



GAIA Dal Tera al Peta

Il "social network" nella Galassia ?

Roberto Morbidelli @ INAF – OATo

Con l'indispensabile supporto e consulenza di

Deborah Busonero e Mario G. Lattanzi @ INAF - OATo

Rosario Messineo; Angelo Fabio Mulone @ ALTEC



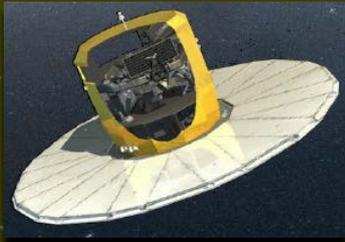
Le «Date» nei «dati» di Gaia



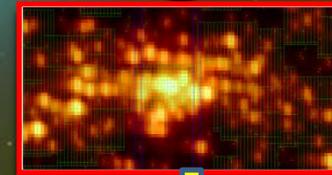
- Inizio missione: 19/12/2013
- Fine missione: 2019 – 2020 (5-6 anni)
- Check points con rilascio di cataloghi intermedi:
 - 1^a versione: giugno 2016
 - 2^a versione: dicembre 2016
 - 3^a versione: dicembre 2017
 - 4^a versione: marzo 2019
 - 5^a versione: + 3 anni da termine missione (2022 ?)



Dall'archivio Gaia al catalogo

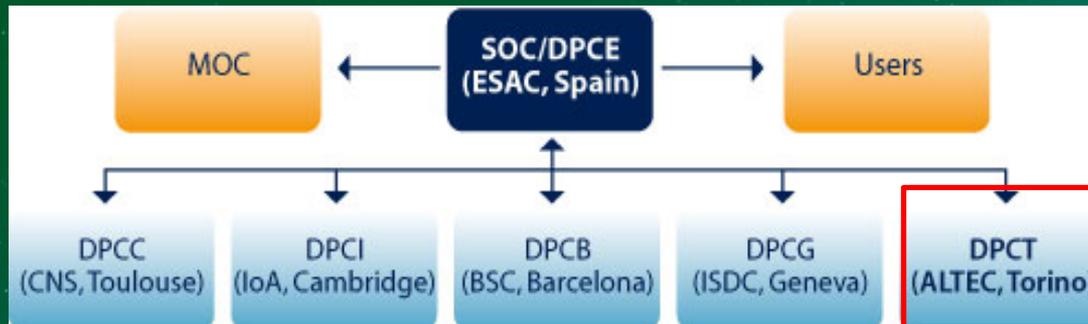


72155956	08971943	089459
11944644	04840895	048946
46886890	16496483	164868
65048483	95185645	951784
52997947	80444504	804479
21889614	41969159	419096
99578539	93858518	938485
87294100	04717437	047841
78786621	18234104	182466
45449806	99959939	999198
54859045	44348858	443990
80548411	99100744	991884
28768990	88461965	884489
35309420	35309420	608230





L'archivio Italiano Gaia@DPCT



- Gaia è una missione spaziale caratterizzata da:
 - All sky survey (customizable scanning law strategy).
 - Europea (pertanto è difficile definire un'univoca collocazione di **tutti i dati della missione**).
 - E' una «Corner Stone» programma ESA Orizon 2000+ intrinsecamente un test bed per il futuro.
 - Che richiede integrazione con osservazioni da Terra e la disponibilità ed accessibilità ai dati ancillari (es: **IGSL** e **GBOT**)
 - Che richiede **diagnostica remota dell'attività del satellite** al fine di gestire i comportamenti «inattesi»
 - Elevato trougput di dati. (trasmette circa 1 MB/s 8 ore/giorno)
- Gaia non è una missione spaziale caratterizzata da:
 - produzione di immagini per tramite di puntamento.
 - risultati «definitivi» in tempi prevedibili (Distanza delle Pleiadi: Hipparcos docet)



L'archivio Italiano Gaia@DPCT



Il DPCT, Data Processing Center di Torino, è il DPC Italiano dedicato alla missione GAIA. Uno dei 6 DPCs previsti dal Gaia Data Processing and Analysis Consortium (DPAC). E' il risultato della sinergia tra il **team scientifico dell'Osservatorio Astronomico di Torino** (INAF- OATo) e il **team industriale di ALTEC**. La realizzazione è finanziata dall'ASI attraverso un contratto scientifico (INAF - OATo) ed un contratto industriale (ALTEC S.p.A).

L'Hosting presso ALTEC (TO) presenta molteplici punti di forza e di garanzia per le attività operative, in questo contesto enfasi per:

- La consolidata esperienza nel mantenimento di un Data Center.
- L'affidabile collegamento di rete verso gli altri DPCs.
- L'affidabile popolamento e mantenimento dell'archivio missione.
- La gestione del processamento dei dati e del "life cycle" connesso.
- Presenza di un supporto logistico a 360° con idonee figure professionali.

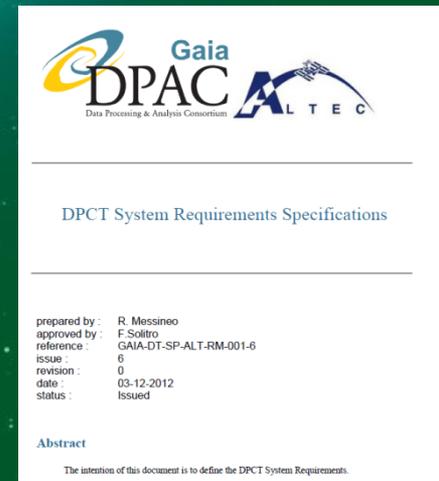
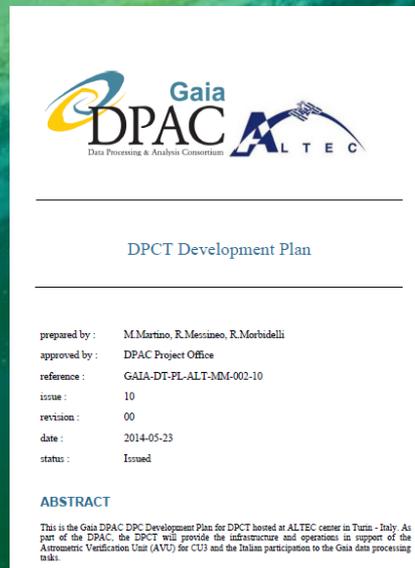
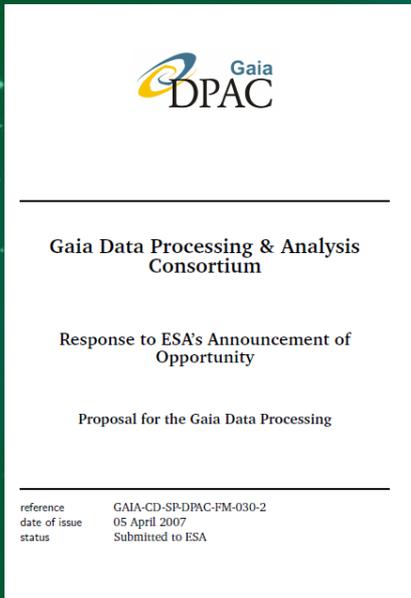


L'archivio Italiano Gaia@DPCT



M.Martino, R.Messineo, R.Morbidelli, Gaia DPAC DPC Development Plan for DPCT , GAIA-DT-PL-ALT-MM-002-10, 2014-05-23 .

Hosted @ ALTEC center in Turin - Italy. As part of the DPAC, the DPCT will provide the infrastructure and operations in support of the Astrometric Verification Unit (AVU) for CU3 and the Italian participation to the Gaia data processing tasks.





L'archivio Italiano Gaia@DPCT



M.Martino, F.Marziani, R.Messineo, R.Morbidelli,
DPCT Operations Plan, GAIA-DT-PL-ALT-MM-010-04



DPCT Procedure Handbook

Prepared by: F. Marziani, R. Morbidelli, ALTEC Technical Staff
 Approved by: DPCT Manager
 reference: GAIA-DT-UG-ALT-FMA-001-01
 issue: 01
 revision: 00
 date: 12 March 2013
 status: Issued

Abstract

The document describes basic procedures and processes for Data Processing Centre Turin (DPCT) developed for Gaia operations activities starting from operation rehearsal number one executed on 25th June 2012.




DPCT Operations Plan

prepared by : M.Martino, F.Marziani, R.Messineo, R.Morbidelli
 approved by : DPAC PO
 reference : GAIA-DT-PL-ALT-MM-010-04
 issue : 4
 revision : 0
 date : 2013-19-12
 status : Issued

ABSTRACT

This is the DPAC DPC Operations Plan for DPCT that is hosted at ALTEC centre in Turin - Italy. As part of the DPAC, the DPCT will provide the infrastructure and operations in support of the Astronomic Verification Unit (AVU) for CCD and the Italian participants to the Gaia data processing tasks.

F. Marziani, R. Morbidelli, ALTEC Technical Staff,
DPCT Procedure Handbook, GAIA-DT-UG-ALT-FMA-001-01



L'archivio Italiano Gaia@DPCT



V. Icardi, DPCT Configuration Item List, GAIA-DT-SP-ALT-VIC-008-11, 2014-03-04
Document identifies the configuration items under control of the DPCT hosted in ALTEC Turin.

#	Name	Description	Det. Desc	Type	Multiplicity	Units	Consumers	Min/Max
1	idSolution	Solution Identifier	View	long				
2	idAlpha	Right Ascension at catalogue epoch	View	double		Angle(arc)		0
3	idDelta	Declination at catalogue epoch	View	double		Angle(deg)		-90
4	idAlphaError	Error in Right Ascension	View	float		Angle(arc)		999
5	idDeltaError	Error in Declination	View	float		Angle(deg)		999
6	idAlphaEpoch	Mean Epoch of Right Ascension	View	float		Time(j2000Time)		
7	idDeltaEpoch	Mean Epoch of Declination	View	float		Time(j2000Time)		
8	idSourcePosition	Source of the position estimate	View	byte				
9	idmuAlpha	Proper motion in RA, μ at catalogue epoch	View	float		Angular Velocity(mas/year)		
10	idmuDelta	Proper motion in DEC, μ at catalogue epoch	View	float		Angular Velocity(mas/year)		
11	idmuAlphaError	Error in proper motion in RA	View	float		Angular Velocity(mas/year)		
12	idmuDeltaError	Error in proper motion in DEC	View	float		Angular Velocity(mas/year)		
13	idsourceRA	Source of the proper motions	View	byte				
14	idgalactl.on	Galactic Longitude	View	float		Angle(deg)		0
15	idgalactl.at	Galactic Latitude	View	float		Angle(deg)		-90
16	idgalactl.on	Galactic Longitude	View	float		Angle(deg)		0
17	idgalactl.at	Galactic Latitude	View	float		Angle(deg)		-90
18	idmagBJ	B mag measure, GSC23 system	View	float		Magnitude(mag)		
19	idmagB>Error	Error in B mag measure	View	float		Magnitude(mag)		
20	idsourceMagBJ	Source B mag	View	byte				
21	idmagRF	R mag measure, GSC23 system	View	float		Magnitude(mag)		
22	idmagRF>Error	Error in R mag measure	View	float		Magnitude(mag)		
23	idsourceMagRF	Source R mag	View	byte				
24	idmagG	G mag estimate	View	float		Magnitude(mag)		
25	idmagG>Error	Error in G mag estimate	View	float		Magnitude(mag)		
26	idsourceMagG	Source G mag	View	byte				
27	idmagGrs	Grs mag estimate	View	float		Magnitude(mag)		
28	idmagGrs>Error	Error in Grs mag estimate	View	float		Magnitude(mag)		
29	idsourceMagGrs	Source Grs mag	View	byte				
30	idclassification	Classification	View	boolean				

CI Number	CI Name	Remark
DB_T02_1	DPCT Main DB	DB containing all the received data from ESAC and all DPCT generated output for daily and DRC processing
DB_T02_2	DPCT Local DB	DB containing all the data that are needed for the software processing at DPCT.
DB_T02_3	IGSL DB	This DB is generated and maintained at DPCT from the catalog release performed by OATo science team.

R. Messineo, DPCT Internal Interface Control Document, GAIA-DT-SP-ALT-RM-002-10 30-06-2014

Contents of the DPCT Internal Data Model and to summarize the interfaces between the DPCT and the DPCE which hosts the Main Database.



L'archivio Italiano Gaia@DPCT



- Nel **Mission Database Italiano** di Gaia confluiscano, **quotidianamente**, i dati di missione scambiati con il DPCE (ESAC - Madrid) e tutti i risultati delle **«on site operations»**. A fine missione **conterrà tutti i dati raw e i risultati delle operations** condotte al DPCT durante la missione.
- Nel **Local Database di Gaia** vengono attualmente elaborati e gestiti i dati «day to day» del «ciclo corrente» (tendenzialmente semestrale) delle «operations» corrispondenti, almeno, ad una scansione completa della volta. **A Luglio è terminata la fase di «commissioning».**
- **L'archivio finale inteso come insieme di dati strutturati e non strutturati dovrebbe contenere circa 1 PB di dati accessibili !**



L'archivio Italiano Gaia@DPAC



Il satellite Gaia è quindi uno **strumento**, osservativo e di acquisizione dati, che produce i dati: osservativi e strumentali (grezzi) della missione.

I s/w delle pipe lines scientifiche di Gaia sono da considerare anch'essi parte dello **strumento**, con valenza logico - matematica, che produce dati astronomici e strumentali .

Il DB di Gaia è anch' esso, in se, parte di questo **strumento**, ne rappresenta la valenza Informatica e metodologica di accesso ai dati strumentali e scientifici, consentendone: **la conservazione, la gestione, l'organizzazione e la manipolazione** secondo schemi funzionali ai due punti precedenti. **Inoltre è l'unico strumento idoneo a mantenere traccia dell'autoconsistenza del contenuto dell'archivio che viene a prodursi.**

L'archivio di missione va quindi considerato, per l'impatto che ha sull'esito della missione, almeno al pari di uno strumento osservativo quando non parte costitutiva, esso stesso, del satellite - strumento !



Gaia@DPCT (accesso ai dati)



Interfacciamento al sistema del DPCT (attuale)

L'accesso per il controllo dello stato del processamento e l'accesso alle informazioni inerenti avviene su collegamento VPN ed interfaccia WEB previa autenticazione su entrambi i contesti

The top screenshot shows the Gaia@DPCT web interface with the following navigation menu: Welcome, Aspera, DTS, RMS, PFS, Output, Data Management, Configuration, AM, BMM, CSR, CACB, Mission Log, Message Board, SOC, Procedures, Decoders. The page displays a 'Sign In' button and a message: 'You are signed in as Roberto Morbidelli'.

The bottom screenshot shows the 'Mission Log' page with a table of mission events. The table has the following columns: ID, Date, Occurrence, Author, Type, Category, Subject, System/Procedure, and Text. The table contains three rows of data:

ID	Date	Occurrence	Author	Type	Category	Subject	System/Procedure	Text
2345	Fri Sep 12 18:22:39 2014	Fri Sep 12 19:20:59 2014	Lorenzo Bramante	Alert	General	EMF OPS OTS_HOOPS itanef	DPCT System	EMF OTSID: 20140912182436771; 803x101050: 275006543804449793; EMF OTSID: 20140912181437393; 803x101050: 275006543804449793
2344	Fri Sep 12 18:19:31 2014		Lorenzo Bramante	Problem Fixed	Workload	HK recover	DPCT System	
2343	Fri Sep 12 18:15:45 2014	Thu Sep 11 18:31:47 2014	Lorenzo Bramante	Problem Fixed	Data Management	HK OTSD: 20140910235912019; 20140910230129486; 803x101050: 275006543804449793	DPCT System	



Gaia@DPCT (accesso ai dati)



Interfacciamento al sistema del DPCT (attuale)

Il processamento dei dati da luogo a catene di documenti e continui log di eventi. Anche questi, per tramite dell'interfaccia WEB organizzata in cartelle e sono accessibili, per la lettura e valutazione dei contenuti, agli Operatori ed ai Manager dei vari sottosistemi.

The image displays three overlapping screenshots of the Gaia@DPCT web interface. The top-left screenshot shows the 'Data Management' section with a table of data processing jobs. The top-right screenshot shows the 'Data Management' section with a table of event reports. The bottom screenshot shows the 'WorkFlow Manager' section with a table of workflow tasks.

ID	STARTTIMESET	ENDTIMESET	ANOMALTYPE	DATASTREAM	ANOMALTYPE	REPORTINDICATE	REPORTER	CYCLE	MITIGATION	REPORT	STARTTIME	ENTIME
274	133	133	None	None	None	None	None	None	None	None	Sat Sep 13 10:24:40 UTC 2014	None
275	133	133	None	None	None	None	None	None	None	None	Fri Sep 12 10:02:24 UTC 2014	None

NAME	ID	STATUS	PERSISTENCEID	STARTTIME	ENTIME
AMCDAIingestAutWorkflo	32295	SUCCESS	28975	Sat Sep 13 19:40:42 UTC 2014	Sat Sep 13 19:42:15 UTC 2014
MCDAIingestAutWorkflo	32076	SUCCESS	28974	Sat Sep 13 19:33:20 UTC 2014	Sat Sep 13 19:34:46 UTC 2014
AMCDAIingestAutWorkflo	32058	SUCCESS	28973	Sat Sep 13 19:33:18 UTC 2014	Sat Sep 13 19:35:01 UTC 2014
MCDAIingestAutWorkflo	31873	SUCCESS	28972	Sat Sep 13 19:30:35 UTC 2014	Sat Sep 13 19:31:39 UTC 2014
AMCDAIingestAutWorkflo	31655	SUCCESS	28971	Sat Sep 13 19:30:13 UTC 2014	Sat Sep 13 19:32:37 UTC 2014
BAHFReportAutWorkflo	22211	SUCCESS	28970	Sat Sep 13 14:37:22 UTC 2014	Sat Sep 13 14:40:03 UTC 2014
BAHFWebProcessingWor	22114	SUCCESS	28969	Sat Sep 13 14:35:41 UTC 2014	Sat Sep 13 14:37:42 UTC 2014
BAHFDataExtractorWork	22094	SUCCESS	28968	Sat Sep 13 14:35:41 UTC 2014	Sat Sep 13 14:37:47 UTC 2014
H0ingestAutWorkflo	31442	SUCCESS	28967	Sat Sep 13 13:37:36 UTC 2014	Sat Sep 13 13:39:09 UTC 2014
MCDAIingestAutWorkflo	31317	SUCCESS	28966	Sat Sep 13 13:36:32 UTC 2014	Sat Sep 13 13:37:58 UTC 2014
AMCDAIingestAutWorkflo	31199	SUCCESS	28965	Sat Sep 13 13:36:30 UTC 2014	Sat Sep 13 13:38:08 UTC 2014
MCDAIingestAutWorkflo	31335	SUCCESS	28964	Sat Sep 13 13:34:36 UTC 2014	Sat Sep 13 13:36:51 UTC 2014

SOLUTIONID	STARTTIMESET	ENDTIMESET	ANOMALTYPE	DATASTREAM	ANOMALTYPE	REPORTINDICATE	REPORTER	CYCLE	MITIGATION	REPORT	STARTTIME	ENTIME
275004502	B	null	OTHER	OTHER	igwrtts	0	Admission	0	Admission	2560201812	0	
163400150	A	null	INSTRUMENT_CH4	INSTRUMENT_CH4	wormulan	0	MIB RED N	0	MIB RED N	8592629890	0	
163400150	A	null	INSTRUMENT_CH4	INSTRUMENT_CH4	wormulan	0	MIB, MAA, F	0	MIB, MAA, F	86105224000	0	
275001052	B	AB	EMPTY	EMPTY	nrtseek	0	DPAC data	0	DPAC data	2329298211	0	
163400150	A	null	INSTRUMENT_CH4	INSTRUMENT_CH4	wormulan	0	Flown Required	0	Flown Required	16 35 10 m	70861224000	708622236
275094902	B	NPFS_MFS_Cable	HARDWARE_PDOI	HARDWARE_PDOI	nrtseek	0	Transitio	0	Transitio	2386889812	2387720923	
163400150	A	null	INSTRUMENT_CH4	INSTRUMENT_CH4	wormulan	0	MIB RED N	0	MIB RED N	85261224000	0	
275094902	B	null	HARDWARE_PDOI	HARDWARE_PDOI	igwrtts	0	On 2014-06	0	On 2014-06	25313078212	0	
275054502	B	File Data from VP	HARDWARE_PDOI	HARDWARE_PDOI	gntas	0	VPUJ was i	0	VPUJ was i	26361108212	2636796955	
275054502	B	null	ORBIT_MANOVR	ORBIT_MANOVR	nrtseek	0	A station	0	A station	2599449212	2598893923	
163400150	A	null	INSTRUMENT_CH4	INSTRUMENT_CH4	wormulan	0	MIB NOAH	0	MIB NOAH	85261224000	0	
275001052	B	File Data from VP	HARDWARE_PDOI	HARDWARE_PDOI	nrtseek	0	Use Paper	0	Use Paper	22603538211	2264859823	



Gaia@DPCT (accesso ai dati)



Interfacciamento al sistema del DPCT (attuale)

Il portale può essere anche un comodo punto di partenza per l'accesso alle risorse del SOC la cui organizzazione delle informazioni è molto simile a quella presente al DPCT.

Gaia Mission Operations
DPCT Morning Briefing Commissioning Phase #001
10.Jan.2014

Written by: F. Marziani
Start at 09:46 End at: 10:15

Participants:

DPCT Role	Persons	Present	Remote
Manager	R. Messineo	X	
Operations Manager	F. Marziani	X	
Scientific Manager	R. Morbidelli	X	
Operator	M. Vaschetto	X	
Infrastructure Engineer	A. Mulone		
Framework Engineer	E. Pigozzi	X	
Interface Engineer	E. Pigozzi	X	
BAM Scientist	A. Riva	X	
AIM Scientist	D. Busonero	X	
PAIGA	V. Icardi	X	
Program Manager	F. Solitto	X	
Other	M. Martino	X	

1. Infrastructure status:

- HW status: Ok. Some issue is still remaining for what is concerning backup on tape.
- SW infrastructure status: BAM patch has been installed on operations environment.
- DB status: Ok

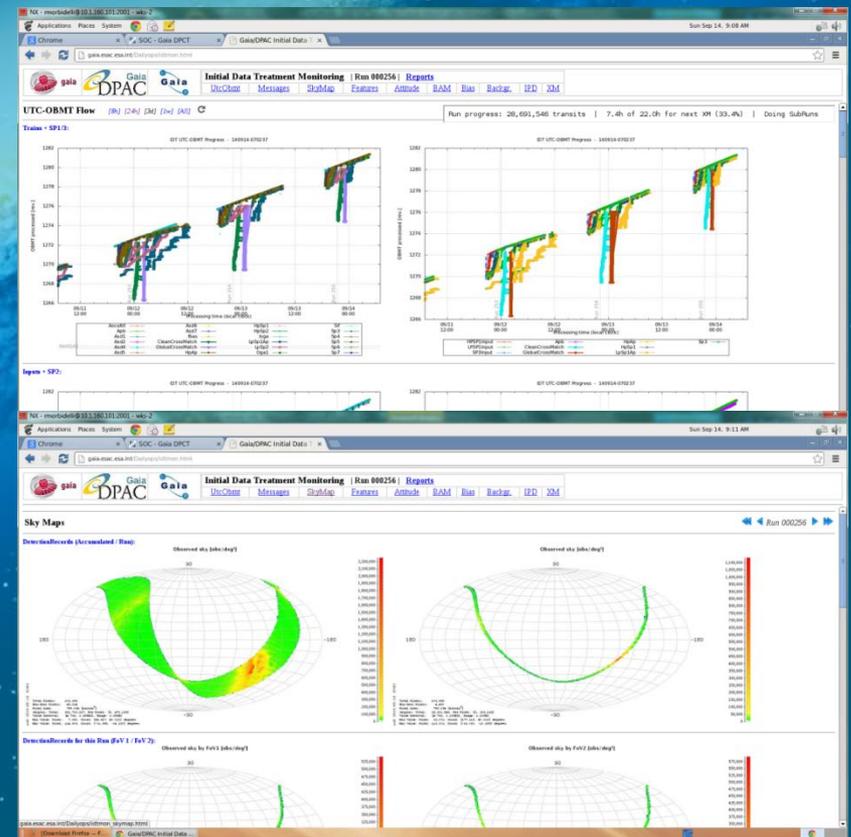
2. Data Transfer status and timeline:

- DPCE contacts and notifications: BAM-redundant laser was switched-on, what can explain all of the errors and warns found (TBC). BAM-ROW 3 is found in the logs (it seems the decoding of the cod Row is wrong for the BAM snapshot packages. TBC). See mission log entry 64 for further details.
- Data transfer to/from DPCT: Data transfer ok; HK and BAM data are being received regularly. No data transfer completed towards DPCE.
- Issues: none.
- Any other topic: Web Interface tool filter is not working correctly.

3. Data processing:

- AIM processing: N/A
- BAM processing: Run 4 has been completed successfully and weekly run 1 started and failed. Therefore no weekly report has been produced. Both weekly and daily

consequences were... The DTS has been... at the same time. The... 2459 and 24788). The...
GaiaMibOm 15.2.0;... structureSW 16.0.2;... ingestion workflow... pending and the... This issue will be... system has been... been completed on... in-going. An email to... issue resolved and... been found in the... gation... EW_IDX index, the... Infrastructure SW...
IDT run 252 has been completed while run 253 is...
contacts and notifications: DPCE Pipeline has been stopped yesterday for... since there was a change in the main satellite software and no scientific telemetry... will be delivered. The DPCE Pipeline will start again today.





Gaia@DPAC (Risorse H/W)



- **HW per il processamento (attuale):**
 - **Storage System: 2 - HP 3PAR StoreServ7400.** [ciascuno: 4 controller nodes, 72 X 500 GB high performance disks and 144 X 2 TB high capacity disks).
 - **Database server: 3 - HP DL570G7 [4 CPU XEON 8-cores e 256 GB RAM]**
 - **Processing Server: 10 - HP DL570G7 [4 CPU XEON 8-cores e 256 GB RAM]**
 - **Processing Server: SGI Altix 450 [8 ITANIUM 64 CPU 2 - Core, 256 GB RAM]**
 - **Managem. Server: 2 - HP DL570G7 [4 CPU XEON 8-cores e 128 GB RAM]**
 - **Backup server: 1 - HP DL570G7 [4 CPU XEON 8-cores e 256 GB RAM]**
 - **Backup System : HP ESL G3 Tape library [4 x LTO-5 Drives 100 tapes] + HP 3PAR StoreServ7400 Snapshot**
 - **Network: Redundant Cisco 4500X switch (con 10 Gbps ports), + un cluster di 4 Cisco 3750 X switches**
 - **User WorkStations: 4 - Hp Z820 (4 monitor ciascuna in sala Gaia)**
 - **Firewall : Pf sense installato su 2xHP DL380G8**
 - **Internet Link : fornito dal GARR (band 300 Mbps max, scalabile fino ad 1 Gbps)**
 - **Link con il CINECA su rete GARR per il trasferimento dati modulo GSR**





Gaia@DPCT (Risorse H/W)



- **HW per sviluppo e test (attuale):**

- **HP EVA 4400 storage array, dual controller, con 48 TB su dischi ad alta capacità, e 18 TB su dischi ad alte prestazioni.**

SAN Switch in comune con la piattaforma operativa.

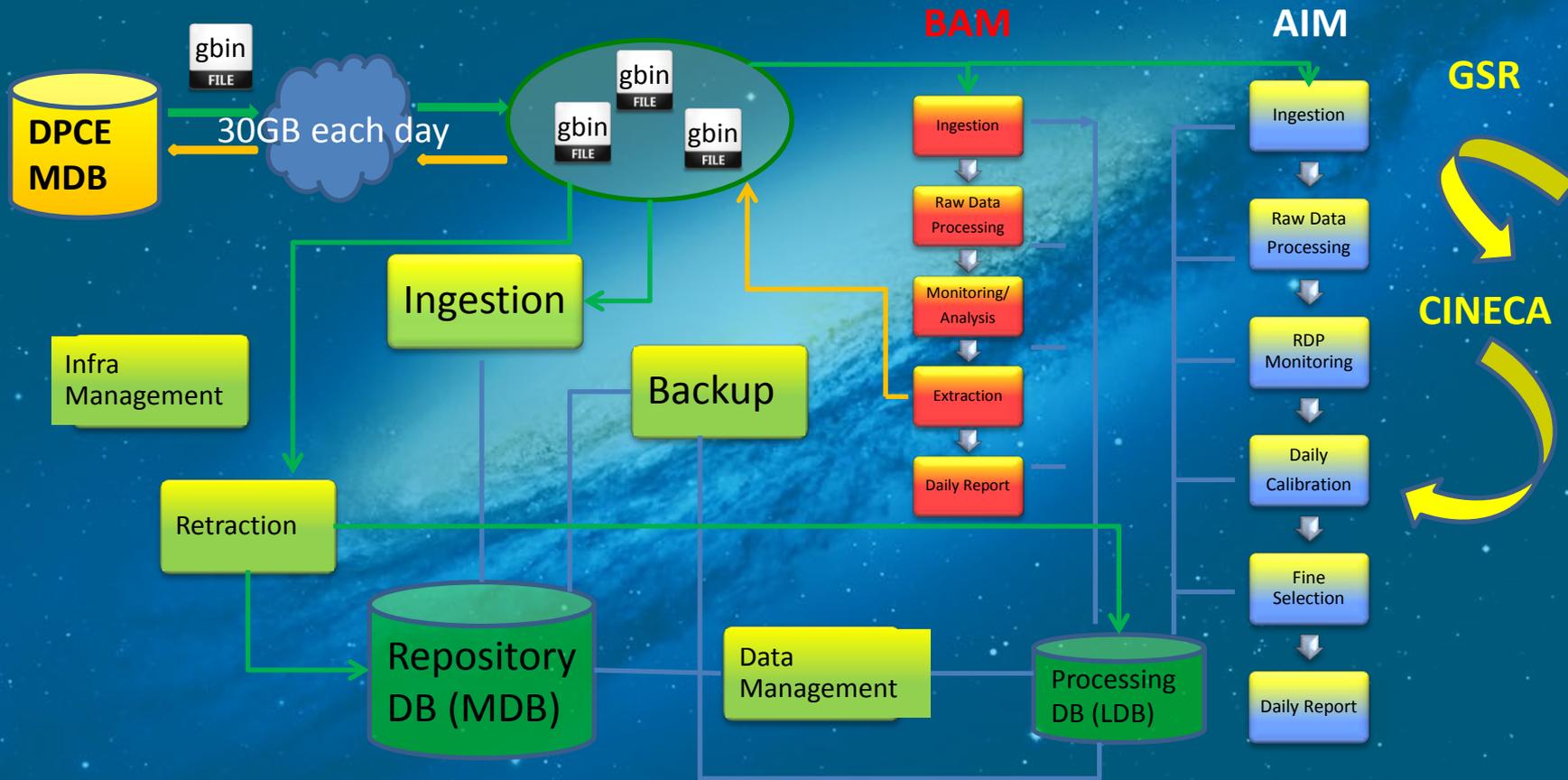
HP Proliant Server (Processing and Database purpose)

- 5 HP DL580 G7 (4 x 4core, 128GB RAM)
- 2 HP DL380 G8 (2 x 8core, 128GB RAM)
- 1 HP DL380 (management server)
- *SGI Altix 450: 8 CPU 2 Core, 256 GB Shared Memory (risorsa condivisa con segmento ops.)*

Backup system, Network, firewall e Internet in comune con la piattaforma operativa



Gaia@DPCT (Processing)



Con l'attuale disponibilità di H/W tutto il processamento dei dati quotidiani avviene nell'arco delle 24 ore (più o meno)



Gaia@DPCT DB (DBMS Oracle)



Collocazione:

- I 2 DB: MDB e LDB sono collocati su 3 server HP DL570G7 (un quarto è off-line per eventuale crash recovery e gestione del backup) per tramite del RAC Oracle presso l'area CED ALTEC - TO
- Sono gestite sotto lo stesso contratto di licenziamento e supporto le attività "ground based" presso il Telescopio Nazionale Galileo nelle Canarie per la parte connesso alla missione Gaia.

SW Oracle coinvolto:

- Real Application Clusters - Named User Plus Perpetual
- Tuning Pack - Named User Plus Perpetual
- Oracle Advanced Compression - Named User Plus Perpetual
- Tuning Pack - Named User Plus Perpetual
- **Spatial and Graph - Named User Plus Perpetual**
- Partitioning - Named User Plus Perpetual
- Diagnostics Pack - Named User Plus Perpetual

Il s/w è disponibile in uso perpetuo e per il progetto viene gestito via Enterprise Manager. Al medesimo è annualmente associato un supporto di base (CSI) ed un supporto avanzato (ACS)



Entità dell'archivio di missione

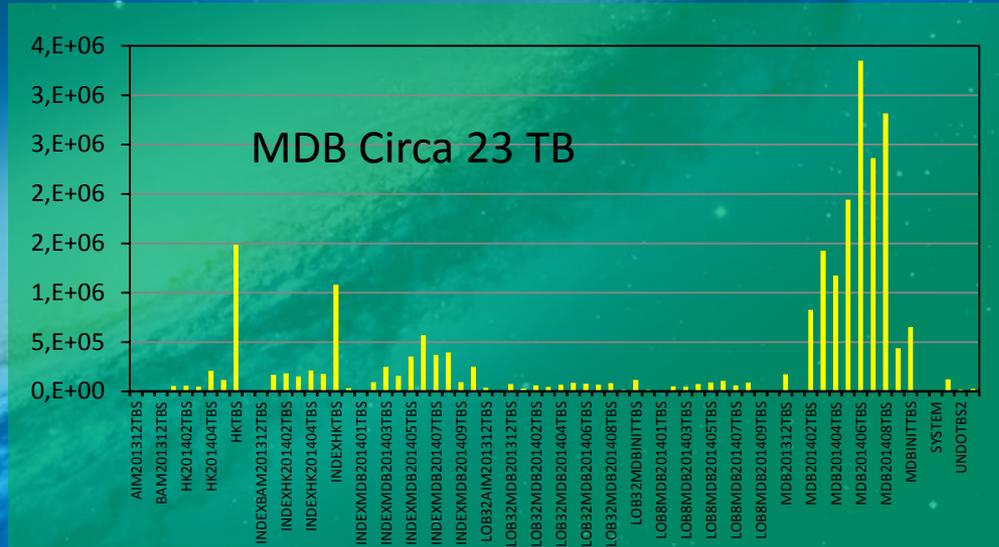


Situazione attuale (4/09/2014):

AIM Schema			
Tablespace	Used MB	Free MB	Total MB
AIMTBS	2.896.825	84.795	2.981.620
INDEXAIMTBS	779.596	32.452	812.048
LOB32AIMTBS	2.277.204	29.216	2.306.420
LOB8AIMTBS	451.773	6.347	458.120
	6.405.398	152.810	6.558.208

BAM Schema			
Tablespace	Used MB	Free MB	Total MB
LOB32BAMTBS	648.152	21.320	669.472
INDEXBAMTBS	677	5.571	6.248
LOB8BAMTBS	18.849	1.023	19.872
BAMTBS	8.063	238.009	246.072
	675.741	265.923	941.664

INFRA Schema			
Tablespace	Used MB	Free MB	Total MB
INFRATBS	24.195	1.129	25.324
INDEXINFRATBS	1.159	65	1.224
LOB32INFRATBS	1.318	106	1.424
	26.672	1.300	27.972



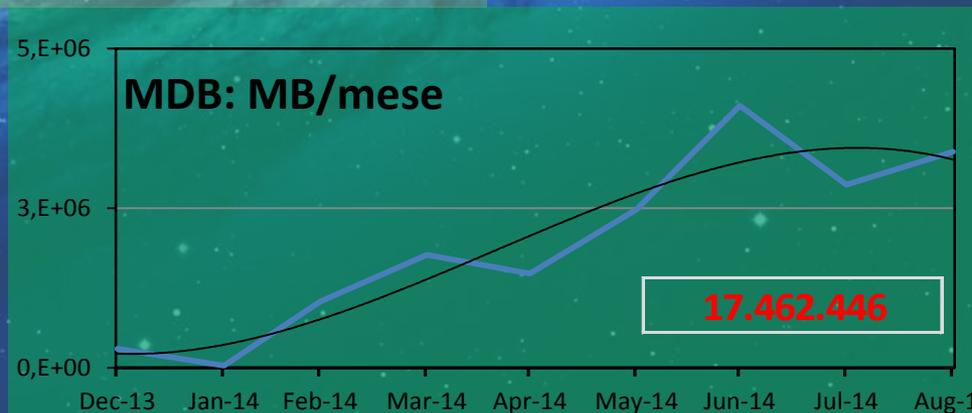
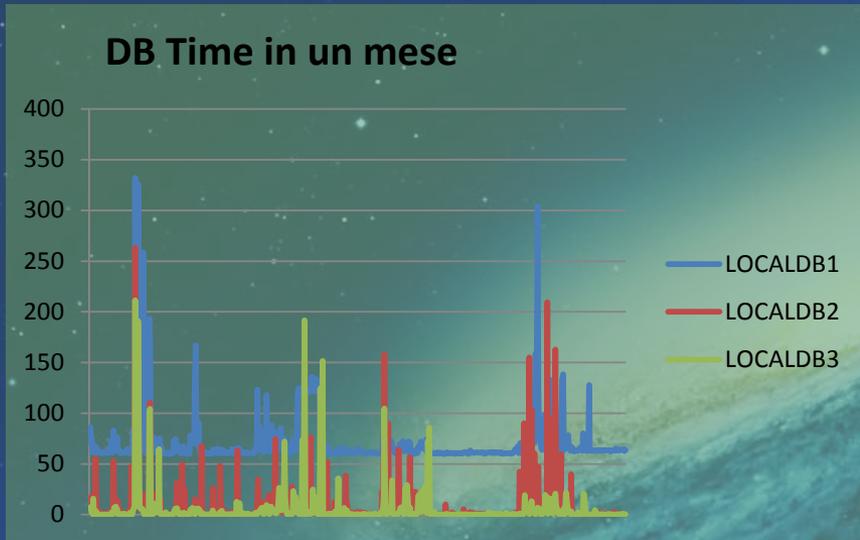
Allocazioni di spazio nelle tablespaces del Repository Database. Sono connesse ai cicli mensili di esecuzione dei task costituenti la pipe line di processamento dei dati del modulo AVU e di tutti i dati prodotti dal s/w di infrastruttura nell'arco di un semestre di operazioni. Da questi conteggi e da quelli che seguono mancano ancora i carichi derivanti dal modulo GSR (Global Sphere Reconstruction). Considerevole produttore anch'esso di «complicazione» della linea di «data processing».



Entità dell'archivio di missione



Caratteristiche «dimensionali» del DB Gaia

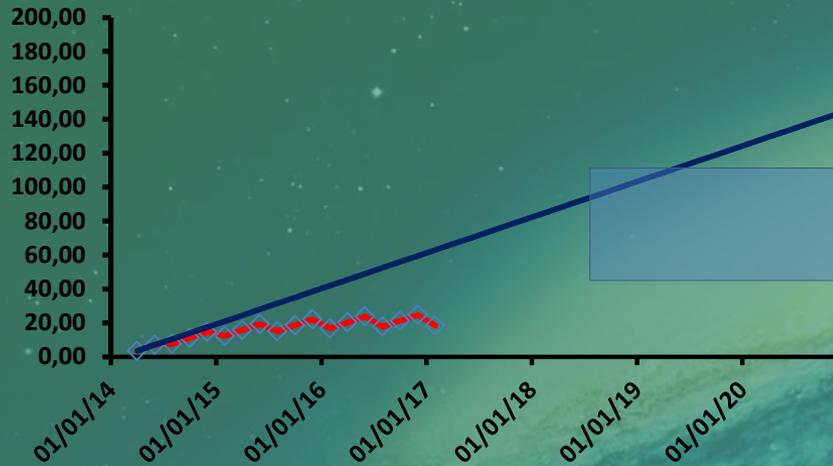




Entità dell'archivio di missione

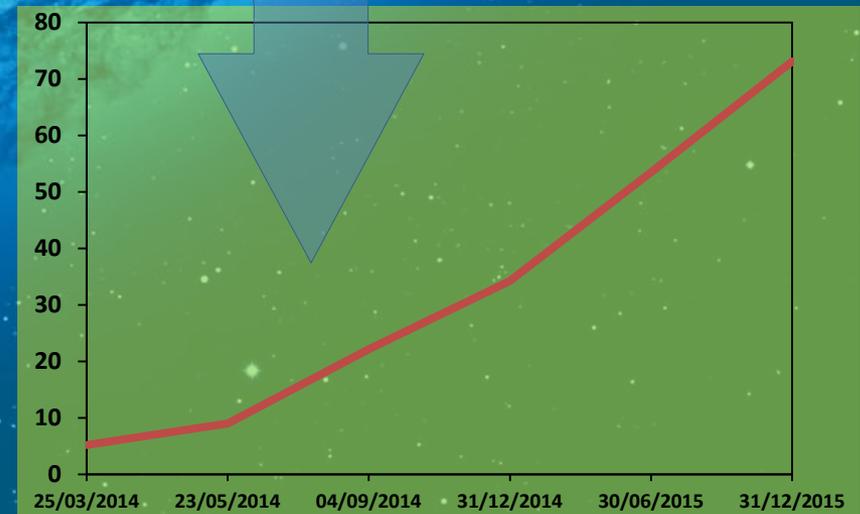


A confronto l'entità dello spazio allocato dal Local Database e dal Mission Database



data	LDB Size in TB	
	LDB Cleaned	LDB uncleaned
30/04/14	3,50	3,50
30/06/14	7,00	7,00
31/08/14	7,70	10,50
31/10/14	11,20	14,00
31/12/14
28/02/15

Mission Database Size in TB		
Date	Size in TB	
25/03/14	5,20	Coll. data
23/05/14	9,00	
04/09/14	22,17	
31/12/14	34,20	Exp. data
30/06/15	53,51	



Dimensione della base dati a fine missione stimata pari ad almeno 250TB

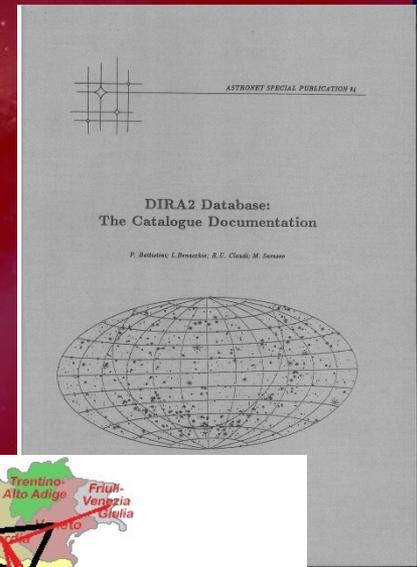


DIRA un VO «ante litteram»



- Prodotto per la gestione informatizzata di cataloghi astronomici.
- Accessibile tramite rete Decnet è sviluppato in Fortran e C.
- Multiplatforma (VMS e UNIX) Prima redazione intorno al '84
- Fruibilità su tutti i nodi della rete Decnet dotati del S/W DIRA
- Possibilità di trasferire in toto l'archivio e i metodi di accesso per tramite di PDB
- Completamento della II release con tools statistici e grafici intorno al 1986
- Circa 100 cataloghi distribuiti su 3 poli:
 - Bologna: Cataloghi di oggetti Radio
 - Torino: Cataloghi Astrometrici (incluso il GSC1)
 - Padova: Spettroscopici ed Extragalattici

Anche se il sistema è stato abbandonato la similitudine sul piano concettuale e metodologico, sia con il VO che con la logica di distribuzione nodale replicabile resta, e ne fa, una sorta di precursore dell'idea di accesso all'informazione distribuita e replicata sia per contenuto dell'archivio che per i metodi di gestione.

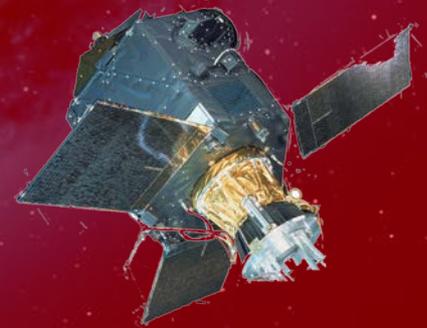




Hipparcos – un consuntivo



- Inizio studio missione (accettazione): 1980.
- Lancio: agosto 1989.
- Fine operatività del satellite: marzo 1993 (3.5 anni).
- Prima redazione catalogo: 1997
- Completamento catalogo Tycho 2 (fine missione): 2000.
- Astrometria al millisecondo di 120,000 stelle.
- Fotometria di 2,539,913 di stelle (completo alla 11^a).
- Dati elaborati in parallelo da 2 consorzi.
- **Rielaborazione di Floor van Leeuwen ultimata nel 2007.**
- Raw data prodotti: < 1 TB.
- Cataloghi prodotti: Hipparcos e Tycho (< 10 GB).



“A VLBI Resolution of the Pleiades Distance Controversy” (Submitted on 27 Aug 2014) è un esempio del problema del ri-utilizzo dei dati raw di una missione a operazioni ultimate.

Ma l'archivio dei dati raw di missione non è più disponibile !



HST (stabilità a lungo termine)



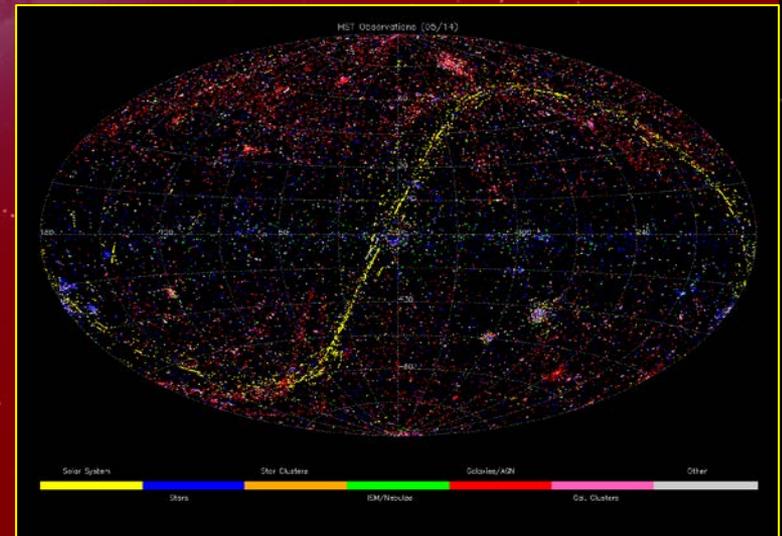
MAST è un archivio «maturo» di circa 100 TB !

Hubble Space Telescope (HST) is an orbiting astronomical observatory operating from the near-infrared into the ultraviolet. Launched in 1990 and scheduled to operate at least through 2015. HST carries and has carried a wide variety of instruments producing: imaging, spectrographic, astrometric, and photometric data.

MAST is the primary archive and distribution center for HST data, distributing science, calibration, and engineering data to HST users and the astronomical community at large. About 1×10^6 observations of more than 30 000 targets are available for retrieval from the Archive.

- Tutti I dati archiviati sono in formato FITS
- 1995 produzione dati: circa 1.5 TB/anno
- Aprile 2014: l'archivio dati vanta ~ 100 TB!

(Fonte STScI)

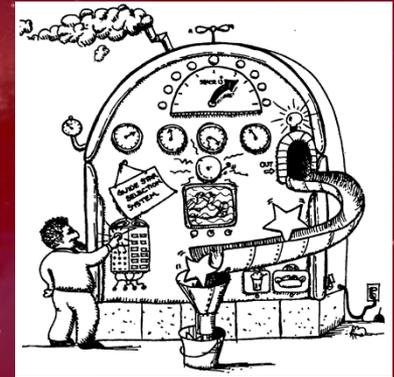




Il GSCII (meta) dati «dormienti»



- ✓ Segue la pubblicazione del GSC I (1989) è prodotto a supporto delle HST operations.
- ✓ Inizio attività per la seconda serie di digitalizzazioni (lastre fotografiche GSCII; 1988)
- ✓ Acquisizione e riduzione di ~ 8000 campi CCD (catalogo fotometrico GSPC II)
- ✓ Ultima revisione catalogo (2.3.2 in uso per il puntamento accurato di HST - 2008)
- ✓ Dimensioni archivio immagini digitali (9137 digitalizzazioni di lastre): ~ 7 TB
- ✓ Dimensioni finali catalogo GSC II (3.2.3): ~120 GB (32768 FITS files)
- ✓ Contenuto: posizioni, classificazione e fotometria BVRI (945,592,683 oggetti)
- ✓ Dimensione finale database misure oggetti: ~ 3.7 TB + 3.4 TB (DBMS: Objectivity 8)



- ✓ Versioni DB: DUE, ciascuna articolata in 32768 data container secondo gerarchia HTM
- ✓ Contenuto del DB: Ogni versione contiene più di 7 miliardi di record (misure raw)

L'archivio dei dati raw di missione è collocato sia presso lo STScI che presso l'OATo - INAF

I DBMS di entrambe le versioni sono operativi (ma dormienti)

Il contenuto del secondo archivio (INAF – OATo) è servito come lista oggetti di base per la produzione della base dati utilizzata per realizzare l'Input Gaia Source List (IGSL)

Su 32768 tabelle ad oggi sono corrotte: 5 (Lo erano da principio !)

Su 9137 plates Schmidt digitalizzati ne sono preservati 8688



Preservation, Data Mining e...



Hipparco di Nicea, si ritiene, abbia quantificato l'entità della precessione degli equinozi in virtù del confronto tra misure di posizione delle stelle raccolte da Caldei, Babilonesi ed Egizi, a lui pervenute. Quelle che confrontò con le sue si presume coprissero più di 2000 anni ed erano inerenti un migliaio di oggetti !

E' arbitrario ritenere che il cardine dell'astronomia, quella del XX e XXI secolo, così come quella della storia remota dell'umanità, debba necessariamente fare affidamento sulla preservazione dell'integrità ed accessibilità dei dati (grezzi ed elaborati) per lungo tempo ?

A mio parere e sono confidente sia il parere di molti: NO!



Backup and Mirror strategies



INAF (fortunatamente) non è un ospedale ma non vorrei mai leggere comunicati del genere

«.....Visto il perdurare dei disagi ai nostri pazienti e il comparire di articoli sulla stampa contenenti informazioni fuorvianti e non puntuali, la Direzione del Presidio Sanitario XXXXXXX, ha ritenuto opportuno fare il punto della situazione ad una settimana dal blocco subito dai nostri server di produzione. Vogliamo assicurare i nostri pazienti: i loro dati sono al sicuro, non sono stati in alcun modo corrotti e non c'è stato alcun scambio di dati tra pazienti. Smentiamo inoltre le dichiarazioni riguardanti prenotazioni di pazienti deceduti: dalle 15.30 di lunedì 20 gennaio.... **il sistema di produzione è semplicemente inaccessibile** pertanto non abbiamo più potuto effettuare alcuna prenotazioni.

Il problema principale, ad oggi, consiste nella **ricostruzione dei dati**, che pur essendo integri non sono accessibili dal nostro sistema. **Per ovviare a questo problema da una settimana abbiamo avviato tutte le procedure necessarie, fino all'invio dei nastri di backup alla filiale tedesca di un'azienda americana (Kroll Ontrack) che sta cercando di ricostruire la struttura logica del database per rendere accessibili i dati. Non siamo ad oggi in grado di garantire che questi saranno nuovamente utilizzabili**, ma ribadiamo che non sono stati in alcun modo diffusi o danneggiati, questo significa che nel momento in cui riusciremo ad accedere nuovamente alle informazioni non ci saranno **sovrapposizioni o perdite**.

Fonte: La Stampa, gennaio 2014

La verifica della fruibilità di un risultato scientifico (il dato) è relativamente semplice (il file c'è o no?). Ma la verifica della consistenza degli Archivi dei dati è altra cosa! Presuppone aver costantemente «cura» del dato! «.. Ieri quella tabella aveva 10 oggetti oggi ne ha 9.. ??!»

L'uso del grid è certamente una parte della soluzione del problema. E' orientata, soprattutto, alla distribuzione dei risultati, talora associata alla replica dei «servizi» che ne garantiscono la continuità d'uso; VO docet! Ma questo, da solo, temo, sia insufficiente a garantire integrità e fruibilità globale quando si parla dell'archivio!
Unica via, a mio parere, è possedere l'archivio.



In quest'ottica potremmo dire che l'equivalente, in INAF, di una «biblioteca Benedettina o Alessandrina» non è stata, ancora, costruita ...



Architettura dati missione Gaia



Allo stato attuale Gaia

- Punti di forza (almeno in Italia):
 - Distribuzione delle competenze per il s/w di processamento (varietà).
 - Affidabilità temporale del DBMS adottato (Oracle).
 - Dedicata linea di distribuzione e tutela dei dati (il tutto è su rete privata).
 - Flessibilità nella scelta delle architetture
 - Centralizzazione delle risorse e informazioni di processamento (ALTEC).
- Punti di debolezza (?):
 - Dati raw scambiati / conservati in formato proprietario (Gbin chi è costui ?).
 - Scelta di far affluire parte dei dati con serializzazione java nel DB (LOB)
 - Limitata ridondanza di processamento (dov'è il disaster recovery ?) .
 - Varie strategie di gestione dei dati nei DPCs «non DPCE».



Un archivio Astronomico Italiano



Concentrare la conoscenza: Tracciabilità di quanto fatto.

Chi sa è perché lo fa (sinergie ricerca – industria quasi un referaggio) !

Decentrare la competenza: Proattività nei processi.

Condividere e realizzare con chi sa fare per non improvvisare (ALTEC, CINECA, EURIX, ACS Oracle, Supporto HP, Supporto SGI ...)!

Concentrare le risorse: La cassaforte !

Meglio condividere che possedere è sempre vero ?



Decentrare gli strumenti: Diversificare i nodi dal forziere ?

*Se questo **duplica** quanto c'è in cassaforte: SIA!*

*Ma un forziere, nel quale conservare, con cura, **tutti i dati** deve esistere !*



Prospettive



Il Gaia DPC Italiano è ad oggi un “key asset” per ASI ed INAF.

Attualmente il data center sta dimostrando capacità di:

- Partecipazione affidabile a tutte le fasi di processamento della missione
- Gestione di archivi di grosse dimensioni
- Supporto logistico alle operazioni scientifiche
- Implementazione di architetture per l'esecuzione di pipeline complesse
- **Adeguato controllo delle «unpredictables» (successo OR* e commissioning Gaia)**
- Sostenibilità delle «High Performance» richieste dalle attività connesse.
- Efficace disegno, calibrazione e monitoraggio delle infrastrutture ICT coinvolte.

Il DPC Italiano d'altro canto mostra evidenti potenzialità per:

- Sviluppare e implementare sistemi avanzati e tecnologie di supporto alla ricerca scientifica basati su **regole di nuova concezione** (es. il DB stesso e un uso dei molti metodi in esso esistenti di accesso divengono strumentali alla missione o di sviluppo per missioni successive)
- Sviluppare metodi avanzati di **conservazione dei dati** per un uso scientifico di lungo termine, con **«reale e materiale»** capacità di **retro - processamento** parziale o globale degli archivi.
- **Fornire servizi** per tramite di infrastrutture dinamicamente **funzionali alla comunità scientifica e tecnologica** mantenendo sincronia con l'evoluzione del mondo dell'ICT.



Prospettive



Sinergie con realtà commerciali il caso: Oracle

Il colloquio instauratosi da più di un lustro tra DPCT e Oracle per Gaia non si è esaurito in una logica di «fornitura» ma continua ad evolversi in una volontà di compartecipazione e costruttiva soluzione dei problemi connessi alla gestione dei dati di questa missione ed oltre. POC congiunte INAF (OATo/TNG) – Oracle in questo senso sono in sviluppo per creare dei «Challenge Cases» che possano essere eseguiti in ambienti Oracle.

Questo nello spirito non solo di affrontare i problemi quando questi si presentano (soluzioni); ma cercando di studiare preventivamente, sul piano tecnico, gli scenari e gli aspetti funzionali per i quali i dati disponibili sono e risulteranno fruibili.

L'INAF può risparmiare risorse umane rivolgendosi a scelte ponderate, perché preliminarmente testate, di soluzioni consolidate dal mercato e d'altro canto **l'azienda produttrice di ICT può trovare nelle tematiche astronomiche stimoli per porsi obiettivi ambiziosi sul piano infrastrutturale e metodologico.**

Uno «telescopio» termina il suo ciclo di utilizzo e finisce in un museo !

Le osservazioni astronomiche effettuate con esso, le misure derivate, l'esperienza che le ha prodotte, possono ambire a perdurare su tempi di gran lunga maggiori !

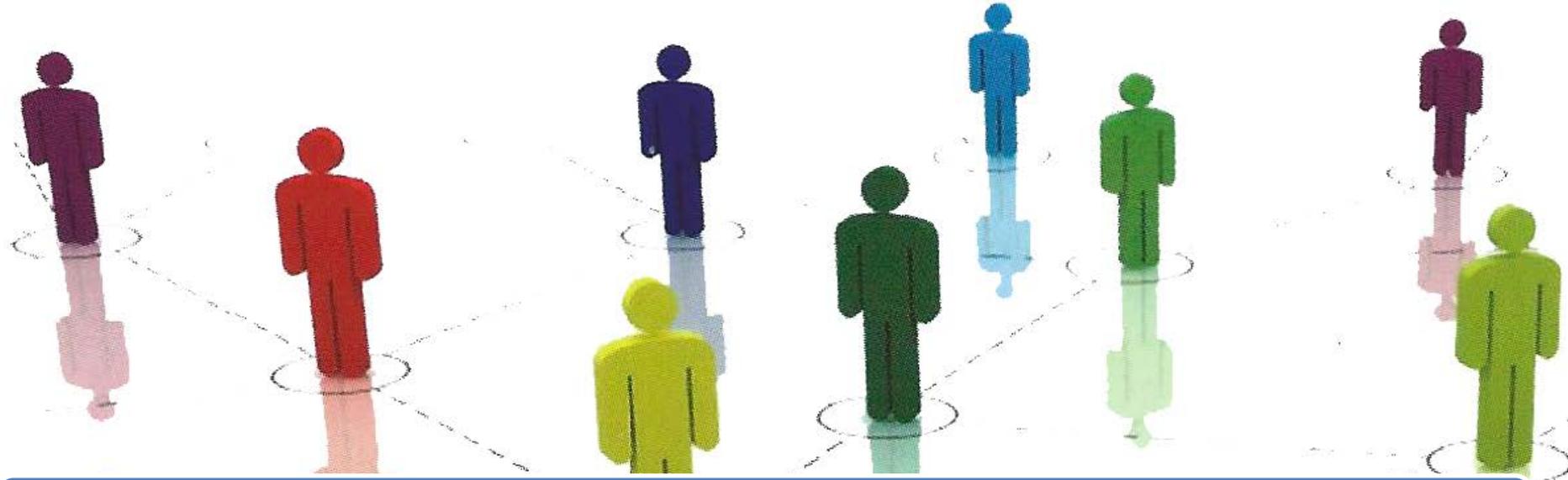


Concludendo



Un «social network galattico» ed oltre ...

- VO ? Sì grazie, ma non solo! Per la **biblioteca Nazionale** qual' è la via ?
- Biblioteche periferiche ? Certo possibili, auspicabilmente sinergiche a costituire un **backup ulteriore** di quella centrale, ma **non intese come «esclusive»** di quella centrale!
- Quando i bit divengono **«dato scientifico»** la genesi che ha prodotto quest'ultimo è **preservata ? Riproducibile (Galileo) !!**
- Avere un occhio attento e propositivo a cosa offre il «mercato ICT» **evita dispendio di energie economiche ed intellettive.**



Non c'è futuro se non si è capaci di serbare la memoria del passato

