascuna d'aquestes CU s'encarrega d'una tasca específica de processament i l'equip Gaia de la Universitat de Barcelona té una participació significativa en diverses d'elles. En particular:

- Ha liderat des dels seus inicis la CU2, encarregada de la generació de simulacions (vegeu «Les simulacions de les dades»).
- Té un paper clau en la CU3, encarregada del processament astromètric, on ha dissenyat i implementat

dos dels sistemes crítics de la missió (IDT i IDU, vegeu «Processament massiu de les dades»).

- Participa en la CU5, encarregada del processament de la fotometria de la missió, on ha definit el mètode de calibració absoluta del sistema fotomètric de Gaia.
- Lidera la CU9, la darrera unitat en constituir-se i que s'encarregarà de dissenyar, implementar i operar l'arxiu de les dades de la missió.

A més, el DPAC inclou sis centres de processament de dades i un d'ells (DPCB) és a Barcelona, on s'usen els supercomputadors del BSC i el CESCA/CSUC per a la generació de simulacions i el processament de les dades de Gaia.

4 Gaia: astrometria global

La principal missió de Gaia és l'obtenció d'astrometria per a mil milions d'estrelles. Clàssicament l'astrometria d'un objecte celest es descriu amb cinc paràmetres (o sis, en cas d'incloure la velocitat radial com en el cas de Gaia):

- La posició angular de l'estrella al cel, usualment en coordenades equatorials:¹ ascensió recta α i declinació δ .

¹ http://ca.wikipedia.org/wiki/Coordenades_equatorials

- La paral·laxi trigonomètrica², el petit canvi angular en la posició d'una estrella degut a l'efecte de perspectiva provocat pel moviment de la Terra entorn del Sol. Està directament relacionat amb la distància a l'estrella i permet, per tant, deduir-la. La mesura precisa de distàncies estel·lars serà una de les grans aportacions de Gaia (vegeu seccions anteriors).
- El moviment propi³, o moviment (angular) aparent de les estrelles al cel degut a la seva velocitat relativa respecte al Sol. En primera aproximació es considera lineal i s'expressa com dues velocitats angulars, μ_α i μ_δ , que indiquen la variació de les coordenades equatorials (α, δ).
- La velocitat radial, o velocitat d'allunyament o acostament de l'estrella en direcció al Sol. En el cas de Gaia es mesurarà de forma separada de la resta de paràmetres mitjançant l'espectròmetre de velocitats radials, RVS.

Quan s'aborda la mesura de l'astrometria es presenta una primera qüestió fonamental: respecte a què (quina referència) mesurem el moviment d'una estrella? Les mesures clàssiques, des de terra, permetien obtenir el que s'anomena *astrometria relativa*: el moviment d'una estrella es mesurava respecte les estrelles de fons, més febles, que es suposaven a distàncies molt més grans i que, per tant, es podien aproximar com a «immòbils». L'adveniment de l'astrometria des de l'espai, amb el satèl·lit HIPPARCOS i ara amb Gaia, permet superar aquesta limitació i obtenir *astrometria absoluta*. Per fer-ho, els satèl·lits observen el cel en dues direccions separades per un angle gran, 106° en el cas de Gaia (l'anomenat angle de base γ).

Això permet obtenir observacions simultànies d'un gran nombre de parells de direccions al cel, de manera que podem imaginar que al final de la missió tindrem un gran nombre d'observacions de la separació angular unidimensional de parells d'estrelles. D'aquesta manera cada estrella estarà lligada a un cert nombre d'altres estrelles (unes 70, de mitjana) mitjançant aquestes mesures de separació angular, en direccions diferents. Podem combinar tota aquesta informació encadenant totes les mesures de separació angular i formar una malla de relacions angulars que cobreixi tot el cel, de manera que al final del procés podem determinar amb precisió la posició relativa (les unes respecte de les altres) i moviment de les estrelles expressats en un sistema de referència propi, global i autoconsistent. Per fer-ho, les observacions de Gaia s'introdueixen en un gran sistema d'equacions que resol simultàniament els paràmetres astromètrics de les estrelles, els paràmetres que descriuen l'apuntament

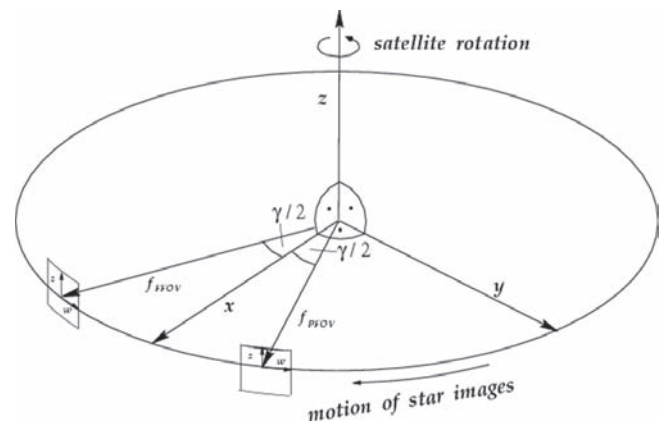


Figura 7: Sistemes de referència bàsics de Gaia

dels telescopis i els paràmetres de calibració geomètrica de l'instrument de Gaia. La resolució d'aquest sistema de milers de milions d'equacions amb milers de milions d'in-cògnites ha estat un dels grans reptes de la preparació del processament de Gaia en què l'equip de la Universitat de Barcelona ha participat. Aquest procés s'ha implementat en l'anomenada Astrometric Global Iterative Solution (AGIS), una implementació de resolució iterativa del sistema d'equacions que està preparada per executar-se sobre les dades reals del satèl·lit.

El darrer pas consisteix a relacionar aquest sistema de referència intern respecte al sistema de referència estàndard en astronomia, l'International Celestial Reference System⁴ (ICRS), el que s'anomena *inercialitzar* l'esfera. Això s'aconsegueix mitjançant les posicions observades d'una llista de quàsars de referència de l'ICRS, que estableixen amb precisió la relació entre els dos sistemes.

5 Les simulacions de les dades

Un pas imprescindible durant el desenvolupament de qualsevol missió espacial és la generació de dades el més realistes possible per tal de poder preparar la cadena de processament de dades amb tots els algorismes que hi intervenen. En el cas de Gaia bona part d'aquestes dades simulades s'han generat als ordinadors del CESCA/CSUC i del BSC.

El simulador de Gaia comprèn tres generadors de dades amb objectius diferents: GASS (Gaia System Simulator) permet generar grans quantitats de la telemetria simulada com la que envia Gaia a terra, amb un grau de realisme prou elevat per poder posar a punt el procés de reducció de les dades; GIBIS (Gaia Instrument and Basic Image Simulator), destinat a l'obtenció de les finestres amb un

² <http://ca.wikipedia.org/wiki/Paral%C2%B7laxi>

³ http://ca.wikipedia.org/wiki/Moviment_propi

⁴ http://www.iers.org/ IERS/EN/Science/ICRS/ICRS.html?__nnn=true

grau de detall més elevat que en el cas de GASS, però amb la limitació de no poder simular intervals de temps massa llargs precisament a causa de la seva complexitat; i GOG (Gaia Object Generator) dissenyat per simular els resultats del catàleg final de la missió.

El simulador és un codi Java (amb gairebé 200.000 línies de codi) que consta de dos mòduls principals: un encarregat de la simulació dels objectes de l'Univers observats per Gaia (amb la seva posició, moviments, distàncies, temperatures, tipus espectral, etc.) i un altre mòdul que implementa les característiques dels instruments (miralls, CCD, posició i moviment del satèl·lit, etc.). La combinació d'ambdós mòduls permet generar simulacions realistes de les dades que Gaia obté, tant les dades en brut com les ja processades.

En el cas de GASS, desenvolupat a la Universitat de Barcelona, el codi és fàcilment executable en un entorn multiprocessador, com ara el supercomputador Mare-Nostrum, gràcies al fet que podem dividir l'interval de simulació en petits subintervals, cadascun d'ells executat en un processador diferent. D'aquesta manera, al llarg del temps de preparació de la missió s'han fet diverses simulacions que comprenen des de simulacions del total de cinc anys de missió amb un nombre reduït d'objectes a observar fins a simulacions d'uns quants dies a zones del cel amb una alta densitat d'estels. Algunes d'aquestes simulacions han requerit centenars de milers d'hores de CPU, amb un volum de dades de fins a 1 TB [2].

6 Processament massiu de dades

El control del satèl·lit i els seus instruments es fa des del centre de control de missió a Darmstadt (Alemanya). Des d'aquest centre es transmeten les instruccions i els paràmetres d'operació, se'n verifica el correcte funcionament (especialment quant a òrbita i actitud) i se'n reben totes les dades, si bé en aquest centre només se n'analitza una petita fracció. El processament de les grans quantitats de dades científiques es duu a terme al centre d'operacions científiques, situat a l'European Space Astronomy Centre (ESAC), vora Madrid.

La cadena de processament diari de dades a l'ESAC està formada principalment per tres sistemes: MIT, IDT i FL. El primer (MOC-SOC Interface Task) s'encarrega de rebre i descomprimir tota la telemetria científica rebuda del satèl·lit a través de les estacions terrestres i del centre de control de missió.

IDT (Initial Data Treatment), liderat per la Universitat de Barcelona, és el primer gran sistema de processament de dades per on passa la ingent quantitat de telemetria científica de Gaia. Aquest sistema duu a terme nombroses i complexes tasques sobre els 18 tipus de paquets de dades que arriben del satèl·lit, començant per la deter-

minació de l'actitud (apuntament) del satèl·lit, acurada fins a uns 0,1 segons d'arc. Aquesta precisió s'aconsegueix combinant les dades d'actitud en brut amb les mesures de les estrelles més brillants, havent-les identificat en un catàleg de referència terrestre. IDT reconstrueix totes les observacions fetes a bord, combinant els paquets de mesures amb altres paquets de referències instrumentals. Després de determinar el nivell de referència de cada CCD i el fons de cel, IDT continua amb la determinació del color, centroide i brillantor de totes les mesures (de mitjana, uns 50 milions d'estrelles cada dia) utilitzant els darrers models instrumentals. Per acabar, IDT determina les posicions de totes les estrelles mesurades (amb una precisió de l'ordre del segon d'arc), les intenta identificar en catàlegs terrestres existents, i crea noves entrades de catàleg per aquelles no identificades. Totes aquestes operacions són monitoritzades i diagnosticades gairebé en temps real pels operadors i científics del sistema.

El darrer gran sistema diari, FL (First Look), duu a terme un gran nombre de diagnòstics i anàlisis sobre els resultats d'IDT, comprova el correcte funcionament de tots els instruments i avisa en cas de necessitar ajustaments o correccions. Així mateix, FL determina la primera solució astromètrica (només sobre la regió observada durant un dia) i diverses calibracions instrumentals retornades a IDT perquè aquest pugui obtenir encara millors resultats.

Totes aquestes tasques diàries requereixen una gran potència de processament, així com una base de dades extremadament eficient i robusta. Calen uns 500 processadors que sumen gairebé 1 TB de RAM (distribuit en uns vint nodes), que aporta l'ESAC, així com uns 100 TB de base de dades (usant un motor Intersystems Caché). Òbviament calen bastants nodes de càlcul addicionals per dur a terme proves i, en cas que sigui necessari, reprocessar les dades que hagin pogut resultar defectuoses. Com que es tracta de dades crítiques, les còpies de seguretat són també obligatòries, no només en l'àmbit local (ESAC) sinó també en còpies remotes al CNES (Toulouse, França).

El CESCA/CSUC ha resultat essencial en el desenvolupament d'IDT, ja que ha permès al nostre equip dur a terme proves realistes abans d'entregar el sistema a l'ESAC, i aportar també recursos per a algunes simulacions especialment complexes. El CESCA/CSUC també està aportant els recursos necessaris per al desenvolupament i prototipatge dels sistemes per a la CU9, per a l'explotació de les dades que generarà Gaia i DPAC. Tornant a l'ESAC, aquest centre també aporta els recursos necessaris per executar el sistema AGIS descrit anteriorment, el qual obté la solució final astromètrica a partir de les dades proporcionades per IDT.

Sense sortir de la unitat de coordinació astromètrica (CU3), el sistema IDU (Intermediate Data Updating)

és el que permet obtenir les precisions finals desitjades. Aquest sistema bàsicament repeteix bona part dels càlculs fets per IDT, però utilitza un més bon coneixement de l'instrument. IDU generarà el catàleg final de Gaia, en el sentit que identificarà les observacions que pertanyen a cada estrella, crearà també les entrades necessàries en el catàleg, eliminarà entrades errònies, corregirà identifications errònies, etc. També obtindrà unes calibracions instrumentals i centroides més acurades, la qual cosa permetrà tornar a executar AGIS sobre dades de millor qualitat i, per tant, obtenir unes posicions, distàncies i velocitats encara millors. En resum, és el sistema que permetrà tancar la iteració global de tot el DPAC (figura 8) i, per tant, assolir les precisions esperades.

Per dur a terme totes aquestes operacions, IDU necessita un entorn de supercomputació. És, probablement, el sistema que necessitarà més recursos d'entre tots els del DPAC. La raó és que IDU haurà de reprocessar repetidament totes les dades acumulades fins a cert moment. La planificació actual és que IDU s'executi cada 6 a 12 mesos de missió sobre totes les dades rebudes. Per tant, la primera execució (prevista per a 2015) haurà de tractar uns 18 mil milions de mesures (sumant uns 10 TB de dades), la segona uns 27 mil milions, i així fins a arribar a les darreres execucions sobre el total de la missió, de manera que sumarà gairebé 100 mil milions de mesures i uns 60 TB de dades. Per sort, IDU, liderat també per la Universitat de Barcelona, compta amb un entorn d'execució idoni, com és el supercomputador MareNostrum. Les previsions indiquen que un 10 % del supercomputador actual hauria de ser suficient per cobrir les necessitats d'IDU.

Conclusions

El processament de les dades generades per la missió Gaia suposa un gran repte computacional. El catàleg final

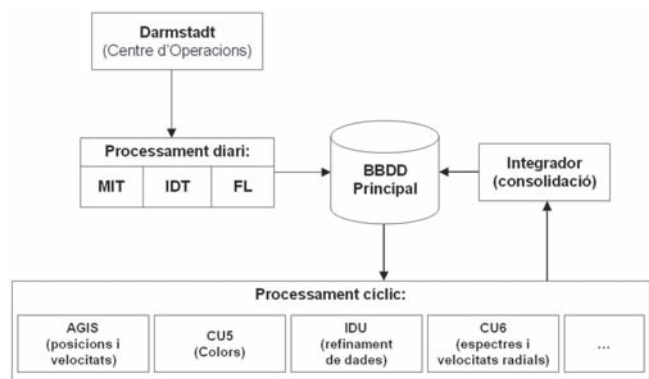


Figura 8: Aproximació iterativa per la reducció de dades de Gaia

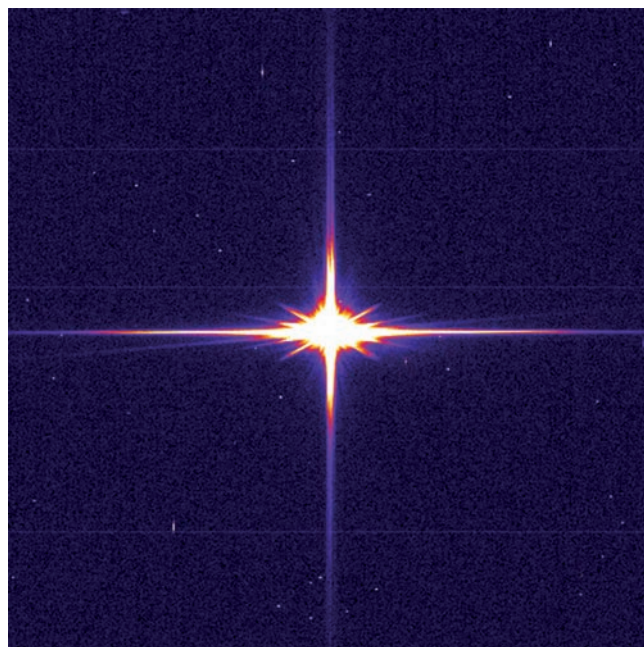


Figura 9: Imatge de l'estrella Sadalmelik obtinguda per Gaia el 15 de gener de 2014 durant el procés de calibratge dels instruments. Els telescopis encara no estaven perfectament enfocats

de la missió, previst per a l'any 2022, serà el resultat d'haver tractat més de 10^{12} observacions individuals, juntament amb milers de paràmetres que descriuen els instruments, incloent-hi el mateix satèl·lit. La solució iterativa que permetrà obtenir les dades científiques finals només és abordable en un entorn de supercomputació, tant pel que fa a les necessitats de potència de càlcul com pel que fa a la quantitat de memòria. Els algorismes per tractar les dades han estat provats abans del començament de la missió fent servir telemetria simulada generada també en supercomputadors com el MareNostrum o els ordinadors del CESCA/CSUC.

Les primeres dades de la missió, que han de servir per calibrar els instruments i posar-los a punt per a la seva tasca científica, van començar a arribar pocs dies després del llançament (figura 9). D'aquesta manera es va engegar un llarg procés de càlcul que finalitzarà d'aquí a vuit anys amb la consecució del catàleg més complet i precís de la Galàxia obtingut fins ara.

Bibliografia

- [1] M.A.C. PERRYMAN i K.S. de BOER et al., *Astronomy and Astrophysics*, 369 (2001)
- [2] A. ROBIN i X. LURI et al., *Astronomy and Astrophysics*, 543A (2012)